Pig Heart Calpastatin: Identification of Repetitive Domain Structures and Anomalous Behavior in Polyacrylamide Gel Electrophoresis†

Emiko Takano,‡ Masatoshi Maki,§ Hirotaka Mori,‡ Masakazu Hatanaka,§ Thomas Marti,∥ Koiti Titani,∥,⊥ Reiji Kannagi,‡ Tatsuo Ooi,# and Takashi Murachi*,‡

Department of Clinical Science and Laboratory Medicine, Faculty of Medicine, and Institute for Virus Research, Kyoto University, Kyoto 606, Japan, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611, Japan, and Department of Biochemistry, University of Washington, Seattle, Washington 98195

Received August 3, 1987; Revised Manuscript Received December 17, 1987

ABSTRACT: Isolation and nucleotide sequencing of the complementary DNA for pig heart calpastatin have been completed. The amino acid sequence of 713 residues predicted from the nucleotide sequence contains five domains, each composed of approximately 140 amino acid residues. A unique N-terminal domain is followed by four mutually homologous domains. The best fit alignment of these four domains gives residue identities between any two domains of 22.5–36.0%. The analysis of the sequence similarities by several methods also suggests the existence of additional shorter repeats at intervals of 60–80 residues. The calculated molecular weight of pig calpastatin of 713 amino acid residues (M_r 77 122) is significantly lower than the value of purified pig heart calpastatin (M_r 107 000) estimated by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (SDS–PAGE). The expression of the calpastatin genes in *Escherichia coli* and the detection of the translation products of 713, 366, and 140 amino acid residues by the specific anti-calpastatin antibody indicate that the products always migrate considerably slow on SDS–PAGE, giving an average of 1.53 for the ratio of the molecular weight estimated by SDS–PAGE to the value calculated from the amino acid sequences. It is most likely that the discrepancy in the molecular weight is caused by an anomalous behavior of calpastatin in SDS–PAGE.

Calpastatin is an endogeneous inhibitor protein acting specifically on calpain (EC 3.4.22.17; Ca^{2+}-dependent cysteine proteinase). Both calpain and calpastatin are known to be widely distributed in mammalian and avian cells (Murachi et al., 1981; Murachi, 1983a). Although physiological roles of calpain have not yet been clarified, the proteinase–proteinase inhibitor system has been suggested to play important roles in various cellular functions coupled with calcium ion mobilization (Murachi, 1983b, 1984; Pontremoli & Melloni, 1986; Murray et al., 1987; Suzuki, 1987).

Calpastatin (70 kDa)[1] was first purified to homogeneity from human erythrocyte (Takano & Murachi, 1982). Using specific anti-calpastatin antibody, we found diversity of calpastatins with different molecular sizes in various organs of human and pig in SDS–PAGE (50–110 kDa) (Takano et al., 1984a,b). A high molecular weight form (107 kDa) of calpastatin was purified from pig heart muscle and compared to a lower molecular weight form (68 kDa) isolated from pig erythrocyte. The biochemical properties of those calpastatins and the stoichiometrical analysis of the calpain inhibition led us to propose that the calpastatin molecule may comprise several functional regions composed of homologous amino acid sequences (Takano et al., 1986a). These hypothetical multidomain structures of pig calpastatin were substantiated by the finding of internally repetitive amino acid sequence deduced from a partial nucleotide sequence of a cloned cDNA (Takano et al., 1986b). Furthermore, the repetitive sequence was demonstrated to be the functional unit of the proteinase inhibitor by producing a calpastatin segment (22 kDa) composed of only one repetitive region in *Escherichia coli* (Maki et al., 1987a).

Recently, Emori et al. (1987) have reported the amino acid sequence of rabbit calpastatin deduced by cDNA cloning. The calculated molecular weight of rabbit liver mature calpastatin (M_r 68 113) was significantly lower than the value estimated by SDS–PAGE (M_r 110 000). The discrepancy in the molecular weight has remained to be solved.

In this paper, we present the primary structure of pig heart calpastatin deduced from the sequences of cloned cDNAs and discuss the multidomain structure. We also observed the molecular weight discrepancy and concluded that it is caused by an anomalous behavior of the protein in SDS–PAGE.

Materials and Methods

Materials. AMV reverse transcriptase was obtained from Seikagaku Kogyo Co. (Tokyo, Japan). 3′-Oligo(dG)-tailed pUC9, RNase H, *E. coli* DNA polymerase I, *E. coli* DNA ligase, and terminal deoxyribonucleotidyl transferase were obtained from Pharmacia (Uppsala, Sweden). Restriction

† This work was supported in part by Grants-in-Aid for Scientific and Cancer Research from the Ministry of Education, Science and Culture (T. Murachi and M.H.), a Basic Research Grant from the Muscular Dystrophy Association (T. Murachi), NIH Grant GM-15731 (to Dr. Kenneth A. Walsh), and a scholarship from the Swiss National Foundation (to T. Marti).

* Correspondence should be addressed to this author.

‡ Department of Clinical Science and Laboratory Medicine, Faculty of Medicine, Kyoto University.

§ Institute for Virus Research, Kyoto University.

∥ Department of Biochemistry, University of Washington.

⊥ Present address: Laboratory of Biomedical Polymer Science, School of Medicine, Fujita-Gakuen Health University, Toyoake, Aichi 470-11, Japan.

Institute for Chemical Research, Kyoto University.

[1] Abbreviations: SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; HPLC, high-performance liquid chromatography; kb, kilobase(s); kDa, kilodalton(s); Av.CC, average correlation coefficient; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; EDTA, ethylenediaminetetraacetic acid.

0006-2960/88/0427-1964$01.50/0

endonucleases, T4 DNA ligase, T4 DNA kinase, and plasmid pUC118/119 (Vieira & Messing, 1987) were obtained from Takara Shuzo Co. (Kyoto, Japan). High molecular weight protein standards were the products of Bio-Rad Laboratories (Richmond, CA). Nitrocellulose filters were from Schleicher & Schuell (Dassel, West Germany). Peroxidase-conjugated goat anti-rabbit IgG was from Miles Laboratories, Inc. (Naperville, IL). Other reagent-grade chemicals were obtained from Wako Pure Chemicals (Osaka, Japan) or Nakarai Chemicals (Kyoto, Japan).

Amino Acid Sequence Analysis. Preparation of peptide fragments and sequence analysis were performed as previously described (Takano et al., 1986b).

Preparation of Synthetic Oligonucleotides. Oligodeoxyribonucleotides were synthesized by the phosphoramidite method (Beaucage & Caruthers, 1981) with an automatic DNA synthesizer (Applied Biosystems Inc., Model 380B). After the synthesis, oligonucleotides were purified with HPLC under the conditions recommended by Applied Biosystems Inc. Oligonucleotides for sequencing primers shorter than 20-mer were used without HPLC purification, but repeatedly precipitated by 2 volumes of ethanol in the presence of 0.1 M sodium acetate.

Isolation of Primer-Extended cDNA Clones. Fifteen micrograms of poly(A+) RNA prepared from pig heart was used as a template for the synthesis of double-stranded cDNA essentially as described by Gubler and Hoffman (1983) except that 20 pmol of 25-mer calpastatin-specific oligonucleotide (see Results) was used as a primer. After dC tailing of cDNA and annealing with dG-tailed pUC9, *E. coli* DH 5 was transformed with the resultant recombinant DNA. Colony hybridization using 26-mer calpastatin-specific oligonucleotide (see Results) was performed as described previously (Takano et al., 1986b).

