SHORT COMMUNICATION

DNA Sequences in the Promoter Region of the *NF1* Gene Are Highly Conserved between Human and Mouse

AMITAV HAJRA, ANTONIA MARTIN-GALLARDO,* SUSAN A. TARLÉ, MATTHEW FREEDMAN, SUSAN WILSON-GUNN, ANDRE BERNARDS,† AND FRANCIS S. COLLINS[1]

*Departments of Human Genetics and Internal Medicine and the Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan 48109-0650; *Centro Nacional de Biotecnologia, 28049 Madrid, Spain; and †Massachusetts General Hospital Molecular Neurogenetics Unit and Cancer Center, Charlestown, Massachusetts 02129*

Received December 2, 1993; revised March 11, 1994

The gene for type 1 neurofibromatosis (*NF1*) is most highly expressed in brain and spinal cord, although low levels of mRNA can be found in nearly all tissues. As a first step in investigating the regulation of *NF1* gene expression, we have cloned and sequenced the promoter regions of the human and mouse *NF1* genes and mapped the transcriptional start sites in both species. We report here that the 5′ ends of the human and murine *NF1* genes are highly conserved. While no discernable TATA or CCAAT box sequences are seen, transcription initiates at identical sites in both species, 484 nucleotides upstream of the ATG initiation codon in the human gene. The human and mouse *NF1* genes share particularly high sequence homology (95%) between nucleotides −33 and +261 and contain several perfectly conserved transcription factor binding site motifs, including a cAMP response element, several AP2 consensus binding sites, and a serum response element. The high conservation of these sequences indicates that they are likely to be significant in the regulation of *NF1* gene expression.

Neurofibromatosis type 1 (NF1) is one of the most common human genetic disorders, with an incidence of approximately 1 in 3000. This disorder is inherited in an autosomal dominant fashion and primarily affects tissues derived from the neural crest (17). The gene responsible for NF1 was identified by a positional cloning approach (6, 18, 19) and was found to encode a protein with GTPase-activating properties (2, 14, 20). The *NF1* gene is highly conserved in vertebrates (4) and encodes an 11 to 13-kb-long mRNA (19). While *NF1* mRNA can be detected in nearly all adult tissues (19), mRNA and protein levels vary in different cell types, with highest levels found in brain, spinal cord, and peripheral nerve (8).

Little is known about the regulatory elements that contribute to *NF1* expression. As a first step toward identifying such elements, we have compared the sequences of the 5′ ends of the human and mouse *NF1* genes to identify regions that have been conserved through evolution. Such an approach has been successful in identifying protein binding sites in the γ-globin promoter (10) and has been used to identify potential protein binding sites in the N-CAM (7) and *RB1* (21) promoters. We report here that the 5′ ends of the human and mouse *NF1* genes have very high sequence identity, with conservation of transcription start sites and potential transcriptional regulatory elements.

The nucleotide sequence for the promoter region of the human *NF1* gene was obtained by sequencing the distal half of a *Not*I linking clone (17L1B) containing the 5′ end of the human *NF1* gene (13). The entire 4.0-kb clone was sequenced with automated DNA sequencing technology using a previously described *Exo*III unidirectional approach (15, 16). The corresponding region of the mouse *NF1* promoter was cloned as previously described (3) and was sequenced from the 3′ end of the *NF1* first exon upstream to the 5′ end of the clone (1011 bp) on both strands by primer walking using Sequenase Version 2.0 (U.S. Biochemicals). This same region of the mouse *NF1* gene has also been sequenced independently by Bernards *et al.* (3).

Comparisons between the mouse and the human sequences were performed using the MacVector 3.5 software (IBI-Kodak). A portion of the human sequence (nt 2943 to 3953) containing the first 60 bp of the coding region and 5′ flanking sequences was compared to the corresponding mouse sequence. We also searched for potential regulatory elements within the mouse and human *NF1* promoter region using MacVector 3.5 with a custom-designed database of transcription control elements, based on a previously compiled list of transcriptional control sequences (12). The mouse and human sequences were also used to search the GenBank database.

Sequence data from this article have been deposited with the GenBank Database under Accession Nos. U09106 (human) and L10367 (mouse).

[1] To whom correspondence and reprint requests should be addressed at: National Center for Human Genome Research, National Institutes of Health, Building 38A, Room 605, 9000 Rockville Pike, Bethesda, MD 20892. Telephone: (301) 496-0844. Fax: (301) 402-0837.

