Rapid Paper

Cloning of cDNA Coding for Guinea Pig Liver Transglutaminase

Koji IKURA, Taka-aki NASU,
Hiroyuki YOKOTA, Ryuzo SASAKI
and Hideo CHIBA

Department of Food Science and Technology,
Faculty of Agriculture,
Kyoto University,
Kyoto 606, Japan

Received November 10, 1986

Seven peptides derived from guinea pig liver transglutaminase were sequenced. Using synthetic oligonucleotides that we designed from the amino acid sequences as probes, we obtained a cDNA clone of transglutaminase from a cDNA library constructed with mRNA partly purified from guinea pig liver. About 60% of the primary structure of enzyme protein was predicted by sequencing of the cDNA. The predicted amino acid sequence included a sequence of pentapeptide containing the active-site cysteine residue [J. E. Folk and P. W. Cole, *J. Biol. Chem.*, **241**, 3238 (1966)]. An expanded region of the sequence surrounding the active-site cysteine residue showed significant homology to that of thiol proteases. Two regions rich in glutamic acid residues were found; they are possible calcium-binding sites of this enzyme.

Transglutaminases (EC 2.3.2.13) catalyze the formation of an amide bond between the γ -carboxamide groups of peptide-bound glutamyl residues and the primary amino groups in a variety of compounds, including the ϵ -amino group of lysine in certain proteins. The enzymes have been found in most tissues and in many different cell types, and several have diverse biological functions (for reviews, see refs. 1~3). Liver transglutaminase is one of the most extensively studied transglutaminases. Slife *et al.* have reported that some of the transglutaminase activity in rat liver is associated with the plasma membrane and may be responsible for forming covalently cross-linked matrices of protein at sites of cell-to-cell contact.⁴⁾ We have reported on possible uses of transglutaminase reactions in applied enzymology.^{5~10)} Obtaining cDNA clones for

enzyme proteins is an important first step to clarify the regulatory mechanism of enzyme action on the basis of enzyme structure and the expression of enzyme gene(s). Furthermore, cDNAs are often key materials for the production of coding proteins in gene-engineering techniques. In this paper, we describe cloning of cDNA and its nucleotide sequence that encodes about 60% of guinea pig liver transglutaminase with a molecular weight of 85,000.¹¹⁾

MATERIALS AND METHODS

Materials. Materials were obtained from the sources indicated: TPCK-treated trypsin (Worthington), guanidinium isothiocyanate (Fulka AG), cesium chloride (Nakarai Chemicals), oligo(dT)-cellulose, oligo(dT)_{12~18}, *Escherichia coli* RNase H, terminal deoxynucleotidyl-transferase, and *E. coli* plasmid pUC9 (Pharmacia), *Staphylococcus aureus* cells (Calbiochem), reverse transcriptase, T4 polynucleotide kinase, restriction enzymes, and the M13 DNA sequencing kit (Takara Shuzo), *E. coli* DNA polymerase I and DNA ligase (New England Biolabs), and [³⁵S]methionine, [α -³²P]dCTP, [α -³²P]dGTP, and [γ -³²P]ATP (Amersham).

Trypsin digestion and amino acid sequence. Transglutaminase purified from guinea pig liver extract¹²⁾ was reduced and *S*-carboxamidemethylated as described elsewhere.¹³⁾ The carboxamidemethylated enzyme protein (~1 mg) was digested with TPCK-treated trypsin (10 μ g) for 24 hr at 25°C in 700 μ l of 0.2 M Tris-HCl buffer (pH 8.0) containing 10 mM CaCl₂. The resulting peptides were separated and purified by reverse-phase HPLC on a Cosmosil 5C₁₈ column (4.6 $\times$ 150 mm, Nakarai Chemicals) with a trifluoroacetic acid/acetonitrile system. We analyzed the amino acid sequence of the purified peptide fractions with a gas-phase protein sequencer (Applied Biosystems model 470A).

