SHORT COMMUNICATION

Synteny between the Loci for a Novel FACIT-like Collagen Locus (D6S228E) and α1(IX) Collagen (COL9A1) on 6q12–q14 in Humans

HIDEKATSU YOSHIOKA,*,† HUI ZHANG,† FRANCESCO RAMIREZ,†,[1] MARIE-GENEVIEVE MATTEI,‡ MAHNAZ MORADI-AMELI,§ MICHEL VAN DER REST,§ AND MARION K. GORDON[||]

*Department of Molecular Biology and Biochemistry, Okayama University Medical School, Okayama 700, Japan; †Brookdale Center for Molecular Biology, Mount Sinai School of Medicine, New York, New York 10029; ‡U.242 INSERM, Hôpital d'Enfants de la Timone, Marseille 13385, France; §Institut de Biologie et Chimie des Proteines, CNRS UPR 412, Villeurbanne 69622, France; and ||Department of Anatomy and Cellular Biology, Tufts Medical School, Boston, Massachusetts 02111

Received September 11, 1991; revised March 13, 1992

A 1.8-kb cDNA encoding portion of a novel collagenous chain was isolated from a human rhabdomyosarcoma cell line by cross-hybridization using a chicken type V collagen probe. Sequence analysis suggests that this chain belongs to the recently discovered group of collagens, termed the FACIT class of macromolecules. This cDNA was used to locate the corresponding gene (D6S228E) to chromosome 6, notably at position 6q12–q14. Interestingly, within this region of human chromosome 6 resides the α1(IX) collagen gene (COL9A1), a member of the FACIT group.

A recently discovered group of collagens (types IX, XII, and XIV) is believed to provide molecular connections between fibrils and/or between fibrils and other components of the extracellular matrix (2). Structurally, these macromolecules exhibit stretches of triple helical sequences interrupted by noncollagenous domains (NC domains), often containing cysteinyl residues (2). Based on these characteristics, these collagens have been named FACIT for *f*ibril-*a*ssociated *c*ollagens with *i*nterrupted *t*riple helices (2). Here, we report the localization of a novel FACIT-like gene collagen to the same region of chromosome 6, where α1(IX) collagen (COL9A1) has been previously mapped.

A human rhabdomyosarcoma cDNA library was screened at low stringency with a chicken α1(V) collagen probe (Gordon, unpublished data) in an attempt to isolate the cognate human sequences. Sequencing one of the several positive clones, HY-67, revealed that this 1.8-kb cDNA codes for a portion of a novel collagen chain, arbitrarily termed α1(Y) (Fig. 1).

More than half of the conceptual amino acid translation of HY-67 includes two stretches of collagenous sequences separated by a 20-amino-acid noncollagenous domain (Fig. 1). Both collagenous sequences exhibit imperfections or short interruptions, typical of nonfibrillar collagens (2). Sequences derived from the 5′ end of the cDNA encode an incomplete N-terminal domain consisting of a globular region characterized by the presence of 10 cysteinyl residues (Fig. 1). Database searches of the sequence encoded by HY-67 revealed that the N-terminal domain of α1(Y) is related somewhat to the N-terminal NC4 and NC3 domains of the α1(IX) and α1(XII) collagens, respectively. A feature of these FACIT molecules is a stretch of about 200 residues that contains related sequences and four conserved cysteines (2). These cysteines are also present in the N-terminal domain of α1(Y) (amino acids 9, 164, 207, and 217 in Fig. 1). Using these four residues as reference points, the N-terminal domain of α1(Y) was aligned with those of chicken α1(IX) and α1(XII) collagens (1, 6). By taking into account conservative changes, this comparative analysis revealed a 43% identity between α1(Y) and the two avian chains (data not shown). Based on these data, α1(Y) is likely to be an additional member of the FACIT group.

