P. Jennings, and B. Posner for purifying some of the proteins used in this study. The support of the Penn State NMR facility is also gratefully acknowledged.

References

Baccanari, D. P., Averett, D., Briggs, C., & Burchall, J. (1977) *Biochemistry 16*, 3566–3572.
Benkovic, S. J., Fierke, C. A., & Naylor, A. M. (1988) *Science (Washington, D.C.) 239*, 1105–1110.
Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.
Birdsall, B., De Graw, J., Feeney, J., Hammond, S., Searle, M. S., Roberts, G. C. K., Colwell, W. T., & Crase, J. (1987) *FEBS Lett. 217*, 106–110.
Birdsall, B., Feeney, J., Tendler, S. J. B., Hammond, S. J., & Roberts, G. C. K. (1989) *Biochemistry 28*, 2297–2305.
Bodenhausen, G., Vold, R. L., & Vold, R. R. (1980) *J. Magn. Reson. 37*, 93–106.
Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 370–388.
Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., & Kraut, J. (1982) *J. Biol. Chem. 257*, 13650–13662.
Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D., & Jeanloz, R. W. (1984) *J. Am. Chem. Soc. 106*, 811–813.
Braunschweiler, L., Bodenhausen, G., & Ernst, R. R. (1983) *Mol. Phys. 48*, 535–560.
Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285–298.
Charlton, P. A., Young, D. W., Birdsall, B., Feeney, J., & Roberts, G. C. K. (1979) *J. Chem. Soc., Chem. Commun.*, 922–924.
Chen, J.-T., Mayer, R. J., Fierke, C. A., & Benkovic, S. J. (1985) *J. Cell. Biochem. 29*, 73–82.
Chen, J.-T., Taira, K., Tu, C.-P. D., & Benkovic, S. J. (1987) *Biochemistry 26*, 4093–4100.
Dalvit, C., & Wright, P. E. (1987) *J. Mol. Biol. 194*, 313–327.
Davis, D. G. (1989) *J. Magn. Reson. 81*, 603–607.
Fierke, C. A., & Benkovic, S. J. (1989) *Biochemistry 28*, 478–486.
Fierke, C. A., Johnson, K. A., & Benkovic, S. J. (1987) *Biochemistry 26*, 4085–4092.
Kramer, W., Drutsa, V., Jansen, H.-W., Kramer, B., Pflugfelder, M., & Fritz, H.-J. (1984) *Nucleic Acids Res. 12*, 9441–9456.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Murphy, D. J., & Benkovic, S. J. (1989) *Biochemistry 28*, 3025–3031.
Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187–193.
Perry, K. M., Onuffer, J. J., Touchette, N. A., Herndon, C. S., Gittelman, M. S., Matthews, C. R., Chen, J.-T., Mayer, R. J., Taira, K., Benkovic, S. J., Howell, E. E., & Kraut, J. (1987) *Biochemistry 26*, 2674–2682.
Rance, M. (1987) *J. Magn. Reson. 74*, 557–564.
Rance, M., & Wright, P. E. (1986) *J. Magn. Reson. 66*, 372–378.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Selinsky, B. S., Perlman, M. E., London, R. E., Unkefer, C. J., Mitchell, J., & Blakley, R. L. (1990) *Biochemistry 29*, 1290–1296.

Heterogeneity of Rat Tropoelastin mRNA Revealed by cDNA Cloning[†,‡]

Richard A. Pierce,[§] Susan B. Deak,[§] Catherine A. Stolle,[∥] and Charles D. Boyd*[,§]
Departments of Surgery and Medicine, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, New Brunswick, New Jersey 08903-0019

Received March 30, 1990; Revised Manuscript Received July 23, 1990

ABSTRACT: A λgt11 library constructed from poly(A+) RNA isolated from aortic tissue of neonatal rats was screened for rat tropoelastin cDNAs. The first screen, utilizing a human tropoelastin cDNA clone, provided rat tropoelastin cDNAs spanning 2.3 kb of carboxy-terminal coding sequence and extended into the 3′-untranslated region. A subsequent screen using a 5′ rat tropoelastin cDNA clone yielded clones extending into the amino-terminal signal sequence coding region. Sequence analysis of these clones has provided the complete derived amino acid sequence of rat tropoelastin and allowed alignment and comparison with published bovine cDNA sequence. While the overall structure of rat tropoelastin is similar to bovine sequence, numerous substitutions, deletions, and insertions demonstrated considerable heterogeneity between species. In particular, the pentapeptide repeat VPGVG, characteristic of all tropoelastins analyzed to date, is replaced in rat tropoelastin by a repeating pentapeptide, IPGVG. The hexapeptide repeat VGVAPG, the bovine elastin receptor binding peptide, is not encoded by rat tropoelastin cDNAs. Variations in coding sequence between rat tropoelastin cDNA clones were also found which may represent mRNA heterogeneity produced by alternative splicing of the rat tropoelastin pre-mRNA.

Elastin is the major connective tissue protein that confers elasticity to vertebrate elastic tissues (Partridge, 1962; Rosenbloom, 1984). Tropoelastin is the soluble secreted precursor of elastin. In the extracellular space, tropoelastin is extensively cross-linked by the copper-requiring enzyme lysyl oxidase, resulting in an insoluble elastic fiber (Partridge, 1962). The amino acid sequence of tropoelastin was first obtained from tryptic peptides derived from the aortic tissue of copper-deficient pigs. Extensive sequencing of these fragments revealed glycine-rich hydrophobic domains and alanine-rich cross-

[†] This work was supported by NIH Grant HL-42798.
[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05292.
* Address correspondence to this author.
[§] Department of Surgery.
[∥] Department of Medicine.

0006-2960/90/0429-9677$02.50/0

linking domains containing lysine residues (Sandberg & Davidson, 1984). Earlier work by several investigators suggested the presence of multiple isoforms of this multidomain tropoelastin (Foster et al., 1980; Davidson et al., 1982a; Chipman et al., 1985). More recently, by use of cDNA sequencing, considerable heterogeneity in the primary sequence of tropoelastins was confirmed in several phylogenetic species, including bovine, chick, and man (Raju & Anwar, 1987a; Indik et al., 1987; Baule & Foster, 1988). It is well established that tropoelastin is synthesized from a single copy gene (Olliver et al., 1987), and several investigators have demonstrated that this sequence heterogeneity arises as a consequence of alternative splicing of the primary transcript of the tropoelastin gene.