Cloning. Poly(A⁺) RNA was isolated from guinea pig liver by the guanidinium isothiocyanate/CsCl method¹⁴⁾ followed by oligo(dT)-cellulose chromatography.¹⁵⁾ Transglutaminase mRNA was enriched 10-fold by centrifugation of the poly(A⁺) RNA through a 5~20% sucrose gradient in 10 mM Tris-HCl buffer (pH 7.5) containing 0.1 M KCl and 1 mM EDTA at 1.3 $\times$ 10⁵ g for 13 hr at 4°C. The template activity of mRNA in the fractionated poly(A⁺) RNAs was assayed by cell-free translation in a message-dependent rabbit reticulocyte system¹⁶⁾ with [³⁵S]methionine as the label. Translation products were analyzed by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate after immunoprecipi-

tation¹⁷⁾ with rabbit antibody against guinea pig liver transglutaminase and *S. aureus* cells. Fractions containing mRNA for transglutaminase were pooled. Double-stranded cDNA was synthesized as described by Gubler and Hoffman¹⁸⁾ with the partly purified mRNA as a template, and inserted into the *Pst*I site of *E. coli* plasmid pUC9 by the dG-dC homopolymer extension method.¹⁹⁾ Here the vector plasmid was dC-tailed and cDNA was dG-tailed. *E. coli* JM83 cells were transformed with the recombinant plasmids as described by Hanahan.²⁰⁾ The constructed cDNA library was screened for the transglutaminase sequence by high-density colony hybridization²¹⁾ at 35°C with ³²P-labelled oligonucleotide probes. The probes were synthesized by a DNA synthesizer (Applied Biosystems model 380A).

DNA sequencing. DNA sequence analysis was done by the dideoxy chain termination method²²⁾ after subcloning of restriction fragments into M13-based cloning vectors.²³⁾

RESULTS

Partial amino acid sequence of transglutaminase

Treatment of transglutaminase with trypsin

TABLE I. AMINO ACID SEQUENCES OF THE SEVEN SELECTED PEPTIDES

Peptide No.	Sequences
8	Ala-Asn-His-Leu- ⁵ Asn-Lys
13	Tyr-Pro-Glu-Gly-Ser ⁵ -Glu-Glu-Glu-Arg
15	Glu-Asp-Ile-Thr ⁵ -His-Thr-Tyr-Lys
16	Glu-Glu-Ala-Gln-Glu ⁵ -Glu-Thr-Gly-Val ¹⁰ -Ala-Met-Arg
20	Ser-Val-Glu-Val ⁵ -Pro-Asp-Pro-Val-Glu ¹⁰ -Ala-Gly-Glu-Gln-Ala-Lys ¹⁵
22	Asp-Ile-Tyr-Leu-Glu ⁵ -Asn-Pro-Glu-Ile ¹⁰ -Lys
35	Trp-Asp-Asn-Asn ⁵ -Tyr-Ser-Asp-Gly-Val ¹⁰ -Ser-Pro-Met-Ser-X ¹⁵ -Gly-Ser-Val-Asp ²⁰ -Ile-Leu-X-Arg

The pentapeptide portions underlined were used for preparation of the oligonucleotides to be used as probes. X, an unidentified amino acid.

resulted in the formation about 50 peptides, which were fractionated by reverse-phase HPLC. After rechromatography of the main peak fractions, seven major peptides (Nos. 8, 13, 15, 16, 20, 22, and 35) were selected and their amino acid sequence was analyzed. The

TABLE II. SYNTHETIC OLIGONUCLEOTIDES USED TO SCREEN cDNA CLONES FOR GUINEA PIG LIVER TRANSGLUTAMINASE

Mixed 14-mer oligonucleotides were complementary to all possible coding sequences corresponding to the two pentapeptides shown in Table I: probe A, residues 1~5 of peptide No. 35; probe B, residues 5~9 of peptide No. 22. The third position of the codon of the amino acid at the C-termini was not included. N represents A, T, G, and C.

A	
Amino acid sequence	Trp Asp Asn Asn Tyr
Possible codons	5' UGG GAU AAU AAU UAU 3' C C C C
Mixed oligonucleotide	3' ACC CTA TTA TTA AT 5' G G G
B	
Amino acid Sequence	Glu Asn Pro Glu Ile
Possible codons	5' GAA AAU CCN GAA AUC 3' G C G A
Mixed oligonucleotide	3' CTT TTA GGN CTT TA 5' C G C

FIG. 1. Restriction Endonuclease Cleavage Map of pLTG20 Insert DNA and Strategy for Nucleotide Sequencing.

