Amino Acid Sequence of Guinea Pig Liver Transglutaminase from Its cDNA Sequence†

Koji Ikura,* Taka-aki Nasu, Hiroyuki Yokota, Yoichi Tsuchiya, Ryuzo Sasaki, and Hideo Chiba

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

Received October 1, 1987; Revised Manuscript Received December 23, 1987

ABSTRACT: Transglutaminases (EC 2.3.2.13) catalyze the formation of ε-(γ-glutamyl)lysine cross-links and the substitution of a variety of primary amines for the γ-carboxamide groups of protein-bound glutaminyl residues. These enzymes are involved in many biological phenomena. In this paper, the complete amino acid sequence of guinea pig liver transglutaminase, a typical tissue-type nonzymogenic transglutaminase, was predicted by the cloning and sequence analysis of DNA complementary to its mRNA. The cDNA clones carrying the sequences for the 5′- and 3′-end regions of mRNA were obtained by use of the sequence of the partial-length cDNA of guinea pig liver transglutaminase [Ikura, K., Nasu, T., Yokota, H., Sasaki, R., & Chiba, H. (1987) *Agric. Biol. Chem. 51*, 957–961]. A total of 3695 bases were identified from sequence data of four overlapping cDNA clones. Northern blot analysis of guinea pig liver poly(A+) RNA showed a single species of mRNA with 3.7–3.8 kilobases, indicating that almost all of the mRNA sequence was analyzed. The composite cDNA sequence contained 68 bases of a 5′-untranslated region, 2073 bases of an open reading frame that encoded 691 amino acids, a stop codon (TAA), 1544 bases of a 3′-noncoding region, and a part of a poly(A) tail (7 bases). The molecular weight of guinea pig liver transglutaminase was calculated to be 76 620 from the amino acid sequence deduced, excluding the initiator Met. This enzyme contained no carbohydrate [Folk, J. E., & Chung, S. I. (1973) *Adv. Enzymol. Relat. Areas Mol. Biol. 38*, 109–191], but six potential Asn-linked glycosylation sites were found in the sequence deduced. The primary structure of guinea pig liver transglutaminase was compared with that of the catalytic a subunit of human factor XIII (zymogenic transglutaminase). Several regions of strong homology, including the region surrounding the active site cysteine residue, were observed. A hydropathy profile of the liver transglutaminase suggested that the active site cysteine residue was at the amino-terminal end of a highly hydrophobic region. Regions clearly having the "E–F-hand structure", a typical calcium-binding site, were not found in the sequence of the liver enzyme.

Transglutaminases (protein-glutamine:amine γ-glutamyltransferase, EC 2.3.2.13) are calcium-dependent acyltransferases that catalyze the formation of an amide bond between the γ-carboxamide groups of peptide-bound glutamine residues and the primary amino groups in a variety of compounds, including the ε-amino group of lysine in certain proteins. These enzymes are widely distributed in most tissues and body fluids, and several are involved in diverse biological functions in which they catalyze the formation of ε-(γ-glutamyl)lysine cross-links with protein substrates [see Folk and Finlayson (1977), Folk (1980), and Lorand and Conrad (1984) for reviews]. These functions include stabilization of the fibrin structure in hemostasis and in endogenously occurring fibrinolysis (Lorand, 1972; Sakata & Aoki, 1980, 1982), formation of the cornifying envelope in epidermal keratinocyte (Rice & Green, 1978), stiffening of the erythrocyte membrane (Siefring et al., 1978), wound healing (Monsher & Schad, 1979), formation of a postejaculation vaginal plug in rodents (Williams-Ashman, 1984), and cellular growth regulation and differentiation (Birckbichler & Patterson, 1978; Birckbichler et al., 1981; Kannagi et al., 1982; Murtaugh et al., 1984).

Liver transglutaminase, a tissue-type nonzymogenic enzyme, is one of the most extensively studied transglutaminases, but its biological role is unknown. Slife et al. (1986) 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. We have reported possible applications of transglutaminase reactions in various areas with the use of guinea pig liver enzyme (Ikura et al., 1980, 1981, 1984, 1985; Yoshikawa et al., 1982; Okumura et al., 1984). Obtaining cDNA clones for enzyme proteins is an important step for the clarification of the regulatory mechanism of enzyme action on the basis of enzyme structure and the expression of enzyme gene. cDNAs are often key materials for the production of their coding proteins in genetic engineering techniques. We have obtained one partial-length cDNA clone of guinea pig liver transglutaminase and predicted about 60% of its primary structure (Ikura et al., 1987). Here we report the complete amino acid sequence of this enzyme, which was deduced from the nucleotide sequence of its cDNA.

Liver transglutaminase is a nonzymogenic single polypeptide chain; however, human factor XIII, a zymogenic transglutaminase, is found in plasma as a tetramer consisting of two catalytic a subunits and two noncatalytic b subunits and in platelets and placenta as a dimer of the a subunit. The zymogenic factor XIII is converted to an active form through the limited proteolysis of the a subunit by thrombin in the presence of Ca^{2+}. cDNA of the catalytic a subunit of human factor XIII has been cloned and its amino acid sequence deduced (Ichinose et al., 1986; Grundmann et al., 1986). The primary structure of the a subunit also was analyzed by the methods of protein chemistry (Takahashi et al., 1986). In this

† This study was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, and Culture of Japan.

* Author to whom correspondence should be addressed.

0006-2960/88/0427-2898$01.50/0

paper, we compare liver transglutaminase and the catalytic a subunit of factor XIII to evaluate the homology of their amino acid sequences.

Materials and Methods

Materials. Materials were obtained commercially from the sources indicated here: guanidinium isothiocyanate (Fluka AG); cesium chloride (Nakarai Chemicals); oligo(dT)–cellulose, *Escherichia coli* plasmid pUC9, and terminal deoxynucleotidyltransferase (Pharmacia); restriction enzymes, reverse transcriptase, T4 polynucleotide kinase, T4 DNA ligase, and an M13 DNA sequencing kit (Takara Shuzo); *Escherichia coli* DNA polymerase I and *Escherichia coli* DNA ligase (New England Biolabs); human placental ribonuclease inhibitor, RNasin (Wako Pure Chemical Ind.); a multiprime DNA labeling system, [α-^{32}P]dCTP, and [γ-^{32}P]ATP (Amersham). Total RNA was extracted from guinea pig liver by the guanidinium isothiocyanate/CsCl method (Chirgwin et al., 1979), and poly(A+) RNA was isolated by oligo(dT)–cellulose chromatography (Aviv & Leder, 1972). Oligonucleotides used for primers and probes were synthesized by a DNA synthesizer (Pharmacia Gene Assembler).