GENOMICS **21**, 649–652 (1994)
0888-7543/94 $6.00

RNase protection assays were performed using the ribonuclease protection assay RPA II kit from Ambion, Inc. following the manufacturer's instructions. The templates for the human and mouse antisense riboprobes were constructed by subcloning the appropriate fragment into pBluescript II KS (Stratagene). The human fragment corresponded to nt −36 to +273 in Fig. 1, while the mouse fragment spanned nt −49 to +234 in Fig. 1. The human template was linearized and then transcribed using an *in vitro* transcription kit with T7 polymerase (Promega). The antisense riboprobe consisted of 309 nt complementary to the human *NF1* gene and 24 nt of vector sequence. The mouse template was also linearized and was transcribed using T3 polymerase (Promega). The resulting antisense riboprobe contained 282 nt complementary to the mouse *NF1* gene and 30 nt complementary to vector sequence. The riboprobes were hybridized to total RNA isolated from either human or mouse brain using RNAzol B (Tel-Test, Inc.).

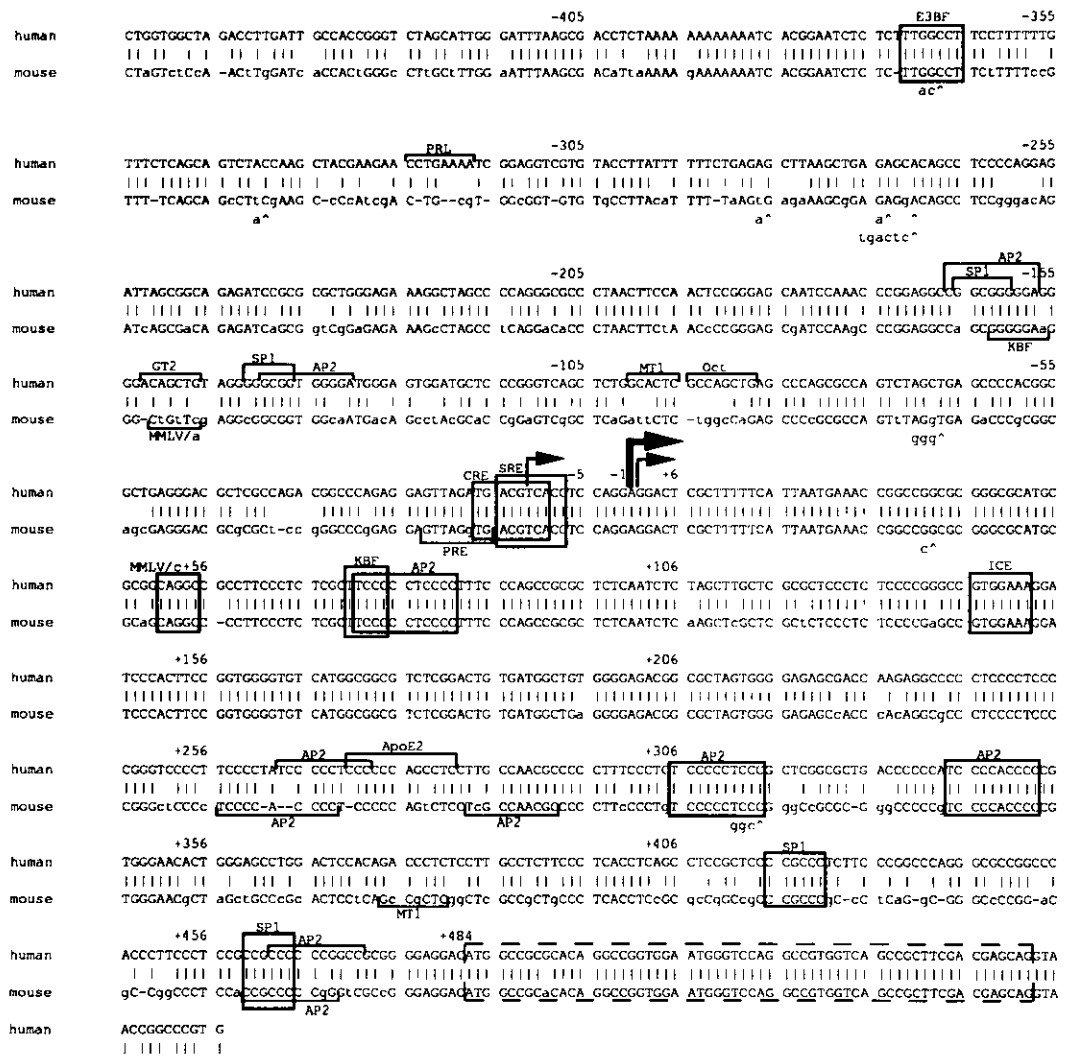

FIG. 1. Optimized alignment of the sequences at the 5′ end of the human and mouse *NF1* genes. This region contains the first 60 bp of the *NF1* coding region (bordered by a dashed box) and approximately 900 bp of upstream flanking sequence, including 5′-untranslated and promoter regions. The mouse sequence has been previously published independently by Bernards *et al.* (3). The human sequence from +274 to +543 has been previously published as the first 270 nucleotides of a human *NF1* cDNA (13). The large arrow indicates the major transcription initiation site identified by RNase protection, while the two smaller arrows indicate minor transcription start sites (see Fig. 2). The human sequence is numbered relative to the major transcription start site. Dashes indicate where gaps have been placed in the mouse sequence for optimal alignment. Nucleotides that have been removed from the mouse sequence for optimized alignment are shown underneath the mouse sequence. Transcription factor binding site motifs found in both the human and the mouse sequences are bordered by solid boxes. Binding site motifs found in only the human sequence are indicated by brackets above the human sequence, while those motifs found in only the mouse sequence are indicated by brackets beneath the mouse sequence. Abbreviations for binding site motifs are AP2, AP2 consensus binding site; ApoE2, enhancer in Apo2 gene; CRE, cAMP response element; E3BF, binding site found in adenovirus E3 gene and cellular genes; GT2, binding site in immunoglobulin and SV40 enhancers; ICE, insulin gene enhancer; KBF, binding site in immunoglobulin enhancer; MMLV/a, c, binding sites in Moloney murine leukemia virus enhancer; MT1, metal response element; Oct, octamer binding motif; PRE, progesterone response element; PRL, enhancer in prolactin gene; SP1, SP1 binding site; SRE, serum response element.

