The Glycine Cleavage System

OCCURRENCE OF TWO TYPES OF CHICKEN H-PROTEIN mRNAs PRESUMABLY FORMED BY THE ALTERNATIVE USE OF THE POLYADENYLATION CONSENSUS SEQUENCES IN A SINGLE EXON*

(Received for publication, July 28, 1989)

Masayuki Yamamoto‡, Hirohisa Koyata, Chihiro Matsui, and Koichi Hiraga§

From the Department of Biochemistry, Tohoku Medical and Pharmaceutical University School of Medicine, Sendai 980, Japan

Several cDNAs encoding H-protein, a constituent of the glycine cleavage system, were cloned from chicken liver cDNA libraries with an antibody against rat H-protein or with a nick-translated cDNA of an immunoreactive clone. The structure of the H-protein cDNA consisting of 910 base pairs was determined using clones with an apparent overlap in the nucleotide sequence. The cDNA encodes the precursor form of H-protein that is comprised of 39 amino acid residues for a mitochondrial presequence and 125 amino acid residues for the mature protein, following a 5' untranslated region of 13 base pairs. There are two genuine consensus sequences for the cleavage/polyadenylation of the precursor H-protein mRNA in the 3' untranslated region of the cDNA sequence. We showed by comparison with the 6-aminolevulinate synthase gene that only one copy of the H-protein cDNA occurs in the haploid genome of the chicken. Nevertheless, two types of H-protein mRNAs, which differ by the length of their 3' untranslated region, are produced in liver.

The chicken H-protein gene extends over 8 kilobase pairs on the genome and includes 5 exons that encode the entire cDNA sequence. Two AATAAA motifs are coded in the last exon of this gene, suggesting that the differently sized H-protein mRNAs are produced by the alternative use of these motifs.

Hydrogen carrier protein (H-protein, (1)) is a component of the glycine cleavage system which contains three additional constituent proteins: glycine decarboxylase (2) (tentatively named as P-protein (3)), tetrahydrofolate-requiring protein (T-protein), and lipoamide dehydrogenase (3). H-Protein binds lipoic acid as the prosthetic group and acts as the carrier of H-protein or with a nick-translated cDNA of an immunoreactive clone. The structure of the H-protein cDNA consisting of 910 base pairs was determined using clones with an apparent overlap in the nucleotide sequence. The cDNA encodes the precursor form of H-protein that is comprised of 39 amino acid residues for a mitochondrial presequence and 125 amino acid residues for the mature protein, following a 5' untranslated region of 13 base pairs. There are two genuine consensus sequences for the cleavage/polyadenylation of the precursor H-protein mRNA in the 3' untranslated region of the cDNA sequence. We showed by comparison with the 6-aminolevulinate synthase gene that only one copy of the H-protein cDNA occurs in the haploid genome of the chicken. Nevertheless, two types of H-protein mRNAs, which differ by the length of their 3' untranslated region, are produced in liver.

The chicken H-protein gene extends over 8 kilobase pairs on the genome and includes 5 exons that encode the entire cDNA sequence. Two AATAAA motifs are coded in the last exon of this gene, suggesting that the differently sized H-protein mRNAs are produced by the alternative use of these motifs.

This enzyme system appears to be unusual in many respects (8). Glycine decarboxylase by itself exhibits extremely low activity and requires H-protein as a regulatory protein for the expression of this activity. The purified glycine decarboxylase and H-protein form an enzyme complex in vitro; 1 mol of glycine decarboxylase (a homodimer) binds 2 mol of monomeric H-protein, resulting in a structural change at the active site of glycine decarboxylase and conversion to the active enzyme (8). This interaction plays a key role in catalyzing the initial step of the glycine cleavage (8).

Contents of the component proteins which were estimated by specific enzyme activities appear to vary from tissue to tissue in vertebrates (9). One can anticipate a closely related mode of biosynthesis of component enzymes. To our knowledge, to date, no study has reported either the biosynthesis or the regulation of this process for the components of the glycine cleavage system. Accordingly, we attempted to clone a cDNA encoding chicken H-protein and use the cloned cDNA as a probe to study the regulation mechanisms involved. We have already isolated human H-protein cDNA (10), and characterization of the human H-protein gene is in progress in our laboratory. In this paper, we report on the cloning of the cDNA and gene encoding chicken H-protein, and their structures, and suggest that two different types of H-protein mRNAs are produced by the alternative use of two AATAAA motifs localized in a single exon.