The restriction map shows only those cleavage sites relevant to DNA sequencing. The open box indicates the coding region. Thin lines and bold lines indicate vector and oligo dC-dG tails, respectively. Horizontal lines with arrows represent the direction and extent of the DNA sequences identified.

results are summarized in Table I. For screening of cDNA clones that contained the transglutaminase sequence, two sets of mixed oligonucleotides, each 14 nucleotides long, were prepared as probes (A and B in Table II).

Cloning of transglutaminase cDNA

From about 7×10^4 transformants, one recombinant clone that was hybridized with

both of the probes A and B was isolated. The plasmid DNA from this clone, designated pLTG20, contained about 1.3-kilobase insert DNA. To confirm that pLTG20 contained the sequence of transglutaminase cDNA, the nucleotide sequence of the insert DNA was identified by the chain termination method. The restriction endonuclease cleavage map of insert DNA and the sequencing strategy are shown in Fig. 1. Figure 2 shows the nucleotide

ATC CTG GAT ATT TGC CTG ATG CTC TTG GAC ACC AAC CCC AAG TTC CTG AAG AAT GCT GGC	60
Ile Leu Asp Ile Cys Leu Met Leu Leu Asp Thr Asn Pro Lys Phe Leu Lys Asn Ala Gly	20
CAA GAC TGC TCG CGC CGC AGC AGA CCT GTC TAC GTG GGC CGG GTG GTG AGC GCC ATG GTC	120
Gln Asp Cys Ser Arg Arg Ser Arg Pro Val Tyr Val Gly Arg Val Val Ser Ala Met Val	40
AAC TGC AAT GAC GAT CAG GGC GTG CTT CAG GGA CGC TGG GAC AAC AAC TAC AGT GAT GGT	180
Asn Cys Asn Asp Asp Gln Gly Val Leu Gln Gly Arg Trp Asp Asn Asn Tyr Ser Asp Gly	60
GTC AGC CCC ATG TCC TGS ATC GGC AGC GTG GAC ATC CTG CGG CGC TGS AAA GAC TAT GGC	240
Val Ser Pro Met Ser Trp Ile Gly Ser Val Asp Ile Leu Arg Arg Trp Lys Asp Tyr Lys	80
TGC CAG CGC GTC AAG TAC GGC CAG TGC TGG GTC TTC GCT GCT GTG GCC TGC ACA GTG CTG	300
Cys Gln Arg Val Lys Tyr Gly Gln Cys Trp Val Phe Ala Ala Val Ala Cys Thr Val Leu	100
CGG TGC CTT GCG ATC CCC ACC CGA GTC GTG ACC AAC TTT AAC TCA GCC CAC GAC CAG AAC	360
Arg Cys Leu Ala Ile Pro Thr Arg Val Val Thr Asn Phe Asn Ser Ala His Asp Gln Asn	120
AGC AAC CTG CTC' ATC GAG TAC TTC CGA AAC GAG TCT GGG GAG ATC GAG GGG AAC AAG AGC	420
Ser Asn Leu Leu Ile Glu Tyr Phe Arg Asn Glu Ser Gly Glu Ile Glu Gly Asn Lys Ser	140
GAG ATG ATC TGG AAC TTC CAC TCA CTG CTG GGT GGA GTC GTG GAT GAC CAG GCC GGA CCT	480
Glu Met Ile Trp Asn Phe His Ser Leu Leu Gly Gly Val Val Asp Asp Gln Ala Gly Pro	160