Nucleotide Sequencing. Nucleotide sequence analysis was carried out by the method of Sanger et al. (1977) with a kit supplied by Takara Shuzo Co. (Kyoto, Japan). Restriction fragments of cDNA were subcloned into pUC118/119 or their derivatives, and single-stranded DNA was prepared as described by Vieira and Messing (1987). After the cloning into pUC119, sequential deletion mutants were also prepared with a kilo-sequenching kit (Takara Shuzo Co.) (Yanisch-Perron et al., 1985).

Quantitative Comparison of Amino Acid Sequences. Amino acid sequence similarity was estimated quantitatively by two methods. A modified program using the mutation data matrix (Dayhoff et al., 1983) was employed for a best fit alignment of two similar sequences including insertions and deletions (Kanehisa, 1982). To estimate more quantitatively structural similarity of segments with consecutive residues, the extent of similarity was computed by the method of Kubota et al. (1982) and expressed as average correlation coefficients. According to this method, segments of identical residues should give a value of 1.00. The length of segments to be computed was 11 residues.

Construction of Recombinant Plasmid Expressing Calpastatin cDNA. A *Sac*II/*Eco*RI fragment of clone pPECS14 was subcloned into the *Hin*cII site of pUC119 after both ends of the cDNA fragment were blunt-ended with T4 DNA polymerase (Wartell & Renznikoff, 1980). The mutant clone pPECS146NK was obtained by creating restriction-cleavage sites of *Nco*I and *Kpn*I by in vitro site-directed mutagenesis (Tayler et al., 1985a) as described previously (Maki et al., 1987a,b) with a kit obtained from Amersham. *Ava*I was used for nicking the double-stranded DNA (Tayler et al., 1985b). An *Nco*I/*Eco*RV fragment of pOBSC1 was subcloned into

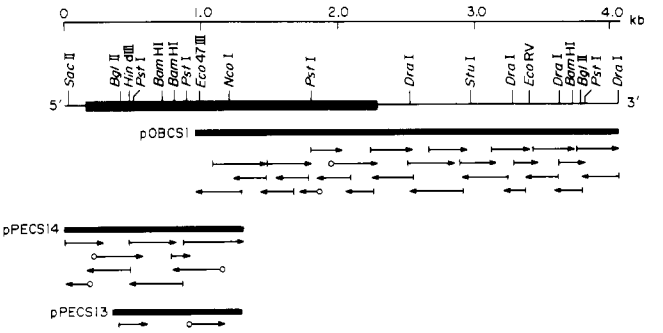

FIGURE 1: Restriction map and sequencing strategy for cDNA clones of pig heart calpastatin. The protein coding region is indicated by a closed box. Clone pOBCS1 was constructed by the method of Okayama and Berg. Clones pPECS13 and pPECS14 were obtained from the primer-extended cDNA library. Arrows indicate the direction and length of sequencing with universal (|→) and calpastatin-specific primers (O→).

the *Nco*I/*Hin*cII site of the pUC118 derivative which contained an *Nco*I site at the initiation methionine codon of *lac Z'* gene. The *Nco*I fragment of pPECS146NK was inserted into the *Nco*I site of the resultant subclone pCSDN234. A recombinant plasmid capable of expressing pig calpastatin cDNA which covers the whole translated region of the gene was termed pCSFL713.

Extraction of Pig Calpastatin Expressed in E. coli. E. coli cell extracts were prepared according to the method of Maki et al. (1987a). Briefly, harvested cells were washed with 20 mM Tris-HCl buffer, pH 7.5, containing 1 mM EDTA, 1 mM ethylene glycol bis(β-aminoethyl ether)-*N,N,N',N'*-tetraacetic acid, 50 mM NaCl, and 0.2 mM phenylmethanesulfonyl fluoride and suspended in the same buffer. A part of the cells was directly dissolved into the sample buffer for SDS–PAGE and boiled. The rest of the cells were treated with lysozyme (50 μg/mL) at 4 °C for 10 min, freeze-thawed 3 times, and centrifuged. The supernatant was used for the inhibitory activity assay and for Western blot analysis.

Assay for Calpastatin Activity. Calpain II (high Ca^{2+} requiring) was purified from pig kidney according to the method of Kitahara et al. (1984). Calpastatin was purified to homogeneity from pig heart by the method of Takano et al. (1986b). The calpastatin activity was measured by the method of Murakami et al. (1981), using pig calpain II as the test enzyme.

Western Blot Analysis. This procedure was carried out as described by Towbin et al. (1979). Briefly, marker proteins, purified pig heart calpastatin, and *E. coli* cell extracts were first subjected to SDS–PAGE according to the procedure of Laemmli (1970) using a 7.5% resolving gel and a 3% stacking gel. The portion of the gel with marker proteins was cut off and stained with Coomassie brilliant blue. The rest of the gel was then transferred to a nitrocellulose membrane. The nitrocellulose filter was first incubated with affinity-purified rabbit anti-calpastatin IgG (Takano et al., 1986a) and then with peroxidase-conjugated goat anti-rabbit IgG as the second antibody. The peroxidase staining was developed with 4-chloro-1-naphthol as the substrate.

RESULTS

Sequencing of cDNA Clone and Amino Acid Sequence. Nucleotide sequences of calpastatin cDNAs were determined according to the strategy shown in Figure 1. Isolation of clone pOBCS1 was previously described (Takano et al., 1986b). Clones pPECS13 and pPECS14 were obtained by extending the 25-mer synthetic oligonucleotide as a primer and by using the 26-mer synthetic oligonucleotide as a probe (comple-

