Structure of the Human Laminin α2-Chain Gene (LAMA2), Which Is Affected in Congenital Muscular Dystrophy*

(Received for publication, June 12, 1996)

Xu Zhang‡‡§§, Reetta Vuoletenaho‡‡§§, and Karl Tryggvason‡‡§§

From the ‡Division of Matrix Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institute, S-171 77 Stockholm, Sweden and the §Biocenter and Department of Biochemistry, University of Oulu, FIN-90570 Oulu, Finland

We have determined the structure and complete exon size pattern of the human laminin α2-chain gene (LAMA2), which has been shown to be affected in congenital muscular dystrophy (Helbling-Leclerc, A., Zhang, X., Topaloglu, H., Cruaud, C., Tesson, F., Weissbach, J., Tomé, F. M. S., Schwartz, K., Fardeau, M., Tryggvason, K., and Guicheney, F. (1995) Nat. Genet. 11, 216–218). The gene is over 260,000 base pairs and contains 64 exons. The sequence of all exon-intron borders was determined. Two of the exons, i.e., exons 43 and 52, are extremely small in size, 6 and 12 base pairs, respectively. Comparison of the exon pattern of the human LAMA2 gene with that of the Drosophila LAMA gene revealed that only 2 of 63 intron locations in the 5′-end of the human gene match the intron locations in the Drosophila gene, which contains 14 introns.

Laminins are a family of large trimeric basement membrane glycoproteins composed of α-, β-, and γ-chains (1–3). The three subunit chains associate at their carboxyl termini in a coiled coil, usually forming a cross-shaped molecule with the long arm contributed to by the coiled coil and the amino termini forming the short arms. To date, five genetically distinct α-chains, three β-chains, and two γ-chains have been identified (1–7). The complete sequence of all of the human chains except for α5 has been determined (6, 8–16). The laminin chains form a variety of isoforms that may vary extensively with respect to their tissue distribution. There is still little knowledge about the regulation of tissue-specific expression of laminin genes, and the structure of the human genes has only been determined for the β1-chain (17), γ1-chain (18), and γ2-chain (19). These studies have revealed considerable structural divergence between the β- and γ-chain genes. As yet, the structure of no mammalian α-chain gene has been reported, but the Drosophila α-chain gene has been shown to contain 15 exons (20).

The human laminin genes are quite dispersed in the genome, but many of them are located in proximity to each other, which, in turn, indicates their evolutionary relationship. Thus, LAMA1 and LAMA3, encoding the α1- and α3-chains, respectively, are located on the same chromosome 18, but distantly from each other at 18p11.3 and 18q11.2, respectively (11, 21). LAMA2 and LAMA4 are located close to each other at 6q22–23 and 6q21, respectively (5, 10). The β-chain genes are all located on different chromosomes, with LAMB1 at 7q22 (12), LAMB2 at 3p21 (13, 22), and LAMB3 at 1q32 (23). The two γ-chain genes, LAMC1 and LAMC2, are located in very close proximity to each other, both genes being located at 1q25–31 (16, 24).

The laminin-2 and laminin-4 isoforms (4), which have the molecular formulas α2:β1:γ1 and α2:β2:γ1, respectively, are characteristically enriched in basement membranes surrounding skeletal muscle fibers (25, 26). The high tissue specificity of these isoforms is provided only by the α2-chain, previously termed merosin (8), as the other component chains of this laminin isoform, the β1- or β2- and γ1-chains, are quite ubiquitous. The α2-chain is expressed widely in basement membranes of skeletal muscle, both at neuromuscular synapses and extrasynaptically as well as in the myotendinous junctions (27–29). The skeletal muscle-specific location of the α2-chain implies a specific function for muscle development and function, and this chain has been shown to bind to the sarcolemma protein complex dystroglycan, which, in turn, interacts with the cytoskeleton component dystrophin and extracellular laminin (30, 31). Mutations in the genes for dystrophin have previously been shown to cause Duchenne’s muscular dystrophy (32), and mutations have also been identified in components of the dystroglycan protein complex in other types of muscular dystrophy (33–37). Recently, mutations were described also in the laminin α2-chain gene (LAMA2) in muscular dystrophy in mice (38), and more recently, we identified mutations in human patients with congenital muscular dystrophy (39, 40). These findings demonstrate the role of the α2-chain in skeletal muscle function.