cDNA Clonings. Two kinds of cDNA libraries were used to obtain cDNA clones covering the entire transglutaminase protein. One was identical with the library used in an earlier study (Ikura et al., 1987), which was constructed from partly purified mRNA of guinea pig liver transglutaminase. cDNA clones coding for the carboxyl-terminal region of transglutaminase were selected with this library. The other libary was constructed to obtain cDNA clones coding for the amino-terminal region of the enzyme. An appropriate synthetic oligonucleotide primer (4.2 pmol) was elongated by reverse transcription of poly(A+) RNA (30 μg) in a volume of 60 μL, and then double-stranded cDNA was synthesized as described by Gubler and Hoffman (1983). The resulting double-stranded cDNA was treated with T4 DNA polymerase to remove the remaining 3'-end overhangs from the first-strand cDNA and inserted into the *Pst*I site of *E. coli* plasmid pUC9 by the (dG)–(dC) homopolymer extension method (Peacock et al., 1981). Here the cDNA insert was dG-tailed and the vector plasmid was dC-tailed. *E. coli* JM83 cells were transformed with the recombinant plasmids as described by Hanahan (1983). The yield of transformant clones was about 10000. These two cDNA libraries were screened for transglutaminase cDNA sequences by colony hybridization (Maniatis et al., 1982) with use of appropriate ^{32}P-labeled oligonucleotides and restriction fragments obtained from the cDNAs as probes. Oligonucleotides were labeled at their 5'-ends by use of [γ-^{32}P]ATP and T4 polynucleotide kinase, and restriction fragments were multiprime-labeled (Feinberg & Vogelstein, 1983) by use of [α-^{32}P]dCTP and the multiprime DNA labeling system according to the procedure described by the supplier.

RNA-Blot Analysis. RNA-blot analysis of guinea pig liver poly(A+) RNA was done essentially by the method of Thomas (1980). RNA was denatured with glyoxal and dimethyl sulfoxide and resolved by agarose gel (1%) electrophoresis in 10 mM sodium phosphate (pH 6.6). The RNA was transferred overnight to a nylon filter (Biodyne from Pall) with 20× standard saline citrate (SSC).[1] The filter was dried at room temperature and then baked for 2 h at 80 °C. RNA blots were

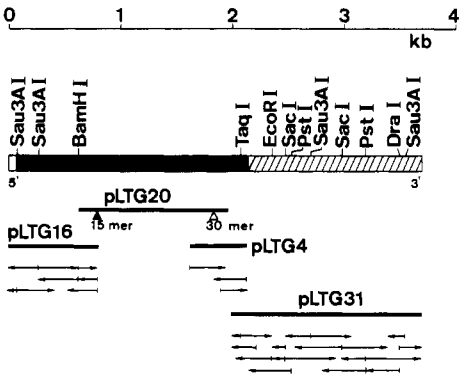

FIGURE 1: Restriction map and sequencing strategy for the cDNA inserts in clones pLTG4, pLTG16, and pLTG31. The restriction map shows only restriction endonuclease sites relevant to DNA sequencing. The open, solid, and hatched boxes represent a portion of the 5'-untranslated region, the protein-coding region, and the 3'-noncoding region, respectively. The closed and open triangles indicate the positions of the synthetic oligonucleotides used as the primer (15-mer) and probe (30-mer), respectively. Horizontal lines with arrows represent the direction and extent of the DNA sequences analyzed. The location of the cDNA insert of clone pLTG20 (Ikura et al., 1987) is also shown.

prehybridized overnight at 42 °C in a hybridization buffer [50% formamide/5× Denhardt's solution (Denhardt, (1966)/0.9 M NaCl/0.05 M sodium phosphate, pH 8.3/5 mM EDTA/0.3% SDS/250 μg/mL denatured calf thymus DNA]. Hybridization was done overnight at 42 °C with the multiprime-labeled cDNA fragment as probe in the hybridization buffer. The filter was washed 4 times at room temperature with 2× SSC/0.1% SDS and twice at 42 °C with 0.1× SSC/0.1% SDS before being dried and autoradiographed.

Primer Extension Analysis. Primer extension analysis was done essentially as described by Williams and Mason (1985). The guinea pig liver poly(A+) RNA and a 5'-labeled 17-mer oligonucleotide were used as a template and a primer, respectively. The conditions used were as follows. A hybridization mixture containing 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.15 M NaCl, 9 units of RNasin, 40 μg of poly(A+) RNA, and 10 fmol of ^{32}P-labeled primer (6×10^4 cpm) in a volume of 10 μL was incubated for 2 h at 50 °C. The incubated hybridization mixture was mixed with 10 μL of an extension buffer containing 0.1 M Tri-HCl (pH 8.3), 10 mM $MgCl_2$, 10 mM dithiothreitol, 4 mM sodium pyrophosphate, 2 mM dCTP, 2 mM dATP, 2 mM dGTP, 2 mM dTTP, and 15 units of reverse transcriptase. After incubation for 1 h at 42 °C, the reaction mixture was extracted with phenol/$CHCl_3$ (1:1 v/v). The aqueous phase was adjusted to 2 M ammonium acetate, and the nucleic acids were precipitated by the addition of 2 volumes of ethanol. The precipitate was rinsed 2 times with 0.5 mL of 70% ethanol, dried under vaccum for 1 min, dissolved in 3.7 μL of 10 mM Tris-HCl (pH 7.5)/1 mM EDTA, and mixed with 3 μL of 95% formamide containing 0.1% xylene cyanol and 0.1% bromphenol blue. This mixture was heated at 95 °C for 3 min and cooled quickly in ice–water. The labeled product of primer extension in the cooled mixture was analyzed by electrophoresis on a 6% polyacrylamide/8 M urea slab gel alongside labeled size markers (*Hae*III digest of ϕX174 DNA) and by following the autoradiography.

DNA Sequence Analysis. DNA was sequenced by the dideoxy chain-termination method (Sanger et al., 1977) after subcloning of restriction fragments into M13-based cloning vectors (Messing & Vieira, 1982). DNA and amino acid sequences were analyzed with the computer programs SDC-GENETYX (Software Developing, Tokyo) and DNASIS (Hitachi Software Engineering, Yokohama) on an NEC personal

[1] Abbreviations: SSC, standard saline citrate (0.15 M NaCl/0.015 M sodium citrate, pH 7.0); EDTA, ethylenediaminetetraacetic acid; SDS, sodium dodecyl sulfate; bp, base pair(s); kb, kilobase(s); Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride.