While previous work had not demonstrated alternative splicing in rat tropoelastin, several differences between rat tropoelastin and other mammalian tropoelastins have been suggested. Rat tropoelastin has been reported, for example, to be 75000–79000 in molecular weight (Chipman et al., 1985; Franzblau et al., 1989), which is larger than reported sizes of bovine tropoelastin (Davidson et al., 1982a; Wrenn et al., 1986). Preliminary evidence has also proposed that rat tropoelastin, in contrast to all other tropoelastins analyzed to date, is a substrate for glycosylation (Mogayzel et al., 1987). In addition, studies have suggested that rat tropoelastin is synthesized in a precursor form of higher molecular weight (Chipman et al., 1985). Franzblau et al. (1989) have shown that discrete elastin peptides isolated from neonatal rat smooth muscle cells contain identical NH_2-terminal amino acid sequences and have suggested that these peptides were generated by proteolytic cleavage of the COOH terminus of rat tropoelastin.

Therefore, to characterize possible rat-specific differences in tropoelastin synthesis, cDNAs coding for rat tropoelastin were isolated from a neonatal rat aorta cDNA library. This paper describes the characterization of overlapping cDNA clones that contain DNA sequence coding for the complete rat tropoelastin sequence. While alignment with bovine tropoelastin sequence demonstrates conserved primary structure, we found numerous differences. The elastin receptor binding hexapeptide repeating element VGVAPG, found in bovine and human tropoelastins, is not encoded by these cDNAs, nor is the pentapeptide repeat VPGVG, found in bovine, human, and chick tropoelastins and used for structure function analyses of tropoelastin. The results obtained from this analysis indicate that rat tropoelastin is a higher molecular weight protein than the bovine equivalent. However, no evidence was found to suggest that rat tropoelastin is a substrate for N-linked glycosylation. In addition, while considerable sequence variability was evident in the rat tropoelastin cDNAs characterized, no evidence was found for a rat-specific protropoelastin sequence. The sequence heterogeneity characterized in these rat tropoelastin cDNAs is very likely due to alternative splicing of the primary transcript of the rat tropoelastin gene.

Materials and Methods

The isolation and characterization of overlapping cDNA clones encompassing 252 bp of translated sequence and 971 bp of the 3′-untranslated sequence of rat tropoelastin have been previously described (Deak et al., 1988). In this study, poly(A+) RNA was isolated from pooled neonatal rat aortas and used to construct a λgt11 cDNA library. Recombinants were selected by screening with the human tropoelastin cDNA clone pcHEL-2 (Indik et al., 1987). Subsequently, the library was rescreened by using a 240-bp 5′ rat elastin cDNA fragment derived from the clone REL 124. After preparative isolation of positive clones, the *Eco*RI cDNA inserts were subcloned into the M13 vector, mp18.

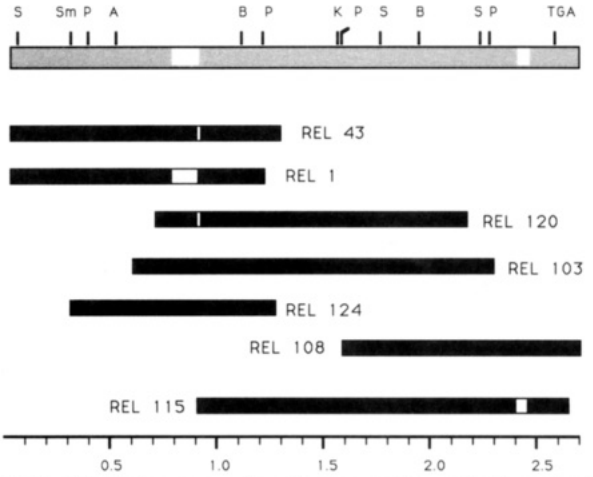


FIGURE 1: Restriction map and overlapping positions of rat tropoelastin cDNAs. The lightly shaded bar represents a composite restriction map. Individual cDNA clones from which both this information and DNA sequence were obtained are indicated as solid bars. Open boxes are regions of DNA sequence that were missing in the indicated clones. The scale indicates size of DNA fragments in kilobases. The symbols are as follows: A, restriction enzyme *Acc*I; B, *Bam*HI; K, *Kpn*I; P, *Pst*I; S, *Sau*3A; Sm, *Sma*I; TGA, termination codon.

DNA sequencing was carried out by using the dideoxy chain termination procedure (Sanger et al., 1980), using a modified T7 DNA polymerase (Sequenase, U.S. Biochemicals, Cleveland, OH) and [^{35}S]dATPγS. Single-stranded DNA was primed with universal primer or, alternatively, with rat elastin specific oligomers. The oligomers were synthesized with an automated Model 381A Applied Biosystems DNA synthesizer using β-cyanoethyl phosphoramadite reagents (Matteucci & Carruthers, 1981); oligomers were finally purified by preparative polyacrylamide gel electrophoreses and ethanol precipitation. Computer analysis of DNA sequence was carried out by using an IBM AT computer and software provided by International Biotechnologies, Inc. (IBI).

Results

From the initial rat aorta cDNA library screening, using a 1.2-kb *Bam*HI/*Hin*dIII fragment of pcHEL-2 (Indik et al., 1987), a number of positive clones were selected for further analysis. These clones range from 0.9 to 1.7 kb in size and include REL 103, REL 108, REL 115, REL 120, and REL 124 (Figure 1). These overlapping cDNAs encompass 2285 bp of coding sequence and extend into the 3′-untranslated portion of the cDNA. They contain a single open reading frame which codes for rat tropoelastin and is contiguous with previously published rat tropoelastin sequences (Deak et al., 1988). A 241-bp *Hpa*II fragment of the rat elastin cDNA REL-124 was subsequently used to screen the rat aorta λgt11 cDNA library. A number of positive clones were isolated including REL 1 and REL 43 (Figure 1). These clones overlap with the clones from the first screening and extend into the signal peptide sequence of rat tropoelastin. The clones isolated in the two screening procedures contain 2593 nucleotides of coding sequence and code for 864 amino acids, including 21 amino acids of the signal sequence (Figure 2). As these cDNAs correspond to sequences from all the bovine tropoelastin exons described to date (Yeh et al., 1989; Indik et al., 1990), it is likely they represent the entire amino acid sequence of secreted tropoelastin and most of the signal peptide.