Detailed characterization of the LAMA2 gene is essential for studies on gene regulation, analysis of mutations, and studies on the pathogenesis of congenital muscular dystrophy. In this work, we have determined the entire exon pattern of the human LAMA2 gene and shown it to exceed 260,000 base pairs in size and to contain 64 exons.

MATERIALS AND METHODS

Isolation of Genomic Clones and Gene Mapping—Three human λ-phage libraries (CLONTECH HU2004j, HL111j, and HL1067j) were screened to isolate genomic LAMA2 clones. The libraries were screened with γ-labeled human laminin α2-chain cDNA inserts (10) according to standard procedures (41). Purified clones were characterized by restriction mapping and hybridization with different laminin α2-chain cDNA fragments or sequence-specific oligonucleotide probes, and appropriate restriction fragments were subcloned into the pBluescript SK vector for further analysis.

DNA Sequencing and Estimation of Intron Sizes—Sequencing of exons and exon-intron boundaries was carried out on purified λ-phage clones directly or, in some cases, on subcloned restriction fragments of genomic clones by using an AmpliCycle kit (Perkin-Elmer) and a Cycle Sequencing kit (Pharmacia Biotech Inc.). Sizes of introns were assessed by determination of the size of electrophoresed fragments polymerase chain reaction-amplified from genomic DNA or by Southern hybridization and mapping of restriction fragments.
FIG. 1. Schematic structure of the human LAMA2 gene and alignment of exons with structural domains in the α2-polyepitope chain. Top, structural domains of the α2-chain. The amino terminus (N) and carboxyl terminus (C) are indicated. Repeats and subdomains of domains IIIa, IIIb, V, and G are indicated by boxes. Middle, schematic structure of the entire gene, with locations of exons shown by red vertical bars and introns shown by the horizontal line. The exons are numbered from the 5'-end of the gene. Four intron sequences of unknown size that are not contained in the genomic clones are indicated by circles. The exons coding for structural domains are indicated by dashed lines. Bottom, alignment of the 32 λ-phage clones characterized in this study. Scales in amino acids (aa) and kilobases (kb) are shown.

TABLE I
Sizes of introns were determined by mapping of genomic clones by restriction enzymes (Fig. 1). The sizes of introns were determined directly from the exon-containing λ-phage clones or alternatively, the nucleotide sequences of exons were determined by direct sequencing of small introns from subcloned restriction fragments or were estimated by polymerase chain reaction amplification of DNA segments between adjacent exons.

