Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 5680–5684, August 1987
Biochemistry

Alternative splicing of human elastin mRNA indicated by sequence analysis of cloned genomic and complementary DNA

(*Alu* repetitive sequences/porcine tropoelastin peptides)

ZENA INDIK*, HELENA YEH*, NORMA ORNSTEIN-GOLDSTEIN*, PAUL SHEPPARD*, NOEL ANDERSON*, JOAN C. ROSENBLOOM*, LEENA PELTONEN†, AND JOEL ROSENBLOOM*‡

*Center for Oral Health Research, School of Dental Medicine, University of Pennsylvania, Philadelphia, PA 19104; and †Recombinant DNA Laboratory, University of Helsinki, Helsinki 38, Finland

Communicated by Robert P. Perry, May 8, 1987 (received for review February 17, 1987)

ABSTRACT **Poly(A)$^+$ RNA, isolated from a single 7-mo fetal human aorta, was used to synthesize cDNA by the RNase H method, and the cDNA was inserted into λgt10. Recombinant phage containing elastin sequences were identified by hybridization with cloned, exon-containing fragments of the human elastin gene. Three clones containing inserts of 3.3, 2.7, and 2.3 kilobases were selected for further analysis. Three overlapping clones containing 17.8 kilobases of the human elastin gene were also isolated from genomic libraries. Complete sequence analysis of the six clones demonstrated that: (*i*) the cDNA encompassed the entire translated portion of the mRNA encoding 786 amino acids, including several unusual hydrophilic amino acid sequences not previously identified in porcine tropoelastin, (*ii*) exons encoding either hydrophobic or crosslinking domains in the protein alternated in the gene, and (*iii*) a great abundance of *Alu* repetitive sequences occurred throughout the introns. The data also indicated substantial alternative splicing of the mRNA. These results suggest the potential for significant variation in the precise molecular structure of the elastic fiber in the human population.**

The elastic properties of many vertebrate tissues including the lung and larger arteries are due to the presence in the extracellular matrix of elastic fibers composed primarily of the protein elastin. The individual polypeptide chains in the elastin fibers are covalently connected by crosslinkages derived from the oxidation of lysine residues by a Cu-requiring enzyme, peptidyl-lysine oxidase (1–4). The extensive crosslinking results in great insolubility, and substantial determination of elastin primary structure occurred only after the isolation of a soluble polypeptide, designated tropoelastin (M_r ≈72,000), from Cu-deficient or lathyritic animals (5, 6). Although many of the tryptic peptides derived from tropoelastin have been sequenced, they have not been ordered (7, 8). Cell-free translation of elastin mRNA has demonstrated that tropoelastin is the primary translation product (9–12).

Sequence analysis of a 1.3-kilobase (kb) ovine elastin cDNA clone, of the corresponding untranslated region of an ovine genomic clone, and of bovine genomic clones have demonstrated the following; (*i*) functionally distinct crosslinking and hydrophobic domains of elastin appear to be encoded by separate exons, (*ii*) the carboxyl terminus of these elastins ends with a cysteine-containing sequence, Gly-Gly-Ala-Cys-Leu-Gly-Lys-Ser-Cys-Gly-Arg-Lys-Arg-Lys, not previously observed in the protein sequencing of tropoelastin, and (*iii*) there is a 1.0- to 1.2-kb untranslated segment at the 3′ message end (13–15). Recent analyses of 6.2 kb of the 3′ portion of the human elastin gene have confirmed these observations but also have suggested that the human gene lacks two exons found in the bovine gene (16). RNA hybridization analyses have demonstrated that chick, ovine, human, and bovine elastin mRNAs contain ≈3500 nucleotides (13, 16–18). Thus, these mRNAs should contain ≈2.4 kb of translated sequence encoding 800 amino acids.

We report the construction and sequence analysis[§] of human elastin cDNA clones and correlation of the sequence with 17.8 kb of the human elastin gene.