EXPERIMENTAL PROCEDURES

Materials—One-month-old White Leghorn hens were used. Radioactive nucleotides were products of Du Pont-New England Nuclear and [3H]anti-rabbit IgG, F(ab')2, was from Amersham, Japan. Several plasmid vectors (pGEM1, pGEM3 (Promega), and Bluescript (Strategene Cloning Systems)) were used. Other materials were commercially obtained.

Cloning of Chicken H-Protein cDNA—In advance, we confirmed that an immunopurified antibody raised against rat H-protein (10) reacted with chicken H-protein. A chicken liver cDNA expression library (11) was screened by the method described previously (11) using this antibody and a cDNA of an immunoreactive clone (pCH3a, about 0.5 kb). A commercially obtained library (CLONTECH Laboratory Inc.) was also used to obtain longer cDNA inserts.

RNA and Genomic DNA Blot Hybridization—Chicken liver total RNA was prepared by the method of Pyrberg et al. (12), and an aliquot (20 μg) were digested with restriction enzymes and subjected to Southern analysis. Nick-translated probes were hybridized to RNA and DNA, which
were immobilized on a nitrocellulose filter, for about 15 h at 32°C or 42°C in a solution composed of 50 mM Tris-HCl buffer (pH 8.0), 1 M NaCl, 5 mM EDTA, 4 mM sodium phosphate, 0.1% SDS, 50% formamide, 100 µg/ml salmon testes DNA, and 5 × Denhardt’s solution. The filter was washed at 37–54°C for 6 h with several changes of a solution containing 10 mM Tris-HCl buffer (pH 8.0), 25 mM NaCl, 0.1% SDS, and 1 mM concentration each of EDTA and sodium phosphate. RNA and DNA were located by autoradiography, and intensities of signals were densitometrically determined using a Shimadzu TLC Scanner CS-910 equipped with an integrator, Chromatopac C-R1A.

Cloning of the Chicken H-Protein Gene—A chicken genomic library was constructed with λEMBL3 as a vector by the method of Choi and Engel (14). Approximately 1 × 10^8 phages in the initial preparation were amplified once, and 6 × 10^6 phages were subject to screening with the nontranslated pCH3a insert.

DNA Sequencing—Variously truncated cDNA and genomic DNA fragments were subcloned. Some of genomic clones were subjected to isolation of ordered serial deletion mutants of both strands by the method of Henikoff (15). Nucleotide sequences were determined on the plasmids by the method of Sanger et al. (16) using 7-deaza-dGTP (17). Several oligonucleotides commercially available were used as primers. In particular cases, oligonucleotides complementary to the sequences in the chicken H-protein gene were synthesized using a DNA Synthesizer, model 381A (Applied Biosystems Inc., Japan) and used as primers.

RESULTS

Characterization of Chicken H-Protein cDNA—Several immunoreactive clones, including pCH3a, were isolated in the primary selection. None of the approximately 80 clones obtained using the pCH3a cDNA probe contained inserts over 800 bases long. The restriction map, relative locations of the cDNA clones, and strategy for DNA sequencing are summarized in Fig. 1, A, B, and C.

The nucleotide sequence of the H-protein cDNA was determined using the cDNA inserts from pCH37e, pCH50, and pCH62 (Fig. 2). The pCH37e cDNA comprises 561 base pairs with a poly(A) tail, and a methionine codon located at nucleotides 14 to 16 precedes an open reading frame. The sequence surrounding this codon matches the expected consensus for the initiator methionine codon (18), and a protein of 164 amino acid residues was deduced in this reading frame. The sequence downstream of this coding region is coded in the mitochondrial codon, indicating that all the cDNAs listed encode chicken H-protein. Initial 39 amino acid residues can be assigned to a mitochondrial presequence. The pCH34d and pCH50 cDNAs were significantly longer than the others and differed in their 3' end sequences from the pCH3a and pCH37e inserts. Although the sequence between nucleotides 506 and 542 (until T shown with # in Fig. 2) is commonly coded in the pCH3a and pCH37e cDNAs and actually followed by a poly(A) sequence, this region of the pCH50 cDNA further extends to the 892nd nucleotide. A clone designated as pCH62 contained a sequence identical with that of the pCH50 insert, but was further followed by an 18-bp sequence containing an AATAAA motif. These results indicate that two types of cDNA clones with structural heterogeneity in the 3' untranslated region were isolated.