| Exon and intron No. | Exon size | Intron size | Exon and intron No. | Exon size | Intron size |
|---------------------|-----------|-------------|---------------------|-----------|-------------|
| 1                   | 112       | 15,000 bp   | 3                    | 113       | 3,500 bp    |
| 2                   | 171       | 700 bp      | 4                    | 243       | 36 bp       |
| 3                   | 270       | 750 bp      | 5                    | 118       | 5,700 bp    |
| 4                   | 179       | 380 bp      | 6                    | 100       | 600 bp      |
| 5                   | 161       | 5,000 bp    | 7                    | 104       | 400 bp      |
| 6                   | 141       | 2,200 bp    | 8                    | 174       | 7,000 bp    |
| 7                   | 102       | 2,000 bp    | 9                    | 121       | 1,100 bp    |
| 8                   | 112       | 4,000 bp    | 10                   | 174       | 7,000 bp    |
| 9                   | 114       | 3,500 bp    | 11                   | 102       | 2,000 bp    |
| 10                  | 114       | 3,400 bp    | 12                   | 212       | 1,100 bp    |
| 11                  | 114       | 3,500 bp    | 13                   | 212       | 1,100 bp    |
| 12                  | 128       | 6,500 bp    | 14                   | 112       | 4,000 bp    |
| 13                  | 127       | 5,100 bp    | 15                   | 114       | 3,500 bp    |
| 14                  | 127       | 5,100 bp    | 16                   | 128       | 6,500 bp    |
| 15                  | 127       | 5,100 bp    | 17                   | 87        | 5,100 bp    |
| 16                  | 127       | 5,100 bp    | 18                   | 212       | 3,700 bp    |
| 17                  | 197       | 13,000 bp   | 19                   | 107       | 13,000 bp   |
| 18                  | 181       | 3,100 bp    | 20                   | 181       | 3,100 bp    |
| 19                  | 137       | 530 bp      | 21                   | 137       | 530 bp      |
| 20                  | 227       | 1,600 bp    | 22                   | 227       | 1,600 bp    |
| 21                  | 234       | 1,100 bp    | 23                   | 234       | 1,100 bp    |
| 22                  | 180       | 150 bp      | 24                   | 180       | 150 bp      |
| 23                  | 189       | 1,100 bp    | 25                   | 189       | 1,100 bp    |
| 24                  | 134       | 4,400 bp    | 26                   | 134       | 4,400 bp    |
| 25                  | 118       | 8,500 bp    | 27                   | 118       | 8,500 bp    |
| 26                  | 135       | 8,000 bp    | 28                   | 135       | 8,000 bp    |
| 27                  | 129       | 3,900 bp    | 29                   | 87        | 6,300 bp    |
| 28                  | 194       | 300 bp      | 30                   | 194       | 300 bp      |
| 29                  | 143       | 3,900 bp    | 31                   | 143       | 3,900 bp    |

* bp, base pairs.

Results

Characterization of Genomic Clones—We isolated and further characterized a total of 32 genomic clones from the λ-phage libraries using our previously isolated cDNA clones (11). These clones, which contain all exons of the gene, were aligned and mapped using restriction enzymes (Fig. 1). The overlapping clones spanned 260,000 base pairs of genomic DNA and the entire gene, with the exception of parts of introns 2, 3, 4, and 21 not contained in the clones.

Exon Size Pattern and Exon-Intron Boundaries—Restriction fragments were subcloned into Bluescript vectors for sequencing, or alternatively, the nucleotide sequences of exons were determined directly from the exon-containing λ-phage clones using cycle sequencing. The sizes of introns were determined by direct sequencing of small introns from subcloned restriction fragments or were estimated by polymerase chain reaction amplification of DNA segments between exons from genomic DNA, followed by size fractionation of the polymerase chain reaction products on agarose gels. The sizes of exons and available introns are summarized in Fig. 2 and Table I, respectively. The 364-residue N-terminal domain VI (264 residues) is encoded by exons 1–5, domain V (241 residues) by exons 5–10, domain IV (196 residues) by exons 10–14, domain IIIb (452 residues) by exons 14–23, domain IVa (204 residues) by exons 23–27, domain IIIa (194 residues) by exons 27–31, domains I + II (580 residues) by exons 32–45, and the largest domain G (957 residues) by exons 45–64.
Figure 2. Exon-intron boundaries in the human LAMA2 gene. Intronic sequences and parts of the 5'- and 3'-end untranslated sequences are depicted by lower-case letters, and exon sequences by upper-case letters, with the amino acid residues shown by the one-letter code below the first base of a codon. The amino acids are numbered according to Ref. 17. The translation stop codon tag is indicated by an asterisk. The coding sequences of the two unusually short exons 43 and 52 are underlined.
The exon sizes, excluding the 5'- and 3'-untranslated regions, vary between 6 and 270 base pairs (Fig. 2). Two of the exons are unusually small; exon 43 has only 6 base pairs, and exon 52 has 12. The remaining 62 exons vary in size between 87 and 270 base pairs. Thirty exons start with a split codon (see Fig. 2).
Assignment of Exons to EcoRI Restriction Fragments—Since the human LAMA2 gene has been shown to be mutated in patients with congenital muscular dystrophy (39, 40), it is useful to know the location of exons in restriction fragments of the gene when analyzing break points of deletions or other rearrangements of the gene in patients. To facilitate such analyses, we have assigned all exons to EcoRI restriction fragments in the gene (Table II).