To determine the chromosomal location of the α1(Y) gene (D6S228E), *in situ* hybridization on spreads of metaphase chromosomes was employed using previously described conditions (5). In the 150 metaphase cells examined, after *in situ* hybridization, there were 201 silver grains associated with chromosomes. Of these, 103 (51.2%) were located on chromosome 6. The distribution of grains on this chromosome was nonrandom. To be precise, 81.5% of the grains (84/103) mapped to the [q12–q14] region of the chromosome 6 long arm with a maximum in the 6q13 band. Specific labeling of chromosome 6 is shown in Fig. 2. These results were essentially identical to those previously obtained with the COL9A1 probe (4).

In conclusion, in this study we have partially characterized a novel gene that exhibits a strong structural homology to the FACIT group of collagens. More importantly, we have also documented the first cytological linkage between two members of this collagen subgroup.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. M63597.

[1] To whom correspondence should be addressed at Brookdale Center for Molecular Biology, Mt. Sinai School of Medicine, Box 1126, 1 Gustave Levy Place, New York, NY 10029.

0888-7543/92 $5.00

```
GTT AGG GAC AAG ACA GAA GAG TCA TGC CCT ATC CTG AGA ATA GAG GGA CAT   51
Val Arg Asp Lys Thr Glu Glu Ser Cys Pro Ile Leu Arg Ile Glu Gly His   17

CAG CTG ACA TAT GAC AAC ATA AAC AAA CTT GAA GTT TCA GGT TTT GAT CTA  102
Gln Leu Thr Tyr Asp Asn Ile Asn Lys Leu Glu Val Ser Gly Phe Asp Leu   34

GGA GAC AGC TTT TCT CTA AGA CGT GCA TTT TGT GAA AGT GAT AAA ACC TGT  153
Gly Asp Ser Phe Ser Leu Arg Arg Ala Phe Cys Glu Ser Asp Lys Thr Cys   51

TTC AAA TTG GGA AGT GCA CTT CTT ATT AGA GAC ACT ATT AAG ATA TTT CCC  204
Phe Lys Leu Gly Ser Ala Leu Leu Ile Arg Asp Thr Ile Lys Ile Phe Pro   68

AAA GGC CTT CCT GAG GAG TAC TCA GTA GCT GCC ATG TTT CGA GTA CGA AGA  255
Lys Gly Leu Pro Glu Glu Tyr Ser Val Ala Ala Met Phe Arg Val Arg Arg   85

AAC GCC AAA AAG GAA CGG TGG TTT CTG TGG CAG GTT TTA AAC CAG CAG AAT  306
Asn Ala Lys Lys Glu Arg Trp Phe Leu Trp Gln Val Leu Asn Gln Gln Asn  102