GGA GCC TGG GTA CGA GGG GTG CAG GCC CTG GAC CCC ACA CCC CAG GAG AAG AGT GAA GGG	540
Gly Ala Trp Val Arg Gly Val Gln Ala Leu Asp Pro Thr Pro Gln Glu Lys Ser Glu Gly	180
ACA TAC TGC TGT GGC CCA GTT CCG GTT CGA GCC ATC AAG GAG GGC CAC CTG AAC GTC AAG	600
Thr Tyr Cys Cys Gly Pro Val Pro Val Arg Ala Ile Lys Glu Gly His Leu Asn Gly Lys	200
TAT GAT GCA CCT TTC GTG TTT GCT GAG GTC AAT GCT GAC GTG GTG AAC TGG ATC CGG CAG	660
Tyr Asp Ala Pro Phe Val Phe Ala Glu Val Asn Ala Asp Val Val Asn Trp Ile Arg Gln	220
AAA GAT GGG TCC CTG CGC AAG TCC ATC AAC CAT TTG GTT GTG GGG CTG AAG ATC AGT ACT	720
Lys Asp Gly Ser Leu Arg Lys Ser Ile Asn His Leu Val Val Gly Leu Lys Ile Ser Thr	240
AAG AGT GTG GGC CGC GAT GAG CGA GAG GAC ATC ACC CAC ACC TAC AAG TAC CCA GAG GGA	780
Lys Ser Val Gly Arg Asp Glu Arg Glu Asp Ile Thr His Thr Tyr Lys Tyr Pro Glu Gly	260
TCT GAA GAG GAG CGG GAA GCT TTT GTT AGG GCC AAC CAC CTA AAT AAA CTG GCC ACA AAG	840
Ser Glu Glu Glu Arg Glu Ala Phe Val Arg Ala Asn His Leu Asn Lys Leu Ala Thr Lys	280
GAA GAG GCT CAG GAG GAA ACG GGA GTG GCC ATG CGG ATC CGT GTG GGC CAG AAC ATG ACT	900
Glu Glu Ala Gln Glu Glu Thr Gly Val Ala Met Arg Ile Arg Val Gly Gln Asn Met Thr	300
ATG GGC AGT GAC TTT GAC ATC TTT GCC TAC ATC ACC AAT GGC ACT GCT GAG AGC CAC GAA	960
Met Gly Ser Asp Phe Asp Ile Phe Ala Tyr Ile Thr Asn Gly Thr Ala Glu Ser His Glu	320
TGC CAA CTC CTG CTC TGT GCA CGC ATC GTC AGC TAC AAT GGA GTC CTG GGG CCC GTG TGC	1020
Cys Gln Leu Leu Leu Cys Ala Arg Ile Val Ser Tyr Asn Gly Val Leu Gly Pro Val Cys	340
AGC ACC AAC GAC CTG CTC AAC CTG ACC CTG GAT CCC TTC TCG GAG AAC AGC ATC CCC CTG	1080
Ser Thr Asn Asp Leu Leu Asn Leu Thr Leu Asp Pro Phe Ser Glu Asn Ser Ile Pro Leu	360
CAC ATC CTC TAT GAG AAG TAC GGT GAC TAC CTG ACT GAG TCC AAC CTC ATC AAG GTG CGA	1140
His Ile Leu Tyr Glu Lys Tyr Gly Asp Tyr Leu Thr Glu Ser Asn Leu Ile Lys Val Arg	380
GGC CTC CTT ATC GAG CCA GCA GCC AAC AGC TAT GTA TTG GCC GAG AGG GAC ATT TAC CTG	1200
Gly Leu Leu Ile Glu Pro Ala Ala Asn Ser Tyr Val Leu Ala Glu Arg Asp Ile Tyr Leu	400
GAG AAT CCA GAA ATC AAG ATC CGG GTC TTG GGG GAG CCC AAG CAG AAC CGC AAG CTG ATT	1260
Glu Asn Pro Glu Ile Lys Ile Arg Val Leu Gly Glu Pro Lys Gln Asn Arg Lys Leu Ile	420
GCT GAG GTG TCT CTG AAG AAT CCG CTC OCT GTG CCG CTG CTG GGT TGT ATC TTC ACC GTG	1320
Ala Glu Val Ser Leu Lys Asn Pro Leu Pro Val Pro Leu Leu Gly Cys Ile Phe Thr Val	440
GAA GGA GCT GGC	
Glu Gly Ala Gly	

FIG. 2. Nucleotide Sequence and Predicted Amino Acid Sequence of the pLTG20 Insert DNA.