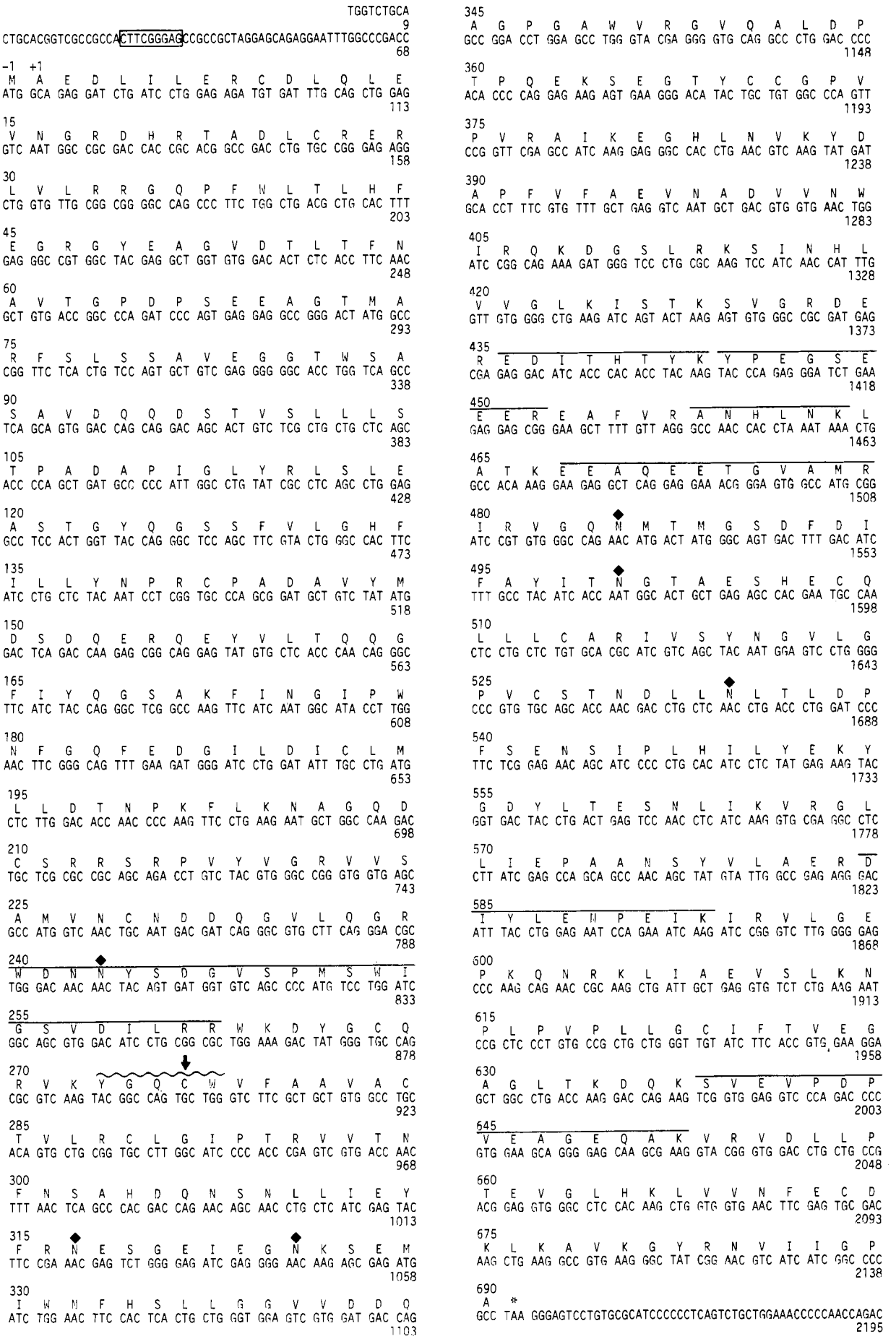
TGGTCTGCA
9
CTGCACGGTCGCCGCCACTTCGGGAGCCGCCGCTAGGAGCAGAGGAATTTGGCCCGACC
68
-1 +1
M A E D L I L E R C D L Q L E
ATG GCA GAG GAT CTG ATC CTG GAG AGA TGT GAT TTG CAG CTG GAG
113
15
V N G R D H R T A D L C R E R
GTC AAT GGC CGC GAC CAC CGC ACG GCC GAC CTG TGC CGG GAG AGG
158
30
L V L R R G Q P F W L T L H F
CTG GTG TTG CGG CGG GGC CAG CCC TTC TGG CTG ACG CTG CAC TTT
203
45
E G R G Y E A G V D T L T F N
GAG GGC CGT GGC TAC GAG GCT GGT GTG GAC ACT CTC ACC TTC AAC
248
60
A V T G P D P S E E A G T M A
GCT GTG ACC GGC CCA GAT CCC AGT GAG GAG GCC GGG ACT ATG GCC
293
75
R F S L S S A V E G G T W S A
CGG TTC TCA CTG TCC AGT GCT GTC GAG GGG GGC ACC TGG TCA GCC
338
90
S A V D Q Q D S T V S L L L S
TCA GCA GTG GAC CAG CAG GAC AGC ACT GTC TCG CTG CTG CTC AGC
383
105
T P A D A P I G L Y R L S L E
ACC CCA GCT GAT GCC CCC ATT GGC CTG TAT CGC CTC AGC CTG GAG
428
120
A S T G Y Q G S S F V L G H F
GCC TCC ACT GGT TAC CAG GGC TCC AGC TTC GTA CTG GGC CAC TTC
473
135
I L L Y N P R C P A D A V Y M
ATC CTG CTC TAC AAT CCT CGG TGC CCA GCG GAT GCT GTC TAT ATG
518
150
D S D Q E R Q E Y V L T Q Q G
GAC TCA GAC CAA GAG CGG CAG GAG TAT GTG CTC ACC CAA CAG GGC
563
165
F I Y Q G S A K F I N G I P W
TTC ATC TAC CAG GGC TCG GCC AAG TTC ATC AAT GGC ATA CCT TGG
608
180
N F G Q F E D G I L D I C L M
AAC TTC GGG CAG TTT GAA GAT GGG ATC CTG GAT ATT TGC CTG ATG
653
195
L L D T N P K F L K N A G Q D
CTC TTG GAC ACC AAC CCC AAG TTC CTG AAG AAT GCT GGC CAA GAC
698
210
C S R R S R P V Y V G R V V S
TGC TCG CGC CGC AGC AGA CCT GTC TAC GTG GGC CGG GTG GTG AGC
743
225
A M V N C N D D Q G V L Q G R
GCC ATG GTC AAC TGC AAT GAC GAT CAG GGC GTG CTT CAG GGA CGC
788
240
W D N N Y S D G V S P M S W I
TGG GAC AAC AAC TAC AGT GAT GGT GTC AGC CCC ATG TCC TGG ATC
833
255
G S V D I L R R W K D Y G C Q
GGC AGC GTG GAC ATC CTG CGG CGC TGG AAA GAC TAT GGG TGC CAG
878
270
R V K Y G Q C W V F A A V A C
CGC GTC AAG TAC GGC CAG TGC TGG GTC TTC GCT GCT GTG GCC TGC
923
285
T V L R C L G I P T R V V T N
ACA GTG CTG CGG TGC CTT GGC ATC CCC ACC CGA GTC GTG ACC AAC
968
300
F N S A H D Q N S N L L I E Y
TTT AAC TCA GCC CAC GAC CAG AAC AGC AAC CTG CTC ATC GAG TAC
1013
315
F R N E S G E I E G N K S E M
TTC CGA AAC GAG TCT GGG GAG ATC GAG GGG AAC AAG AGC GAG ATG
1058
330
I W N F H S L L G G V V D D Q
ATC TGG AAC TTC CAC TCA CTG CTG GGT GGA GTC GTG GAT GAC CAG
1103
345
A G P G A W V R G V Q A L D P
GCC GGA CCT GGA GCC TGG GTA CGA GGG GTG CAG GCC CTG GAC CCC
1148
360
T P Q E K S E G T Y C C G P V
ACA CCC CAG GAG AAG AGT GAA GGG ACA TAC TGC TGT GGC CCA GTT
1193
375
P V R A I K E G H L N V K Y D
CCG GTT CGA GCC ATC AAG GAG GGC CAC CTG AAC GTC AAG TAT GAT
1238
390
A P F V F A E V N A D V V N W
GCA CCT TTC GTG TTT GCT GAG GTC AAT GCT GAC GTG GTG AAC TGG
1283
405
I R Q K D G S L R K S I N H L
ATC CGG CAG AAA GAT GGG TCC CTG CGC AAG TCC ATC AAC CAT TTG
1328
420
V V G L K I S T K S V G R D E
GTT GTG GGG CTG AAG ATC AGT ACT AAG AGT GTG GGC CGC GAT GAG
1373
435
R E D I T H T Y K Y P E G S E
CGA GAG GAC ATC ACC CAC ACC TAC AAG TAC CCA GAG GGA TCT GAA
1418
450
E E R E A F V R A N H L N K L
GAG GAG CGG GAA GCT TTT GTT AGG GCC AAC CAC CTA AAT AAA CTG
1463
465
A T K E E A Q E E T G V A M R
GCC ACA AAG GAA GAG GCT CAG GAG GAA ACG GGA GTG GCC ATG CGG
1508
480
I R V G Q N M T M G S D F D I
ATC CGT GTG GGC CAG AAC ATG ACT ATG GGC AGT GAC TTT GAC ATC
1553
495
F A Y I T N G T A E S H E C Q
TTT GCC TAC ATC ACC AAT GGC ACT GCT GAG AGC CAC GAA TGC CAA
1598
510
L L L C A R I V S Y N G V L G
CTC CTG CTC TGT GCA CGC ATC GTC AGC TAC AAT GGA GTC CTG GGG
1643
525
P V C S T N D L L N L T L D P
CCC GTG TGC AGC ACC AAC GAC CTG CTC AAC CTG ACC CTG GAT CCC
1688
540
F S E N S I P L H I L Y E K Y
TTC TCG GAG AAC AGC ATC CCC CTG CAC ATC CTC TAT GAG AAG TAC
1733
555
G D Y L T E S N L I K V R G L
GGT GAC TAC CTG ACT GAG TCC AAC CTC ATC AAG GTG CGA GGC CTC
1778
570
L I E P A A N S Y V L A E R D
CTT ATC GAG CCA GCA GCC AAC AGC TAT GTA TTG GCC GAG AGG GAC
1823
585
I Y L E N P E I K I R V L G E
ATT TAC CTG GAG AAT CCA GAA ATC AAG ATC CGG GTC TTG GGG GAG
1868
600
P K Q N R K L I A E V S L K N
CCC AAG CAG AAC CGC AAG CTG ATT GCT GAG GTG TCT CTG AAG AAT
1913
615
P L P V P L L G C I F T V E G
CCG CTC CCT GTG CCG CTG CTG GGT TGT ATC TTC ACC GTG GAA GGA
1958
630
A G L T K D Q K S V E V P D P
GCT GGC CTG ACC AAG GAC CAG AAG TCG GTG GAG GTC CCA GAC CCC
2003
645
V E A G E Q A K V R V D L L P
GTG GAA GCA GGG GAG CAA GCG AAG GTA CGG GTG GAC CTG CTG CCG
2048
660
T E V G L H K L V V N F E C D
ACG GAG GTG GGC CTC CAC AAG CTG GTG GTG AAC TTC GAG TGC GAC
2093
675
K L K A V K G Y R N V I I G P
AAG CTG AAG GCC GTG AAG GGC TAT CGG AAC GTC ATC ATC GGC CCC
2138
690
A *
GCC TAA GGGAGTCCTGTGCGCATCCCCCCTCAGTCTGCTGGAAACCCCCAACCAGAC
2195