Figure 1 shows an optimized alignment of sequences from the 5′ end of the human and mouse *NF1* genes. The nucleotides are numbered relative to the major transcription start site, which is described below. Both promoters are very G+C-rich and contain a high proportion of CpG dinucleotides, with no obvious TATA or CCAAT boxes present. These features are characteristic of promoters of typical housekeeping genes (9). The region of greatest sequence similarity between the human and mouse genes occurs between nt −33 and +261, with a sequence identity of 95%. A search of GenBank using the human and mouse sequences did not reveal any significant homologies to other genes.

To determine whether sequences conserved between human and mouse were located in transcribed or nontranscribed (regulatory) regions, we mapped the transcriptional start sites of the human and mouse *NF1* genes by RNase protection. As shown in Fig. 2, the full-length human riboprobe (333 nt, lane 1) protected a major fragment of 272 nt (lane 2). Since 24 of the 61 digested nucleotides of the riboprobe are from the vector sequence, the major transcriptional start site is located 36 nt downstream from the 5′ end of the riboprobe, corresponding to 484 nt upstream of the translation initiation codon. The location of this site is in agreement with data from previous primer extension studies (13). The full-length mouse riboprobe (lane 4) protects a major fragment of 234 nt (lane 5), indicating that the transcription start site lies 48 nucleotides downstream from the 5′ end of the probe. This start site corresponds to the same nucleotide as the human transcriptional start site and is located 476 nucleotides upstream of the ATG initiation codon in the mouse gene. Both the human and the mouse genes also show minor transcription start sites 11 nucleotides upstream and 1 nucleotide downstream of the major start site (Fig. 2, arrowheads).

The region of highest conservation in the *NF1* promoter contains several perfectly conserved potential regulatory elements (Fig. 1). Among these conserved elements are Sp1 binding sites (+416 and +460), several consensus AP2 binding sites (+72, +264, +306, +335, and +463), a cAMP response element (−16), and a serum response element (−14). In addition, elements found in the cores of the insulin (+137) and immunoglobulin (+71) enhancers are also conserved in this region. No perfectly conserved regulatory elements are located farther than 20 bp upstream of the transcription start site, except for an element found in the adenovirus E3 gene (−372). An interesting feature of the predicted 5′-untranslated region in both human and mouse is the presence of a conserved upstream ATG (+189) followed closely by a stop codon (TAG, +210), a motif that is common in mRNAs that are poorly translated (5).

