Volume 17 Number 22 1989 Nucleic Acids Research

Nucleotide sequence of the full length cDNA encoding for human type II procollagen

Ming-Wan Su[1,2], Brendan Lee[1,2], Francesco Ramirez[1]*, Mirta Machado[3] and William Horton[3]

[1]Brookdale Center for Molecular Biology, Mt Sinai School of Medicine, PO Box 1128A, New York, NY 10029, [2]Department of Microbiology and Immunology, SUNY Health Science Center at Brooklyn, Brooklyn, NY 11209 and [3]Department of Pediatrics, University of Texas Medical School, PO Box 20708, Houston, TX 77025, USA

Submitted October 18, 1989 EMBL accession no. X16468

Type II collagen is a homotrimeric molecule composed of three identical α1(II) collagen chains that are synthesized as precursor procollagens with amino- and carboxy-terminal propeptides that are removed extracellularly.[1] Mutations in the α1(II) chain result in disproportionate dwarfing disorders in the spectrum of skeletal dysplasias.[2,3] Hitherto, only the propeptides' sequences have been published.[4,5] Here, we report the isolation of the full-length α1(II) cDNA and derived from it, the nucleotide sequence of the entire mature α1(II) chain. The availability of these sequences will expedite the molecular characterization of other type II collagen disorders in the skeletal dysplasias.

```
   1 CAGATGGCTGGAGGATTTGATGAAAAGGCTGGTGGCGCCCAGTTGGGAGTAATGCAAGGACCAATGGCCCCATG
  76 GGACCTCGAGGACCTCCAGGCCCTGCAGGTGCTCCTGGGCCTCAAGGATTTCAAGGCAATCCTGGTGAACCTGGT
 151 GAACCTGGTGTCTCTGGTCCCATGGGTCCCCGTGGTCCTCCTGGTCCCCCTGGAAAGCCTGGTGATGATGGTGAA
 226 GCTGGAAAACCTGGAAAAGCTGGTGAAAGGGGTCCGCCTGGTCCTCAGGGTGCTCGTGGTTTCCCAGGAACCCCA
 301 GGCCTTCCTGGTGTCAAAGGTCACAGAGGTTATCCAGGCCTGGACGGTGCTAAGGGAGAGGCGGGTGCTCCTGGT
 376 GTGAAGGGTGAGAGTGGTTCCCCGGGTGAGAACGGATCTCCGGGCCCAATGGGTCCTCGTGGCCTGCCTGGTGAA
 451 AGAGGACGGACTGGCCCTGCTGGCGCTGCGGGTGCCCGAGGCAACGATGGTCAGCCAGGCCCCGCAGGGCCTCCG
 526 GGTCCTGTCGGTCCTGCTGGTGGTCCTGGCTTCCCTGGTGCTCCTGGAGCCAAGGGTGAAGCCGGCCCCACTGGT
 601 GCCCGTGGTCCTGAAGGTGCTCAAGGTCCTCGCGGTGAACCTGGTACTCCTGGGTCCCCTGGGCCTGCTGGTGCC
 676 TCCGGTAACCCTGGAACAGATGGAATTCCTGGAGCCAAAGGATCTGCTGGTGCTCCTGGCATTGCTGGTGCTCCT
 751 GGCTTCCCTGGGCCACGGGGCCCTCCTGACCCTCAAGGTGCAACTGGTCCTCTGGGCCCGAAAGGTCAGACGGGT
 826 AAACCTGGTATTGCTGGCTTCAAAGGTGAACAAGGCCCCAAGGGAGAACCTGGCCCTGCTGGCCCCCAGGGAGCC
 901 CCTGGACCCGCTGGTGAAGAAGGCAAGAGAGGTGCCCGTGGAGAGCCTGGTGGCGTTGGGCCCATCGGTCCCCCT
 976 GGAGAAAGAGGTGCTCCCGGCAACCGCGGTTTCCCAGGTCAAGATGGTCTGGCAGGTCCCAAGGGAGCCCCTGGA
1051 GAGCGAGGGCCCAGTGGTCTTGCTGGCCCCAAGGGAGCCAACGGTGACCCTGGCCGTCCTGGAGAACCTGGCCTT
1126 CCTGGAGCCCGGGGTCTCACTGGCCGCCCTGGTGATGCTGGTCCTCAAGGCAAAGTTGGCCCTTCTGGAGCCCCT
1201 GGTGAAGATGGTCGTCCTGGACCTCCAGGTCCTCAGGGGGCTCGTGGGCAGCCTGGTGTCATGGGTTTCCCTGGC
1276 CCCAAAGGTGCCAACGGTGAGCCTGGCAAAGCTGGTGAGAAGGGACTGCCTGGTGCTCCTGGTCTGAGGGGTCTT
1351 CCTGGCAAAGATGGTGAGACAGGTGCTGAAGGCCCCCCTGGCCCTGCTGGACCTGCTGGTGAACGAGGCGAGCAG
1426 GGTGCTCCTGGGCCATCTGGGTTCCAGGGACTTCCTGGCCCTCCTGGTCCCCCAGGTGAAGGTGGAAAACCAGGT
1501 GACCAGGGTGTTCCCGGTGAAGCTGGAGCCCCTGGACTAGTGGGTCCCAGGGGTGAACGAGGTTTCCCAGGTGAA
1576 CGTGGCTCTCCCGGTGCCCAGGGCCTCCAGGGTCCCCGTGGCCTCCCCGGCACTCCTGGCACTGATGGTCCCAAA
1651 GGTGCATCTGGCCCAGCAGGCCCCCCTGGGGCTCAGGGCCCTCCAGGTCTTCAGGGAATGCCTGGCGAGAGGGGA
1726 GCAGCTGGTATCGCTGGGCCCAAAGGCGACAGGGGTGACGTTGGTGAGAAAGGCCCTGAGGGAGCCCCTGGAAAG
1801 GATGGTGGACGAGGCCTGACAGGTCCCATTGGCCCCCCTGGCCCAGCTGGTGCTAATGGCGAGAAGGGAGAAGTT
1876 GGACCTCCTGGTCCTGCAGGAAGTGCTGGTGCTCGTGGCGCTCCGGGTGAACGTGGAGAGACTGGGCCCCCCGGG
1951 ACCAGCGGGATTGCTGGGCCTCCTGGTGCTGATGGCCAGCCTGGGGCCAAGGGTGAGCAAGGAGAGGCCGGCCAG
2026 AAAGGCGATGCTGGTGCCCCTGGTCCTCAGGGCCCCTCTGGAGCACCTGGGCCTCAGGGTCCTACTGGAGTGACT
2101 GGTCCTAAAGGAGCCCGAGGTGCCCAAGGCCCCCCGGGAGCCACTGGATTCCCTGGAGCTGCTGGCCGCGTTGGA
2176 CCCCCAGGCTCCAATGGCAACCCTGGACCCCCTGGTCCCCCTGGTCCTTCTGGAAAAGATGGTCCCAAAGGTGCT
2251 CGAGGAGACAGCGGCCCCCCTGGCCGAGCTGGTGAACCCGGCCTCCAAGGTCCTGCTGGACCCCCTGGCGAGAAG
2326 GGAGAGCCTGGAGATGACGGTCCCTCTGGTGCCGAAGGTCCACCAGGTCCCCAGGGTCTGGCTGGTCAGAGAGGC
2401 ATCGTCGGTCTGCCTGGGCAACGTGGTGAGAGAGGATTCCCTGGCTTGCCTGGCCCGTCGGGTGAGCCCGGCCAG
2476 CAGGGTGCTCCTGGAGCATCTGGAGACAGAGGTCCTCCTGGCCCCGTGGGTCCTCCTGGCCTGACGGGTCCTGCA
2551 GGTGAACCTGGACGAGAGGGAAGCCCCGGTGCTGATGGCCCCCCTGGCAGAGATGGCGCTGCTGGAGTCAAGGGT
2626 GATCGTGGTGAGACTGGTGCTGTGGGAGCTCCTGGAGCCCCTGGGCCCCCTGGCTCCCCTGGCCCCGCTGGTCCA
2701 ACTGGCAAGCAAGGAGACAGAGGAGAAGCTGGTGCACAAGGCCCCATGGGACCCTCAGGACCAGCTGGAGCCCGG
2776 GGAATCCAGGGTCCTCAAGGCCCCAGAGGTGACAAAGGAGAGGCTGGAGAGCCTGGCGAGAGAGGCCTGAAGGGA
2851 CACCGTGGCTTCACTGGTCTGCAGGGTCTGCCCGGCCCTCCTGGTCCTTCTGGAGACCAAGGTGCTTCTGGTCCT
2926 GCTGGTCCTTCTGGCCCTAGAGGTCCTCCTGGCCCCGTCGGTCCCTCTGGCAAAGATGGTGCTAATGGAATCCCT
3001 GGCCCCATTGGGCCTCCTGGTCCCCGTGGACGATCAGGCGAAACCGGCCCTGCTGGTCCTCCTGGAAATCCTGGA
3076 CCCCCTGGTCCTCCAGGTCCCCCTGGCCCTGGCATCGACATGTCCGCCTTTGCTGGCTTAGGCCCGAGAGAGAAG
3151 GGCCCCGACCCCCTGCAGTACATGCGGGCC
```

*To whom correspondence should be addressed

References

1) Miller,E.J. and Gay,S. (1987) Meth. in Enz., 144,3-40. 2) Lee,B. et al. (1989) Science, 244,978-980. 3) Vissing,H. et al. (1989) J. Biol. Chem., in press. 4) Cheah,K., et al. (1985) Proc. Natl. Acad. Sci. USA, 82,2555-2559. 5) Su,M.W. et al. (1989) Genomics, 4,438-441.