**DISCUSSION**

This work, describing the entire exon sequence pattern of the human LAMA2 gene, provides the first structure of a mammalian laminin α-chain gene. The presence of 64 exons in this gene, which encodes a 9500-nucleotide transcript and the 3110-residue α2-polypeptide chain, shows that the gene is considerably larger and more complex than the human genes coding for the smaller β- and γ-chains. Thus, the LAMB1 gene of the 1786-residue β1-chain (12) has 34 exons (17); the LAMC1 gene of the 1609-residue γ1-chain (15) has 28 exons (18); and the LAMC2 gene of the 1172-residue γ2-chain (16) has 23 exons (19). In contrast, the LAMA gene of the 3712-residue α-chain of Drosophila is considerably more compact, containing only 15 exons (20). Our previous work (17, 18) has indicated that the structures of the genes for the β- and γ-type laminin chains have diverged extensively. For example, the exon size pattern of the LAMB1 gene differs significantly from that of LAMC1 and LAMC2, with the latter two showing considerable similarity in gene structure between themselves. The present results show that the α-chain genes have diverged extensively from the β- and γ-chain genes since the LAMA2 gene exon size pattern is very different from that of the LAMB1, LAMC1, and LAMC2 genes. It is likely, however, that the different α-chain genes exhibit considerable structural homology between themselves, as do the two human γ-chain genes (18, 19).

It was of interest to compare the structural relationship of the >260,000-base pair multiexon human LAMA2 gene with that of the more compact 14,000-base pair Drosophila gene. This comparison (Fig. 3) revealed that the location of intervening sequences in the two genes is poorly conserved as only two locations of intervening sequences are conserved between the two genes (dashed lines).

The complete size of the COLAA6 gene for the type IV collagen α6-chain has been shown to be 425,000 base pairs and to contain 46 exons (42). In that gene, intron 2 was estimated to be 340,000 base pairs. The human LAMC2 gene is 55,000 base pairs with 23 exons (19). The complete sizes of other type IV collagen or laminin genes have not been elucidated because of the presence of large introns not contained in the clones isolated. However, the sizes of such genes whose entire exon pattern has been determined are all large. Thus, the LAMB1 gene is >80,000 base pairs (17); LAMC1 is >60,000 base pairs (18); COLA1 is >100,000 base pairs (43); and COLA5 is >250,000 base pairs (44, 45).

The human laminin α2-chain gene has been shown to be mutated in patients with congenital muscular dystrophy. All three mutations reported to date are single base changes leading to splice site, nonsense, or missense mutations (39, 40). The present work provides the basis for detailed analysis of the gene in the disease. In the case of large gene rearrangements, the assignment of exons to specific EcoRI restriction fragments facilitates the analysis of break points. However, as has turned out to be the case for most genetic diseases, the majority of mutations are small gene changes such as deletions or insertions or single base changes within or adjacent to exons. Therefore, determination of the sequences of exon-intron boundaries is essential for mutational analysis.

In conclusion, this study has provided the first exon-intron structure of a mammalian laminin α-chain gene. This work has direct practical applications as the gene has been shown to be mutated in congenital muscular dystrophy; and therefore, this work provides the basis for mutational analysis and even future gene therapy. Also, knowledge about gene structure is a prerequisite for studies on the regulation of gene expression.