FIGURE 2: Nucleotide sequence of guinea pig liver transglutaminase cDNA and the amino acid sequence deduced. The nucleotide sequence of the cDNA was deduced from the sequences of the cDNA inserts in the clones given in Figure 1. Nucleotide residues are numbered in the 5′ to 3′ direction beginning with the 5′-end residue. The amino acid sequence deduced is given above the nucleotide sequence, and the amino acid residues are numbered beginning with the Ala (position 1) next to the initiator Met (position −1). The nucleotide sequences corresponding to a putative ribosome binding site and a potential 3′-polyadenylation signal are boxed. the poly(A) tail is underlined. The asterisk indicates a stop codon. Amino acid sequences of seven peptides isolated from a trypsin digest of guinea pig liver transglutaminase (Ikura et al., 1987) are overlined. The wavy line indicates the amino acid sequence of pentapeptide containing the active site cysteine residue (Folk & Cole, 1966a). The solid arrow indicates the active site cysteine residue. Potential glycosylation sites (Asn-X-Ser or Asn-X-Thr) are shown by solid diamonds.

computer, Model PC-9801VX.

Results and Discussion

Isolation of cDNA Clones. We have already obtained one partial-length cDNA clone of guinea pig liver transglutaminase (pLTG 20, harboring nucleotide residues 633–1964 in Figure 2) and predicted about 60% of its primary structure, not including the amino or carboxyl termini (Ikura et al., 1987). To obtain cDNA clones carrying the sequences for the 5′-end region of mRNA, another cDNA library was constructed by elongation of a synthetic oligonucleotide primer (5′-GACACCATCACTGTA-3′), the sequence of which corresponded to nucleotide residues 801–815 near the 5′-end of the pLTG20 cDNA insert (Figures 1 and 2). This library was screened by hybridization with the ^{32}P-labeled oligonucleotide used as the primer. Of 10000 transformants, 20 hybridization-positive clones were isolated. One of them, named pLTG16, harbored the longest cDNA insert, and its nucleotides were sequenced. To obtain cDNA clones covering the 3′-end region of mRNA, the cDNA library, from which pLTG20 was isolated (Ikura et al., 1987), was screened again by hybridization with a ^{32}P-labeled probe of a 30-mer oligonucleotide corresponding to the sequence near the 3′-end of the pLTG20 cDNA insert (nucleotide residues 1821–1850, Figure 2). One hybridization-positive clone (pLTG4) that contained an insert longer than that in pLTG20 with respect to the 3′-end region was obtained (Figure 1). However, the cDNA insert (510 bp) of pLTG4 extended only 171 bp beyond the 3′-end of the pLTG20 cDNA insert and did not carry a translational stop codon. Therefore, the same cDNA library was screened again by hybridization with a multiprime-labeled probe prepared from the cDNA insert of pLTG4. Four hybridization-positive clones were isolated. One of them, named pLTG31, harbored the longest cDNA insert, and its nucleotides were sequenced (Figure 1).

Nucleotide Sequence of cDNA and Amino Acid Sequence Deduced. The nucleotide sequence of the cDNA coding for guinea pig liver transglutaminase was determined from the overlapped nucleotide sequences of the cloned cDNAs (pLTGs 4, 16, 20, and 31; Figure 2). The sequence contained 3695 nucleotide residues. A single long open reading frame was present beginning with an initiation codon (ATG) at residues 69–71 and ending with a stop codon (TAA) at residues 2142–2144; it included information for 691 amino acids. The sequence deduced for these amino acids is also shown in Figure 2. It included all sequences of seven peptides obtained from the tryptic digest of guinea pig liver transglutaminase (Ikura et al., 1987) and the sequence of pentapeptide containing the active site cysteine residue (Folk & Cole, 1966a), thus indicating that the open reading frame found here is coding for transglutaminase protein. This open reading frame can extend beyond the first ATG sequence to the 5′-end of the cDNA, and there were not any in-frame stop codons in the region upstream from the first ATG. But this ATG was assigned to the translation initiation codon, because the sequence ACCATGG is optimal for initiation on eukaryotic ribosomes (Kozak, 1986) and the sequence 27–35 (5′-CTTCGGGAG-3′) found in the untranslated region was highly complementary to the sequence (3′-GAAGGCGUC-5′) in the 3′-end region of the eukaryotic 18S rRNA (Hagenbuchle et al., 1978). The stop codon (TAA) was followed by 1544 bases of noncoding region, excluding the 3′-poly(A) sequence. This relatively long 3′-noncoding region is characteristic of mRNA of the catalytic a subunit of human factor XIII (Ichinose et al., 1986; Grundmann et al., 1986). The putative polyadenylation signal, AATAAA, was 23 bases upstream from the poly(A) tail. Northern blot hybridization of guinea pig liver poly(A+) RNA with a probe prepared from a restriction fragment of the pLTG20 cDNA insert gave a single band corresponding to about 3.7–3.8 kb (Figure 3). The sequence of cDNA determined here was 3695 bases long. It seemed that the 5′-untranslated region of intact guinea pig liver transglutaminase mRNA has about 110 nucleotide residues, because a primer extension product with about 280 nucleotides was detected when a 17-mer oligonucleotide corresponding to nucleotide residues 224–240 was used as a primer (data not shown). Thus, we have identified almost all of the liver transglutaminase mRNA sequence containing the coding region, but some nucleotides in the 5′-untranslated region were not identified.

We concluded that the amino-terminal amino acid residue of guinea pig liver transglutaminase was Ala (position 1, Figure

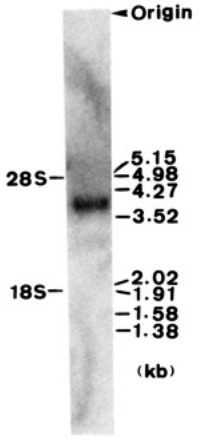

FIGURE 3: Identification of transglutaminase mRNA by blot hybridization. A 6-μg sample of poly(A+) RNA from guinea pig liver was analyzed. The ^{32}P-labeled probe was prepared from *Hin*cII–*Aat*I fragment (residues 750–1772) derived from the pLTG20 cDNA insert by multiprime DNA labeling (Feinberg & Vogelstein, 1983). The size markers used were 28S and 18S rRNAs of guinea pig liver and *Hin*dIII/*Eco*RI double digests of λDNA. Numbers on the right-hand side show the size of the digests of λDNA.

2) and that it was probably acetylated, for the following reasons. (i) In many eukaryotic proteins, Ser, Ala, and Gly in the position next to the initiator Met favor removal of the initiator Met and their N^{α}-acetylation (Boissel et al., 1985). (ii) The α-amino group of the amino-terminal amino acid residue of guinea pig liver transglutaminase is blocked (Connellan et al., 1971). The carboxyl-terminal amino acid residue should be Ala (position 690), because the Ala codon is followed with the stop codon. Connellan et al. (1971) reported that the amino- and carboxyl-terminal sequences of guinea pig liver transglutaminase are pyroGlu-Ala-Asp-Leu and Ser-Gly, respectively. However, we could not identify sequences corresponding these in the deduced amino acid sequence except for Ser-Gly (positions 319–320). The amino-terminal sequence assumed here (Ala-Glu-Asp-Leu) is similar but not identical with that reported by Connellan et al. (1971). The reasons for these discrepancies are not known.

The results of amino acid sequencing show that guinea pig liver transglutaminase consists of 690 amino acid residues with the following composition, excluding the initiator Met: Ala_{45}, Arg_{40}, Asn_{37}, Asp_{41}, $^1/_2$-Cys_{17}, Gln_{27}, Glu_{50}, Gly_{57}, His_{12}, Ile_{34}, Leu_{69}, Lys_{29}, Met_9, Phe_{23}, Pro_{32}, Ser_{43}, Thr_{31}, Trp_{10}, Tyr_{24}, Val_{60}. The composition ratio is in good agreement with that of the composition data reported by Folk and Cole (1966b). The molecular weight was calculated to be 76 620 from the composition obtained here. This agreed well with the values of 76 900 ± 5000 estimated by sedimentation and diffusion procedures (Folk & Cole, 1966b) and of 77 000 estimated by polyacrylamide gel electrophoresis in the presence of SDS (data not shown). The transglutaminase protein translated from the guinea pig liver poly(A+) RNA in the rabbit reticulocyte lysate system (Pelham & Jackson, 1976), which was identified in the SDS–polyacrylamide gel electrophoresis after immunoprecipitation, had the same molecular weight as the purified enzyme protein (data not shown). These findings suggest that the liver transglutaminase is not processed proteolytically after its translation except for minor modifications such a removal of the initiator Met at the amino terminal. The 17 $^1/_2$-Cys residues deduced in the amino acid sequence were consistent with the results of chemical modification experiments in which guinea pig liver transglutaminase was found to contain 16–18 sulfhydryl groups and no disulfide bonds (Folk & Cole, 1966b; Boothe & Folk, 1969).

The amino acid sequence deduced from the cDNA sequence showed the presence of six potential Asn-linked glycosylation sites (Figure 2), but purified guinea pig liver transglutaminase contains no carbohydrate (Folk & Chung, 1973).

The active site cysteine residue of the liver enzyme was at position 276 (Figure 2). Previous reports (Ikura et al., 1987; Takahashi et al., 1986) have indicated that the amino acid sequences surrounding the active site cysteine residues of guinea pig liver transglutaminase (Folk & Cole, 1966a) and human factor XIII (Holbrook et al., 1973) are highly homologous to the active site sequences of thiol proteases containing functional cysteine residues.

This is the first complete sequence determination of tissue-type nonzymogenic transglutaminase. The amino acid sequence of guinea pig liver transglutaminase was compared with that of the a subunit of human factor XIII, zymogenic transglutaminase. The latter is activated through the cleavage of the activation peptide of 37 residues from its amino-terminal region. The sequence alignment for maximum homology (Figure 4) had an overall homology of 36%, or 55% when both identical and conservative residues were taken into account, and a continuous stretch of gaps was counted as one substitution regardless of its length; this shows that the genes encoding these enzymes may originate from a common ancestor. Several regions of marked homology, including the region surrounding the active site cysteine residue, were found. These highly conserved regions may be important for the catalytic functions of transglutaminase. The results of the sequence alignment also showed that the cleavage of the amino-terminal activation peptide of the zymogenic a subunit of factor XIII yielded the active enzyme, which is more similar than the zymogenic a subunit to the liver transglutaminase in both primary structure and molecular weight. Results of the homology matrix analysis were consistent with those described here (data not shown). In this analysis, the amino acid sequence of the liver tansglutaminase was compared with that of the a subunit of human factor XIII by use of the computer program to generate diagonal lines indicating segments of 15 residues long that show sequence homology of 11/15 and over. Here both identical and conservative residues were taken into account. Several continuous diagonal lines were found, and the sequences of two enzymes indicated by these continuous lines corresponded well with the highly homologous regions shown in Figure 4. Internal homology was not found in the sequence of the liver transglutaminase by the homology matrix analysis, in which the amino acid sequence of this enzyme was compared with that of itself.

A hydropathy analysis (Kyte & Doolittle, 1982) was done for the sequence of guinea pig liver transglutaminase with use of the computer program (Figure 5). The active site cysteine residue was located at the amino-terminal end of a highly hydrophobic segment (positions 275–290, Figure 5). The same situation was observed in the a subunit of factor XIII (Takahaski et al., 1986). This result is compatible with the finding of Folk and Gross (1971), obtained with the use of a "reporter" group labeled halomethyl ketone, that the calcium-activated liver transglutaminase contains a hydrophobic region near its active site thiol group. Hydropathy profiles of the liver enzyme and the a subunit of factor XIII excluding the part of the activation peptide were similar, and the similarity seemed to be higher when the analysis was done at the long span of 19 residues than at the span of 9 residues (data not shown).

Some of the results described above suggest that there is a structural similarity between guinea pig liver transglutaminase and the activated a subunit of human factor XIII,

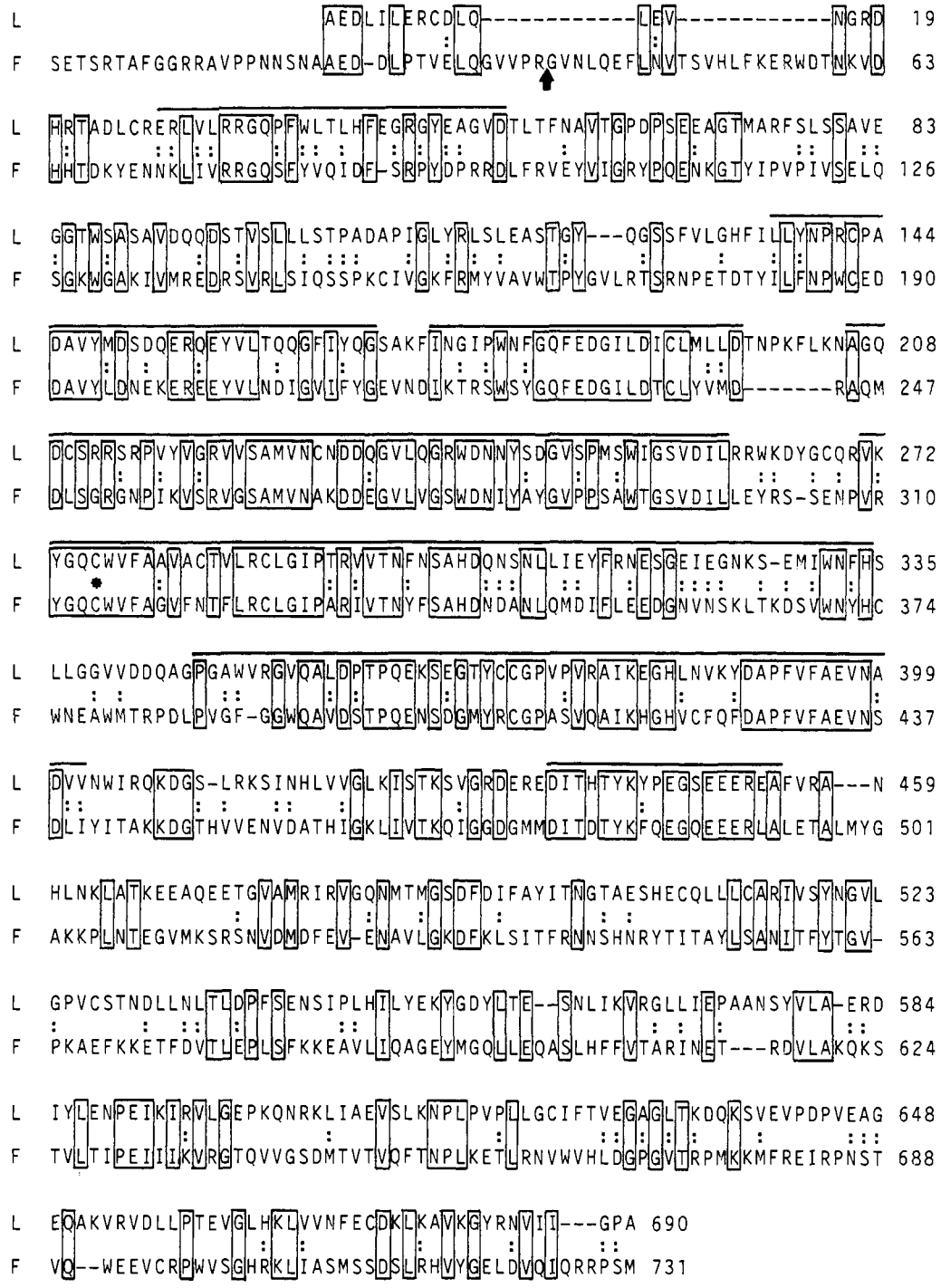

FIGURE 4: Sequence alignment for homology between guinea pig liver transglutaminase and the a subunit of human factor XIII. One-letter amino acid notation is used. L and F represent the sequences of liver transglutaminase and the a subunit of factor XIII, respectively. Gaps are inserted to achieve maximum homology. Amino acid residues are numbered beginning at the assumed amino-terminal Ala residue for liver enzyme and at the amino-terminal Ser residue for the a subunit of factor XIII. Sets of identical amino acid residues are boxed, and sets of residues that are thought to be the results of conservative substitution (Dayhoff et al., 1978) are indicated with colons. Highly homologous regions, which are longer than 15 residues and have more than 70% homology when both identical and conservative residues are taken into account, are indicated by horizontal lines. The solid arrow indicates the thrombin-activation cleavage site of the a subunit of factor XIII. The asterisk indicates the active site cysteine residues of both enzymes. Sequence data of the a subunit were taken from the reports of Ichinose et al. (1986), Grundmann et al. (1986), and Takahashi et al. (1986). For the residues (positions 77, 78, 88, 650, and 651) for which polymorphism was observed, the consensus residue at each position is given.

but marked differences in substrate specificity have been observed between them (Folk & Finlayson, 1977; Folk, 1983). The determination of regional sequences ruling the catalytic properties is necessary.