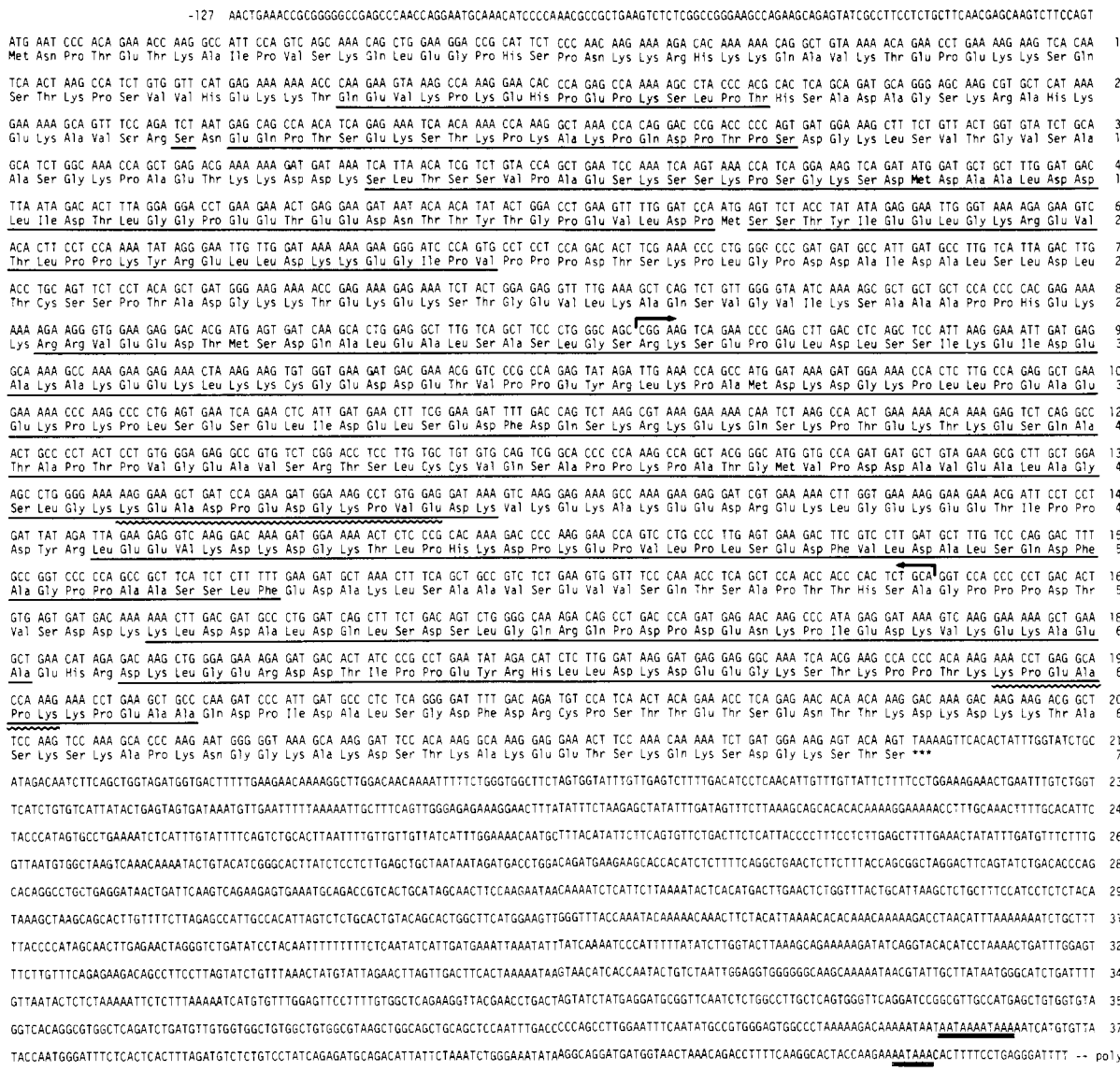

FIGURE 2: Nucleotide sequences of cloned cDNAs and predicted amino acid sequence of pig heart calpastatin. The deduced amino acid sequence is shown below the nucleotide sequence. Amino acid sequences of pig heart calpastatin determined by the Edman degradation method are underlined. Wavy lines indicate the partial amino acid sequence of pig erythrocyte calpastatin by the Edman degradation method. Double underlines indicate poly(A) signal sequences (AATAAA). The partial nucleotide sequence (residues between 910 and 1662), previously reported by Takano et al. (1986b), is marked by arrows.

mentary to nucleotide residues 1128–1152 and residues 925–950 in Figure 2), respectively. The nucleotide sequence presented in Figure 2 was obtained by superimposing the two overlapping sequences of clone pPECS14 (residues −127 to 1149) and clone pOBCS1 (residues 797–3911). The sequence in the overlapping region was identical between the two clones except for residue 982, being C as CTA (Leu) in clone pPECS14 and G as GTA (Val) in clone pOBCS1.

The sequence contains 4038 nucleotides comprising a 127-bp 5′-untranslated region, a 2139-bp translated region (corresponding to 713 amino acid residues), and a 1772-bp 3′-untranslated region (Figure 2). Initiation methionine was assigned by comparing the sequence with that of rabbit calpastatin cDNA recently published (Emori et al., 1987). Partial amino acid sequences of pig calpastatins from heart muscle (107 kDa) and from erythrocyte (68 kDa) were also determined by the Edman degradation method (underlined with solid lines and wavy lines, respectively, in Figure 2). Inconsistency of the amino acid sequences was observed at two positions. Amino acid residue 88 was determined as Ser by the Edman degradation method, whereas it was deduced as Asn from the nucleotide sequences (ATT) of both clones pPECS13 and pPECS14. As described above, residue 328 was deduced as Val (GTA) from clone pOBSC1, while it was deduced as Leu (CTA) from clone pPECS13 as well as from clone pPECS14. Leu at residue 328 was consistent with the result from the Edman degradation method. These variations in the amino acid sequences may be due to polymorphism of the calpastatin gene in the allele or the individual, if not due to error during cDNA cloning, or to protein sequencing. The appearance of more than one poly(A) signal sequence (Figure 2) may imply the occurrence of polymorphism in the 3′-noncoding region of mRNA.

Internally Repetitive Structure of Calpastatin. As shown in Figure 3A, the primary structure of pig calpastatin is homologous to that of reported rabbit calpastatin throughout the sequence except a part proximal to the N-terminus (residues 66–140). The aligned sequence of pig calpastatin, compared with that of rabbit calpastatin, had three insertions of one residue each and five deletions of a total of eight residues (aligned sequence not shown). Sequence homology between the two species as a whole was 68.1% (491 identical residues in 721 residues including gaps). The comparison matrix with its own sequence (Figure 3B) shows several segmented lines

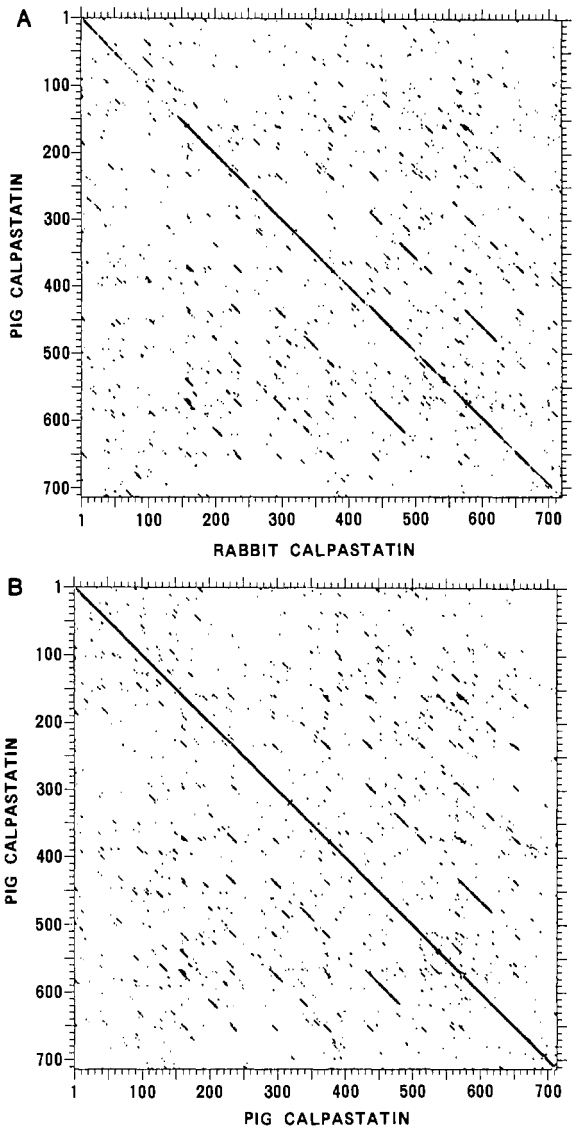

FIGURE 3: Comparison matrix of calpastatin. Similarities of amino acid sequence (A) between pig and rabbit and (B) within pig sequence were quantitatively estimated as described by Kubota et al. (1982). Residues with correlation coefficients greater than 0.50 were dotted. Segments of similar sequences give diagonal lines.

parallel to the central diagonal line, suggesting occurrence of internally repetitive sequences. Intervals of the four prominent lines including the main diagonal line are approximately 140 residues. Similar patterns can also be observed in the comparison with the rabbit sequence (Figure 3A).