**Acknowledgment**—We thank Pascale Guicheney for critical comments on the manuscript.

**REFERENCES**

1. Engel, J. (1992) Biochemistry 31, 10643–10651
2. Tryggvason, K. (1993) Curr. Opin. Cell Biol. 5, 877–882
3. Timpl, R., and Bottino, I. (1994) Matrix Biol. 14, 275–281
4. Burgesson, R. E., Chiapet, M., Deutzmann, R., Eckelmann, P., Engel, J., Kleinman, H., Martin, G. R., Meneguzzi, G., Paulsson, M., Sanes, J., Timpl, R., Tryggvason, K., Yamada, Y., and Yurchenco, P. D. (1994) Matrix Biol. 14, 209–211
5. Richards, A. J., Al-Imarah, L., Carter, N. P., Lloyd, J. C., Leversha, M. A., and Pope, F. M. (1994) Genomics 22, 237–239
6. Ivanainen, A., Sainio, K., Sariola, H., and Tryggvason, K. (1995) FEBS Lett. 365, 183–188
7. Miner, J. H., Lewis, R. M., and Sanes, J. R. (1995) J. Biol. Chem. 270, 28523–28526
8. Nissinen, R., Vuolteenaho, R., Boot-Handford, R., Kallunki, P., and Tryggvason, K. (1991) Biochem. J. 276, 369–379
9. Harapapanta, U., Uitto, J., Ruoslahti, E., and Engvall, E. (1991) Matrix Biol. 11, 151–169
10. Vuolteenaho, R., Nissinen, M., Sainio, K., Byers, M., Eddy, M., Hirvonen, H., Showa, T. B., Sariola, H., Engvall, E., and Tryggvason, K. (1994) J. Cell Biol. 124, 381–394

**Fig. 3. Comparison of locations of intervening sequences in the human LAMA1 and Drosophila LAMA genes.** Exons coding for the different structural domains of the human laminin α2- and Drosophila α-chains are depicted by different colors, with the corresponding domains indicated at the top. Only two locations of intervening sequences are conserved between the two genes (dashed lines).
Human LAMA2 Gene