Solid lines, amino acid sequences of peptides selected from tryptic digest of transglutaminase; wavy line, amino acid sequence containing the active-site cysteine residue;²⁴⁾ broken lines, regions rich in glutamic acid residues. Numbering of nucleotides and amino acids is tentative.

sequence of the insert DNA, together with the amino acid sequence predicted from it. The predicted amino acid sequence covered the sequences of six of the peptides (Nos. 8, 13, 15, 16, 22, and 35) obtained from the tryptic digest of transglutaminase (Table I and Fig. 2). The open reading frame including the amino acid sequences of these six peptides expanded all over the insert DNA except the dC-dG tails. From these results, we concluded that the insert DNA of pLTG20 is cDNA partly encoding guinea pig liver transglutaminase. The insert cDNA contained 1332 base pairs excluding the dC-dG tails.

DISCUSSION

A sequence of 444 amino acid residues of guinea pig liver transglutaminase was deduced (Fig. 2). The cDNA thus obtained probably covered about 60% of the entire amino acid sequence of guinea pig liver transglutaminase, if we assume that the enzyme consists of 750~800 amino acid residues.

The predicted amino acid sequence of residues 86~90 corresponded exactly to the sequence surrounding the active-site cysteine residue of this enzyme, Tyr·Gly·Gln·Cys·Trp²⁴⁾ (Fig. 2). This pentapeptide sequence is highly homologous to the active site sequences of thiol proteases containing a functioning cysteine residue.^{24,25)} Figure 3 compares a more expanded region of amino acid

sequence around the active-site cysteine residue of transglutaminase with that of various thiol proteases.²⁶⁾ Homology was seen in a sequence much longer than previously noted.²⁴⁾ Both reactions of thiol proteases and transglutaminase proceed *via* common catalytic steps, formation of an acyl-enzyme intermediate at the active-site cysteine residue and later acyl-transfer,^{1,28)} and therefore the homologous sequence surrounding the cysteine residue seems to be important for these catalytic steps. Thiol proteases have another consensus region containing active-site histidine residue, which usually is located on the carboxyl-terminal side of the active-site cysteine residue at a distance of about 150 amino acid residues.²⁶⁾ In the transglutaminase amino acid sequence predicted here, however, we could not find the corresponding sequence containing a histidine residue. These findings may reflect the similar but different catalytic actions of transglutaminase and protease.

Transglutaminase is a Ca^{2+} -dependent enzyme and the guinea pig liver enzyme has several calcium-binding sites that seem to have different binding affinities.²⁹⁾ The presence of an "E-F hand structure," a typical structure of calcium-binding site of some proteins with a high affinity, was searched for by the test sequence method,³⁰⁾ but no regions in the predicted sequence of transglutaminase had high enough scores to represent this structure.

FIG. 3. Comparison of Predicted Amino Acid Sequence Containing the Active-site Cysteine Residue of Transglutaminase with That of Thiol Proteases.

One-letter amino acid notation is used. Residues identical to those in thiol proteases²⁶⁾ are boxed; circled residues represent favoured amino acid substitutions, defined as described previously.²⁷⁾ The asterisk indicates the active-site cysteine residue. Gaps are inserted to achieve maximum homology. The number above the sequence corresponds to the sequence of transglutaminase (Fig. 2).

There are two regions rich in glutamic acid residues, residues 259~266 and 281~286 (Fig. 2); they may be involved in the formation of calcium binding sites of transglutaminase.

REFERENCES

- 1) J. E. Folk and J. S. Finlayson, "Advances in Protein Chemistry," Vol. 31, ed. by C. B. Anfinsen, J. T. Edsall and F. M. Richards, Academic Press, New York, 1977, p. 1.
- 2) J. E. Folk, "Annual Review of Biochemistry," Vol. 49, ed. by E. E. Snell, Annual Reviews Inc., Palo Alto, California, 1980, p. 517.
- 3) L. Lorand and S. M. Conrad, *Mol. Cell. Biochem.*, **58**, 9 (1984).
- 4) C. W. Slife, M. D. Dorsett and M. L. Tillotson, *J. Biol. Chem.*, **261**, 3451 (1986).
- 5) K. Ikura, T. Kometani, M. Yoshikawa, R. Sasaki and H. Chiba, *Agric. Biol. Chem.*, **44**, 1567 (1980).
- 6) K. Ikura, M. Yoshikawa, R. Sasaki and H. Chiba, *Agric. Biol. Chem.*, **45**, 2587 (1981).
- 7) K. Ikura, M. Goto, M. Yoshikawa, R. Sasaki and H. Chiba, *Agric. Biol. Chem.*, **48**, 2347 (1984).
- 8) K. Ikura, K. Okumura, M. Yoshikawa, R. Sasaki and H. Chiba, *Agric. Biol. Chem.*, **49**, 1877 (1985).
- 9) M. Yoshikawa, M. Goto, K. Ikura, R. Sasaki and H. Chiba, *Agric. Biol. Chem.*, **46**, 207 (1982).
- 10) K. Okumura, K. Ikura, M. Yoshikawa, R. Sasaki and H. Chiba, *Agric. Biol. Chem.*, **48**, 2435 (1984).
- 11) J. M. Connellan, S. I. Chung, N. K. Whetzel, L. M. Bradley and J. E. Folk, *J. Biol. Chem.*, **246**, 1093 (1971).
- 12) K. Ikura, H. Sakurai, K. Okumura, R. Sasaki and H. Chiba, *Agric. Biol. Chem.*, **49**, 3527 (1985).
- 13) M. Noda, S. Shimizu, T. Tanabe, T. Takai, T. Kayano, T. Ikeda, H. Takahashi, H. Nakayama, Y. Kanaoka, N. Minamino, K. Kangawa, H. Matsuo, M. A. Raftery, T. Hirose, S. Inayama, H. Hayashida, T. Miyata and S. Numa, *Nature*, **312**, 121 (1984).
- 14) J. M. Chirgwin, A. E. Przybyla, R. J. MacDonald and W. J. Rutter, *Biochemistry*, **18**, 5294 (1979).
- 15) H. Aviv and P. Leder, *Proc. Natl. Acad. Sci. U.S.A.*, **69**, 1408 (1972).
- 16) H. R. B. Pelham and R. J. Jackson, *Eur. J. Biochem.*, **67**, 247 (1976).
- 17) S. W. Kessler, "Methods in Enzymology," Vol. 73, ed. by J. J. Langone and H. V. Vunakis, Academic Press, New York, 1981, p. 442.
- 18) U. Gubler and B. J. Hoffman, *Gene*, **25**, 263 (1983).
- 19) S. L. Peacock, C. M. McIver and J. J. Monahan, *Biochim. Biophys. Acta*, **655**, 243 (1981).
- 20) D. Hanahan, *J. Mol. Biol.*, **166**, 557 (1983).
- 21) D. Hanahan and M. Meselson, *Gene*, **10**, 63 (1980).
- 22) F. Sanger, S. Nicklen and A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 5463 (1977).
- 23) J. Messing and J. Vieira, *Gene*, **19**, 269 (1982).
- 24) J. E. Folk and P. W. Cole, *J. Biol. Chem.*, **241**, 3238 (1966).
- 25) K. Suzuki, H. Hayashi, T. Hayashi and K. Iwai, *FEBS Lett.*, **152**, 67 (1983).
- 26) S. Ohno, Y. Emori, S. Imajoh, H. Kawasaki, M. Kisaragi and K. Suzuki, *Nature*, **312**, 566 (1984).
- 27) R. M. Schwartz and M. O. Dayhoff, "Atlas of Protein Sequence and Structure," Vol. 5, suppl. 3, ed. by M. O. Dayhoff, National Biochemical Research Foundation, Washington, 1978, p. 345.
- 28) A. N. Glazer and E. L. Smith, "The Enzymes," Vol. 3, ed. by P. D. Boyer, Academic Press, New York, 1971, p. 501.
- 29) J. E. Folk and S. I. Chung, "Advances in Enzymology," Vol. 38, ed. by A. Meister, John Wiley, New York, 1973, p. 109.
- 30) R. M. Tufty and R. H. Kretsinger, *Science*, **187**, 167 (1975).