Occurrence of the Two Types of H-Protein mRNAs—We also detected multiple H-protein mRNAs in chicken liver total RNA by Northern analysis. As shown in Fig. 3, the cDNA insert from pCH3a, most of which corresponds to the protein coding region, gives two bands of about 1.3 kb and 0.9 kb, respectively (lane 1). The 0.9-kb RNA seems to be twice as intense in density as the 1.3-kb RNA on the autoradiogram. In contrast, the pCH50 EcoT22/EcoRI cDNA fragment, most of which encompasses the 3' untranslated region, hybridizes to the longer RNA, showing a significantly intense signal. This probe, however, reveals a much weaker intensity of the signal for the 0.9-kb RNA (lane 2), compared to that of the longer RNA. These results indicate that two types of H-protein mRNAs produced by the differently sized 3' untranslated region occur in the chicken liver.

Single Genomic Locus Specifying Chicken H-Protein—Two possible mechanisms which confer such heterogeneity in the 3' noncoding region can be postulated. One is the alternative use of multiple poly(A) sites and the other involves the differential use of separate exons which encode different structures for the 3' untranslated regions. To distinguish which of these mechanisms specifies the two H-protein mRNAs, we first examined the copy number for the H-protein cDNA in the chicken genome, and then the structure of the chicken H-protein gene, especially in the exon for the 3' untranslated region.

Fig. 4 shows results of Southern analysis. The pCH3a probe hybridizes to the 11- and 1.5-kb EcoRI fragments (lane 2). The HindIII fragment of about 6 kb is also hybridized with this probe (lane 3). In contrast, the pCH50 EcoT22/EcoRI probe hybridized to the 1.5-kb EcoRI fragment alone (lane 5) and also gives a 5.6-kb band in the HindIII-treated DNA (lane 6). The 5' region of the H-protein cDNA (the 5' EcoRI/ EcoRV fragment of pCH37e insert) gives a single band of 11 kb in the EcoRI-treated genomic fragments (data not shown). These results demonstrate that the sequence downstream from the EcoT22I site of the H-protein cDNA is coded in the 1.5-kb EcoRI fragment of genomic DNA. The integrated sizes of the fragments revealed in the lanes for the EcoRI- and HindIII-treated DNAs suggest that the signal in lane 3 comprises two fragments with similar sizes. (Note: one of the HindIII sites necessary for production of the 5.8-kb fragment in lane 6 seems to be beyond the range of Fig. 5B.) The pCH3a probe hybridizes to far longer fragments in the genomic DNA treated with BamHI (lane 1) or KpnI (lane 4).

The copy number for the H-protein cDNA was determined by Southern analysis using the 6-aminolevulinate synthase gene as a standard, because a single locus specifies this gene in the chicken genome (20). The pCH34d cDNA (769 bp formed by EcoRI) was ligated with the 790-bp EcoRI/SphI fragment of 6-aminolevulinate synthase cDNA, pAL10 (14), and subcloned at the EcoRI and SphI sites of the pGEM3 vector (pAHLink). This insert was nick-translated and hybridized to the chicken genomic DNA treated with EcoRI with HindIII. As shown in lane 7 of Fig. 4C, the 11- and 1.5-kb fragments originate from the H-protein gene, and the bands indicated as A are from the 6-aminolevulinate synthase gene (20). The integrated intensity of the fragments from the H-protein gene (98 in arbitrary units) is similar to that from the marker gene (112 units). The genomic DNA treated with
**Chicken H-Protein cDNA and Gene**

**FIG. 2.** Nucleotide sequence of chicken H-protein cDNA and its deduced primary structure. The pCH37e, pCH50, and pCH62 inserts were sequenced as described in the text. The deduced primary structure is shown under the nucleotide sequence, and the 40th Ser corresponds to the amino terminus of the mature H-protein. Numbering begins at the nucleotide sequence for a mitochondrial presequence and Exon 1. The termination codon is indicated with, and two AATAAA motifs are doubly underlined. Boundaries for the 5' and 3' ends of exons are indicated as A and H in panel C.