Transglutaminase is a calcium-dependent enzyme, and the guinea pig liver enzyme has several calcium-binding sites with different binding affinities (Folk & Chung, 1973). We have described the absence of an E–F-hand structure, a typical calcium-binding site characteristic of calmodulin, in the partial amino acid sequence of guinea pig liver enzyme deduced from the cDNA sequence of the clone pLTG20 (Ikura et al., 1987). Even in the entire amino acid sequence deduced here, we did not identify any region having an E–F-hand structure by use of the test sequence method (Tufty & Kretsinger, 1975) or by a search for homology to the consensus sequence of the calcium-binding site of calmodulins (Klee & Vanaman, 1982).

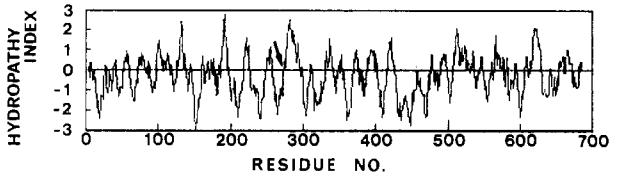

FIGURE 5: Hydropathy profile of guinea pig liver transglutaminase. The averaged hydropathy index (Kyte & Doolittle, 1982) of a nonapeptide composed of amino acid residues $i - 4$ to $i + 4$ is plotted against i, the amino acid number. The arrow indicates the position of the active site cysteine residue. Amino acid residues are numbered as described in the legend of Figure 4.

However, Takahashi et al. (1986) have identified potential calcium-binding sites in the primary structure of the a subunit of human factor XIII by homology to the high-affinity calcium-binding sites of calmodulins. It is difficult to locate the calcium-binding sites in the primary structure of guinea pig liver transglutaminase, although we have presumed that two regions rich in Glu residues (positions 446–453 and 468–473) are involved in the formation of calcium-binding sites of enzyme (Ikura et al. 1987). The region rich in Glu residues also can be found in the amino acid sequence of the a subunit of human factor XIII at positions 485–490 (Takahashi et al., 1986; Grundmann et al., 1986; Ichinose et al., 1986). The Glu-rich regions of both enzymes were mutually located on the carboxyl-terminal side of their active site cysteine residues at a distance of about 175 residues.

A genetic polymorphism of the liver transglutaminase was not found when a comparison of the cDNA inserts was made in regions where overlapping sequences were obtained (see Figure 1). The possibility that there are two differently charged transglutaminases in guinea pig liver is reported by Lorand et al. (1979). We do not know whether or not this finding and the discrepancies of results obtained with the amino- and carboxyl-terminal sequences of the liver enzyme, which are described above, are due to a genetic polymorphism. Further studies are necessary for the analysis of a polymorphism of liver transglutaminase gene. With the a subunit of human factor XIII, the existence of genetic polymorphism is suggested (Takahashi et al. 1986; Grundmann et al., 1986; Ichinose et al., 1986).

Registry No. Transglutaminase, 80146-85-6; DNA (guinea pig liver precursor protein moiety reduced), 113507-18-9; transglutaminase (guinea pig liver precursor protein moiety reduced), 113507-19-0; transglutaminase (guinea pig liver protein moiety reduced), 113507-20-3; blood coagulation factor XIII, 9013-56-3.

REFERENCES

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Birckbichler, P. J., & Patterson, M. K., Jr. (1978) *Ann. N.Y. Acad. Sci. 312*, 354–365.

Birckbichler, P. J., Orr, G. R., Patterson, M. K., Jr., Conway, E., & Carter, H. A. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 5005–5008.

Boissel, J.-P., Kasper, T. J., Shah, S. C., Malone, J. I., & Bunn, H. F. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 8448–8452.

Boothe, R. L., & Folk, J. E. (1969) *J. Biol. Chem. 244*, 399–405.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

Connellan, J. M., Chung, S. I., Whetzel, N. K., Bradley, L. M., & Folk, J. E. (1971) *J. Biol. Chem. 246*, 1093–1098.

Dayhoff, M. O., Schwartz, R. M., & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 3, pp 345–352, National Biomedical Research Foundation, Silver Spring, MD.

Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun. 23*, 641–646.

Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6–13.

Folk, J. E. (1980) *Annu. Rev. Biochem. 49*, 517–531.

Folk, J. E. (1983) *Adv. Enzymol. Relat. Areas Mol. Biol. 54*, 1–56.

Folk, J. E., & Cole, P. W. (1966a) *J. Biol. Chem. 241*, 3238–3240.

Folk, J. E., & Cole, P. W. (1966b) *J. Biol. Chem. 241*, 5518–5525.

Folk, J. E., & Gross, M. (1971) *J. Biol. Chem. 246*, 6683–6691.

Folk, J. E., & Chung, S. I. (1973) *Adv. Enzymol. Relat. Areas Mol. Biol. 38*, 109–191.

Folk, J. E., & Finlayson, J. S. (1977) *Adv. Protein Chem. 31*, 1–133.

Grundmann, U., Amann, E., Zettlmeissl, G., & Kupper, H. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8024–8028.

Gubler, U., & Hoffman, B. J. (1983) *Gene 25*, 263–269.

Hagenbuchle, O., Santer, M., Steitz, J. A., & Mans, R. J. (1978) *Cell (Cambridge, Mass.) 13*, 551–563.

Hanahan, D. (1983) *J. Mol. Biol. 166*, 557–580.

Holbrook, J. J., Cooke, R. D., & Kingston, I. B. (1973) *Biochem. J. 135*, 901–903.

Ichinose, A., Hendrickson, L. E., Fujikawa, K., & Davie, E. W. (1986) *Biochemistry 25*, 6900–6906.

Ikura, K., Kometani, T., Yoshikawa, M., Sasaki, R., & Chiba, H. (1980) *Agric. Biol. Chem. 44*, 1567–1573.

Ikura, K., Yoshikawa, M., Sasaki, R., & Chiba, H. (1981) *Agric. Biol. Chem. 45*, 2587–2592.

Ikura, K., Goto, M., Yoshikawa, M., Sasaki, R., & Chiba, H. (1984) *Agric. Biol. Chem. 48*, 2347–2354.

Ikura, K., Okumura, K., Yoshikawa, M., Sasaki, R., & Chiba, H. (1985) *Agric. Biol. Chem. 49*, 1877–1878.

Ikura, K., Nasu, T., Yokota, H., Sasaki, R., & Chiba, H. (1987) *Agric. Biol. Chem. 51*, 957–961.

