The Human Tenascin-R Gene*

Alessandra Leprini, Roberto Gherzi, Annalisa Siri, Germano Querze, Francesca Viti, and Luciano Zardi‡

From the Laboratory of Cell Biology, Istituto Nazionale per la Ricerca sul Cancro, Largo R. Benzi, 10, 16132 Genova, Italy

The human tenasin-R gene encodes a multidomain protein belonging to the tenasin family, until now detected only in the central nervous system. During embryonic development, tenasin-R is presumed to play a pivotal role in axonal path finding through its adhesive and repulsive properties. Recently, the primary structure of human tenasin-R has been elucidated (Carnemolla, B., Leprini, A., Borsi, L., Querze, G., Urbini, S., and Zardi, L. (1996) J. Biol. Chem. 271, 8157–8160). As a further step to investigate the role of human tenasin-R, we defined the structure of its gene. The gene, which spans a region of chromosome 1 approximately 85 kilobases in length, consists of 21 exons, ranging in size from 90 to >670 base pairs. The sequence analysis of intron splice donor and acceptor sites revealed that the position of introns in human tenasin-R are precisely conserved in the other two tenasin family members, tenasin-C and tenasin-X. The determination of intronic sequences flanking two tenascin family members, tenasin-C and tenasin-X. The determination of intronic sequences flanking the exons boundaries will allow investigation of whether mutations may be responsible for altered function of the gene product(s) leading to central nervous system development defects.

The tenasin (TN)1 family includes at least four distinct genes in mammalians: tenasin-C, tenasin-X, tenasin-R, and tenasin-Y (1, 2). These genes are expressed in distinct tissues at different times during embryonic development and adult life, and their expression is presumed to be selectively altered during distinct pathological conditions (reviewed in Refs. 1–3). In the human, the genes encoding the tenasins are located on distinct chromosomes, and the expression of each tenasin gene is likely controlled by independent mechanisms. The structure of the genes coding for human TN-C and TN-X has recently been elucidated (4–6).

The extracellular matrix protein TN-R (previously denominated restrictin in chicken and janusin or J1-160/180 in rats) is the smallest member of the tenasin family and, like the three other TN family members, is a modular glycoprotein composed of four structural motifs. From the NH2 to the COOH terminus, human TN-R contains: (i) a cysteine-rich segment unique to all the TN family members; (ii) 4.5 epidermal growth factor-like repeats; (iii) 9 fibronectin type III-like (TNfn) repeats; and (iv) a fibrinogen-like sequence (7–9; for review, see Ref. 10). Differently from TN-C and TN-X, which have been found in many tissues and organs (11), TN-R seems to display a much more selective expression pattern restricted to the nervous system. More specifically, unlike TN-C, which is expressed early during development and in both the central and the peripheral nervous systems, TN-R has so far been detected only in the central nervous system (CNS) (7–9). In chickens and rats, TN-R has been localized in the cerebellum, hippocampus, olfactory bulb, retina, and spinal cord (10, 12, 13). TN-R has been reported to be an oligodendrocyte-derived molecule (14); it is maximally expressed during the time of active myelination, and it has been found to be associated with the surface of myelinating oligodendrocytes, neurons, and type 2 astrocytes (12–15). In the chicken brain, TN-R has been detected at embryonic days 6–16 but is barely detectable after birth (7). Finally, TN-R is also synthesized by subpopulations of neurons at early developmental stages, when active dendritic and axonal outgrowth and synaptogenesis occurs (12).

The transient expression of TN-R in axon-rich regions of selected areas of the developing nervous system suggested an involvement of the molecule in axon growth (16–18). Furthermore, the expression of TN-R by oligodendrocytes that are nonpermissive for axonal growth in vitro (19) implied that TN-R may contribute to confine axonal pathways during development. Studies aimed at clarifying the functions of TN-R using recombinant fragments indicated that the molecule possesses both nonpermissive and adhesive properties, as has also been proposed for TN-C (16). Altogether, these findings suggested an involvement of TN-R in CNS plasticity during development (10).

Despite the extensive and rapid accumulation of data on both the expression pattern and the cellular activation mechanisms mediated by TN-R, little is known about the molecular genetics of the protein. The spatiotemporal pattern of expression of the protein during CNS development suggested that mutations at the TN-R locus, mapped to the q region of chromosome 1 (9), could be implicated in alterations of CNS morphogenesis.