**FIG. 3.** Occurrence of two types of H-protein mRNAs which arise from the differently processed 3' untranslated regions. Poly(A)* RNA (4 µg for lane 1) and total RNA (40 µg for lane 2) were subjected to Northern analysis using the nick-translated pCH3a (lane 1) and pCH50EcoT22/EcoRI (lane 2) inserts as probes. The sizes of two mRNAs were determined by comparison with the known values of E. coli ribosomal RNAs.

**FIG. 4.** Occurrence of a single gene locus of chicken H-protein. High molecular weight DNA from chicken liver (20 µg) was digested with BamHI (lane 1), EcoRI (lanes 2, 5, and 7), HindIII (lanes 3, 6, and 8), and KpnI (lane 4), separated on a 0.8% agarose gel, and transferred to nitrocellulose filters. The nick-translated inserts from pCH3a (lanes 1, 2, 3, and 4), pCH50EcoT22/EcoRI (lanes 5 and 6), and pAHlink (lanes 7 and 8) were used as probes. Fragments from the 5'-aminolevulinate synthase and H-protein genes are indicated as A and H in panel C.

HindIII likewise revealed similar values (lane 8). The observed results indicate that there is only one copy of the H-protein cDNA in the chicken haploid genome.

Cloning and Characterization of the Chicken H-Protein Gene—Genomic clones were obtained from a chicken genomic library by using the nick-translated pCH3a DNA. Some of the 14 possible clones revealed that two EcoRI fragments (11 and 1.5 kb) shown in Fig. 4, lane 2, when the pCH34d cDNA probe was employed (data not shown), and λCH2A, λCH4C, and λCH4D were characterized.

The currently elucidated cDNA sequence is coded for by the 5 exons which are distributed in a stretch of about 8 kb in the H-protein gene. The structure of the 5' region of this gene, however, is still uncertain, because none of the cDNA clones contained a 5' untranslated region longer than that of pCH37e. Therefore, the genomic region corresponding to nucleotides 1–134 of the cDNA was designated as Exon A and the subsequent exons as B, C, D, and E (Fig. 5, A, B, and C). The exact boundaries of these exons are shown in Fig. 2, and the genomic sequences for the exon/intron boundaries and those flanking the 5' and 3' ends of the cDNA structure are presented in Fig. 6.

Exon A encodes the 13-bp 5' untranslated region followed by the nucleotide sequence for a mitochondrial presequence and Exon D includes the Lys residue for the lipoic acid binding site (Lys-98, in the precursor form). One of the HindIII sites, from which the 5.6-kb fragment was exhibited in Fig. 4B, is not found in the genomic fragments cloned and analyzed, suggesting the location of this HindIII site at the downstream...
Fig. 5. Restriction map and partial structure of the chicken H-protein gene. A, relative positions of the cloned genomic fragments and the sites for restriction endonucleases are shown. The EcoRI fragments between asterisks were subcloned from XCH4C and analyzed. B, structural organization of the H-protein gene is shown with the sites for the additional restriction enzymes. The unknown 5' boundary of Exon A is depicted by a dotted line, and the confirmed regions of the exons are shown with closed boxes. The 3' end of Exon E corresponds to the 3' end of the cDNA sequence. C, strategy for DNA sequencing of the precise exon/intron boundary sequences is shown. Each arrow indicates the strand and length of the sequence determined. The abbreviations of the restriction endonucleases used are: E, EcoRI; ET, EcoT221; EV, EcoRV; S, SalI; Sa, SacI; Sm, SmaI; A, ApaI; B, BamHI; H, HindIII; P, PstI; and X, XbaI.

Fig. 6. The exon/intron boundary sequences and the genomic sequences flanking the 5' and 3' ends of the cDNA structure. The boundaries of the exons and introns were determined from both cDNA and genomic sequences. The entire sequences for the exons are indicated in Fig. 2. In this figure, the boundary sequences are shown with 10 nucleotides for exons (upper case letters) and for introns with 20 nucleotides (lower case letters). The positions of the nucleotides at both ends in each row indicate those in the cDNA sequence. The sequences flanking the 5' and 3' ends of the cDNA sequence are presented by lower case letters from -1 to the 5' upstream, and from 911 to the 3' downstream. The 3' EcoRI site corresponds to that of the 1.5-kb EcoRI fragment shown in Fig. 4.