Taking the patterns of these comparison matrices (Figure 3) as the guide, the pig calpastatin molecule was broken down into an N-terminal 150-residue portion, which is to be referred to as domain L, and four approximately 140-residue portions that follow, which are to be called domains 1, 2, 3, and 4. Figure 4 shows the best fit alignment of these latter four domains, indicating that these constitute four internal repeats. All possible pairs out of the four repeats shown in Figure 4 were compared so as to calculate the residue identities from both N-termini to the C-terminus of the shorter counterpart. Percent identity figures thus obtained ranged from 22.5% for domain 1 vs domain 2 to 36.0% for domain 2 vs domain 3.

Closer examination of the comparison matrices (Figure 3) also suggests that, in domains 1–4, shorter repetitive sequences at periodical intervals of about 60 and 80 residues may exist. In order to explore such a possibility, similarities of sequences were computed by a modified program using the mutation data matrix (Kanehisa, 1982). Figure 5 lists three top-scored alignments (A, B, and C) thus obtained, excluding the alignments given in Figure 4, and an alignment with a higher score (less similar) but containing a high content of identical residues (D). Some similarity exists between larger parts of domains L and 4 and between domain 3 and parts of domains L and 1 (Figure 5A,B). Besides, both between domains 1 and 2 to 3 (Figure 5C) and between domains 1 and 3 to 4 (Figure 5D), considerable similarity can be seen when the N-terminal half of one domain is aligned in parallel with the C-terminal half of the other domain. This led us to adopt, for the convenience of comparison, the assignment of regions I, II, III, and IV within each domain as are utilized in Figures 4 and 5. Combined regions I and II make an approximately 80-residue stretch, while combined regions III and IV represent an approximately 60-residue sequence. Several identical or conserved residues between regions I and III can also be found in Figure 4, supporting the notion of half-domain repeats of 60–80 residues.

Quantitative analysis was conducted for further examination of the structural similarity of short repetitive sequences. Average correlation coefficient (Av.CC) was computed for a larger number of combinations. The pairs of the segments that were found to give values of Av.CC of 0.50 or above and have residue lengths of at least 12 are listed in Table I. The table indicates that those pairs which gave higher values of Av.CC are the pairs either between the corresponding regions I and I, II and II, etc. or between noncorresponding regions I and III, with a few exceptions for domain L. Thus, the longest (52-residue) correlative segments giving the greatest value of Av.CC (0.73) are located between region I of domain 3 and region I of domain 4. Similarity between noncorresponding segments of region III of domain 1 and region I of domain 3 (15 residues; Av.CC, 0.65) is greater than the similarities between corresponding segments of region I of domain 1 and region I of domain 3 (12 residues; Av.CC, 0.54) and of region III of domain 1 and region III of domain 3 (13 residues; Av.CC, 0.55). Although the pair of region III of domain 1 and region I of domain 4 had a smaller Av.CC value of 0.53, which is largely due to the presence of a gap, it should be noted that the first part of aligned sequence presented in Figure 5D, 219–240 and 556–577, had a residue identity of 61.9%. All these findings lend support to postulating the correspondence of high degrees between regions I and III throughout the four domains 1–4.

Anomalous Mobility in SDS–PAGE of Translation Products of Calpastatin cDNA. Plasmids pCSFL713 and pCSDN234, capable of expressing pig calpastatin cDNA for 713 and 366 amino acid residues, respectively, in *E. coli*, were used. Plasmid pCSFL713 was constructed by the recombination of the two overlapping clones pPECS146NK and pCSDN234, as shown in Figure 6. The extracts from *E. coli* cells harboring pCSFL713 and pCSDN234 were found to show inhibitory activities against calpains I and II, as this was the case for *E. coli* expressed domain 3 (Maki et al., 1987a). When the electrophoresis samples were prepared by directly dissolving *E. coli* cells with the sample buffer containing SDS and boiling for 3 min, the major bands detected by Western blot analysis with anti-pig heart calpastatin antibody corresponded to molecular mass values of 120 kDa for clone pCSFL713 and 64 kDa for clone pCSDN234 (Figure 7, lanes 2 and 3), both being anomalously larger than the respective molecular weight values calculated. The numerical data are summarized in Table II, showing abnormal migration of calpastatins, regardless of whether they are of full length or truncated. Figure 7 also

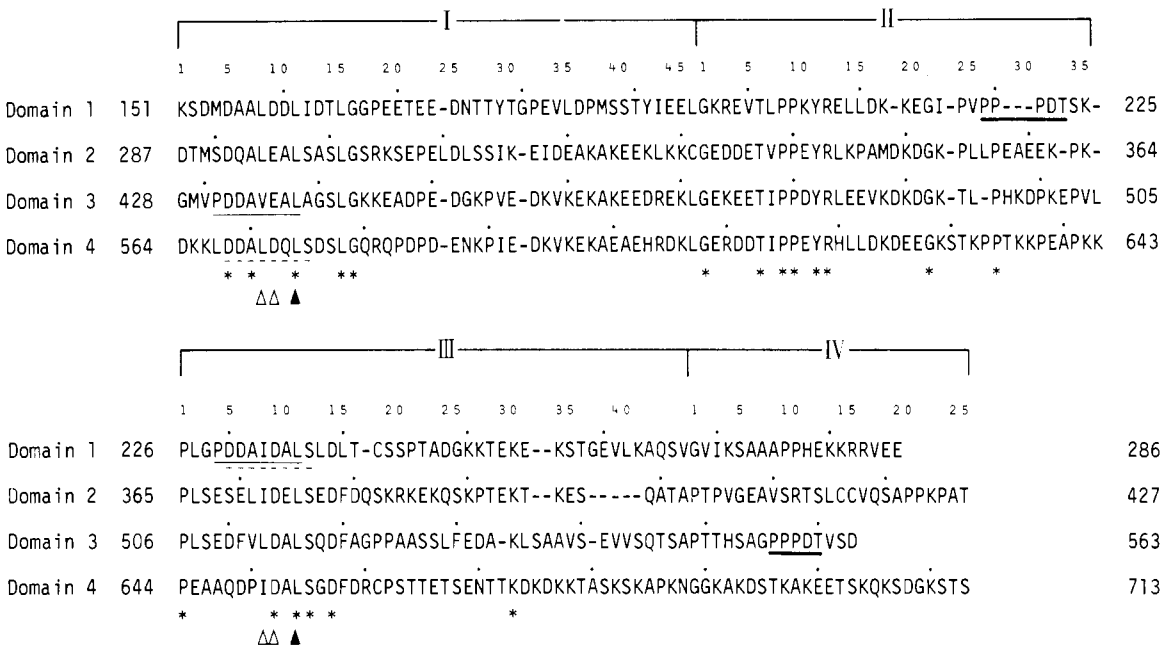

FIGURE 4: Sequence alignment of internally repetitive domains of pig calpastatin. Boundaries of domains 1–4 and regions I–IV were arbitrarily set by relative positions of conserved or identical residues: residues in all four domains are identical at position 1 in regions II and III, G and P, respectively. Residues at position 1 in regions I and IV were assigned by referring to the corresponding residues in regions III and II, respectively, as shown in Figures 5C,D. Asterisks indicate identical residues in all four domains. Closed and open triangles indicate identical and conserved residues, respectively, between regions I and III. Segments with identical residues between regions II and IV are marked with thick underlines. Segments with residues similar between regions I and III are marked with thin or dotted underlines.

```
A    26-152 versus 564-692          Aligned sequence length  129    Identical a.a.  18.6%   Score -81
     (Domain L)
 26  HKKQAVKTEPEKKSQSTKPSVVHEKKTQEVKPKEHPEPKSLPTHSADAGSKRAHKEKAVSRSNEQPTSEKSTKPKAKPQDPTPSDGKL-SVTGVSAASGK  124
564  DKKLDDALDQLSDSLGQRQPDPDENKPIEDKVKEKAEAEHRDKLGERDDTIPPEYRHLLDKDEEGKSTKPPTKKPEAPKKPEAAQDPIDALSGDFDRCPS  663
     I(Domain 4)                                  II                                        III

125  PAETKKDD-KSLTSSVPAESKSSKPSGKS  152
664  TTETSENTTKDKDKKTASKSKAPKNGGKA  692
                                IV

B    114-211 versus 434-531         Aligned sequence length   99    Identical a.a.  15.2%   Score -72
     (Domain L)                          I(Domain 1)                                   II
114  SVTGVSAASGKPAETKKDDKSLTSSVPAESKSSKPSGKSDMDAALDDLIDTLGGPEETEEDNTTYTGP-EVLDPMSSTYIEELGKREVTLPPKYRELLD  211
434  AVEALAGSLGKKEADPEDGKPVEDKVKEKAKEEDREKLGEKEETIPPDYR-LEEVKDKDGKTLPHKDPKEPVLPLSEDFVLDALSQDFAGPPAASSLFE  531
     (Domain 3)                          II                                        III

C    216-278 versus 418-481         Aligned sequence length   65    Identical a.a.  29.2%   Score -70
     (Domain 1)III                                          IV
216  IPVPPPDTSKPLGPDDAIDALSLDLTCSSPTA-DGKKTE-KEKSTGEVLKAQSVGVIKSAAAPPH  278
418  VQSAPPKPATGMVPDDAVEALAGSLGKKEADPEDGKPVEDKVKEKAKEEDREKLGE-KEETIPPD  481
     (Domain 2)I(Domain 3)                                  II

D    219-278 versus 556-617         Aligned sequence length   63    Identical a.a.  36.5%   Score -58
          III(Domain 1)                                     IV
219  PPPDT-SKPLGPDDAIDALSLDLTCSSPTADGKKT-E-KEKSTGEVLKAQSVGVIKSAAAPPH  278
556  PPPDTVSDDKKLDDALDQLSDSLGQRQPDPDENKPIEDKVKEKAEAEHRDKLGE-RDDTIPPE  617
          I(Domain 4)                                       II
```

FIGURE 5: Best fit sequence alignment of additional internally similar regions. Similarity of sequences was computed by a program using Dayhoff's mutation data matrix. Longer aligned sequences with more identical (+) or preferably substituted residues (;) and with less gaps give lower scores (negative numbers). Three top-scored alignments (A, B, and C), excluding alignments shown in Figure 4, and an alignment with a higher score (less similar) but containing a higher content of identical residues (D) are presented.

includes the data showing that proteolytic degradation of the translation products occurred when the extraction from *E. coli* was performed by lysozyme treatment and freeze-thawing (lanes 5 and 6).

Table I: Quantitative Comparison of Similar Sequences

	compared sequences[a]			location of center residue[b]				
	residues		no. of residues	domain-region		position[c]		av correlation coeff
	A	B		A	B	A	B	
domains 1–4								
corresponding region	430–481	566–617	52	3-I	4-I	29	29	0.73
	286–309	563–586	24	2-I	4-I	11	11	0.62
	333–355	473–495	23	2-II	3-II	13	13	0.67
	154–171	567–584	18	1-I	4-I	12	12	0.58
	289–306	430–447	18	2-I	3-I	11	11	0.64
	362–378	503–519	17	2-III	3-III	6	6	0.62
	197–212	609–624	16	1-II	4-II	9	9	0.67
	331–345	607–621	15	2-II	4-II	7	7	0.64
	225–239	364–378	15	1-III	2-III	7	7	0.62
	368–381	647–660	14	2-III	4-III	10	10	0.61
	229–241	647–659	13	1-III	4-III	10	10	0.57
	227–239	507–519	13	1-III	3-III	8	8	0.55
	159–170	436–447	12	1-I	3-I	14	14	0.54
noncorresponding region	225–239	427–441	15	1-III	3-I	7	7	0.65
	526–540	558–572	15	3-III	4-I	28	2	0.55
	155–168	536–549	14	1-I	3-III	11	39	0.59
	225–238	563–576	14	1-III	4-I	6	6	0.53
	307–319	479–491	13	2-I	3-II	27	14	0.58
	157–169	367–379	13	1-I	2-III	13	9	0.56
	156–168	511–523	13	1-I	3-III	12	12	0.53
	432–443	648–659	12	3-I	4-III	10	10	0.57
	291–302	648–659	12	2-I	4-III	10	10	0.54
domain L	126–138	558–570	13	L	4-I	132	1	0.60
	60–71	489–500	12	L	3-II	65	24	0.53

[a] The range indicates the start and end numbers of the middle of the segments (11 residues) to be computed; i.e., the additional five residues at both sides were also taken for computation (Kubota et al., 1982). [b] Center of a segment with an even number of residues was assigned for the last residue of the first half of the segment. [c] Position number in the corresponding region shown in Figure 4.

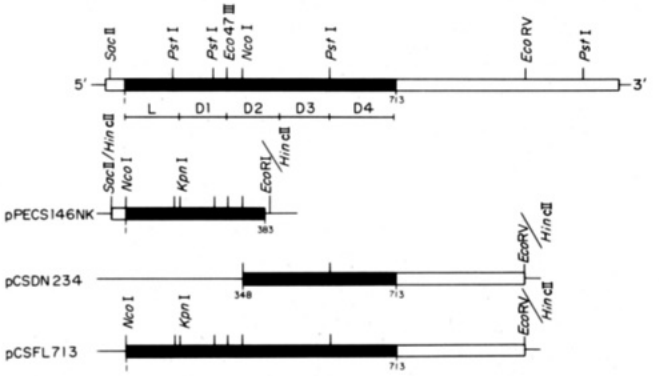

FIGURE 6: Structure of cDNA for producing pig heart calpastatin in *E. coli*. The top bar indicates the cDNA structure of pig heart calpastatin. The protein coding region and untranslated regions are indicated by a closed box and open boxes, respectively. Construction procedure of clone pCSFL713 is described in the text.

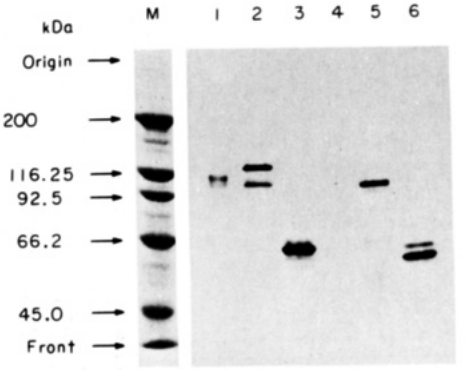

FIGURE 7: Western blot analysis of purified authentic pig heart calpastatin and the extracts from *E. coli* cells. After SDS–PAGE, the marker protein column was cut off and stained with Coomassie brilliant blue. The rest of the gel was transferred to a nitrocellulose membrane. The membrane was immunoblotted with anti-calpastatin antibody. (Lane 1) Purified authentic pig heart calpastatin; (other lanes) extracts from *E. coli* cells harboring clones pCSFL713 (lanes 2 and 5), pCSDN234 (lanes 3 and 6), and pUC119 (lane 4, negative control). For lanes 2–4, the extracts were prepared by directly dissolving the harvested and washed cells into the sample buffer for SDS–PAGE and boiling. For lanes 5 and 6, the extracts were prepared by treating the cells with lysozyme followed by freeze-thawing and centrifugation.

DISCUSSION

In this paper we present the primary structure of pig heart calpastatin (Figures 1 and 2). It has four internally repetitive sequences that are separated at intervals of about 140 amino acid residues (Figure 3). This structural feature is similar to that of rabbit calpastatin, recently reported by Emori et al. (1987), who independently cloned and sequenced cDNA. Although a part of the N-terminal domain (domain L) of pig calpastatin is slightly similar to regions I–III of domain 4 and region I of domain 3 (Figure 5A,B), the conserved residues in the aligned sequences are not identical with those among domains 1–4 in most cases. A second half of domain L is less conserved with the rabbit sequence (Figure 3A). Since domain L is also rich in acidic and basic residues as other domains are, relatively good scores obtained for the aligned sequences (Figure 5A,B) may be due to such a structural feature. Recently we reported that all four of internally repetitive domains 1–4, but not domain L, possess inhibitory activities against calpains (Maki et al., 1987b).

In addition to the four prominent repetitive sequences at intervals of about 140 residues in pig calpastatin, we found shorter similar sequences at intervals of about 80 residues within each repetitive domain or at intervals of about 60 residues between the neighboring domains (Figures 3B and 4). Aligned sequence have revealed that regions I and III are similar with each other (Figure 5C,D). Quantification of similarities of the segments in these regions indicates that in a few cases the segments in noncorresponding regions I and III are even more correlated than those between the corre-

Table II: Comparison of the Molecular Mass Values Estimated by SDS–PAGE to the Molecular Weight Values Calculated from Amino Acid Residues As Deduced from cDNA

plasmid	amino acid residues[a]	molecular mass/weight values (A) SDS–PAGE (kDa)	(B) calculated ($\times 10^{-3}$)[b]	A/B
pCSFL713	1–713	120	77.1	1.56
pCSDN234	348–713	64	39.7	1.61
pCSD3	426–555	22[c]	15.0	1.47

[a] Residue numbers of calpastatin shown in Figure 2. pCSD3 has 10 extra non-calpastatin residues (Maki et al., 1987a). Asn-2 was mutated to Asp in clone pCSFL713 by the creation of an *Nco*I site containing the translation initiation codon (see Materials and Methods). [b] Translation initiation methionine is included for the calculation of molecular weight. [c] Taken from Maki et al. (1987a).

sponding region I or III in different domains (Table I). Comparison of calpastatin sequences between pig and rabbit suggests that the shorter repetitive sequences at intervals of 60 or 80 residues are not unique to pig calpastatin but are also present in rabbit calpastatin (Figure 3A). A segment of residues, 554–562, in region IV of domain 3 of pig calpastatin, AGPPPDTVS, which includes the underlined residues PPPDT in Figure 4, is more similar to a segment of residues, 216–224, AGPPPDSVT, of rabbit calpastatin (Emori et al., 1987) than to a corresponding segment of pig calpastatin, residues 217–225, PVPPPDTSK, which is located in region II of domain 1.

A large number of cysteine proteinase inhibitors including kininogens are classified as the "cystatin superfamily", having the consensus sequence Gln-Val-Val-Ala-Gly (Barrett et al., 1986; Barrett, 1987). The absence of such a sequence in a 251-residue portion of pig calpastatin was earlier reported (Takano et al., 1986b). The present study confirmed the earlier result and has furthermore demonstrated that the cystatin-consensus sequence is present neither in any of the internally homologous domains 1–4 nor in a unique N-terminal domain L. Since domain 2 of low molecular weight kininogen is known to possess the inhibitory activity against calpain (Salvesen et al., 1986), we have most carefully searched if the sequence homology should exist between any portions of pig and rabbit calpastatins (Figure 2) and domain 2 of the kininogen. However the result was negative. Therefore, we conclude that calpastatins constitute a unique family of cysteine proteinase inhibitor quite independent from the cystatin superfamily.

It is reasonable to speculate that calpastatins have evolved from an ancestral calpastatin-type inhibitor which possesses only one repeat unit. It is not evident yet, however, that a half-domain structure, i.e., composed of about 60–80 residues, is a prototype of a 140-residue repeat unit. Recently we found that domain 3 protein which lacks region III and IV still possess calpastatin activity (Maki et al., unpublished results). It would be of interest to see whether there exists a half-domain structure gene in some other living organisms such as lower vertebrates. Two-step gene duplication including internally homologous recombination has been postulated for the three-times repeat structure of kininogen heavy chain which must have evolved from an ancestral cystatin or stefin polypeptide (Müller-Esterl et al., 1986). A possibility remains that domain L of calpastatin was derived from an ancestral active domain but it lost its function during its evolution.

The calculated molecular weight of the pig calpastatin protein (M_r 77 122) is significantly lower than the value of pig heart calpastatin (M_r 107 000) estimated by SDS–PAGE. The discrepancy of the molecular weight estimated by SDS–PAGE (110 000) from that based on the amino acid sequence (68 000) was also reported in the case of rabbit liver calpastatin (Imajoh et al., 1987; Emori et al., 1987). These authors explained the discrepancy as the result of posttranslational modification. According to Imajoh et al. (1987), an *E. coli* expressed, truncated calpastatin of 373 amino acid residues behaved normally on SDS–PAGE, giving a value of 40 kDa. Our study did not support these authors' observation in that not only the naturally occurring calpastatin but also the *E. coli* expressed calpastatin molecules showed abnormal behaviors in SDS–PAGE. As shown in Figure 7, the expression of the calpastatin gene in *E. coli* and the detection of translation products by the antibody revealed that the proteins migrate considerably slow on SDS–PAGE. A similar result had been obtained in our previous work where a portion of the calpastatin gene containing only domain 3 (clone pCSD3) was expressed in *E. coli* (Maki et al., 1987a). The ratios of the molecular weight of the translated products estimated by SDS–PAGE to the value calculated from the translated amino acid sequence were 1.56, 1.61, and 1.47 for clones pCSFL713, pCSDN234, and pCSD3, respectively (Table II). These three figures given an averaged ratio of 1.53.

Regardless of the different sizes of the translation products expressed in *E. coli*, they must have similar amino acid compositions to one another due to the occurrence of repetitive sequences. Since posttranslational modification responsible for an increase in apparent molecular weight is unlikely to occur in *E. coli* cells, the retardation of the electrophoretic migration should be attributable to the nature of the calpastatin polypeptide. It is thus concluded that calpastatin molecule or the repetitive domain(s) thereof show(s) an anomalous electrophoretic behavior in SDS–PAGE, giving a large positive bias over the true molecular weight. The physicochemical properties of calpastatin responsible for such an anomalous behavior are yet to be defined, but they must reflect rather unique amino acid compositions of calpastatins, being very poor in aromatic amino acids (1.5–2.8 mol %) and rich in proline (9.2–11.1 mol %) and charged amino acids (Glu, 11.5–12.1 mol %; Asp, 9.4–11.4 mol %; Lys, 10.6–14.7 mol %). The numerical values shown above in parentheses were taken from the sequence-deduced amino acid compositions of the translation products of pCSFL713, pCSDN234, and pCSD3 (Figure 6 and Table II) and from the chemically analyzed data on purified pig heart calpastatin (Takano et al., 1986a). High contents of negatively charged amino acids may restrict the binding of SDS to protein molecules. Remarkable overestimation of the apparent molecular weight by SDS–PAGE has been reported for nuclear histone-binding protein N1/N2 which has high contents of acidic amino acids (Glu, 18.0 mol %; Asp, 7.8 mol %) (Kleinschmidt et al., 1986). This protein of a M_r of 64 774 as calculated from its amino acid sequence was estimated to be approximately 110 kDa on SDS–PAGE. Anomalous electrophoretic behaviors have also been reported for several other proteins which are rich in proline or acidic residues (Yoshikawa et al., 1981; Zakut-Houri et al., 1983; Iino et al., 1986).

Two tryptic peptides were isolated from pig erythrocyte calpastatin and identified by Edman degradation to be located in domains 3 and 4 (Figures 2 and 4). Pig erythrocyte calpastatin gave a protein band of 68 kDa in SDS–PAGE (Takano et al., 1986a), while the *E. coli* expressed product of clone pCSDN234 gave a value of 64 kDa (Table II). Since clone pCSDN234 encodes domains 3 and 4 and a portion of domain 2 (Figure 6), erythrocyte calpastatin is likely to contain three internally repetitive domains, 2–4.

Overestimation of the calpastatin molecular weight by SDS–PAGE should elicit reconsideration on the previously reported stoichiometry of calpain inhibition. We reported that pig heart and erythrocyte calpastatins could bind approximately eight and five calpain molecules per molecule of calpastatin, respectively (Takano et al., 1986a). The calculation was based on the assumption that the molecular mass values obtained by SDS–PAGE, 107 kDa for heart calpastatin and 68 kDa for erythrocyte calpastatin, should directly represent the respective molecular weights. However, now that the overestimation by SDS–PAGE has become evident, the calculation of the calpain-binding capacities of calpastatin species should be corrected accordingly. For such correction, one may utilize an averaged ratio of the overestimation, 1.53, obtained from the data shown in Table II. It follows that 1 molecule of pig heart and erythrocyte calpastatins can bind approximately 5 and 3 calpain molecules, respectively. These corrected numbers may correspond to the numbers of repetitive domains, four in heart calpastatin and three in erythrocyte calpastatin, confirmed by the present study.

The diversity of molecular weight of calpastatin in various tissues is well-known (Takano et al., 1984b; Parkes, 1986). Calpastatin molecules greater than 120 kDa in SDS–PAGE were reported: e.g., 145 kDa by Mellgren and Carr (1983) and 172 kDa by Lepley et al. (1985). However, the present results as well as the results reported by Emori et al. (1987) have shown that such larger molecules may not appear as the primary product of the calpastatin gene. Unlike previous reports that showed mostly cytosolic localization of calpastatin, Mellgren et al. (1987) reported a tight membrane association of calpastatin in bovine cardiac sarcolemma and sarcoplasmic vesicles. The possibility of posttranslational modification of calpastatin which may cause an apparent molecular weight increase cannot be excluded in eukaryotic cells. On the other hand, no calpastatin molecule greater than 68–70 kDa could be found in pig and human erythrocytes (Takano et al., 1984b). It has not been known whether such a shorter molecule is formed at the transcriptional or at the posttranslational stage of the biosynthesis of calpastatin protein. Apparent proteolytic processing of *E. coli* expressed calpastatin into shorter fragments was found to occur readily during the course of extraction procedures (Figure 7). Similarly shorter fragments were also earlier reported to appear on the Western blots of various tissue extracts (Takano et al., 1984b, 1986a). Proteolysis of rabbit skeletal muscle calpastatin during the purification procedures was also reported (Nakamura et al., 1985). These lines of information indicate that calpastatin is a highly proteinase-susceptible protein. The calpastatin preparation isolated from pig heart migrates faster (107 kDa) on SDS–PAGE than the *E. coli* expressed translation product of 713 amino acid residues does (120 kDa) (Figure 7). The rabbit liver calpastatin isolated was reported to have 79 residues less at the N-terminus than the polypeptide corresponding to the total open reading frame deduced from the nucleotide sequence (Emori et al., 1987). Such shortening may or may not represent an artifact. The true mechanism for and implication of molecular diversity of calpastatins must still await further elucidation.

ACKNOWLEDGMENTS

We thank Akihiko Sato for technical assistance and Takako Saiga and Nobuko Ohmori for preparation of the manuscript.

REFERENCES

Barrett, A. J. (1987) *Trends Biochem. Sci. (Pers. Ed.) 12*, 193–196.

Barrett, A. J., Fritz, H., Gruff, A., Isemura, S., Jarvinen, M., Katunuma, N., Machleidt, W., Müller-Esterl, W., Sasaki, M., & Turk, V. (1986) *Biochem. J. 236*, 312.

Beaucage, S. L., & Caruthers, M. H. (1981) *Tetrahedron Lett. 22*, 1859–1862.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524–545.

Emori, Y., Kawasaki, H., Imajoh, S., Imahori, K., & Suzuki, K. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 3590–3594.

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

Iino, T., Takeuchi, K., Nam, S. H., Siomi, H., Sabe, H., Kobayashi, N., & Hatanaka, M. (1986) *J. Gen. Virol. 67*, 1373–1379.

Imajoh, S., Kawasaki, H., Emori, Y., Ishiura, S., Minami, Y., Sugita, H., Imahori, K., & Suzuki, K. (1987) *FEBS Lett. 215*, 274–278.

Kanehisa, M. I. (1982) *Nucleic Acids Res. 10*, 183–196.

Kitahara, A., Sasaki, T., Kikuchi, T., Yumoto, N., Hatanaka, M., Yoshimura, N., & Murachi, T. (1984) *J. Biochem. (Tokyo) 95*, 1759–1766.

Kleinschmidt, J. A., Digwall, C., Maier, G., & Franke, W. W. (1986) *EMBO J. 5*, 3547–3552.

Kubota, Y., Nishikawa, K., Takahashi, S., & Ooi, T. (1982) *Biochim. Biophys. Acta 701*, 242–252.

Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.

Lepley, R. A., Pampusch, M., & Dayton, W. R. (1985) *Biochim. Biophys. Acta 828*, 95–103.

Maki, M., Takano, E., Mori, H., Kannagi, R., Murachi, T., & Hatanaka, M. (1987a) *Biochem. Biophys. Res. Commun. 143*, 300–308.

Maki, M., Takano, E., Mori, H., Sato, A., Murachi, T., & Hatanaka, M. (1987b) *FEBS Lett. 223*, 174–180.

Mellgren, R. L., & Carr, T. C. (1983) *Arch. Biochem. Biophys. 225*, 779–786.

Mellgren, R. L., Lane, R. D., & Kakar, S. S. (1987) *Biochem. Biophys. Res. Commun. 142*, 1025–1031.

Müller-Esterl, W., Iwanaga, S., & Nakanishi, S. (1986) *Trends Biochem. Sci. (Pers. Ed.) 11*, 336–339.

Murachi, T. (1983a) *Trends Biochem. Sci. (Pers. Ed.) 8*, 167–169.