As a further step to gain insight into possible roles of TN-R, we have now determined the structure of its gene. The tenasin-R gene consists of 21 exons. Exon boundaries and flanking intronic sequences observed suggest a remarkable evolutionary similarity between the members of the human tenasin family. The detailed knowledge of the TN-R gene structure may serve as a prerequisite to investigate whether mutations may be responsible for altered functions of the gene product(s) leading to CNS malformations.

EXPERIMENTAL PROCEDURES

Molecular Cloning—A human genomic Aflix II phage library (Stratagene, La Jolla, CA) prepared from human placenta was screened with 32P-labeled probes derived from human TN-R cDNA clones according to standard procedures (20). We identified 10 positive clones, and purified plaques were initially analyzed by hybridization on replica filters with 32P-labeled human TN-R cDNA fragments corresponding to different TN-R domains. DNA were prepared from homogeneous populations of...
positive phages (20), and further mapping of the genomic clones was achieved by DNA sequencing and polymerase chain reaction (PCR) amplification. The size of TN-R exon 1 was determined by S1 nuclease analysis using a 32P-labeled human TN-R genomic fragment as previously reported (6). The genomic clones further characterized are schematically represented in Fig. 1.

The location of exons in the genomic clones was determined by PCR amplification using specific primers (purchased from TIB MolBiol, Genova, Italy) that annealed to known sequences within the human TN-R cDNA. Similarly, we amplified the introns and estimated their size on agarose gels. The PCR products were sequenced to verify their identity.

DNA Sequencing and PCR Amplification—Genomic clones were partially sequenced by the dideoxy-nucleotide chain termination method (20) using a Sequenase version 2.0 kit (U. S. Biochemical Corp.) according to the manufacturer's instructions and both vector- and sequence-specific primers. DNA sequences were analyzed and compared using Microgenie (Beckman Instruments) for IBM PCs and GeneWorks 2.3.1® software for Macintosh computers. Multiple sequence alignments were performed using both the specific program in the GeneWorks 2.3.1 software and the Megalign program of DNASTar (Madison, WI).

To obtain completely overlapping genomic sequences, three TN-R genomic fragments were obtained by PCR amplification (Fig. 1). PCR amplifications were performed using human genomic DNA (Boehringer Mannheim) as the template and different sets of primers localized in TN-R gene exons 2 and 3, in exons 7 and 9 and in exons 15 and 18, respectively. PCR amplifications were carried out using the Expand long template PCR system (Boehringer Mannheim) according to the manufacturer’s instructions and both vector- and sequence-specific primers; PCR products were purified on agarose gels before sequencing.

RESULTS AND DISCUSSION

We have recently described the cDNA sequence of the entire coding region of human TN-R and localized the gene to the q region of chromosome 1 (9). Here we report on the molecular characterization of the human gene encoding TN-R: the gene consists of 21 exons that range in size from 90 to >670 bp (Fig. 1).

The analysis of the 10 independent λFix II clones indicated
that they contain the entire coding region of the human TN-R gene (Fig. 1 and data not shown). As presented in Fig. 1, three gaps between AFix II clones present inside introns II, VII, and XVI were filled in amplifying human genomic DNA by PCR. In total, the recombinant phages that were isolated and the genomic PCR products contain overlapping human genomic DNA fragments spanning approximately 85 kb (Fig. 1). Comparison of the genomic sequences to that of the cDNA revealed four single base differences: (i) G replaces A at position 581 (in the sequence reported in 9) thus producing an amino acid substitution (Gly instead of Glu at position 167); (ii) G replaces A at position 753 (without altering the deduced amino acid sequence); (iii) A replaces G at position 2009, producing an amino acid substitution (Lys instead of Arg at position 643); and (iv) T replaces C at position 3318 (without altering the deduced amino acid sequence).

The genomic sequence of TN-R exons was completely determined in the course of this study to rule out the possibility of undetected small introns within blocks of the coding sequence. In particular, the genomic sequence analysis failed to reveal any intron in the context of exon 1, in which a small alternatively spliced region has been detected in chickens (7).

The location of introns was determined by comparing the genomic and cDNA sequences. There are a total of 20 intervening sequences ranging in size from 600 bp to approximately 19 kb (Figs. 1 and 2). The distribution of sizes of introns within the gene is asymmetric: the 2,116 5'-bp of the cDNA are encoded by about 38 kb of genomic DNA; the 575 3'-bp of the cDNA are distributed over 32 kb of the genome; and the central 1329 bp of the cDNA are encoded by only 12 kb of genomic DNA (Figs. 1 and 2).

The human TN-R exons, intron-exon boundaries, and portions of flanking introns were sequenced, and the sequences of the splice junctions are shown in Fig. 2. All of the splice donor and acceptor sites conform well to the known consensus sequences (21). Sixteen of the 20 intervening sequences interrupt codons, whereas 4 occur between codons (Fig. 2). The location of the 20 introns of the human TN-R gene are shown graphically in relation to the protein functional domains in Fig. 1. As for human TN-C (4), the NH2-terminal knob of human TN-R is encoded by two distinct exons (Fig. 1). It is noteworthy that five of nine TNfn repeats are contained in two distinct exons (Fig. 1). Similarly to the human TN-C gene and many other genes encoding TNfn domains, the introns that separate the sequences encoding each TNfn unit are all phase I introns. This finding further supports the hypothesis that intronic recombination is responsible for the internal duplication of these domains (22).

The structure of the genes coding for both chicken and rat TN-R have not been determined yet. However, the comparison of the exon-exon boundaries of the human TN-R gene with the hypothetical exon-exon junctions of chicken and rat TN-R cDNAs revealed a relevant degree of conservation (Fig. 3).

A homology comparison of intron splice donor and acceptor sites revealed that intron positions are precisely conserved in human TN-R, TN-C, and TN-X and in the COOH-terminal part of pig TN-X (4, 5, 23; data not shown). This finding extends to all the introns up to TNfn repeat 1 and from TNfn repeat 6 through the fibrinogen-like domains for all tenascins. TN-R and TN-C are also conserved in the region from TNfn repeat 1 to TNfn repeat 5 (which, on the contrary, are not matched by TN-X). The sole exception pertains to human TN-R intron XIX (the sequence GTAGDSL in Fig. 2), which is split at the base G in TN-R and in TN-X, whereas its equivalent in human TN-C is split at the base A (4). This is probably a true shift, because the nucleotide sequence of human TN-C cannot support a split in the G. It is important to note that these intron positions have apparently been in place and completely conserved for 800,000–1,000,000 years, since the gene duplication leading to the divergence of TN-C, TN-R, and TN-X occurred approximately that long ago (24).

Tenascin-R is expressed in morphoregulatory relevant areas of the CNS during development (10), and its sequence displays a high degree of conservation between species and with the other tenascin family members. These observations allow the hypothesis that mutations in either coding or noncoding regions of the gene led to the synthesis of a nonfunctional protein in CNS pathologies. The physical map of the human TN-R gene

| Intron Number | Position | Sequence | Intron Number | Position | Sequence | Intron Number | Position | Sequence |
|---------------|----------|----------|---------------|----------|----------|---------------|----------|----------|
| I             | 164      | AATGQLD  | VIII          | 682      | ARTELDS  | XV            | 1081     | SRKELIV  |
|               | 164      | ---------|               | 681      | ---------|               | 1079     | ---------|
|               | 169      | ---------|               | 703      | ---------|               | 1112     | ---------|
| II            | 323      | CSAVAPP  | IX             | 770      | AFTGFRP  | XVI           | 1125     | FTTGGRV  |
|               | 323      | ---------|               | 769      | ---------|               | 1123     | ---------|
|               | 322      | ---------|               | 795      | ---------|               | 1158     | ---------|
| III           | 411      | VATHLST  | X              | 860      | ITTGLDP  | XVII          | 1176     | GWIVFQR  |
|               | 410      | ---------|               | 859      | ---------|               | 1174     | ---------|
|               | 424      | ---------|               | 886      | ---------|               | 1211     | ---------|
| IV            | 450      | IPKNEEG  | XI             | 900      | PTQVGLR  | XVIII         | 1208     | FWGLDN   |
|               | 449      | ---------|               | 898      | ---------|               | 1206     | ---------|
|               | 463      | ---------|               | 918      | ---------|               | 1243     | ---------|
| V             | 500      | VSTVZDG  | XII            | 949      | VHTAMDN  | XIX           | 1262     | GTAGDSL  |
|               | 499      | ---------|               | 947      | ---------|               | 1260     | ---------|
|               | 515      | ---------|               | 978      | ---------|               | 1299     | ---------|
| VI            | 590      | FTTXEDA  | XIII           | 989      | HFAVAGE  | XX            | 1317     | HSQGINW  |
|               | 589      | ---------|               | 987      | ---------|               | 1315     | ---------|
|               | 609      | ---------|               | 1025     | ---------|               | 1356     | ---------|
| VII           | 652      | TLLDLVP  | XIV            | 1037     | FSTLLDP  |               |          |          |
|               | 651      | ---------|               | 1035     | ---------|               |          |          |
|               | 673      | ---------|               | 1066     | ---------|               |          |          |
we have here reported will permit the development of rapid assays to detect whether point mutations or small deletions may be involved in CNS defects during development.

Acknowledgments—We thank Dr. Harold Erickson for many discussions and help with sequence alignments, Dr. Laura Borsi for technical assistance, and Thomas Wiley for assistance in editing. We are indebted to Prof. Leonardo Santi for encouragement.

REFERENCES
1. Erickson, H. P. (1993) Curr. Opin. Cell Biol. 5, 869–876
2. Chiquet-Ehrismann, R., Hagios, C., and Matsumoto, K. (1994) Perspect. Dev. Neurobiol. 2, 3–7
3. Erickson, H. P., and Bourdon M. A. (1989) Annu. Rev. Cell. Biol. 5, 71–92
4. Gulcher, J. R., Nies, D. E., Alexakos, M. J., Ravikant, N. A., Sturgill, M. E., Marton, L. S., and Stefansson, K. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 9438–9442
5. Bristow, J., Tee, M. K., Gitelman, S. E., Mellon, S. H., and Miller, W. L. (1993) J. Cell Biol. 122, 265–278
6. Gherzi, R., Carnemolla, B., Siri, A., Ponassi, M., Balza, E., and Zardi, L. (1995) J. Biol. Chem. 270, 3429–3434
7. Nörenberg, U., Wille, H., Wolff, J. M., Frank, R., and Rathjen, F. G. (1992) Neuron 8, 849–863
8. Fuss, B., Wintergerst, E.-S., Bartsch, U., and Schachner, M. (1993) J. Cell Biol. 120, 1237–1249
9. Carnemolla, B., Leprini, A., Borsi, L., Querzé, G., Urbini, S., and Zardi, L. (1996) J. Biol. Chem. 271, 8157–8160
10. Schachner, M., Taylor, Y., Bartsch, U., and Pesheva, P. (1994) Perspect. Dev. Neurobiol. 2, 33–41
11. Matsumoto, K., Saga, Y., Ilemura, T., Sakakura, T., and Chiquet-Ehrismann, R. (1994) J. Cell Biol. 125, 483–489
12. Wintergerst, E. S., Fuss, B., and Bartsch, U. (1993) Eur. J. Neurosci. 5, 299–310
13. Bartsch, U., Pesheva, P., Raff, M., and Schachner, M. (1993) Glue 9, 57–69
14. Pesheva, P., Spiess, E., and Schachner, M. (1989) J. Cell Biol. 109, 1765–1778
15. Rathjen, F. G., Wolff, J. M., and Chiquet-Ehrismann, R. (1991) Development (Camb.) 113, 151–164
16. Nörenberg, U., Hubert, M., Brümmendorf, T., Tarnok, A., and Rathjen, F. G. (1995) J. Cell Biol. 130, 473–484
17. Taylor, J., Pesheva, P., and Schachner, M. (1993) J. Neurosci. Res. 35, 347–362
18. Lochter, A., and Schachner, M. (1993) J. Neurosci. 13, 3986–4000
19. Schwab, M. E., Kapfhammer, J.-P., and Bandtlow, C. E. (1993) Annu. Rev. Neurosci. 16, 565–595
20. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
21. Mount, S. M. (1982) Nucleic Acids Res. 10, 459–472
22. Patthy, L. (1990) Cell 61, 13–14
23. Geffrotin, C., Garrido, J. J., Tremet, L., and Vaiman, M. (1995) Eur. J. Biochem. 231, 81–92
24. Erickson, H. P. (1994) Perspect. Dev. Neurobiol. 2, 9–19