ATT CCA CAG ATT TCT ATA GTA GTT GAT GGT GGA AAG AAG GTG GTG GAA TTT  357
Ile Pro Gln Ile Ser Ile Val Val Asp Gly Gly Lys Lys Val Val Glu Phe  119

ATG TTT CAA GCC ACA GAG GGA GAT GTG TTG AAC TAC ATT TTT AGA AAT CGA  408
Met Phe Gln Ala Thr Glu Gly Asp Val Leu Asn Tyr Ile Phe Arg Asn Arg  136

GAA CTC CGT CCT TTG TTT GAT CGT CAG TGG CAC AAA CTT GGC ATT AGT ATA  459
Glu Leu Arg Pro Leu Phe Asp Arg Gln Trp His Lys Leu Gly Ile Ser Ile  153

CAA TCC CAG GTC ATT TCA CTT TAT ATG GAT TGT AAT TTA ATT GCG AGG AGG  510
Gln Ser Gln Val Ile Ser Leu Tyr Met Asp Cys Asn Leu Ile Ala Arg Arg  170

CAG ACT GAT GAA AAG GAC ACT GTG GAT TTC CAT GGA CGG ACA GTT ATT GCT  561
Gln Thr Asp Glu Lys Asp Thr Val Asp Phe His Gly Arg Thr Val Ile Ala  187

ACG CGA GCT TCA GAT GGC AAG CCT GTG GAT ATT GAA CTT CAC CAA CTT AAA  612
Thr Arg Ala Ser Asp Gly Lys Pro Val Asp Ile Glu Leu His Gln Leu Lys  204

ATC TAC TGC AGT GCA AAC CTC ATA GCT CAA GAA ACA TGT TGT GAA ATA TCA  663
Ile Tyr Cys Ser Ala Asn Leu Ile Ala Gln Glu Thr Cys Cys Glu Ile Ser  221

GAT ACT AAG TGC CCA GAG CAG GAT GGC TTT GGA AAT ATT GCA TCA TCA TGG  714
Asp Thr Lys Cys Pro Glu Gln Asp Gly Phe Gly Asn Ile Ala Ser Ser Trp  238

GTA ACT GCT CAT GCC AGT AAA ATG TCT TCA TAT CTG CCA GCA AAG CAG GAA  765
Val Thr Ala His Ala Ser Lys Met Ser Ser Tyr Leu Pro Ala Lys Gln Glu  255

CTT AAA GAC CAG TGC CAG TGC ATT CCA AAC AAG GGA GAA GCA GGA TTA CCA  816
Leu Lys Asp Gln Cys Gln Cys Ile Pro Asn Lys Gly Glu Ala Gly Leu Pro  272

GGA GCT CCG GGT TCA CCT GGG CAG AAA GGG CAT AAA GGA GAG CCG GGT GAA  867
Gly Ala Pro Gly Ser Pro Gly Gln Lys Gly His Lys Gly Glu Pro Gly Glu  289

AAT GGT TTA CAT GGT GCT CCA GGA TTC CCT GGT CAA AAG GGA GAG CAA GGT  918
Asn Gly Leu His Gly Ala Pro Gly Phe Pro Gly Gln Lys Gly Glu Gln Gly  306

TTT GAA GGC AGC AAA GGA GAA ACT GGT GAA AAG GGT GAA CAA GGA GAA AAA  969
Phe Glu Gly Ser Lys Gly Glu Thr Gly Glu Lys Gly Glu Gln Gly Glu Lys  323

GGA GAT CCA GCT CTT GGC TGC CTT AAT GGA GAA AAT GGT TTG AAA GGT GTC 1020
Gly Asp Pro Ala Leu Gly Cys Leu Asn Gly Glu Asn Gly Leu Lys Gly Val  340

TTG GGT CCT CAT GGT CCA CCT GGC CCA AAA GGA GAA AAG GGA GAT ACA GGA 1071
Leu Gly Pro His Gly Pro Pro Gly Pro Lys Gly Glu Lys Gly Asp Thr Gly  357