Murachi, T. (1983b) *Calcium Cell Funct. 4*, 377–410.

Murachi, T. (1984) *Biochem. Soc. Symp. 49*, 149–167.

Murachi, T., Tanaka, K., Hatanaka, M., & Murakami, T. (1981) *Adv. Enzyme Regul. 19*, 407–424.

Murakami, T., Hatanaka, M., & Murachi, T. (1981) *J. Biochem. (Tokyo) 90*, 1809–1816.

Murray, A. W., Fournier, A., & Hardy, S. J. (1987) *Trends Biochem. Sci. (Pers. Ed.) 12*, 53–54.

Nakamura, M., Inomata, M., Hayashi, M., Imahori, K., & Kawashima, S. (1985) *J. Biochem. (Tokyo) 98*, 757–765.

Parkes, C. (1986) in *Proteinase Inhibitors* (Barrett, A. J., & Salvesen, G., Eds.) pp 571–587, Elsevier Science, Amsterdam.

Pontremoli, S., & Melloni, E. (1986) *Annu. Rev. Biochem. 55*, 455–481.

Salvesen, G., Parkes, C., Abrahamson, M., Grubb, A., & Barrett, A. J. (1986) *Biochem. J. 234*, 429–434.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Suzuki, K. (1987) *Trends Biochem. Sci. (Pers. Ed.) 12*, 103–105.

Takano, E., & Murachi, T. (1982) *J. Biochem. (Tokyo) 90*, 889–892.

Takano, E., Kitahara, A., Kannagi, R., & Murachi, T. (1984a) *J. Appl. Biochem. 6*, 117–125.

Takano, E., Yumoto, N., Kannagi, R., & Murachi, T. (1984b) *Biochem. Biophys. Res. Commun. 122*, 912–917.
Takano, E., Kitahara, A., Sasaki, T., Kannagi, R., & Murachi, T. (1986a) *Biochem. J. 235*, 97–102.
Takano, E., Maki, M., Hatanaka, M., Mori, H., Zenita, K., Sakihama, T., Kannagi, R., Marti, T., Titani, K., & Murachi, T. (1986b) *FEBS Lett. 208*, 199–202.
Tayler, J. W., Ott, J., & Eckstein, F. (1985a) *Nucleic Acids Res. 13*, 8765–8785.
Tayler, J. W., Schmidt, W., Cosstick, R., Okruszek, A., & Eckstein, F. (1985b) *Nucleic Acids Res. 13*, 8749–8763.
Towbin, H., Staehelin, T., & Gordon, T. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4350–4354.
Vieira, J., & Messing, J. (1987) *Methods Enzymol.* (in press).
Wartell, R. M., & Reznikoff, W. S. (1980) *Gene 9*, 307–319.
Yanisch-Perron, C., Vieira, J., & Messing, J. (1985) *Gene 33*, 103–119.
Yoshikawa, M., Mizukami, T., Omori, K., Sasaki, R., & Chiba, H. (1981) *Agric. Biol. Chem. 45*, 177–183.
Zakut-Houri, R., Oren, M., Bienz, B., Lavie, V., Hazum, S., & Givol, D. (1983) *Nature (London) 306*, 594–597.

Conservation of Primary Structure in the Lipoyl-Bearing and Dihydrolipoyl Dehydrogenase Binding Domains of Mammalian Branched-Chain α-Keto Acid Dehydrogenase Complex: Molecular Cloning of Human and Bovine Transacylase (E2) cDNAs†

Kim S. Lau, Thomas A. Griffin, Chii-Whei C. Hu, and David T. Chuang*
Genetics Division, Department of Medicine, Veterans Administration Medical Center, and Departments of Medicine and Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106
Received September 4, 1987; Revised Manuscript Received November 18, 1987

ABSTRACT: The subunit structures and conservation of the dihydrolipoyl transacylase (E2) components of bovine and human branched-chain α-keto acid dehydrogenase complexes were investigated by Western blotting, peptide sequencing, and cDNA cloning methods. Rabbit antiserum prepared against the sodium dodecyl sulfate (SDS) denaturated bovine E2 subunit recognized the inner E2 core, and the first hinge region of the E2 chain, but failed to react with the lipoyl-bearing domain as determined by Western blot analysis. The lack of antigenicity in the lipoyl-bearing domain was confirmed with antibodies directed against the native E2 component. A human E2 cDNA (1.6 kb) was isolated from a human liver cDNA library in λgt11 with a combination of the above anti-native and anti-SDS-denatured E2 immunoglobulin G's as a probe. The fidelity of the human E2 cDNA was established by nucleotide sequencing which showed the determined peptide sequences of the amino terminus and tryptic fragments of bovine E2. A bovine E2 cDNA (0.7 kb) was also isolated from a bovine liver cDNA library in λZAP with the human E2 cDNA as a probe. Northern blot analysis using the human E2 cDNA probe showed that E2 mRNAs in bovine liver and human kidney mesangial cells are 3.3 and 4.6 kb in size, respectively. Primary structures derived from human and bovine E2 cDNAs show leader sequences including the initiator methionine and the homologous mature peptides consisting of complete lipoyl-bearing and dihydrolipoyl dehydrogenase (E3) binding domains and two hinge regions. In addition, the human E2 cDNA contains a portion of the inner E2 core sequence, a 3′-untranslated region, and a poly(A+) tail. Deduced amino acid sequences of the mammalian E2's were compared with those of *Escherichia coli* transacetylase and transsuccinylase and bovine kidney transacetylase. The results indicate a high degree of conservation in the sequence flanking the lipoyl-attachment site and in the E3-binding domain. Models are presented to discuss implications for the conserved structure–function relationship in the lipoyl-bearing and E3-binding domains of α-keto acid dehydrogenase complexes.

The mammalian branched-chain α-keto acid dehydrogenase (BCKAD)[1] complex catalyzes the oxidative decarboxylation of α-keto acids that are derived by transamination from the branched-chain amino acids: valine, leucine, and isoleucine. This mitochondrial multienzyme complex is both structurally and mechanistically analogous to pyruvate and α-ketoglutarate dehydrogenase complexes (Reed et al., 1980). The BCKAD complex consists of three catalytic components: a branched-

† This work was supported by Grants DK 37373 and DK 26758 from the National Institutes of Health, by Grant 1-796 from the March of Dimes Birth Defects Foundation, and by the Veterans Administration. K.S.L. was supported by a Postdoctoral Research Fellowship from the American Heart Association/Northeast Ohio Affiliate. T.A.G. is a Medical Scientist Trainee of Grant GM 07250 from the National Institutes of Health.
* Author to whom correspondence should be addressed at the Department of Medicine (151-W), Veterans Administration Medical Center, 10701 East Blvd., Cleveland, OH 44106.

[1] Abbreviations: BCKAD, branched-chain α-keto acid dehydrogenase; E1, branched-chain α-keto acid decarboxylase; E2, dihydrolipoyl transacylase; E3, dihydrolipoyl dehydrogenase; SDS, sodium dodecyl sulfate; IPTG, isopropyl β-D-thiogalactopyranoside; HPLC, high-performance liquid chromatography; TBS, 50 mM Tris-HCl, pH 8.1/150 mM NaCl; SSC, 15 mM sodium citrate, pH 7.4/150 mM NaCl; kb, kilobase pair(s); bp, base pair(s); TLCK, N^{α}-*p*-tosyl-L-lysine chloromethyl ketone; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride.

0006-2960/88/0427-1972$01.50/0