MATERIALS AND METHODS

Identification and Characterization of cDNA and Genomic Clones. The isolation and partial characterization of the human elastin genomic clones, HEL1 and HEL2, have been described (16). A third genomic clone was identified by screening a library in λ EMBL3 with a nonrepetitive restriction fragment near the 5′ end of clone HEL2. The three overlapping clones encompass 17.8 kb of the gene (Fig. 1).

Poly(A)$^+$ RNA was isolated from a single aorta of a 7-mo human fetus (19), and used to synthesize cDNA by the RNase H method (20). The cDNA was inserted into λgt10 (Vector Cloning Systems, San Diego, CA) with linkers, and desired recombinant phage was identified by screening with two probes: (*i*) pcHEL1, a cDNA containing 421 base pairs (bp) of untranslated sequence (21) and (*ii*) a 250-bp genomic fragment containing exon 18. Twenty-six clones were identified, and three clones that hybridized to both probes were purified for detailed study.

DNA Sequencing and Synthesis of Oligonucleotides. Restriction fragments of the cDNA and genomic clones were isolated by electroelution after electrophoresis on 1% agarose gels and were sequenced by the Sanger dideoxynucleotide chain-termination method (22, 23) using a universal primer of 17 nucleotides (Collaborative Research, Waltham, MA) or oligonucleotides of 17–22 bases synthesized in our laboratory by a modification of the phosphite method (24), using a MilliGen (Bedford, MA) programmable machine.

RESULTS

DNA Sequence Analysis of cDNA. Analysis of both strands of the three largest cDNA clones (3.3-, 2.7- and 2.3-kb inserts) revealed that their overlapping portions were identical except for three segments indicated in the bar diagrams in Fig. 1 as open portions (approximate coordinates 1.4, 1.9, and 2.22) to signify that the particular segment was missing in the clone. The cDNA sequences up to the termination codon have been combined into a single sequence in Fig. 2, along with the encoded amino acid sequence. The composite 2358-bp translated sequence begins with a 21-nucleotide 5′ untranslated segment, includes a single methionine initiation

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡To whom reprint requests should be addressed.

§This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J02948).