Kannagi, R., Teshigawara, K., Noro, N., & Masuda, T. (1982) *Biochem. Biophys. Res. Commun. 105*, 164–171.

Klee, C. B., & Vanaman, T. C. (1982) *Adv. Protein Chem. 35*, 213–321.

Kozak, M. (1986) *Cell (Cambridge, Mass.) 44*, 283–292.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Lorand, L. (1972) *Ann. N.Y. Acad. Sci. 202*, 6–30.

Lorand, L., & Conrad, S. M. (1984) *Mol. Cell. Biochem. 58*, 9–35.

Lorand, L., Siefring, G. E., Jr., Tong, Y. S., Bruner-Lorand, J., & Gray, A. J., Jr. (1979) *Anal. Biochem. 93*, 453–458.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Mannual*, pp 309–328, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Messing, J., & Vieira, J. (1982) *Gene 19*, 269–276.

Mosher, D. F., & Schad, P. E. (1979) *J. Clin. Invest. 64*, 781–787.

Murtaugh, M. P., Arend, W. P., & Davies, P. J. A. (1984) *J. Exp. Med. 159*, 114–125.

Okumura, K., Ikura, K., Yoshikawa, M., Sasaki, R., & Chiba, H. (1984) *Agric. Biol. Chem. 48*, 2435–2440.

Peacock, S. L., McIver, C. M., & Monahan, J. J. (1981) *Biochim. Biophys. Acta 655*, 243–250.

Pelham, R. B., & Jackson, R. J. (1976) *Eur. J. Biochem. 67*, 247–256.

Rice, R. H., & Green, H. (1978) *J. Cell. Biol. 76*, 705–711.
Sakata, Y., & Aoki, N. (1980) *J. Clin. Invest. 65*, 290–297.
Sakata, Y., & Aoki, N. (1982) *J. Clin. Invest. 69*, 536–542.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Siefring, G. E., Apostol, A. B., Velasco, P. T., & Lorand, L. (1978) *Biochemistry 17*, 2598–2604.
Slife, C. W., Dorsett, M. D., & Tillotson, M. L. (1986) *J. Biol. Chem. 261*, 3451–3456.
Takahashi, N., Takahashi, Y., & Putnam, F. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8019–8023.
Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5201–5205.
Tufty, R. M., & Kretsinger, R. H. (1975) *Science (Washington, D.C.) 187*, 167–169.
Williams, J. G., & Mason, P. J. (1985) in *Nucleic Acid Hybridisation: A Practical Approach* (Hames, B. D., & Higgins, S. J., Eds.) pp 139–160, IRL, Eynsham, Oxford.
Williams-Ashman, H. G. (1984) *Mol. Cell. Biochem. 58*, 51–61.
Yoshikawa, M., Goto, M., Ikura, K., Sasaki, R., & Chiba, H. (1982) *Agric. Biol. Chem. 46*, 207–213.

Proximity of Thiol Esters and Bait Region in Human α_2-Macroglobulin: Paramagnetic Mapping†

Peter Gettins,*,‡ Albert H. Beth,§ and Leon W. Cunningham‡

Department of Biochemistry and Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, Tennessee 37232

Received October 5, 1987; Revised Manuscript Received November 20, 1987

ABSTRACT: The two key structural features of α_2-macroglobulin (α_2M) involved in inhibitory caging of proteases are the thiol ester and the bait region. This paper examines the environment of the hydrolyzed thiol ester in methylamine-treated human α_2M and the separation between the bait region and the thiol ester and between the four thiol esters in the tetramer to try to further our understanding of how bait region proteolysis triggers thiol ester cleavage. The sulfhydryl groups of Cys-949, formed upon cleavage of the thiol ester by methylamine, were specifically labeled with the nitroxide spin-labels 3-(2-iodoacetamido)-PROXYL (iodo-I) (PROXYL = 2,2,5,5-tetramethylpyrrolidine-1-oxyl), 3-[2-(2-iodoacetamido)acetamido]-PROXYL (iodo-II), and 4-(2-iodoacetamido)-2,2,6,6-tetramethylpiperidine-1-oxyl (iodo-III). ESR spectra of these α_2M derivatives showed that label I is firmly held and label II has limited freedom of rotation consistent with location of the cysteine residue in a narrow cavity. Label III has much greater motional freedom. From the absence of dipole–dipole splittings in the ESR spectra, it is concluded that the four nitroxide groups in the tetramer are more than 20 Å apart for both label I and label II. Label I broadens ^{1}H NMR signals from one phenylalanyl, one tyrosyl, and four histidyl residues in the bait region. Separations of 11–17 Å are estimated between the nitroxide of label I and these residues. Label II is further away and only broadens resonances from one of the histidines. The bait region is thus shown to be quite close to the thiol ester in the methylamine-treated form of α_2M and in addition must have a compact structure. It is suggested that the bait region exists as a loop of two strands in β conformation in the native structure (as well as in the methylamine-modified form of α_2M) and that proteolytic cleavage disrupts the loop and triggers thiol ester cleavage.

Human α_2-macroglobulin (α_2M)[1] is a tetrameric protease inhibitor with a number of unusual features. Rather than binding to the active sites of proteases and thereby inhibiting them, α_2M traps enzymes in a cagelike manner (Barrett & Starkey, 1973; Crews et al., 1987) and thus restricts access to high molecular weight substrates and inhibitors. The trapping or caging occurs as a result of a large-scale conformational change (Branegard et al., 1982; Björk & Fish, 1982; Dangott & Cunningham, 1982; Dangott et al., 1983), which is triggered by proteolytic cleavage of one or more of the four subunits of α_2M by the protease. These cleavages occur in the middle of the polypeptide chain, within a restricted segment of at most 30 amino acid residues (Sottrup-Jensen et al., 1981; Mortensen et al., 1981) termed the bait region (Salvesen & Barrett, 1980). Different proteases cleave at different sites within the bait region, at positions consistent with their specificity (Sottrup-Jensen et al., 1984). A consequence of proteolytic cleavage is the scission of the internal β-cysteinyl-γ-glutamyl thiol ester between Cys-949 and Glx-952, 260 residues removed from the bait region (Sottrup-Jensen et al., 1984), and a subsequent conformational change that results in trapping of the protease (Sottrup-Jensen et al., 1980; Howard, 1981; Salvesen et al., 1981). Since a similar, if not identical, conformational change can be brought about by thiol

†This work was supported by Grants BRSG RR05424-24 and HL34737 from the National Institutes of Health.
*Address correspondence to this author.
‡Department of Biochemistry.
§Department of Molecular Physiology and Biophysics.

[1] Abbreviations: α_2M, α_2-macroglobulin; DTNB, 5,5′-dithiobis(2-nitrobenzoic acid); PROXYL, 2,2,5,5-tetramethylpyrrolidine-1-oxyl; TEMPO, 2,2,6,6-tetramethylpiperidine-1-oxyl; CPMG, Carr–Purcell–Meiboom–Gill; SDS, sodium dodecyl sulfate; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate.

0006-2960/88/0427-2905$01.50/0