Rat Tropoelastin cDNAs Contain Typical Tropoelastin Domains. The derived amino acid sequence of these clones

contains the expected tropoelastin domains. As previously described (Deak et al., 1988), rat tropoelastin is strongly homologous to other tropoelastins in the carboxy-terminal portion of the molecule. In sequences corresponding to exon 36 of the bovine gene, two cysteine residues separated by four amino acids are followed by the characteristic basic, hydrophilic sequence GRKRK found in bovine (Raju & Anwar, 1987a), human (Indik et al., 1987), and chick (Bressan et al., 1987) tropoelastin cDNAs. As we described previously (Deak et al., 1988) rat tropoelastin cDNA contained sequences coding for exons 34 and 35, which are absent in the human gene. The remainder of the coding sequence consists of signal peptide sequences followed by alternating glycine-rich hydrophobic domains and alanine-rich domains containing lysine residues. Thus, the overall structure is similar to bovine and other tropoelastin cDNA sequences.

Rat Tropoelastin cDNAs Demonstrate Further Variation between Species. Previous papers have noted variation between tropoelastin cDNAs isolated from bovine, human, and chick cDNA libraries (Raju & Anwar, 1987b). The composite sequence reported here demonstrated further variability of tropoelastin between species. While alignment of the majority of the sequence with the bovine tropoelastin sequence is unequivocal, numerous conservative amino acid substitutions, deletions, and insertions in the rat tropoelastin cDNA occur. In sequences corresponding to bovine exons 2, 7, 20, 22, 24, 28, and 30, the rat sequence contains apparent insertions relative to the bovine sequence. The coding region corresponding to bovine exon 24 in rat tropoelastin contains seven more amino acids than the corresponding exon in bovine tropoelastin, while exons 2 and 20 contain apparent insertions of eight amino acids each. Relatively large insertions of 13, 21, and 23 amino acids occur in regions corresponding to bovine exons 28, 22, and 30, respectively (Figure 2). All these insertions occur in hydrophobic domains and are often part of repeating peptides in these exons.

Cross-linking domains of rat tropoelastin are strongly conserved when compared to bovine tropoelastin. Many of these domains consist of lysine residues separated by alanines and followed by an aromatic amino acid. Typical is the sequence encoded by nucleotides 1256–1309 in Figure 2. This sequence, GAVSPAAAKAAAKAAKYG, is nearly identical with bovine exon 19 derived amino acid sequence with the tyrosine in rat substituted for phenylalanine in bovine tropoelastin. This pattern continues in the COOH-terminal half of the molecule, with tyrosine residues commonly replacing phenylalanine following potential cross-linking regions.

Repeating Nature of Rat Tropoelastin. As noted in chick tropoelastin cDNAs (Bressan et al., 1987), the hydrophobic domains of tropoelastin contain numerous peptide repeats. In the rat tropoelastin cDNAs, repetitive amino acid sequences are found in regions corresponding to bovine exons 2, 7, 18, 20, 22, 24, 26, 28, and 30 (Table I). The pentapeptide repeat VPGVG, repeated many times in bovine exon 18, has been used in structure–function relationship studies of elastin (Urry & Long, 1976). In rat tropoelastin cDNAs the closely related sequence IPGVG is repeated five times in the corresponding area of the molecule. The tropoelastin hexapeptide VGVAPG, encoded by exon 24 of the human and bovine tropoelastin genes and shown to bind to the bovine elastin receptor (Mecham et al., 1989), is not encoded by these rat tropoelastin cDNAs. Instead, we find related hydrophobic sequences VGGVPG, GGVGPG, and IGTGPG in this region. Other, similar sequences, including VGVLPG, are encoded by other regions of these cDNAs. Interestingly, sequence corresponding to exon 30 contains the amino acid sequence GAGGL repeated six times. This same repeat occurs only five times in a rat elastin cDNA sequence previously reported (Baule & Foster, 1988). This may represent allelic variation in the rat tropoelastin gene.

Table I: Peptide Repeats in the Hydrophobic Domains of Rat Tropoelastin Sequence Derived from cDNA Clones

exon	repeating peptide[a]	exon	repeating peptide[a]
2	PGGV (3×)	22	PGAVPGAL (3×)
7	GVGGVPGA (2×)	24	PGGV (3×)
7	GGIGG (2×)	26	GAGVPG (3×)
18	IPGVG (5×)	28	GGLGGP (3×)
20	YGVGAG (2×)	30	GAGGL (6×)

[a] The sequence of the derived peptide repeat is indicated together with the number of such repeats (in parentheses). The location of these rat tropoelastin peptide domains relative to exons within the bovine tropoelastin gene is indicated.

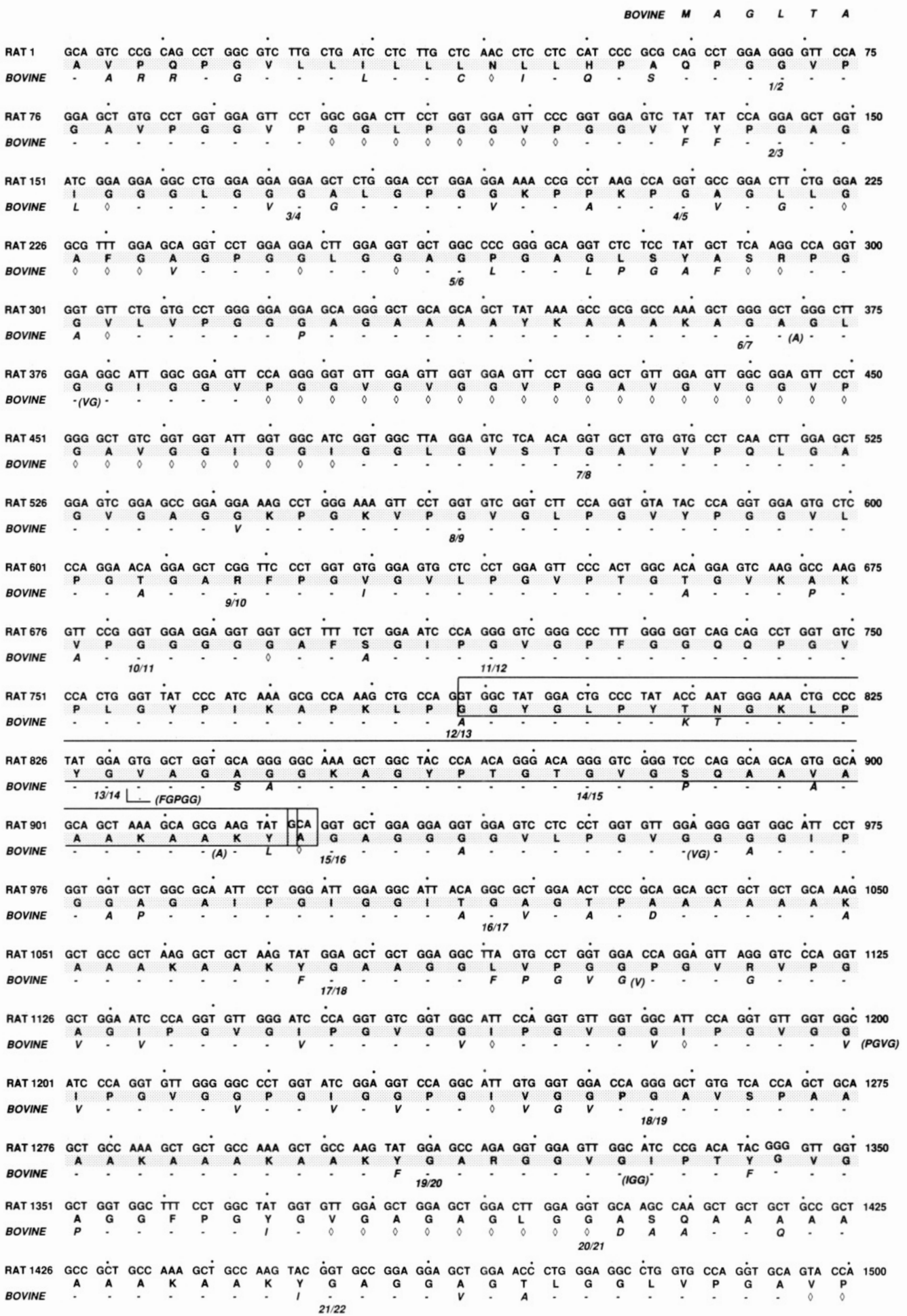
Heterogeneity in Rat Tropoelastin cDNA Sequence. One of the clones analyzed in the present study, REL 115, lacks nucleotides 2426–2491 while spanning this area of the composite cDNA sequence (Figure 2). This apparent deletion, which is present in clone REL 108, corresponds almost exactly to the bovine sequence of exon 33, with one amino acid residue, phenylalanine, replaced by tyrosine in the rat. This 15 amino acid hydrophobic exon has previously been noted to be alternatively spliced in bovine (Yeh et al., 1987) and human (Fazio et al., 1988) cDNAs in a cassette fashion, with the entire exon present or absent in a cDNA.

A more 5′ clone, REL 1, lacks nucleotides 788–924 (Figure 2). These sequences, coding for 45 amino acids, are represented in clones REL 43, REL 103, REL 120, and REL 124. The missing sequence codes for a hydrophobic, tyrosine-rich region and a cross-linking domain and corresponds to bovine tropoelastin exons 13, 14, and 15. Exons 13 and 14 have been shown to be alternatively spliced in bovine cDNAs (Yeh et al., 1987). This rat cDNA may be the first example, therefore, of three contiguous exons spliced alternatively in tropoelastin mRNA.

Another area of heterogeneity in the rat cDNAs analyzed is the variable presence of the alanine codon GCA, nucleotides 922–924 (Figure 2). This codon is present in the clones REL 103 and REL 124, and absent in the clones REL 120, REL 1, and REL 43. This heterogeneity is at the proposed boundary between exons 15 and 16 in the rat tropoelastin gene. Analysis of the 5′ end of exon 16 in the rat tropoelastin gene has revealed two adjacent splice acceptor sites separated by three nucleotides (unpublished data). The alternate usage of these acceptor sites would yield the alternate inclusion or exclusion of nucleotides 923–925 in the cDNA sequence and explain the variation observed.

Rat Tropoelastin Lacks an Asn-X-Ser/Thr Sequence. While it has previously been reported that rat tropoelastin is a substrate for glycosylation (Mogayzel et al., 1987), the present study has not revealed a potential N-linked glycosylation site. These rat tropoelastin cDNAs code for two asparagine residues, at nucleotides 40–42 in the signal peptide and at nucleotides 821–823 in exon 13 (Figure 2). These asparagine residues are not, however, part of the glycosylation sequence Asn-X-Ser/Thr.

Calculated Molecular Weight of Rat Tropoelastin. The molecular weight of rat tropoelastin isoforms estimated by SDS–PAGE analysis varies in the literature. Previous papers have demonstrated that rat tropoelastin isoforms are larger


BOVINE M A G L T A
RAT 1 GCA GTC CCG CAG CCT GGC GTC TTG CTG ATC CTC TTG CTC AAC CTC CTC CAT CCC GCG CAG CCT GGA GGG GTT CCA 75
A V P Q P G V L L I L L L N L L H P A Q P G G V P
BOVINE - A R R - G - - - L - - C ◊ I - Q - S - - - - - -
1/2
RAT 76 GGA GCT GTG CCT GGT GGA GTT CCT GGC GGA CTT CCT GGT GGA GTT CCC GGT GGA GTC TAT TAT CCA GGA GCT GGT 150
G A V P G G V P G G L P G G V P G G V Y Y P G A G
BOVINE - - - - - - - - ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ - - - F F - - - -
2/3
RAT 151 ATC GGA GGA GGC CTG GGA GGA GGA GCT CTG GGA CCT GGA GGA AAA CCG CCT AAG CCA GGT GCC GGA CTT CTG GGA 225
I G G G L G G G A L G P G G K P P K P G A G L L G
BOVINE L ◊ - - - - V - G - - - - V - - A - - - V - G - ◊
3/4 4/5
RAT 226 GCG TTT GGA GCA GGT CCT GGA GGA CTT GGA GGT GCT GGC CCC GGG GCA GGT CTC TCC TAT GCT TCA AGG CCA GGT 300
A F G A G P G G L G G A G P G A G L S Y A S R P G
BOVINE ◊ ◊ ◊ V - - - ◊ - - ◊ - - L - - L P G A F ◊ ◊ - -
5/6
RAT 301 GGT GTT CTG GTG CCT GGG GGA GGA GCA GGG GCT GCA GCA GCT TAT AAA GCC GCG GCC AAA GCT GGG GCT GGG CTT 375
G V L V P G G G A G A A A A Y K A A A K A G A G L
BOVINE A ◊ - - - - - P - - - - - - - - - - - - - - - (A) - -
6/7
RAT 376 GGA GGC ATT GGC GGA GTT CCA GGG GGT GTT GGA GTT GGT GGA GTT CCT GGG GCT GTT GGA GTT GGC GGA GTT CCT 450
G G I G G V P G G V G G V G G V P G A V G V G G V P
BOVINE -(VG)- - - - - ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊
RAT 451 GGG GCT GTC GGT GGT ATT GGT GGC ATC GGT GGC TTA GGA GTC TCA ACA GGT GCT GTG GTG CCT CAA CTT GGA GCT 525
G A V G G I G G I G G L G V S T G A V V P Q L G A
BOVINE ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ - - - - - - - - - - - - - - - -
7/8
RAT 526 GGA GTC GGA GCC GGA GGA AAG CCT GGG AAA GTT CCT GGT GTC GGT CTT CCA GGT GTA TAC CCA GGT GGA GTG CTC 600
G V G A G G K P G K V P G V G L P G V Y P G G V L
BOVINE - - - - - V - - - - - - - - - - - - - - - - - - -
8/9
RAT 601 CCA GGA ACA GGA GCT CGG TTC CCT GGT GTG GGA GTG CTC CCT GGA GTT CCC ACT GGC ACA GGA GTC AAG GCC AAG 675
P G T G A R F P G V G V L P G V P T G T G V K A K
BOVINE - - A - - - - - - I - - - - - - - - - A - - - P -
9/10
RAT 676 GTT CCG GGT GGA GGA GGT GGT GCT TTT TCT GGA ATC CCA GGG GTC GGG CCC TTT GGG GGT CAG CAG CCT GGT GTC 750
V P G G G G G A F S G I P G V G P F G G Q Q P G V
BOVINE A - - - - - ◊ - - A - - - - - - - - - - - - - - -
10/11 11/12
RAT 751 CCA CTG GGT TAT CCC ATC AAA GCG CCA AAG CTG CCA GGT GGC TAT GGA CTG CCC TAT ACC AAT GGG AAA CTG CCC 825
P L G Y P I K A P K L P G G Y G L P Y T N G K L P
BOVINE - - - - - - - - - - - A - - - - - - K T - - - - -
12/13
RAT 826 TAT GGA GTG GCT GGT GCA GGG GGC AAA GCT GGC TAC CCA ACA GGG ACA GGG GTC GGG TCC CAG GCA GCA GTG GCA 900
Y G V A G A G G K A G Y P T G T G V G S Q A A V A
BOVINE - - - - - S A - - - - - - - - - - - - - P - - - A -
13/14 (FGPGG) 14/15
RAT 901 GCA GCT AAA GCA GCG AAG TAT GCA GGT GCT GGA GGA GGT GGA GTC CTC CCT GGT GTT GGA GGG GGT GGC ATT CCT 975
A A K A A K Y A G A G G G G V L P G V G G G G I P
BOVINE - - - - -(A)- L ◊ - - - - - A - - - - - - -(VG)- A - - -
15/16
RAT 976 GGT GGT GCT GGC GCA ATT CCT GGG ATT GGA GGC ATT ACA GGC GCT GGA ACT CCC GCA GCA GCT GCT GCT GCA AAG 1050
G G A G A I P G I G G I T G A G T P A A A A A K
BOVINE - A P - - - - - - - - A - V - A - D - - - - - - A
16/17
RAT 1051 GCT GCC GCT AAG GCT GCT AAG TAT GGA GCT GCT GGA GGC TTA GTG CCT GGT GGA CCA GGA GTT AGG GTC CCA GGT 1125
A A A K A A K Y G A A G G L V P G G P G V R V P G
BOVINE - - - - - - - F - - - - - F P G V G (V) - - - G - - -
17/18
RAT 1126 GCT GGA ATC CCA GGT GTT GGG ATC CCA GGT GTC GGT GGC ATT CCA GGT GTT GGT GGC ATT CCA GGT GTT GGT GGC 1200
A G I P G V G I P G V G G I P G V G G I P G V G G
BOVINE V - V - - - - V - - - - V ◊ - - - - V ◊ - - - - V (PGVG)
RAT 1201 ATC CCA GGT GTT GGG GGC CCT GGT ATC GGA GGT CCA GGC ATT GTG GGT GGA CCA GGG GCT GTG TCA CCA GCT GCA 1275
I P G V G G P G I G G P G I V G G P G A V S P A A
BOVINE V - - - - V - - V - V - - ◊ V G V - - - - - - - -
18/19
RAT 1276 GCT GCC AAA GCT GCT GCC AAA GCT GCC AAG TAT GGA GCC AGA GGT GGA GTT GGC ATC CCG ACA TAC GGG GTT GGT 1350
A A K A A A K A A K Y G A R G G V G I P T Y G V G
BOVINE - - - - - - - - - - F - - - - - - - - - - F - - -
19/20 (IGG)
RAT 1351 GCT GGT GGC TTT CCT GGC TAT GGT GTT GGA GCT GGA GCT GGA CTT GGA GGT GCA AGC CAA GCT GCT GCT GCC GCT 1425
A G G F P G Y G V G A G A G L G G A S Q A A A A A
BOVINE P - - - - - I - ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ ◊ D A A - - Q - -
20/21
RAT 1426 GCC GCT GCC AAA GCT GCC AAG TAC GGT GCC GGA GGA GCT GGA ACC CTG GGA GGC CTG GTG CCA GGT GCA GTA CCA 1500
A A A K A A K Y G A G G A G T L G G L V P G A V P
BOVINE - - - - - - - I - - - - V - A - - - - - - - - ◊ ◊
21/22

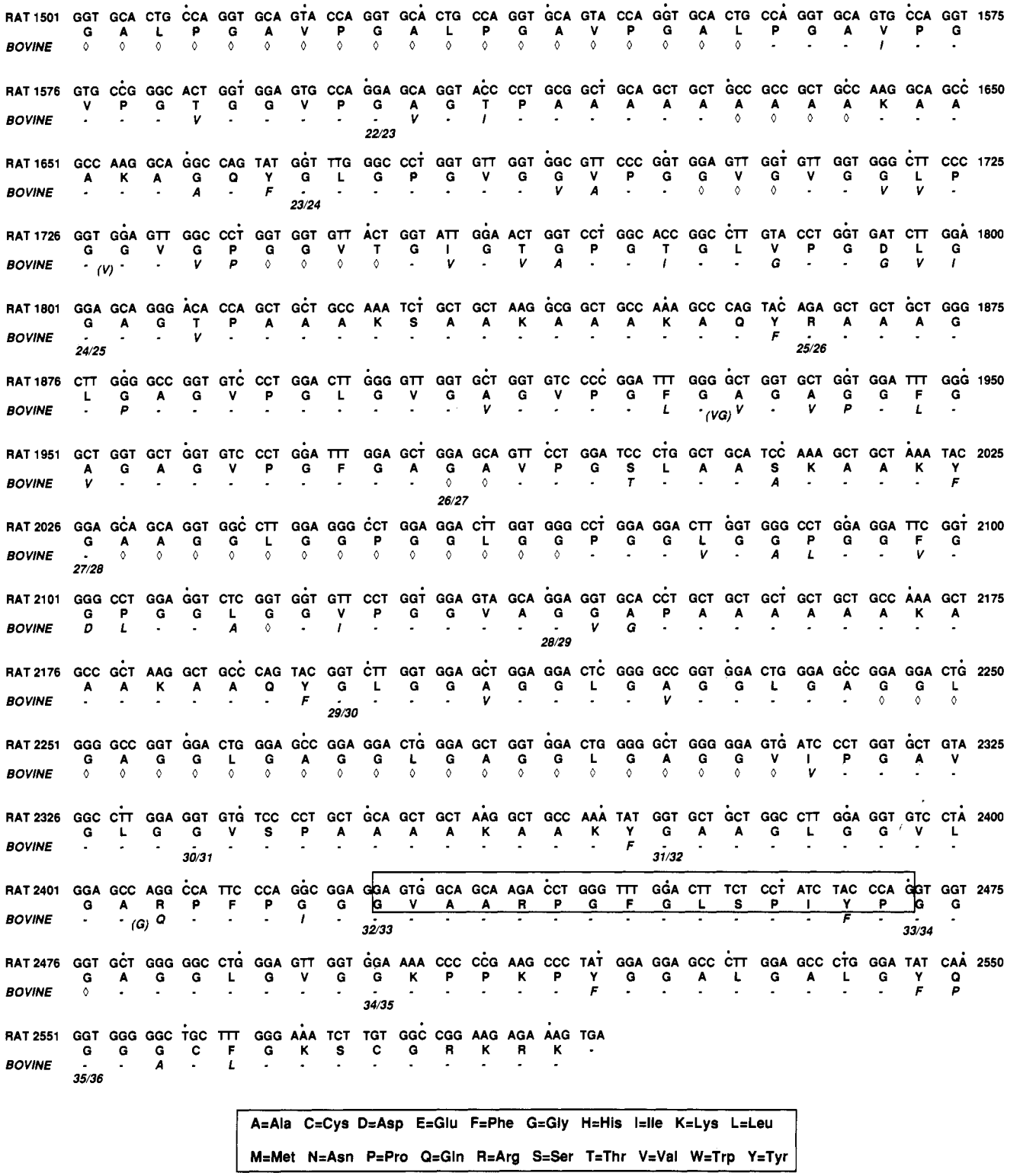


FIGURE 2: cDNA sequence and derived amino acid sequence of rat tropoelastin. (Line 1) DNA sequence derived from rat tropoelastin cDNAs. The sequence is numbered from the 5′-end. (Line 2) Amino acid sequence derived from rat tropoelastin cDNA sequence. The single letter amino acid code is used. The boxed table explains the abbreviations. (Line 3) Aligned bovine tropoelastin amino acid sequence. Amino acid substitutions are indicated; a dash has been used when the amino acid in a particular position is identical with the derived rat tropoelastin sequence, and a diamond indicates the lack of an amino acid residue in bovine tropoelastin in the analogous position of the rat sequence. Amino acids present in the bovine sequence but absent in the equivalent position in rat tropoelastin are indicated in parentheses. (Line 4) Positions of the exon boundaries within the bovine tropoelastin gene. The numbering of the exons is from the 5′-end of the gene. The boxed regions of sequence correspond to the open boxes presented in Figure 1 and represent DNA sequence missing in certain cDNA clones.

in molecular weight than bovine tropoelastin (Chipman et al., 1985; Campagnone et al., 1987). Raju and Anwar (1987a) reported the isolation of bovine tropoelastin cDNAs which potentially code for polypeptides of 64 171, 62 649, and 60 942 daltons. The calculated molecular weight of the derived amino acid sequence of secreted rat tropoelastin is 70 562 daltons. The expected molecular weight of the precursor, signal sequence containing polypeptide should exceed 73 000 daltons. The patterns of alternative splicing in rat tropoelastin mRNA are unknown, and it is not known if all tropoelastin mRNA

species are translated and secreted. The variant cDNA sequences described here demonstrate deletions of 1, 15, and 46 amino acids and could represent decreases of 98, 1484, and 4242 daltons, respectively, in the tropoelastin polypeptide. Therefore, the rat tropoelastin protein has a potential maximum molecular weight of 70 562 and a potential minimum molecular weight of 64 738 as calculated from the cDNAs studied here.

Discussion

The present study has provided the primary sequence of secreted rat tropoelastin and much of the corresponding signal peptide. As expected, rat tropoelastin contains domains similar to human, bovine, and chick tropoelastin. We report the sequence of overlapping rat tropoelastin cDNAs corresponding to exons 1–36 of the bovine gene and align the rat tropoelastin sequence with the bovine sequence (Indik et al., 1990). The most 5′ sequence reported here codes for 21 amino acids contained in the signal peptide. By comparison with the signal peptide of 26 residues sequenced from sheep tropoelastin (Davidson et al., 1982a,b) and the derived signal sequence of 26 amino acid residues in human tropoelastin cDNA (Indik et al., 1987), the majority of the rat tropoelastin signal peptide has been derived. Following the signal peptide are regions coding for alternating hydrophobic and cross-linking domains and the characteristic carboxy terminus containing two cysteine residues and terminating with the amino acid sequence GRKRK. Therefore, the overall primary structure of rat tropoelastin is similar to that of bovine and other tropoelastins.

Significant variability between tropoelastin cDNAs from various species has been reported (Indik et al., 1990). Human tropoelastin cDNAs lack exons 34 and 35 (Indik et al., 1987), and chick tropoelastin cDNAs exhibit significant rearrangements when compared to cDNAs coding for human and bovine tropoelastins. Comparisons between species have revealed conservative amino acid substitutions, as well as deletions and insertions. Rat tropoelastin cDNAs demonstrate further interspecies heterogeneity. In addition to small insertions, deletions, and conservative amino acid substitutions, a number of large insertions were demonstrated in the rat cDNAs compared to bovine. A number of highly repetitive sequences were observed in the hydrophobic domains of rat tropoelastin, and the large insertions occur in these repetitive domains. At present, the significance of these repetitive coding sequences is unknown. It is possible that highly repetitive domains associated with insertions and deletions may occur in the tropoelastin gene by recombination or unequal crossover. Alu repetitive sequences are prevalent in the bovine and human tropoelastin genes (Yeh et al., 1987; Indik et al., 1987) and may contribute to the generation of diversity in tropoelastin sequences.

Specific hydrophobic peptides of bovine tropoelastin have been used in chemotactic assays, studies of elastin binding proteins, and structure–function studies of tropoelastin. The hexapeptide repeat VGVAPG, encoded by exon 24 in the bovine and human tropoelastin genes, has been shown to be a binding site for the 67-kDa membrane-associated component of the elastin receptor. Mecham et al. (1989) demonstrate that the 67-kDa elastin binding protein has distinct binding sites for elastin and carbohydrate and interacts with 61- and 55-kDa integral membrane proteins. They propose the elastin receptor complex may be involved in signal transduction across the cell membrane or may mediate elastin fiber assembly. Rat tropoelastin does not contain the hexapeptide repeat VGVAPG, but does encode similar hydrophobic sequences. These differences suggest that an elastin receptor may bind to closely related sequences or may be species specific in its binding-site recognition. The repetitive sequence VPGVG, present in human, bovine, and chick tropoelastin, has been used in studies of the structure of tropoelastin. Urry et al. (1976) synthesized repeating peptides corresponding to tropoelastin sequences and demonstrated an ordered β-spiral conformation. In rat tropoelastin, the repeat IPGVG replaces the repetitive sequence VPGVG found in other tropoelastins. While various models of elastin structure have been proposed, the considerable variability of hydrophobic sequences observed between species is consistent with the proposed random coil structure of tropoelastin (Fleming et al., 1980).

Fragmentation of elastin peptides in neonatal rat smooth muscle cells in culture has recently been described (Franzblau et al., 1989). They propose a model in which proteolytic cleavage occurs at aromatic residues following α-helical cross-linking domains. It is interesting to note that the aromatic residues following cross-linking domains are almost invariably tyrosine in rat tropoelastin and phenylalanine in bovine tropoelastin. The functional significance of this substitution remains unknown.

Alternative splicing of tropoelastin pre-mRNA has been reported in bovine (Raju & Anwar, 1987a), human (Indik et al., 1987), and chick (Baule & Foster, 1988) tropoelastin cDNAs. In rat tropoelastin, variations between clones are observed in two regions of the composite sequence. Tropoelastin is a single copy gene (Olliver et al., 1987) and the variations observed correspond to an entire exon and contiguous exons. Therefore, alternative splicing of the primary transcript of the rat tropoelastin gene is a probable mechanism for the generation of these cDNA variants. The variability in cDNA sequence reported in this paper offers a possible origin for multiple isoforms of rat tropoelastin, and the derived amino acid sequence and patterns of cDNA sequence heterogeneity strongly suggests that these isoforms are due not to glycosylation or the existence of a secreted tropoelastin precursor but rather to alternative splicing of the transcriptional product of the rat tropoelastin gene.

In addition to the production of multiple tropoelastin isoforms that may be mediated by alternative splicing of the tropoelastin pre-mRNA, variations in rat tropoelastin cDNA sequence which may be due to allelic variation are reported here. It is evident from this sequence heterogeneity within a species, and from phylogenetic comparisons, that the hydrophobic domains of rat and, in general, vertebrate tropoelastins are not subject to rigorous amino acid sequence constraints. It is clear from the comparison of primary amino acid sequence derived from several vertebrate tropoelastins that the cross-linking domains and the cysteine-containing region at the carboxy-terminal end of the protein are rigorously conserved. The glycine-, valine-, and proline-rich hydrophobic domains, however, vary considerably with respect to sequence per se and the number of repeating peptide units. This sequence variation may be significant when considering the possible role of elastin in a variety of inherited disorders of elastic tissue which may involve mutations within the tropoelastin gene. We speculate that mutations within the carboxy terminus or in cross-linking domains of tropoelastin may elicit a clinically relevant alteration in the function of elastic tissue such as would be expected in a monogenic disorder, while the hydrophobic domains may be more tolerant of variation. Subtle mutations within these hydrophobic domains may be neutral or may manifest an altered phenotype only in concert with other genetic or environmental factors. Further work will be necessary to identify and ascertain the role of such mutations or variants in influ-

encing the normal functioning of elastic fibers in elastic tissue.

ACKNOWLEDGMENTS

We thank Dr. Joel Rosenbloom for kindly providing the human tropoelastin cDNA clone pcHEL-2, Leigh Schely and Andrea Carstens for typing the manuscript, and Jeff Friedman (Graphica, Inc., Plainsboro, NJ) for his graphic expertise.

REFERENCES

Baule, V. J., & Foster, J. A. (1988) *Biochem. Biophys. Res. Commun. 154*, 1054–1060.

Bressan, G. M., Argos, P., & Stanley, K. K. (1987) *Biochemistry 26*, 1497–1503.

Campagnone, R., Regan, J., Rich, C. B., Miller, M., Keene, D. R., Sakai, L., & Foster, J. A. (1987) *Lab. Invest. 56*, 224–230.

Chipman, S. D., Faris, B., Barone, L. M., Pratt, C. A., & Franzblau, C. (1985) *J. Biol. Chem. 260*, 12780–12785.

Davidson, J. M., Smith, K., Shibahara, S., Tolstoshev, P., & Crystal, R. G. (1982a) *J. Biol. Chem. 257*, 747–754.

Davidson, J. M., Leslie, B., Wolt, T., Crystal, R. G., & Sandberg, L. B. (1982b) *Arch. Biochem. Biophys. 218*, 31–37.

Davidson, J. M., Shibahara, S., Boyd, C., Mason, M. L., Tolstoshev, P., & Crystal, R. G. (1984) *Biochem. J. 220*, 653–663.

Deak, S. B., Pierce, R. A., Belsky, S. A., Riley, D. J., & Boyd, C. D. (1988) *J. Biol. Chem. 263*, 13504–13507.

Fazio, M. J., Olsen, D. R., Kauh, E. A., Baldwin, C. T., Indik, Z., Ornstein-Goldstein, N., Yeh, H., Rosenbloom, J., & Uitto, J. (1988) *J. Invest. Dermatol. 91*, 458–464.

Fleming, W. W., Sullivan, C. E., & Torchia, D. A. (1980) *Biopolymers 19*, 597–618.

Foster, J. A., Rich, C. B., Fletcher, S., Karr, S. R., & Przybyla, A. (1980) *Biochemistry 19*, 857–864.

Franzblau, C., Pratt, C. A., Faris, B., Colannino, N. M., Offner, G. D., Mogayzel, P. J., & Troxler, R. F. (1989) *J. Biol. Chem. 264*, 15115–15119.

Indik, Z., Yeh, H., Ornstein-Goldstein, N., Sheppard, P., Anderson, N., Rosenbloom, J. C., Peltonen, L., & Rosenbloom, J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5680–5684.

Indik, Z., Yeh, H., Ornstein-Goldstein, N., & Rosenbloom, J. (1990) in *Extracellular Matrix Genes* (Sandell, L. J., & Boyd, C. D., Eds.) Academic Press, New York (in press).

Matteucci, M. D., & Carruthers, M. H. (1981) *J. Am. Chem. Soc. 103*, 3185–3191.

Mecham, R. P., Hinek, A., Entwistle, R., Wrenn, D. S., Griffin, G. L., & Senior, R. M. (1989) *Biochemistry 28*, 3716–3722.

Mogayzel, P. J., Faris, B., Pratt, C. A., & Franzblau, C. (1987) *J. Cell Biol. 105*, 322a.

Olliver, L., Luvalle, P. A., Davidson, J. M., Rosenbloom, J., Mathew, C. G., Bester, A. J., & Boyd, C. D. (1987) *Collagen Relat. Res. 7*, 77–89.

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

Raju, K., & Anwar, R. A. (1987a) *J. Biol. Chem. 262*, 5755–5762.

Raju, K., & Anwar, R. A. (1987b) *Biochem. Cell. Biol. 65*, 842–845.

Rich, C. B., & Foster, J. A. (1989) *Arch. Biochem. Biophys. 268*, 551–558.

Rosenbloom, J. (1984) *Lab. Invest. 51*, 605–623.

Sandberg, L. B., & Davidson, J. M. (1984) in *Peptide and Protein Reviews* (Hearn, M. T. W., Ed.) Vol. 3, pp 169–226, Marcel Dekker, New York.

Sanger. F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., & Roe, B. A. (1980) *J. Mol. Biol. 143*, 161–178.

Urry, D. W., & Long, M. M. (1976) *CRC Crit. Rev. Biochem. 4*, 1–45.

Wrenn, D. S., Griffin, G. L., Senior, R. M., & Mecham, R. P. (1986) *Biochemistry 25*, 5172–5176.

Yeh, H., Ornstein-Goldstein, N., Indik, Z., Sheppard, P., Anderson, N., Rosenbloom, J. C., Cicilia, G., Yoon, K., & Rosenbloom, J. (1987) *Collagen Relat. Res. 7*, 235–247.

Yeh, H., Anderson, N., Ornstein-Goldstein, N., Bashir, M. M., Rosenbloom, J. C., Abrams, W., Indik, Z., Yoon, K., Parks, W., Mecham, R., & Rosenbloom, J. (1989) *Biochemistry 28*, 2365–2370.