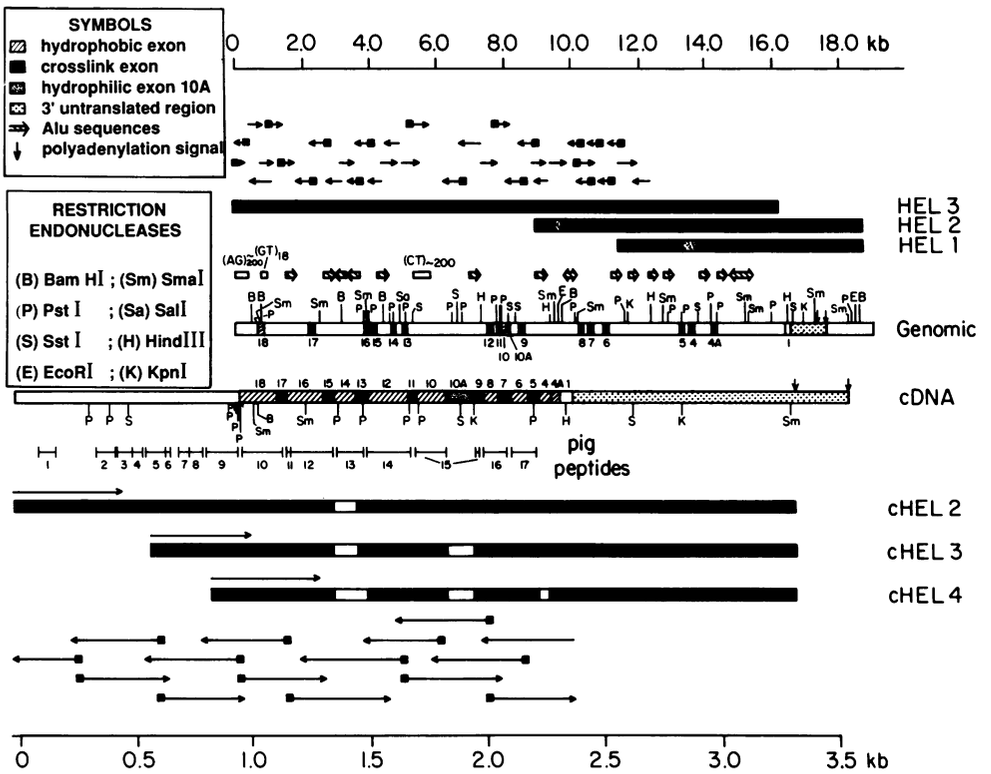

FIG. 1. Restriction maps, structures, and sequencing strategies for human elastin cDNA and genomic clones. Restriction fragments were sequenced by the dideoxynucleotide chain-termination method using either a universal primer (→) or specific oligonucleotides (➡). HEL1 and HEL2 segments (stippled blocks) were used for genomic screening. In the cDNA clones (cHEL2, cHEL3, and cHEL4), open spaces designate exons subject to alternative splicing. The pig peptides designated 1–17 correspond to the pig peptide W series (8): 1, (W13); 2, (W8b); 3, (W11b); 4, (W11c); 5, (W10); 6, (W14b); 7, (W12); 8, (W11a); 9, (W7); 10, (W4); 11, (W16); 12, (W8a); 13, (W3); 14, (W6); 15, (W14); 16, (W5); and 17, (W9).

codon, and extends through the entire translated region to the termination codon. Not included in the figure is the previously published 1.2-kb 3′ untranslated region (16). An amino-terminal signal sequence (Fig. 3) can be identified by comparison with signal sequences determined in ovine (12) and chick-pretropoelastin (11) and by comparison with the amino-terminal sequences of porcine (8) and chick tropoelastin (6).

The encoded amino acid sequence of human tropoelastin shows good homology with the sequenced tryptic peptides of porcine tropoelastin, which have been entirely identified and are here ordered (Figs. 1 and 2). In addition to the sequences homologous to the hydrophobic and crosslinking peptides, several unusual sequences not identified in the porcine peptides were encoded in the human cDNA. These include the following: (*i*) the previously described, highly basic, cysteine-containing carboxyl-terminal region (16), (*ii*) a sequence, Leu-Pro-Gly-Gly-Tyr-Gly-Leu-Pro-Tyr-Thr-Thr-Gly-Lys (residues 213–225) that is part of a tyrosine-rich region (other tyrosines are at residues 206, 228, 230, and 244), and (*iii*) a very hydrophilic sequence (residues 605–644) containing four arginines, one histidine, two aspartates, two glutamates, eight serines, and one threonine.

Structure of the Gene. We have previously reported the sequence of 6.2 kb of the 3′-most portion of the gene (16) in which three exons, numbered 1, 4, and 5, were identified (numbered so that homologous bovine and human exons have consistent designation). As determined by the previous genomic sequencing and confirmed by the present cDNA sequencing, bovine exons 2 and 3 have no counterpart in the human gene. In order to define the remaining exon:intron structure, the human genomic sequence (sequencing strategy diagrammed in Fig. 1) was correlated with the human cDNA sequence and with sequences of porcine peptides. This analysis identified 17 exons numbered 1, 4–18, and 4A. A sequence homologous to exon 4A was not found in the ovine cDNA nor in the porcine peptides (unpublished observations indicate an homologous sequence in the bovine gene). Except at the carboxyl-terminal region, exons encoding hydrophobic domains (even-numbered) alternate with those encoding crosslinking domains (odd-numbered) (Fig. 1). In every case, exon splice junctions conform to the canonical sequence AG–exon–GT (26).

Alternative Splicing of Elastin mRNA. It is clear from the present analyses that the variation in sequence among the human cDNAs is due to the variable presence of several exons. Thus, clone cHEL4 lacks the sequences corresponding to exons 4 and 13, and clones cHEL3 and cHEL4 lack the sequence corresponding to exon 10A. All three cDNA clones lack the sequence corresponding to exon 14. This sequence has been included in the composite sequence because of the extremely strong homology to the porcine peptide W-3 (8) (Figs. 1 and 2) and because it preserves the alternation of hydrophobic and crosslinking domains. Presumably, as more cDNA clones are analyzed, exon 14 will be identified in human cDNA. As discussed below, we believe the most likely explanation for the variability in cDNA sequence is alternative splicing of the primary mRNA transcript.

In an effort to determine the basis for the alternative splicing, we have analyzed the sequences surrounding the splice sites (Fig. 4). Stretches of pyrimidines were found adjacent to the AG dinucleotide, and possible branch-points, deviating from the consensus branch-point sequence, YNYTRAY (27), at no more than two positions, were identified. However, aside from exons 10 and 10A the 5′ intron borders of which were GTGCA instead of the consensus sequence GTRAG, and exon 14 the 3′ intron border of which was GAG instead of YAG, there was very little in the surrounding sequences to distinguish exons that were spliced out from those that were not (for review, see ref. 28).

Hydrophilicity/Hydrophobicity Analysis. Graphical analysis (29) of the distribution of hydrophilic and hydrophobic segments of human tropoelastin demonstrates that segments of potential crosslinking project as relatively hydrophilic regions that are not uniformly distributed and occur at shorter intervals in the first 200 and the last 150 residues of the molecules (Fig. 5). Furthermore, 4 of the 5 potential crosslinking sequences in the first 200 residues contain a proline or other residues between the two lysines instead of the usual alanines. Because the segments containing proline will differ conformationally from those containing alanine, the types of crosslink formed with these two segmental types may differ.

FIG. 2. Nucleotide sequence and translated amino acid sequence of the complete coding portion of human elastin cDNA. Segments corresponding to known exons in the elastin gene are indicated by numbers. Amino acid sequences determined by sequencing peptides derived by tryptic digestion of porcine tropoelastin (7, 8) have been aligned where possible. See also Fig. 1.

```
                Signal Peptide           ↓  N Terminus
Human   MAGLTAAA PRPGVL L LLSILHP   SRPGGVPGAIPGG VPGGVFYPGAGLGALG
Sheep   MKXLTAAAARRPGVLF/LLLLCILFC/PSFPGGVPGAVPGG VPGGVFFXXXXLXXLG
Chick   MRQ  AAAPLLPGVL L LFSILPA   SQQGGVPGAIPGGGVPAGVFFPGAGVGG
Pig                                    GGVPGAVPGG VPGGVFFPGAGLGGLG
```

FIG. 3. Amino acid sequence of the amino terminal region of human pretropoelastin deduced from cDNA. The human sequence has been aligned with sequences determined by sheep (12) and chick (11) pretropoelastins and pig (8) and chick (6) tropoelastins. (↓) Presumptive site of signal sequence cleavage for human pretropoelastin.

This 200-amino acid segment of the protein ends in a tyrosine-rich region of unknown function, but which may be involved in the alignment of the tropoelastin molecules within the fiber before crosslinking occurs. The two crosslinking sequences that contain three lysine residues, KAAAKAAK (exon 17) and KSAAKVAAK (exon 11), occur near the center of the molecule and may have a unique role in crosslinking the tropoelastin molecules. When exon 14 is omitted, crosslinking exons 13 and 15 are placed adjacent to one another so that the sequence becomes KAAKYGVGT-PAAAAAKAAAK; the consequence of this merging of crosslinking domains remains to be determined.

Structure of Intervening Sequences. In the human genome, *Alu* repetitive sequences, which consist of two head-to-tail 130- and 160-bp monomeric units, constitute 3–6% of the total DNA mass (30, 31). *Alu* sequences have some features of transposable elements and pseudogenes in that they are often flanked by direct repeats and have A-rich sequences at their 3′ ends (32). A striking feature of the introns of the human elastin gene is that *Alu* sequences are found at a frequency of about four times the expected value (Fig. 1). Although each *Alu* repeat in the elastin gene contains some features of the classical *Alu* sequence, all sequences do not contain direct repeats—the monomer units are not exactly the same size and are not found strictly in pairs of left and right units. It is perhaps significant that whereas most repeats are oriented in a 5′-to-3′ direction, several are in opposite orientation, resulting in inverted repeats.

In addition to *Alu* repeats, rather long stretches composed of either alternating purines (genomic coordinate 0.2 in Fig. 1) or alternating pyrimidines (coordinate 5.6) were observed. Although the alternation is not strict, many potential loop structures could be constructed between the two segments. Segments of the remaining intervening sequences are of high G+C content and could encode elastin-like sequences, which may be found in some mature mRNA molecules.

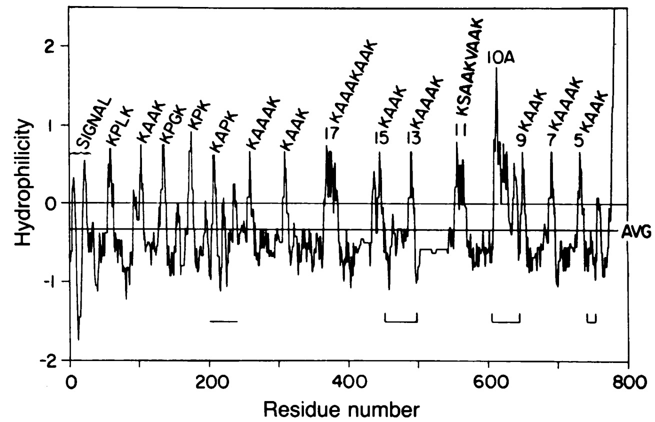

FIG. 5. Hydrophilicity/hydrophobicity analysis of human tropoelastin done by the method of Hopp and Woods (29). The hydrophilic peaks correspond for the most part to potential crosslinking domains. Amino acid sequences of potential crosslinks are given, and crosslinking exons, where known, are numbered. Exons subject to alternative splicing are indicated by brackets (⊔). The bar (—) indicates position of the tyrosine-rich segment.

DISCUSSION

The present studies have elucidated the primary structure of human tropoelastin and have ordered the previously sequenced porcine tryptic peptides (7, 8). The deduced maximum size of the human protein, 786 amino acids, is consistent with previous estimates of the M_r of 72,000. The size of the message is also consistent with previous estimates of 3.5 kb. These sequence results show that, in general, the protein consists of two types of alternating domains, one rich in hydrophobic amino acids and the other rich in lysine residues. The hydrophobic domains are undoubtedly responsible for the elastic properties of elastin, whereas the lysine residues form covalent crosslinks between molecules. This functional organization of the protein is reflected in the portion of the gene thus far analyzed because hydrophobic and crosslinking domains are encoded in separate exons.

An interesting and provocative feature of these results is the observation that the mature elastin mRNA molecules, as determined by cDNA analysis, vary in sequence. It is extremely unlikely that these differences are due to cloning

```
CONSENSUS              yNyTuAy                yyyyyyyyyyyNCAG G           AG GTuAG
                    branch point                      intron 3'   exon        5' intron
  50        40        30        20        10
    . -+++++- .      -+++-++      .         .
AGCCGAAACTGAGAGGGGCCGGACTCACAGTGATGTGCACCTCCTCCCGTCCAG GT   1         TERM
    .         .  ++++-++.    +++-+++        .
GCTGGAGTCAGTTTCCACCCCTACCAACCCACCAACCTGAAATCTCTCCTGCAG GA   4A     AG GTATGCCAGG
    .         .         .   +++++++         .
AGGGCCTCTTCCCGATGGGGGTGTCTTATCCTGACCCCACCTGCCTCTTCTCAG GT   4      AG GTAGGGGTGG
-++++++       .        ++++++-    .         .
GCCTGACCAGGTGGCATTGGCATTCCTGAGCCGTCATGTGCCTCATCTCCCCAG GT   5      CG GTGAGTXCCC
    .         .+++++++  .         .         .
GCTTCAGTCCCACCTTTCTGACCAGCGGAGTCTAATGCTCAGCTGTCTCCACAG GC   6      AG GTGAGAGTTG
    .  -++++++.         .    ++++++-        .
GGAGGGAATCTAACCAGTACAGAGTGCCTCCCTGAACTCGGTCTGTGTTCCCAG GA   7      TG GTGAGCACTG
    .         .         .  ++++-++.         .
AGGGAGACCCATCGTTCAGAAATGGAACACTCATTTTCCCTCCTCTCCCCGCAG GA   8      GG GTGAGTTGAT
    .         .     +++++++       .         .
TCCCAGGCACAGAGCTCGGCTCCTGACCACTCCCCAACTTTTCTTTCTCCCCAG TA   9      TG GTGAGTGCAC
++++-+-       .         .         .         .
TCCTTAGGGGCATGCTCCCTGCCTGCTGTCGCCACCACTGCCCTCTGTCTGCAG GA   10     AG GTGCAGATGA
                                                            10A    GG GTGCATAGTA
    .         .         +++++++   .         .
AGCCTCCATGGGCCCCGCCTCCATCTCTAATCCCCCTCTCTCTCCCTCCCTCAG CT   11     CC GTGAGTGCCT
    .   +++++++         .         .         .
TCTGTCCTCTTTGATCAGGTCTTGGTTAATGATCAGCTCTTCTCAATCTTGCAG GG   12     AG GTGAGTTTCA
    .         .         . ++++-++ .    ++++-++
AGCAGGGAGGGGTGTGAGAGATTACTCTCTCACCCCTTCTCTTCACACCTCCAG GA   13     TG GTAAGTCCCC
    .         .         .    -+++++-   +++--++
CCCCCAAAAAGTGAGTACTGGAGGGGCAAGGCTGAAAGTTCTCCACTCCCCGAG GT   14     AG GTGAGCTGTG
    .         .    ++++++-        .         .
GAGGAGACCCAGGCACGGCTTCTGAGGGTCTCTATCTTTCTCGTTTCCTTGTAG CC   15     CG GTAAGTGCCC
    .        ++++++-    .         .         .
GCCCAGCCTCTCTCACTGAGGCTTCTTTTCTACTTGGCTCCCTTCCCTCTGCAG GG   16     CC GTGAGCCTTA
    . +++-+++ .         .         .         .
TGCTGCCTCCAATGCTGCTGCCTGAGCATGTTGTGTCCCTTTTGGTCTCTCCAG GG   17     CG GTGAGTGCTA
    .  ++++++-.         .         .         .
CTAGCCCCTCTGAGGTTCCCATAGGTTAGGGGAACAATGCTTTTTCTTCCACAG GA   18     AG GTGAGCTGGG

  ELASTIN CONSENSUS SEQUENCE          yyyyyyyyyyyyyyNCAG G             NG GTuAGy
```

FIG. 4. Intron sequences flanking exons of human elastin gene. Possible branch-point sequences with best fit to the consensus sequence, YNYTRAY (27), are indicated. In the figure, y = Y, pyrimidine; N = pyrimidine or purine; and u = R, purine.