27669

11. Ryan, M. C., Tizard, R., VanDevanter, D. R., and Carter, W. G. (1994) J. Biol. Chem. 269, 22779–22787
12. Pikkarainen, T., Eddy, R., Fukushima, Y., Byers, M., Shows, T., Pihlanjami, T., Saraste, M., and Tryggvason, K. (1997) J. Biol. Chem. 272, 10454–10462
13. Ivanainen, A., Vuolteenaho, R., Sainio, K., Janit-Sait, S., Eddy R., Shows, T. B., Sariola, H., and Tryggvason, K. (1994) Matrix Biol. 13, 489–497
14. Gerecke, D. R., Wagman, D. W., Champliaud, M.-F., and Burgeson, R. E. (1994) J. Biol. Chem. 269, 11073–11080
15. Pikkarainen, T., Kallunki, T., and Tryggvason, K. (1988) J. Biol. Chem. 263, 6751–6758
16. Kallunki, F., Sainio, K., Eddy, R., Byers, M., Kallunki, T., Sariola, H., Beck, K., Hirvonen, H., Shows, T. B., and Tryggvason, K. (1992) J. Cell Biol. 119, 679–693
17. Vuolteenaho, R., Chow, L. T., and Tryggvason, K. (1990) J. Biol. Chem. 265, 15611–15616
18. Kallunki, T., Ikonen, J., Chow, L., Kallunki, P., and Tryggvason, K. (1994) Genomics 11, 22779–22787
19. Airenne, T., Haakana, H., Sainio, K., Kallunki, T., Kallunki, P., Sariola, H., Beik, K., Hirvonen, H., Shows, T. B., and Tryggvason, K. (1994) J. Biol. Chem. 269, 675–679
20. Kusche-Gullberg, M., Garrison, K., MacKrell, A.-J., Fessler, L. I., and Fessler, J. H. (1992) J. Biol. Chem. 267, 4519–4527
21. Nagayoshi, T., Mattei, M.-G., Passage, E., Knowlton, R., Chu, M.-L., and Uitto, J. (1989) Genomics 5, 932–935
22. Wewer, U. M., Gerecke, D. R., Durkin, M. E., Kurtz, K. S., Mattei, M.-G., Champliaud, M.-F., Burgeson, R. E., and Albrechtsen, R. (1994) Genomics 24, 243–252
23. Vailly, J., Szepetowski, F., Mattei, M.-G., Pedeutour, F., Burgeson, R., Ortonne, J. P., and Meneguzzi, G. (1994) Genomics 21, 286–288
24. Fukushima, Y., Pikkarainen, T., Kallunki, T., Eddy, R. L., Byers, M. G., Haley, L. L., Henry, W. M., Tryggvason, K., and Shows, T. B. (1988) Cytogenet. Cell Genet. 48, 137–141
25. Leivo, I. and Engvall, E. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 1544–1548
26. Ehrig, K., Leivo, I., Argraves, W. S., Rusoalhti, E., and Engvall, E. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 3264–3268
27. Engvall, E., Earwicker, D., Haaparanta, T., Rusoalhti, E., and Sanes, J. R. (1990) Cell Regul. 1, 731–740
28. Sanes, J. R., Engvall, E., Butkowski, R., and Hunter, D. D. (1990) J. Cell Biol. 111, 1685–1699
29. Haio, L. L., Engvall, E., Peltonen, J., and Uitto, J. (1993) Lab. Invest. 68, 100–108
30. Campbell, K. P. (1995) Cell 80, 675–679
31. Wortman, R. (1995) Science 270, 755–756
32. Monaco, A. P., Neve, R. L., Colletti-Fenner, C., Bertelson, C. J., Kurnit, D. M., and Kunkel, L. M. (1986) Nature 323, 646–650
33. Robbers, S. L., Leturcq, F., Allamand, V., Piccolo, F., Jeanpierre, V., Andersson, R. D., Lim, I. E., Lee, J. C., Tomé, P. M., Romero, N. B., Fardeau, M., Beckmann, J. S., Kaplan, J.-C., and Campbell, K. P. (1994) Cell 78, 625–633
34. Binneweg, C. G., Modi, R., Neguchi, S., Mizuno, Y., Yoshida, M., Gussoni, E., McNally, E. M., Duggan, D. J., Hoffman, E. P., Ozawa, E., and Kunkel, L. M. (1995) Nat. Genet. 11, 266–272
35. Lim, L. E., Ducloux, F., Broix, O., Bourg, N., Sunada, Y., Allamand, V., Meyer, J., Richard, I., Moomaw, C., Slaughter, C., Tomé, F. M. S., Fardeau, M., Jackson, C. E., Beckmann, J. S., and Campbell, K. P. (1995) Nat. Genet. 11, 257–264
36. Neguchi, S., McNally, E. M., Othamane, K. B., Hagiwara, Y., Mizuno, Y., Yoshida, M., Yamamoto, H., Bönemann, C. G., Gussoni, E., Denton, P. H., Kyriakides, T., Middleton, L., Hentati, F., Hamida, B. H., Nonaka, I., Vance, J. M., Kunkel, L. M., and Ozawa, K. (1995) Science 270, 819–822
37. Piccolo, F., Robbers, S. L., Jeanpierre, M., Leturcq, F., Azibi, K., Beldjord, C., Carrié, A., Récan, D., Chao, M., Reghis, A., El Kerch, F., Stefani, A., Vait, T., Merline, I., Collin, H., Eymard, B., Beckmann, J. S., Romero, N. B., Tomé, F. M. S., Fardeau, M., Campbell, K. P., and Kaplan, J.-C. (1995) Nat. Genet. 10, 243–245
38. Xu, H., Wu, X. R., Wewer, U. M., and Engvall, E. (1994) Nat. Genet. 8, 297–302
39. Heibing-Leclerc, A., Zhang, X., Topaloglu, H., Cruaud, C., Tesson, F., Weissenbach, J., Tomé, F. M. S., Schwartz, K., Fardeau, M., Tryggvason, K., and Guicheney, P. (1995) Nat. Genet. 11, 216–218
40. Niesinski, M., Heibing-Leclerc, A., Zhang, X., Evangelista, T., Topaloglu, H., Cruaud, C., Weissenbach, J., Fardeau, M., Tomé, F. M. S., Schwartz, K., Tryggvason, K., and Guicheney, P. (1996) Am. J. Hum. Genet., 58, 1177–1184
41. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY