Biochemistry

 Volume 26, Number 6 March 24, 1987

Accelerated Publications

Repeating Structure of Chick Tropoelastin Revealed by Complementary DNA Cloning†

Giorgio M. Bressan,*,‡ Patrick Argos, and Keith K. Stanley
European Molecular Biology Laboratory, 6900 Heidelberg, FRG
Received November 13, 1986; Revised Manuscript Received January 15, 1987

ABSTRACT: A cDNA library was constructed from chick aorta poly(adenylic acid)-containing RNA in the expression vector pEX1. Several clones were identified by screening the library with a polyclonal antiserum raised against chick tropoelastin and confirmed by DNA sequencing. Analysis of the deduced amino acid sequence, corresponding to the mature tropoelastin and most of the signal peptide, revealed that the molecule is composed of at least 8, and possibly 13, repeating units. The common features of each unit include an N-terminal region composed largely of alanines and lysines and ending with an aromatic amino acid, followed by a GAG span and then a C-terminal region consisting mostly of valines, prolines, and glycines often present in several copies of the sequence (VPGV). This structure is discussed in terms of the functional properties of the molecule.

The elastic component of connective tissue, elastin, is formed by covalent cross-linking of a soluble precursor, tropoelastin, in the extracellular matrix (Sandberg, 1976). Once secreted, tropoelastin is modified by the enzyme lysyl oxidase, which oxidizes the ε-amino group of most lysines to aldehyde groups (Kagan, 1986). These groups spontaneously condense, forming several types of derivatives that link together different molecules (Partridge, 1962; Franzblau, 1971; Gallop et al., 1972).

The primary structure of tropoelastin has been the subject of several studies in the past (Sandberg et al., 1971; Foster et al., 1973, 1975; Rucker et al., 1975; Smith et al., 1981), the most extensive of which concerns pig tropoelastin [reviewed by Sandberg and Davidson (1984)]. These studies have identified two types of regions: alanine- and lysine-rich regions where cross-links are established and hydrophobic sequences which are probably responsible for the elastic behavior of the mature fibers. More recently, additional information has been obtained on the structure of the signal peptide of chick and sheep tropoelastin by protein sequencing (Karr & Foster, 1981; Sandberg & Davidson, 1984) and on the carboxy-terminal end of sheep and bovine tropoelastin by sequencing of cDNA of genomic clones (Yoon et al., 1985; Cicila et al., 1985). In both species the carboxy-terminal end is highly basic and contains two cysteine residues. Until now, however, the complete sequence of tropoelastin from any source has not been determined.

In this paper we report the isolation and characterization of cDNA clones coding for chick tropoelastin. The open reading frame covered by these clones spans most of the signal peptide and the entire sequence of the mature protein. The data show that the molecule has a repeating structure, being formed by 8, and possibly 13, regions with variable degrees of homology.

Experimental Procedures

cDNA Cloning. Poly(A+)[1] RNA was purified (Chirgwin et al., 1979) from 2-day-old chick aortas and cDNA synthesized (Gubler & Hoffman, 1983) by using a random hexanucleotide (Pharmacia) as a primer. A library was constructed in the expression vector pEX1 (Stanley & Luzio, 1984) by use of adaptors (Haymerle et al., 1986). Screening of the library and expression of tropoelastin clones were performed by established procedures (Stanley, 1983; Stanley & Luzio, 1984).

Radioactive probes were prepared as described by Feinberg and Vogelstein (1983). Southern blot analysis was effected

† This work was partially supported by a grant from the Commission of the European Communities (to G.M.B.).
* Author to whom correspondence should be addressed.
‡ On leave from the Institute of Histology, University of Padova, 35100 Padova, Italy.

[1] Abbreviations: poly(A+), poly(adenylic acid) containing; bp, base pairs.

0006-2960/87/0426-1497$01.50/0

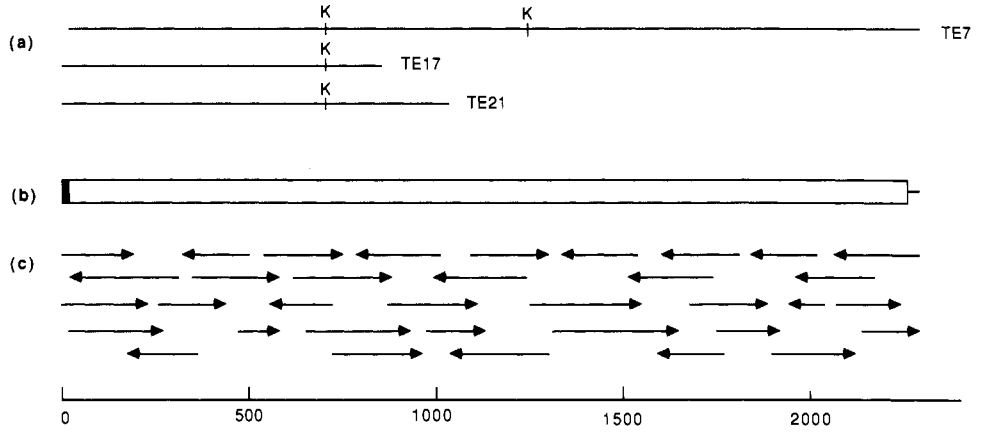

FIGURE 1: Cloning of tropoelastin cDNA: (a) clones that have been characterized; (b) assembled structure of tropoelastin cDNA; (c) M13 clones used for sequencing the cDNA. K = *Kpn*I restriction site; filled box = signal peptide; open box = region coding for mature tropoelastin; continuous line = 3′ untranslated region.

by standard procedures (Maniatis et al., 1982).

DNA Sequencing. Random fragments (Deininger, 1983) or restriction fragments from different tropoelastin clones were subcloned in M13mp19. Sequencing was executed with the method of Sanger et al. (1977) as modified by Biggin et al. (1983).

Sequence Comparison. The sequence start positions of the tropoelastin repeats were delineated by intercomparing all possible tropoelastin spans, 20 residues in length, by a combination of two scoring procedures: the Dayhoff relatedness odds matrix and correlations of several residues' physical characteristics (Argos, 1985, 1986). The mean and the standard deviation (σ) of the scores for all the segment intercomparisons were calculated; the scores were then tabulated in a search matrix (bounded by the tropoelastin sequence positions) as a number of standard deviations above or below the mean score (henceforth referred to as standard deviation fractions or σ level). The self-sequence comparison scores as well as all those below 3.0 σ were set to zero. Scores along a vertical column in the search matrix were then summed and plotted at each sequence position as described previously (Altruda et al., 1985). The sequence positions at the beginning of successive significant peaks represent the suggested start sites for repeating units in the tropoelastin sequence.

Results and Discussion

cDNA Library Screening and Sequence Determination. An aliquot (about 1.2×10^5 independent clones) of the chick aorta library, which contained about 7.5×10^5 clones, was screened with an antiserum obtained in rabbits against chick tropoelastin (Bressan et al., 1983). Of the 30 reactive clones that were chosen, 29 were positive in a second screening with the same antibody. When tested in a "retroblotting" assay (Hall et al., 1984), all the clones gave a positive result. The clone with the longest insert, TE7 (about 2250 bp), was further studied by partial sequencing of its three *Kpn*I fragments in both directions. Comparison with published tropoelastin sequences established that the clone extended from the signal peptide coding region to 32 bases past the termination codon. The sequence of the entire clone was determined by a shotgun approach as summarized in Figure 1.

To gain further information on the region 5′ to TE7, the *Kpn*I digests of the remaining clones were analyzed on a Southern blot with a radioactive probe prepared from the 5′ *Kpn*I fragment of TE7. Clones with a reacting fragment comparable in size to the 5′ *Kpn*I fragment of TE7 were subcloned in M13 and sequenced from the 5′ end. Clones TE17 and -21 were found to extend for an additional 29 bases in the 5′ direction (Figure 1). The sequence data obtained from the three clones are shown in Figure 2.

The sequence does not include the start codon, and the first codon specifies an alanine in a position where protein sequencing of chick pre-tropoelastin has placed the starting methionine (Karr & Foster, 1981). This difference might be due to chick strain variability. The signal peptide shows significant homology with the sheep sequence (Sandberg & Davidson, 1984), and the comparison suggests that two codons might be missing in our cDNA clones.

Assuming an amino terminus as determined by protein sequencing (Foster et al., 1975; Rucker et al., 1975), mature tropoelastin comprises 726 amino acids with a molecular weight of 61 246. The protein does not contain methionine, tryptophan, histidine, aspartic acid, and glutamic acid. It contains one asparagine (residue 196) which is not in a canonical sequence for N-glycosylation.

Repeating Units in Tropoelastin. Figure 3 shows a smoothed plot of the tropoelastin sequence position vs. the summed fractional standard deviation scores of the self-comparison search matrix. It is clear that peaks 3 and 6–12 strongly indicate start positions for repeating sequence units. A visual examination of these segments showed several common features that allowed a facile alignment. The sequence repeat plot also indicated three other possible start positions for repeating units (peaks 1, 2, and 4 in Figure 3). These units were then aligned visually with the previous eight segments in the light of the observed shared organization. Only two spans were left unaligned in the molecule (units 5 and 13); they were subsequently added to the alignment list as they contained certain of the conserved features (Figure 4).

The significance of the putative 13 repeat homologies was tested by a control experiment in which the sequences in the 13 tropoelastin segments were randomly shuffled, resulting in 182 different residue spans. All possible pairwise alignments were made from the N-terminal most ends, the alignment length corresponding to the shorter of the two segments. The number of amino acid identities was collated for each of the 16 471 trials, and their mean and standard deviation were calculated. The mean number of identical residues in these comparisons was 9, with a standard deviation of 4. Table I lists the number of identities for all interrepeat alignments as well as the number of standard deviations (σ) above the mean of the number of identities in the random-shuffled sequences. In agreement with the peak heights of the search plot, units 3 and 6–12 show highly significant relationships. The smallest σ level needed to relate them is 6.0, either by direct means or through implication (i.e., two pairs with a strong relationship

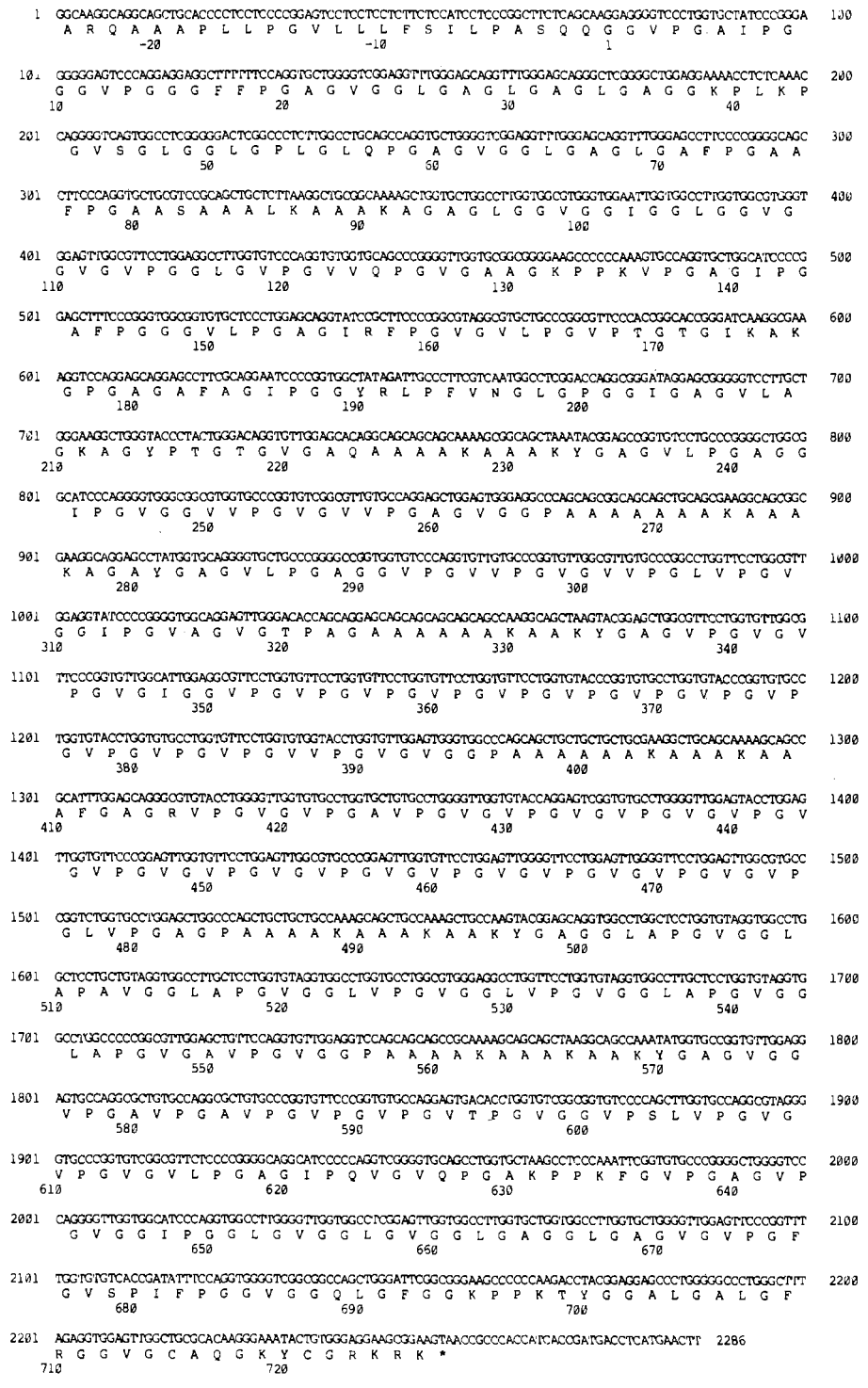

FIGURE 2: Nucleotide and deduced amino acid sequence of tropoelastin. Amino acids are numbered from the start of the mature protein.

implying a third pair). Similarly, units 1, 2, 4, and 5 can be added to the previous cluster with the smallest σ level required at 2.8. Repeat 13 is by far the weakest, showing as its highest σ value only 1.0 with unit 4. Nonetheless, unit 13 does contain the marker sequences KPPK, Y, and GAL (Figure 4). Given the level of statistical significance, it is suggested that at least 8 and possibly 13 of the repeats have evolved from a common ancestral sequence.

The alignment of the repeat units clearly points to several characteristic regions within the repeats, which are depicted in Figure 5 and which probably have functional significance. N-Terminal most is the lysine-containing, cross-link region, which ends with an aromatic amino acid. This is followed by the GAG segment (the hinge region) and then the C-terminal hydrophobic region with its HyPGHy (Hy = hydrophobic residue) repeats, which are generally separated by one to three residues of which one is glycine. The partial exon characterization of bovine tropoelastin, which has been recently reported (Cicila et al., 1985), supports the repeating model proposed here. The cross-link segments comprise one type of exon, while the hinge plus hydrophobic segments are together contained in a separate exon. The only pecularity exists for the weakest repeat, 13, where the homologous bovine sequence shows two exons, the first containing the KPPK span, a hinge region, and part of a hydrophobic segment and the second containing the cysteine pair and a short C-terminal basic span.

Comparison with Other Species. Available partial sequences of tropoelastin [summarized in Sandberg and Davidson (1984)

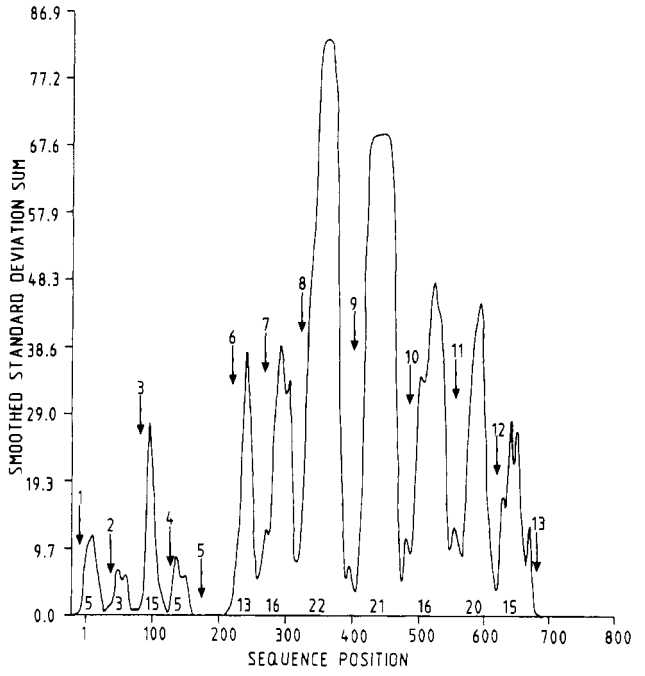

FIGURE 3: Plot of the tropoelastin sequence position vs. the sum of the standard deviation (σ) fractions in the self-comparison search matrix based on a search probe of 20 residues. The numbers at the bottom of each peak refer to the number of search matrix values greater than or equal to 3.0 σ found in determining the sum at the sequence position corresponding to the maximum value of a given peak. The numbered arrows refer to the sequence start positions used for each of the correspondingly numbered repeats in the alignments of Figure 4.

Table I: Comparison of the 13 Tropoelastin Sequence Repeats[a]

	1	2	3	4	5	6	7	8	9	10	11	12	13
1		2.8	0.5	1.3	-0.5	0.8	0.3	1.5	1.0	1.8	0.8	1.8	-0.8
2	20		2.3	2.0	1.3	0.8	1.8	2.5	2.0	3.8	1.8	2.3	0.5
3	11	18		1.5	1.5	3.5	5.0	6.3	5.5	6.0	4.0	1.5	-0.3
4	14	17	15		2.3	2.5	2.3	3.5	2.8	3.3	2.8	1.8	1.0
5	7	14	15	18		0.8	1.8	2.8	2.0	2.5	1.8	1.8	-0.3
6	12	12	23	19	12		6.3	4.8	5.3	5.5	4.5	2.3	0.0
7	10	16	29	18	16	34		7.8	8.8	7.5	7.3	3.5	-0.3
8	15	19	34	23	20	28	40		12.3	9.8	10.5	5.5	0.5
9	13	17	31	20	17	30	44	58		9.3	11.5	7.5	-0.3
10	16	24	33	22	19	31	39	48	46		8.5	5.0	0.5
11	12	16	25	20	16	27	38	51	55	43		4.5	0.0
12	16	18	15	16	16	18	23	31	39	29	27		0.3
13	6	11	8	13	8	9	8	11	8	11	9	10	

[a] The lower left portion of the matrix lists the number of amino acid identities in the alignments (Figure 4) of the respective repeats, while the upper right displays the number of standard deviations above or below the mean number of amino acid identities in the control test. Values over 6 σ are underlined, showing the strong relationship among units 3 and 6–12.

for pig; Yoon et al., 1985; Cicila et al., 1985] allow an extensive comparison between the avian and the mammalian molecules. However, alignment of most sequences is difficult, due to the lack of knowledge of their order. Several tryptic peptides from pig can be aligned at the amino-terminal end of the chick sequence without ambiguity. Homology at the carboxy-terminal end is comprised of about 50 amino acids, which includes the very basic terminal stretch RKRK and the two cysteines of chick tropoelastin. A conserved region in the middle of the protein includes the hydrophobic domain containing the sequence $(VPGVG)_{10}$.

A common feature of tropoelastin from different species is the presence of several tandem copies of short sequences. The repeated short sequences in chick and mammals are only partially identical. Common repeats are the polytripeptide $(GGX)_5$ (amino acids 97–111) and the polypentapeptides $(VPGVG)_{10}$ (amino acids 425–474) and $(GGLGV/A)_n$, where n is 4 in chick (amino acids 650–669) and 2 in pig. Repeated sequences characteristic of chick are the polytripeptide $(VPG)_{12}$ (amino acids 351–386) and the polyheptapeptide $(GGLV/A\ PGV/A)_7$ (amino acids 500–548). On the other hand, the polyhexapeptide $(PGVGVA)_4$ and the polynonapeptide $(AGVPGFGVG)_3$ have been detected only in pig.

A search of the entire National Biomedical Research Foundation Protein Sequence Database (Barker et al., 1986) for sequences related to the chick tropoelastin was performed by using the Lipman–Pearson algorithm (Lipman & Pearson, 1985). Although statistically significant homologies were detected, especially with collagenous proteins, they were rejected because none of the proteins displayed the sequence pattern as found in the overall organization of the unit repeat in tropoelastin.

Structure–Function Relationship in Tropoelastin. Functional aspects of tropoelastin include cross-linking and elastic behavior (Gray et al., 1973), chemotaxis (Senior et al., 1984), and possible assembly into fibers (Bressan et al., 1986), binding to other proteins as lysyl oxidase and microfibrils (Ross & Bornstein, 1969), and cell binding (Netland & Zetter, 1986; Hornebeck et al., 1986).