artifacts because they involve the omission of exons encoded between canonical splice junctions. Such differences could be due to the presence of more than one human elastin gene, to allelic variation in a single gene, or to alternative splicing of the primary transcript of a single gene. All available evidence suggests that there is only one human elastin gene. This evidence includes the identification of a single locus (long arm of chromosome 2) by *in situ* hybridization (21), Southern hybridization analysis of restriction enzyme digests of genomic DNA (16), and the present finding of complete sequence identity between the cDNAs except for the omitted sequences. If more than one gene were involved, we would expect to observe some sequence differences—at least, in wobble positions that entail no amino acid substitutions. Because the cDNA was constructed from mRNA isolated from a single individual with, at most, two alleles, and because three distinct cDNAs were observed, the most likely explanation for the present results is alternative splicing of a single primary transcript (28).

Two types of variable splicing were observed. In the first type, there is complete excision of exons as seen with exons 4, 13, and 14. In the second type, there is excision of a portion of exon 10 in which a single intron 3′ splice site utilizes alternative intron 5′ sites to generate two types of transcripts. Similar excisions have been described previously, including those involving the extracellular matrix protein, fibronectin (28). The elastin gene appears unusual in that two types of alternative splicing mechanisms are used. Obviously the cDNAs analyzed here represent only a small fraction of the total available elastin mRNA, and other types of analysis such as S1 mapping must be done to determine the frequency of such alternative splicing and whether other exons are involved. In RNA hybridization experiments, elastin mRNA is found as a rather diffuse band centered at 3.5 kb (13, 16–18), suggesting that multiple species may be present and that the cDNAs are not the result of rare splicing events.

If all the mRNA molecules are translated, there will be significant variation in the amino acid sequence and length of the tropoelastin. This could explain the finding of at least two forms of tropoelastin in several species (10, 12, 33). At present we do not know whether the splicing pattern is developmentally or tissue regulated and whether there are functional differences between the molecules. The variable expression of exon 10A is particularly intriguing. This domain, which is highly hydrophilic and atypical for elastin in amino acid sequence, may be involved in interaction with other matrix macromolecules.

The presence of an abundance of repetitive sequences within the introns raises questions concerning the stability of the elastin gene within the human population. In other human genes, such as those for the low density lipoprotein receptor (34) and human α hemoglobin (35), deletions apparently mediated by recombination between repeated sequences have occurred, resulting in hereditary diseases. In addition, evidence for genomic instability in regions of human DNA enriched in *Alu* repeat sequences has been presented by Calabretta *et al.* (25). Further studies of the elastin gene are warranted to determine whether significant polymorphism, possibly mediated by similar mechanisms, is found in the population. Because the conformation of elastin is, at least in part, that of a random coil (36), the stringency for conservation of the amino acid sequence may be less than that for other proteins with specific conformations, and elastin molecules that vary in sequence may exist and be compatible with relatively normal life. However, genetic variation in the structure of elastin could result in an ensemble of molecules varying in fitness, with potential adverse consequences on the properties of vital cardiovascular and pulmonary tissues over individual lifetimes.

We thank Dr. Louise C. Showe for the λ EMBL3 human genomic library and Dr. Andy Ross for help in computer analysis. This work was supported by National Institutes of Health Research Grants Am-20553 and Am-35229 and Grant 1-989 from the National Foundation–March of Dimes.