CCC CCA GGA CCA CCA GCC TTA CCT GGT TCC CTG GGG ATA CAA GGC CCC CAA 1122
Pro Pro Gly Pro Pro Ala Leu Pro Gly Ser Leu Gly Ile Gln Gly Pro Gln  374

GGT CCA CCT GGA AAA GAG GGT CAG AGG GGA AGA CGA GGG AAA ACA GGA CCT 1173
Gly Pro Pro Gly Lys Glu Gly Gln Arg Gly Arg Arg Gly Lys Thr Gly Pro  391

CCC GGA AAA CCA GGA CCC CCA GGA CCA CCT GGA CCT CCT GGA ATA CAA GGA 1224
Pro Gly Lys Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Ile Gln Gly  408

ATA CAC CAA ACT CTT GGT GGA GAT GAT AAC AAG GAT AAC AAG GGA AAT GAT 1275
Ile His Gln Thr Leu Gly Gly Asp Asp Asn Lys Asp Asn Lys Gly Asn Asp  425

GAA CAT GAA GCT GGA GGC CTG AAA GGA GAC AAG GGT GAA ACT GGA CTA CCA 1326
Glu His Glu Ala Gly Gly Leu Lys Gly Asp Lys Gly Glu Thr Gly Leu Pro  442

GGA TTT CCA GGG TCT GTT GGC CCT AAA GGA CAA AAA GGA GAA CCT GGA GAG 1377
Gly Phe Pro Gly Ser Val Gly Pro Lys Gly Gln Lys Gly Glu Pro Gly Glu  459

CCT TTT ACA AAA GGA GAA AAA GGA GAT AGA GGA GAA CCT GGG GTA ATA GGA 1428
Pro Phe Thr Lys Gly Glu Lys Gly Asp Arg Gly Glu Pro Gly Val Ile Gly  476

TCA CAG GGA GTA AAG GGT GAA CCT GGA GAT CCC GGA CCC CCT GGT TTA ATA 1479
Ser Gln Gly Val Lys Gly Glu Pro Gly Asp Pro Gly Pro Pro Gly Leu Ile  493

GGA AGC CCA GGA CTA AAG GGT CAG CAA GGA TCT GCA GGC TCC ATG GGA CCC 1530
Gly Ser Pro Gly Leu Lys Gly Gln Gln Gly Ser Ala Gly Ser Met Gly Pro  510

AGA GGA CCG CCA GGA GAT GTT GGA TTG CCA GGA GAA CAT GGT ATC CCA GGA 1581
Arg Gly Pro Pro Gly Asp Val Gly Leu Pro Gly Glu His Gly Ile Pro Gly  527

AAA CAA GGC ATT AAA GGA GAA AAG GGA GAT CCA GGT GGG ATC ATA GGC CCT 1632
Lys Gln Gly Ile Lys Gly Glu Lys Gly Asp Pro Gly Gly Ile Ile Gly Pro  544

CCC GGG CTT CCA GGT CCA AAA GGT GAG GCT GGT CCT CCA GGG AAA AGC CTG 1683
Pro Gly Leu Pro Gly Pro Lys Gly Glu Ala Gly Pro Pro Gly Lys Ser Leu  561

CCA GGG GAA CCA GGA TTA GAT GGA AAT CCT GGA GCA CCT GGT CCA CGT GGG 1734
Pro Gly Glu Pro Gly Leu Asp Gly Asn Pro Gly Ala Pro Gly Pro Arg Gly  578

CCA AAG GGT GAA AGA GGA CTT CCA GGT GTT CAC GGT TCC CCA GGG GAC ATA 1785
Pro Lys Gly Glu Arg Gly Leu Pro Gly Val His Gly Ser Pro Gly Asp Ile  595

GGC CAA AGG GAT
Gly Gln Arg Asp   599
```

FIG. 1. DNA sequence and deduced amino acid sequence of HY-67. Both DNA strands of HY-67 were sequenced several times according to the protocol of (7), after generating nested deletions by the exon III/mung bean nucleases method (3). Collagenous sequences are underlined with dotted lines highlighting short interruptions. Cysteinyl residues are boxed. Nucleotides of the HY-67 insert are numbered on the right.

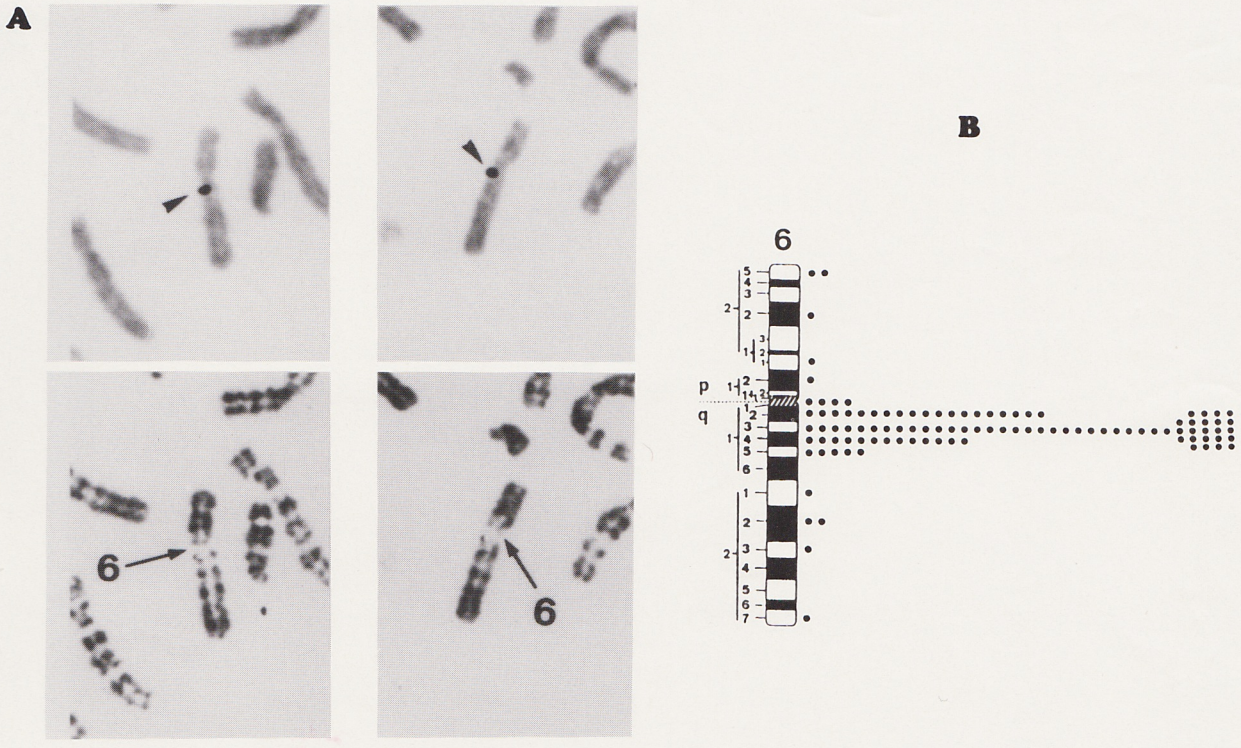

FIG. 2. Localization of human (COLYA1) to chromosome 6. (**A**) Two partial human metaphases showing the specific site of hybridization to chromosome 6. (**Top**) Arrowheads indicate silver grains on Giemsa-stained chromosomes, after autoradiography. (**Bottom**) Chromosomes with silver grains were subsequently identified by R-banding (F. P. G. Technique). (**B**) Idiogram of the human G-banded chromosome 6 illustrating the distribution of labeled sites for the HY-67 probe.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (AR-38648 and EY-09056). The authors thank Drs. B. R. Olsen and T. F. Linsenmayer for many helpful discussions and Ms. R. Lingeza for typing the manuscript. This is article number 53 from the Brookdale Center for Molecular Biology at the Mt. Sinai School of Medicine.

REFERENCES

1. Gordon, M. K., Gerecke, D. R., Dublet, B., van der Rest, M., and Olsen, B. R. (1989). Type XII collagen: a large multidomain molecule with partial homology to type IX collagen. *J. Biol. Chem.* **264:** 19772–19778.
2. Gordon, M. K., and Olsen, B. R. (1990). The contribution of collagenous proteins to tissue-specific matrix assemblies. *Curr. Op. Cell Biol.* **2:** 833–838.
3. Henikoff, S. (1987). Unidirectional digestion with exonuclease III in DNA sequence analysis. *Meth. Enzym.* **155:** 156–165.
4. Kimura, T., Mattei, M-G., Stevens, J. W., Goldring, M. B., Ninomiya, Y., and Olsen, B. R. (1989). Molecular cloning of rat and human type IX collagen cDNA and localization of the α1(IX) gene on the human chromosome 6. *Eur. J. Biochem.* **179:** 71–78.
5. Mattei, M-G., Philip, N., Passage, E., Moisan, J. P., Mandel, J. L., and Mattei, J. F. (1985). DNA probe localization at 18p113 band by *in situ* hybridization and identification of a small supernumerary chromosome. *Hum. Genet.* **69:** 268–271.
6. Vasios, G., Nishimura, I., Konomi, H., van der Rest, M., Ninomiya, Y., and Olsen, B. R. (1988). Cartilage type IX collagen-proteoglycan contains a large amino-terminal globular domain encoded by multiple exons. *J. Biol. Chem.* **263:** 2324–2329.
7. Zagursky, R. J., Berman, M. L., Baumeister, K., and Lomax, N. (1986). Rapid and easy sequencing of large linear double stranded DNA and supercoiled plasmid DNA. *Gene Anal. Techn.* **2:** 232–238.