Table 1
Sequences homologous to the second consensus for the poly(A) site

The sequences similar to the second consensus of the adenovirus E2A gene, SV40 early gene, and β-globin gene are listed for the chicken H-protein gene. The conserved nucleotides are underlined. Positions are expressed by the numbering for the chicken H-protein cDNA.

| Candidate | Position | Wild type |
|-----------|----------|-----------|
| Upstream  | AATGTAGT | 545-552   | GTTGGTGG | SV40   |
|           | CTGGTTTT | 586-593   | TTTGTGTT | E2A    |
| Downstream| TTTGGTTT | 933-940   | TTTGGTTT | E2A    |
|           | GTTGGTTT | 948-955   | YGTGTYYY | β-globin|
|           | GCTGTTTTT| 1016-1025 | SV40 and β-globin |
**Types of H-Protein mRNAs, in a Single Exon**—The last exon, Exon E, is preceded by a 0.6-kb intron and, unlike the other exons, appreciably longer in size. The cDNA sequence for the carboxyl-terminal region of H-protein (32 amino acid residues) and the 3′ untranslated region matches the genomic sequence and follows a 167-bp stretch extending to the EcoR1 site, the 3′ end of the 1.5-kb fragment shown in Fig. 4. It appears that both AATAAA motifs are in a single exon. Given the fact that only one copy for the H-protein cDNA exists in the genome, the alternative use of the different exons for the production of two H-protein mRNAs can be excluded. Conversely, these results indicate that the two types of H-protein mRNAs are formed by the alternative use of the separate poly(A) sites in the last exon.

Several G,T-rich sequences homologous to the second consensus, which promotes the efficiency in producing the poly(A) site of a precursor mRNA (22, 23), are found in close proximity to two AATAAA motifs in the H-protein gene. As listed in Table I, TGTAGT and CTGTTTTT are 23 and 64 bp downstream from the upstream AATAAA, and TGTGTTTTT, CTGTGCTTTT, and GCTGTGTTTT are likely candidates for the second consensus of the downstream AATAAA.

**DISCUSSION**

The chicken H-protein cDNA clones have been isolated and well characterized. Only the pCH37e cDNA encodes an ATG codon which fulfills the requirements for the initiator methionine codon. Considering the size difference between the shorter H-protein mRNA and pCH37e cDNA, and the normal 100-200-bp length of the poly(A) tail, an additional 5′ untranslated region with an unknown length may exist in this gene.

Two types of H-protein cDNAs were cloned. In addition, chicken liver total RNA contains two types of H-protein mRNAs which can be distinguished by the 3′ untranslated region. In the present study, we provide two lines of evidence for the possible mechanisms by which the multiple H-protein mRNAs can be produced. Firstly, H-protein is encoded by a single locus in the chicken genome. Secondly, structural analysis of the cloned H-protein cDNA and gene demonstrates that the two polyadenylation consensus sequences are coded in a span of 371 bp in the H-protein cDNA and last exon, Exon E of the H-protein gene. These observations strongly suggest the idea that the alternative use of two AATAAA motifs of a precursor mRNA produces the differently sized H-protein mRNAs in the chicken. In this context, it is possible that several T- or G,T-rich sequences in the amino-terminal side of Lys-98 is rich in hydrophilic amino acids, but contains no basic amino acid in the proximal 10 residues. In contrast, the carboxyl-terminal side is composed of neutral or hydrophobic amino acids. We designated the structure between the 2 conserved glycine residues as a subdomain for the lipoyl acid binding site. This subdomain is coded for by Exon C and D in the chicken H-protein gene. However, there has been no report on the structural organization of the genes encoding other lipoyl enzymes of eukaryotes. It will be of interest to compare exon structures of the limited number of proteins that participate in the metabolism of α-keto acids and glycine and that can utilize lipoyl acid as a prosthetic group (19, 25–31). It is conceivable that a series of proteins contains a subdomain which evolved from a common ancestral protein.

**Acknowledgment**—We thank Dr. Eiji Tsukamoto, Tsukamoto Hospital (Toyama), for his kind encouragement.

**REFERENCES**

1. Motokawa, Y., and Kikuchi, G. (1969) Arch. Biochem. Biophys. 135, 402–409
2. Hiraga, K., and Kikuchi, G. (1980) J. Biol. Chem. 255, 1164–11670
3. Kikuchi, G. (1973) Mol. Cell. Biochem. 1, 169–187
4. Hiraga, K., Koshi, H., Hisasaka, K., Kikuchi, G., and Nyhan, W. L. (1981) J. Clin. Invest. 68, 555–554
5. Fujiwara, K., Okamura, K., and Motokawa, Y. (1979) Arch. Biochem. Biophys. 197, 456–462
6. Robinson, J. R., Klein, S. M., and Sagers, R. D. (1973) J. Biol. Chem. 248, 5319–5325
7. Kochi, H., and Kikuchi, G. (1976) Arch. Biochem. Biophys. 173, 71–81
8. Hiraga, K., and Kikuchi, G. (1980) J. Biol. Chem. 255, 11671–11676
9. Yoshida, T., and Kikuchi, G. (1979) J. Biochem. (Tokyo) 73, 1012–1022
10. Hiraga, K., Kure, S., Yamamoto, M., Ishiguro, Y., and Suzuki, T. (1988) Biochem. Biophys. Res. Commun. 151, 758–762
11. Yamamoto, M., Kure, S., Engel, J. D., and Hiraga, K. (1980) J. Biol. Chem. 245, 15973–15979
12. Fyberg, E. A., Kindle, R. L., Davidson, N., and Sodja, A. (1980) Cell 19, 365–378
13. Lehrsch, H., Diamond, D., Wozney, J., and Boedeker, H. (1977) Biochemistry 16, 4743–4751
14. Choi, O.-R., and Engel, J. D. (1986) Nature 323, 731–734
15. Henikoff, F. (1984) Gene (Amst.) 28, 351–359
16. Song, P., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5468
17. Mizusawa, S., Nisimura, S., and Seela, F. (1986) Nucleic Acids Res. 14, 1319–1324
18. Koizumi, M. (1986) Cell 44, 283–292
19. Fujiwara, K., Okamura-Ikeda, K., and Motokawa, Y. (1986) J. Biol. Chem. 261, 8836–8841
20. Maguire, D. J., Day, A. R., Borthwick, I. A., Srivastava, G., Wigley, P. L., May, B. B., and Elliott, W. H. (1986) Nucleic Acids Res. 14, 1375–1381
21. Kramer, A. R., and Masaitis, T. (1988) in Transcription and Splicing (Hames, B. D., and Glover, D. M., eds) pp. 131–206, IRL Press Ltd., Oxford, UK
22. McDevitt, M. A., Hart, R. P., Wong, W. W., and Nevins, J. R. (1980) EMBO J. 5, 2907–2912
23. McCracken, J., Gaffney, D., Wharton, J. L., and Clements, J. B. (1986) Nucleic Acids Res. 13, 1347–1348
24. Setzer, D. R., McGrogan, M., and Schimke, R. T. (1982) *J. Biol. Chem.* 257, 5143–5147
25. Hummel, K. B., Litwer, S., Bradford, A. P., Aitken, A., Danner, D. J., and Yeaman, S. J. (1988) *J. Biol. Chem.* 263, 6165–6168
26. Stephens, P. E., Darlison, M. G., Lewis, H. M., and Guest, J. R. (1983) *Eur. J. Biochem.* 133, 491–499
27. Spencer, M. E., Darlison, M. G., Stephens, P. E., Duckenfield, I. K., and Guest, J. R. (1984) *Eur. J. Biochem.* 141, 361–374
28. Bradford, A. P., Aitken, A., Beg, F., Cook, K. G., and Yeaman, S. (1987) *FEBS Lett.* 222, 211–214
29. Lau, K. S., Griffin, A., Hu, C.-W., and Chuang, D. T. (1988) *Biochemistry* 27, 1972–1981
30. Gershwin, M. E., Mackay, I. R., Sturgess, A., and Coppel, R. L. (1987) *J. Immunol.* 138, 3525–3531
31. Van de Water, J., Gershwin, M. E., Leung, P., Ansari, A., and Coppel, R. L. (1988) *J. Exp. Med.* 167, 1791–1799