1. Franzblau, C., Sinex, F. M., Faris, B. & Lampidis, R. (1965) *Biochem. Biophys. Res. Commun.* **21,** 525–531.
2. Partridge, S. M., Elsden, D. F., Thomas, J., Dorfman, A., Telser, A. & Ho, P. (1966) *Nature (London)* **20,** 399–400.
3. Miller, E. J., Martin, G. R., Mecca, C. H. & Piez, K. A. (1965) *J. Biol. Chem* **240,** 3623–3627.
4. Pinnell, S. R. & Martin, G. R. (1968) *Proc. Natl. Acad. Sci. USA* **61,** 708–716.
5. Sandberg, L. B., Weissman, N. & Smith, D. W. (1969) *Biochemistry* **8,** 2940–2945.
6. Foster, J. A., Shapiro, R., Voynow, R. C., Crombie, G., Faris, B. & Franzblau, C. (1975) *Biochemistry* **14,** 5343–5347.
7. Foster, J. A., Burenger, E., Gray, W. R. & Sandberg, L. B. (1973) *J. Biol. Chem.* **248,** 2875–2879.
8. Sandberg, L. B., Leslie, J. G., Leach, C. T., Alvares, V. A., Torres, A. R. & Smith, D. W. (1985) *Pathol. Biol.* **33,** 266–274.
9. Burnett, W. & Rosenbloom, J. (1979) *Biochem. Biophys. Res. Commun.* **86,** 478–484.
10. Foster, J. A., Rich, C. B., Fletcher, S., Karr, S. R. & Przybyla, A. (1980) *Biochemistry* **19,** 857–864.
11. Karr, S. R. & Foster, J. A. (1981) *J. Biol. Chem.* **256,** 5946–5949.
12. Davidson, T. M., Leslie, B., Wolt, T., Crystal, R. G. & Sandberg, L. B. (1982) *Arch. Biochem. Biophys.* **218,** 31–37.
13. Yoon, K., May, M., Goldstein, N., Indik, Z. K., Oliver, L., Boyd, C. & Rosenbloom, J. (1984) *Biochem. Biophys. Res. Commun.* **118,** 261–269.
14. 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.* **185,** 684–691.
15. 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.
16. Indik, Z., Yoon, K., Morrow, S. D., Cicila, G., Rosenbloom, J. C., Rosenbloom, J. & Ornstein-Goldstein, N. (1987) *Connect. Tissue Res.* **16,** 197–211.
17. Burnett, W., Finnigan-Bunick, A., Yoon, K. & Rosenbloom, J. (1982) *J. Biol. Chem.* **257,** 1569–1572.
18. Davidson, J. M., Shibahara, S., Schafter, M. P., Harrison, M., Leach, C., Tolstoshev, P. & Crystal, R. G. (1984) *Biochem. J.* **220,** 643–652.
19. Chirgwin, J. M., Przybyla, A. E., McDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
20. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25,** 263–269.
21. Emanuel, B. S., Cannizzaro, L., Ornstein-Goldstein, N., Indik, Z., Yoon, C., May, M., Oliver, L., Boyd, C. & Rosenbloom, J. (1985) *Am. J. Hum. Genet.* **37,** 873–882.
22. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
24. Matteucci, M. D. & Caruthers, M. H. (1981) *J. Am. Chem. Soc.* **103,** 3185–3191.
25. Calabretta, B., Robberson, D. L., Berrera-Saldana, H. A., Lambrou, T. P. & Saunders, G. F. (1982) *Nature (London)* **296,** 219–225.
26. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
27. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. (1984) *Cell* **38,** 317–331.
28. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1985) *Annu. Rev. Biochem.* **55,** 1119–1150.
29. Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3824–3828.
30. Schmid, C. W. & Jelinek, W. R. (1981) *Science* **216,** 1065–1069.
31. Deininger, D. L., Jolly, D. J., Rubin, C. M., Friedman, T. & Schmid, C. W. (1981) *J. Mol. Biol.* **151,** 17–33.
32. Sharp, P. A. (1983) *Nature (London)* **301,** 471–472.
33. Mecham, R. P., Levy, B. D., Morris, S. L. & Wrenn, D. S. (1985) *J. Biol. Chem.* **259,** 12414–12418.
34. Lehrman, M. A., Schneider, W. J., Sudhof, T. C., Brown, M. S., Goldstein, J. L. & Russell, D. W. (1985) *Science* **227,** 140–146.
35. Orkin, S. H. & Michelson, H. (1980) *Nature (London)* **286,** 538–540.
36. Fleming, W. W., Sullivan, C. E. & Torchia, D. A. (1980) *Biopolymers* **19,** 597–618.