Domains participating in cross-link formation have been studied more thoroughly and are better defined in mammals (Foster et al., 1974; Juricova et al., 1975; Gerber & Anwar, 1975). The results show that the alanine- and lysine-rich regions are cross-link sites where lysine residues forming desmosines are found in pairs, separated by two or three alanines. The secondary structure predicted for these regions is α-helix. Since the lysines would lie on the same side of such a structure, kinetic arguments suggested that the first step in the formation of desmosine is the condensation of several adjacent lysines into lysinonorleucine or allysine aldol, followed by the condensation of lysinonorleucine and allysine aldol of different molecules into desmosine (Gray, 1977). A corollary of this biosynthetic pathway is that one desmosine, which is a tetrafunctional cross-link, can bind together a maximum of two molecules.

Chick tropoelastin contains several alanine-rich regions with exactly two lysine residues separated by two or three alanines (Figures 2 and 4). However, a group of three lysines can also be found in the sequence AAAAKAAAKAAK (amino acids 485–496 and 558–569). These lysines would be on the same side of an α-helix and could therefore condense into merodesmosine as an intermediate step in desmosine biosynthesis, a pathway suggested by Francis et al. (1973). In this case desmosine would also cross-link two molecules. One other lysine residue (lysine-211) is not an element of a couple, the nearest lysine being 17 amino acids downstream. Studies on cross-link-containing peptides have shown that a lysine followed by a sequence identical with the amino acids that follow lysine-211 contributes to desmosine in mammals (Gerber & Anwar, 1975). Finally, lysine-176 is expected to reside on the opposite site of the second lysine of the couple in an α-helical conformation of the polypeptide chain, since the two residues are separated by only one alanine. These two last observations not only emphasize the hypothesis that the formation of some desmosines procedes through the condensation of three lysine residues into merodesmosine but also open the possibility that some desmosines cross-link more than two molecules.

Four of the lysine couples are separated by prolines in chick. These sites occupy a position correspondent to the KAAK

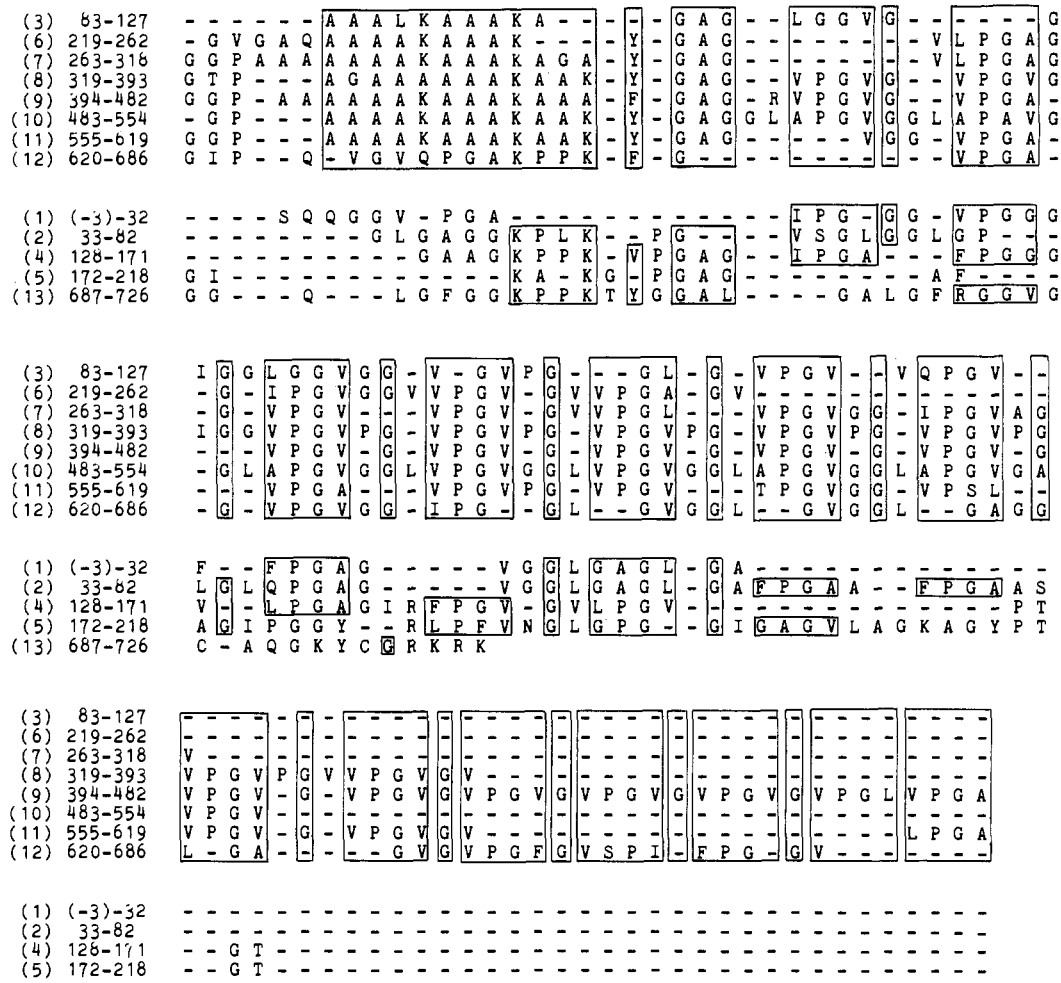

FIGURE 4: Alignment of the 13 repeat structures proposed for chick tropoelastin. The numbers in parentheses at the beginning of each line refer to the repeat number (arrowed numbers in Figure 3). The numbers that follow refer to the inclusive sequence positions for each repeating unit. The repeats are clustered according to the strongly related repeats (units 3 and 6–12) and then the remaining units. The residues are boxed in the first eight units to emphasize the domains and repeating patterns in each unit. Amino acids are boxed in units 1, 2, 4, 5, and 13 if at least two residues within a boxed span match any of the corresponding boxed residues in the eight clustered repeats above them or if boxed single glycines match.

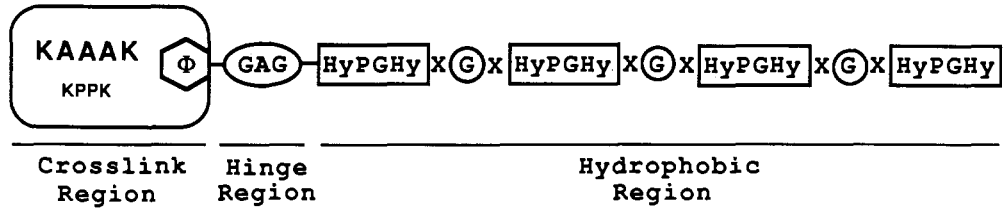

FIGURE 5: Common structural features of the repeat unit of chick tropoelastin. N-Terminal most is the lysine-containing, cross-link region, which ends with an aromatic amino acid (Φ). This is followed by the span GAG (the hinge region) and a hydrophobic region, which includes the repeating HyPGHy-XGX spans. Hy indicates a hydrophobic residue, most frequently valine, while X varies among glycine, proline, valine, leucine, and isoleucine.

regions in the basic repeat of tropoelastin and share sequence homologies with them, like the presence of an aromatic amino acid and the group GAG following the lysines (Figure 4). Moreover, the two types of sequences are likely to be evolutionally related, as the second lysine-containing region from the carboxy-terminal end is AKAAKF in sheep and bovine and AKPPKF in chick. The two sequences are homologous since the amino acids surrounding them are also similar. These observations favor the suggestion by Yoon et al. (1985) that KPPK sequences are potential cross-linking sites, although an α-helical conformation is unlikely.

The cross-link sequence ends with an aromatic group followed by GAG. It is speculated that the aromatic residue may affect the lysine modification (Foster et al., 1974) and that the GAG hinge region may act as a flexible link to the hydrophobic portion of a repeating unit.

Direct information on the function of the hydrophobic region of tropoelastin units is limited to the polyhexapeptide $(PGVGVA)_4$ and the polypentapeptide $(VPGVG)_{10}$ sequences. The first has been shown to be chemotactic for fibroblasts and macrophages (Senior et al., 1984) and is found in pig but not in chick. Therefore, either a chemotactic function is not present in chick or a different sequence performs it. As to the second repeating structure, a synthetic polypeptide $(VPGVG)_n$ mimics some of the temperature-dependent aggregation properties of tropoelastin in vitro (Volpin et al., 1976; Bressan et al., 1983), which have been suggested to control proper alignment of the molecules during fiber formation (Urry, 1976; Bressan et al., 1986). Furthermore, a synthetic polypeptide $(VPGVG)_n$, in which cross-links have been introduced between

chains, shows elastic properties similar to those of elastic fibers (Urry et al., 1976). This is certainly not the only domain with such characteristics, since most of the polypeptide chains of elastin show high mobility (Torchia & Piez, 1973), a property required for the entropic-type of elasticity observed in elastin (Hoeve & Flory, 1974). Examination of the alignments of Figure 4 shows the elemental repeating unit of the chick hydrophobic segment to be HyPGHy, where Hy is mostly valine, but can also be alanine, leucine, and isoleucine. Urry and his colleagues (Urry et al., 1976, 1981; Cook et al., 1980), through various biophysical studies as NMR and X-ray crystallography, have suggested a β-spiral architecture for the $(VPGVG)_n$ repeating peptide, where VPGV adopts a type II β-turn conformation and the connecting glycines allow the formation of the spiral. In chick tropoelastin, besides the variability of the HyPGHy structure in the first and fourth positions, the linker regions between the repeats vary in length from one to three residues, of which one is glycine and the others generally vary among proline, glycine, valine, and leucine. Alanine is rarely found in this hinge connection in the chick sequences; in fact, Urry and collaborators have found that substitution of an alanine in position 5 of the VPGVG repeats abolishes the elastic properties of the polypentapeptide (Urry et al., 1983).

Given the above-mentioned observations, we propose the following rough model for the chick tropoelastin repeat unit. Most cross-link regions would adopt an α-helical conformation, with lysines exposed on the same side of the helix and available for cross-linking where appropriate, as suggested by Gray et al. (1973). The aromatic residue may exist for purposes of structural stability and packing in a hydrophobic environment and/or for protecting the modification of nearby lysines (Foster et al., 1974). The hinge region, with its two glycines, would provide a flexible link to the hydrophobic region. The hydrophobic sequence, with its repeating, relatively rigid β-turns, can be viewed as a series of beads connected by spans of variable length that allow at least some of the elastic properties of the fibrous molecule, presumably through the use of the structurally flexible glycine.

REFERENCES

Altruda, F., Poli, V., Restagno, G., Argos, P., Cortese, R., & Silengo, L. (1985) *Nucleic Acids Res. 13*, 3841–3859.

Argos, P. (1985) *EMBO J. 4*, 1351–1355.

Argos, P. (1986) *J. Mol. Biol.* (in press).

Barker, W. C., Hunt, L. T., George, D. G., Yeh, L. S., Chen, H. R., Blomquist, M. C., Seibel-Ross, E. I., Hong, M. K., Bair, J. K., Chen, S. L., & Ledley, R. S. (1986) Protein Identification Resource, National Biomedical Research Foundation, Washington, DC.

Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Bressan, G. M., Castellani, I., Giro, M. G., Volpin, D., Fornieri, C., & Pasquali-Ronchetti, I. (1983) *J. Ultrastruct. Res. 82*, 335–340.

Bressan, G. M., Pasquali-Ronchetti, I., Fornieri, C., Mattioli, F., Castellani, I., & Volpin, D. (1986) *J. Ultrastruct. Res.* (in press).

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5298.

Cicila, G., May, M., Ornstein-Goldstein, N., Indik, Z., Morrow, S., Yeh, H. S., Rosenbloom, J., Boyd, C., Rosenbloom, J., & Yoon, K. (1985) *Biochemistry 24*, 3075–3080.

Cook, W. J., Einspahr, H., Trapane, T. L., Urry, D. W., & Bugg, C. E. (1980) *J. Am. Chem. Soc. 102*, 5502–5505.

Deininger, P. L. (1983) *Anal. Biochem. 129*, 216–223.

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

Foster, J. A., Bruenger, E., Gray, W. R., & Sandberg, L. B. (1973) *J. Biol. Chem. 248*, 2876–2879.

Foster, J. A., Rubin, L., Kagan, H. M., Franzblau, C., Bruenger, E., & Sandberg, L. B. (1974) *J. Biol. Chem. 249*, 6191–6199.

Foster, J. A., Shapiro, R., Voynow, P., Crombie, G., Faris, B., & Franzblau, C. (1975) *Biochemistry 14*, 5343–5347.

Francis, G., John, R., & Thomas, J. (1973) *Biochem. J. 136*, 45–55.

Franzblau, C. (1971) *Compr. Biochem. 26c*, 659–712.

Gallop, M. P., Blumenfeld, O. O., & Seifter, S. (1972) *Annu. Rev. Biochem. 41*, 617–672.

Gerber, G. E., & Anwar, R. A. (1975) *Biochem. J. 149*, 685–695.

Gray, W. R. (1977) *Adv. Exp. Med. Biol. 79*, 285–290.

Gray, W. R., Sandberg, L. B., & Foster, J. A. (1973) *Nature (London) 246*, 461–466.

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

Hall, R., Hyde, J. E., Goman, M., Simmons, D. L., Hope, I. A., Mackay, M., & Scaife, J. (1984) *Nature (London) 311*, 379–382.

Haymerle, H., Herz, J., Bressan, G. M., Frank, R., & Stanley, K. K. (1986) *Nucleic Acids Res. 14*, 8615–8624.

Hoeve, C. A. J., & Flory, P. J. (1974) *Biopolymers 13*, 667–686.

Hornebeck, W., Tixier, J. M., & Robert, L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5517–5520.

Juricova, M., Franzblau, C., Faris, B., Deyl, Z., & Adam, M. (1975) *Biochim. Biophys. Acta 386*, 239–243.

Kagan, H. M. (1986) in *Biology of the Extracellular Matrix* (Mecham, R. P., Ed.) Vol. 1, pp 321–398, Academic, Orlando, FL.

Karr, S. R., & Foster, J. A. (1981) *J. Biol. Chem. 256*, 5946–5949.

Lipman, D. J., & Pearson, W. R. (1985) *Science (Washington, D.C.) 227*, 1435–1441.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Netland, P. A., & Zetter, B. R. (1986) *Biochem. Biophys. Res. Commun. 139*, 515–522.

Partridge, S. M. (1962) *Adv. Protein Chem. 17*, 227–297.

Ross, R., & Bornstein, P. (1969) *J. Cell Biol. 40*, 366–381.

Rucker, R. B., Tom, K., Tanaka, M., Haniu, M., & Yasunobu, K. T. (1975) *Biochem. Biophys. Res. Commun. 66*, 287–292.

Sandberg, L. B. (1976) *Int. Rev. Connect. Tissue Res. 7*, 159–210.

Sandberg, L. B., & Davidson, J. M. (1984) *Pept. Protein Rev. 3*, 168–226.

Sandberg, L. B., Weissamn, N., & Gray, W. R. (1971) *Biochemistry 10*, 52–56.

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

Senior, R. M., Griffin, G. L., Mecham, R. P., Wrenn, D. S., Prasad, K. V., & Urry, D. W. (1984) *J. Cell Biol. 99*, 870–874.

Smith, D. W., Sandberg, L. B., Leslie, B. H., Wolt, T. B., Minton, S. T., Myers, B., & Rucker, R. B., (1981) *Biochem. Biophys. Res. Commun. 103*, 880–885.

Stanley, K. K. (1983) *Nucleic Acids Res. 11*, 4077–4092.

Stanley, K. K., & Luzio, P. P. (1984) *EMBO J. 3*, 1429–1434.

Torchia, D. A., & Piez, K. A. (1973) *J. Mol. Biol. 76*, 419–424.
Urry, D. W. (1976) *Faraday Discuss. Chem. Soc. 61*, 205–216.
Urry, D. W., Okamoto, K., Harris, R. D., Hendrix, C. F., & Long, M. M. (1976) *Biochemistry 15*, 4083–4089.
Urry, D. W., Trapane, T. L., Sugano, H., & Prasad, K. U. (1981) *J. Am. Chem. Soc. 103*, 2080–2089.
Urry, D. W., Trapane, T. L., Long, M. M., & Prasad, K. V. (1983) *J. Chem. Soc., Faraday Trans. 1 79*, 853–868.
Volpin, D., Pasquali-Ronchetti, I., Urry, D. W., & Gotte, L. (1976) *J. Biol. Chem. 251*, 6871–6873.
Yoon, K., Davidson, J. M., Boyd, C. May, M., LuValle, P., Ornstein-Goldstein, N., Smith, J., Indik, Z., Ross, A., Golub, E., & Rosenbloom, J. (1985) *Arch. Biochem. Biophys. 241*, 684–691.

Articles

Crystal Structure of Cytochrome *c* Peroxidase Compound I[†]

Steven L. Edwards,[*,‡] Nguyen huu Xuong,[‡,§,∥] Ronald C. Hamlin,[§] and Joseph Kraut[‡]
Departments of Chemistry, Physics, and Biology, University of California, San Diego, La Jolla, California 92093
Received July 17, 1986; Revised Manuscript Received October 24, 1986

ABSTRACT: We have compared the 2.5-Å crystal structure of yeast cytochrome *c* peroxidase (CCP) with that of its semistable two-equivalent oxidized intermediate, compound I, by difference Fourier and least-squares refinement methods. Both structures were observed at −15 °C. The difference Fourier map reveals that formation of compound I causes only small positional adjustments of a few tenths of an angstrom. The map's most pronounced feature is a pair of positive and negative peaks bracketing the heme iron position. Least-squares refinement shows that the iron atom moves about 0.2 Å toward the distal side of the heme. No significant difference density is evident near the side chains of Trp-51 or Met-172, each of which has been proposed to be the site of the electron paramagnetic resonance (EPR) active radical in compound I. However, the second most prominent feature of difference density is a negative peak near the side chain of Thr-180, which, according to the results of least-squares refinement, moves by 0.15 Å in the direction of Met-230. These observations, together with the results of mutagenesis experiments [Fishel, L. A., Villafranca, J. E., Mauro, J. M., & Kraut, J. (1987) *Biochemistry 26*, 351–360; Goodin, D. B., Mauk, A. G., & Smith, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1295–1299] in which Trp-51 and Met-172 have been replaced without loss of the EPR radical signal in compound I, lead us to consider the possibility that the radical site lies within a cluster composed of the side chains of Met-230, Met-231, and Trp-191. This cluster is contiguous with Thr-180 and about 10 Å from the heme plane on the proximal side.

Compound I (also termed compound ES) is a semistable enzyme intermediate that results from the oxidation of cytochrome *c* peroxidase (ferrocytochrome *c*:H_2O_2 oxidoreductase, EC 1.11.1.5; CCP)[1] by its substrate, peroxide (Abrams et al., 1942). Various peroxides, including H_2O_2, react with CCP to remove two reducing equivalents from the enzyme to form compound I, which is then rereduced in two successive one-electron steps by ferrocytochrome *c*. The reaction scheme may be written (Yonetani & Ray, 1966)

$$\text{CCP} + \text{ROOH} \rightarrow \text{compound I} + \text{ROH}$$

$$\text{compound I} + \text{cyt-}c(\text{Fe}^{2+}) + \text{H}^+ \rightarrow \text{compound II} + \text{cyt-}c(\text{Fe}^{3+})$$

$$\text{compound II} + \text{cyt-}c(\text{Fe}^{2+}) + \text{H}^+ \rightarrow \text{CCP} + \text{cyt-}c(\text{Fe}^{3+}) + \text{H}_2\text{O}$$

where ROOH is an alkyl peroxide or hydrogen peroxide. Compound I is not an enzyme–substrate complex in the usual Michaelis–Menten sense, but rather, it is an intermediate species in which the first step of the enzyme-catalyzed reaction, reductive cleavage of the peroxide bond, has already occurred.

The importance of CCP as an object of study derives in part from certain functional similarities to respiratory cytochrome oxidase. The latter is a heme-containing multisubunit complex that catalyzes the reduction of molecular oxygen to water by ferrocytochrome *c* as the ultimate step in the electron-transport chain. CCP similarly contains heme as a prosthetic group, catalyzes the reduction of an oxygen–oxygen bond by ferrocytochrome *c*, and has been shown to substitute as a terminal oxidase in yeast when the respiratory oxidase is inhibited (Erecinska et al., 1973). Thus, as a simpler model for the more complicated and experimentally elusive cytochrome oxidase, CCP has become the focus of much study and speculation.

The initial crystal structure of CCP was reported by our laboratory in 1980 (Poulos et al., 1980), and the 1.7-Å refinement in 1984 (Finzel et al., 1984). A structure-based mechanism suggested by Poulos and Kraut (1980a,b) incor-

[†] This work was supported by National Science Foundation Grant DMB-8511656 awarded to J.K. Ng.h.X. acknowledges support from National Institutes of Health Grant RR 01644 for the UCSD Data Collection Resource for Protein Crystallography in addition to a special grant from the NSF Biological Instrument Program (PCM-8400547) and a grant for low-temperature experimentation (DMB-8510860).
[‡] Department of Chemistry.
[§] Department of Physics.
[∥] Department of Biology.

[1] Abbreviations: CCP, cytochrome *c* peroxidase; ENDOR, electron nuclear double resonance; EPR, electron paramagnetic resonance; EXAFS, extended X-ray absorption fine structure; HRP, horseradish peroxidase.

0006-2960/87/0426-1503$01.50/0