In addition to the conserved binding site motifs, there are several binding site motifs in the *NF1* promoter region that are not conserved between human and mouse. For example, the human sequence has nonconserved AP2 sites (−166 and −139), SP1 sites (−141 and −165), and an octamer binding site (−94). The murine sequence has a nonconserved progesterone response element (−22), a MMLV/a site (−151), and a KBF site (−162).

Recently, Bernards *et al.* (3) have independently reported that the 5′ ends of the human and mouse *NF1* cDNAs share high sequence homology. This analysis, however, was performed using only a partial sequence from the 5′ end of the human *NF1* gene. Using additional sequence from the human *NF1* gene, we have confirmed and extended the observations of Bernards *et al.* and have determined that the mouse–human homology in this region is actually much higher than previously reported. Based on our structural comparison of the human and murine *NF1* promoter, the region from −13 to +261 seems to be especially interesting for further investigation. This region is almost identical in human and mouse, contains several elements that have been shown to be important in the regulation of other genes, and also contains the transcriptional start site of the *NF1* gene. Since a large portion of this highly conserved sequence lies within the predicted 5′-untranslated region of the *NF1* gene, it may be important in regulating the stability and translation efficiency of mRNAs (1). For example, myelin basic protein genes, which are expressed in similar tissues as *NF1,* have been shown to contain elements in their 5′-untranslated regions that control the efficiency of mRNA translation and mediate post-transcriptional steroid induc-

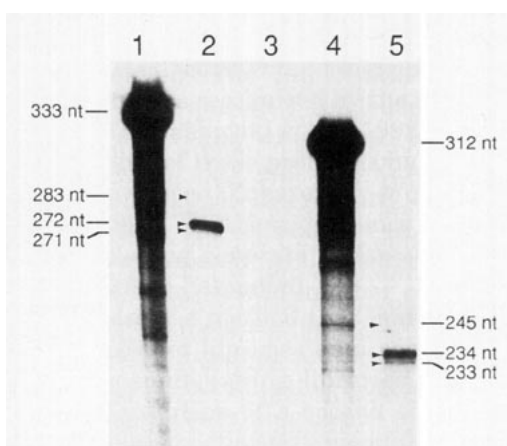

FIG. 2. RNase protection mapping of the transcription start sites of the human and mouse *NF1* genes. Lane **1:** human antisense riboprobe not treated with RNase. Lane **2:** human antisense riboprobe hybridized to 30 μg total RNA from human brain and treated with RNase. Lane **3:** human antisense riboprobe treated with RNase. Lane **4:** mouse antisense riboprobe not treated with RNase. Lane **5:** *mouse antisense riboprobe hybridized to 15 μg total RNA from mouse* and treated with RNase. Lane 5 was exposed on film for approximately twice as long as the other lanes (7 days). Arrowheads indicate the protected mouse and human fragments. On the left are indicated the sizes of the full-length human antisense riboprobe (333 nt) and fragments protected by this probe (283, 272, and 271 nt). The same are indicated on the right for the mouse antisense riboprobe (312 nt full-length probe; 245, 234, and 233 nt protected fragments). Fragment sizes were determined by comparison to a sequencing ladder of the human and mouse riboprobe templates.

tion of protein expression (5). In addition, the cAMP response element in this conserved region is also interesting, since there is evidence that *NF1* expression in Schwann cells increases in response to high doses of cAMP (11).

In summary, we have uncovered rather dramatic sequence conservation between the mouse and human *NF1* promoters. This sequence analysis now sets the stage for a thorough functional analysis of this region.

ACKNOWLEDGMENTS

We thank J. Koh for helpful discussion. This work was supported by Grant NS23410 (F.S.C.). F.S.C. was previously an investigator with the Howard Hughes Medical Institute.

REFERENCES

1. Atwater, J. A., Wisdom, R., and Verma, V. M. (1990). Regulated mRNA stability. *Annu. Rev. Genet.* **24:** 519–541.
2. Ballester, R., Marchuk, D., Boguski, M., Saulino, A., Letcher, R., Wigler, M., and Collins, F. S. (1990). The NF1 locus encodes a protein functionally related to mammalian GAP and yeast IRA proteins. *Cell* **63:** 851–859.
3. Bernards, A., Snijders, A. J., Hannigan, G. E., Murthy, A. E., and Gusella, J. F. (1993). Mouse neurofibromatosis type 1 cDNA sequence reveals a high degree of conservation of both coding and non-coding mRNA segments. *Hum. Mol. Genet.* **2:** 645–650.
4. Buchberg, A. M., Cleveland, L. S., Jenkins, N. A., and Copeland, N. G. (1990). Sequence homology shared by neurofibromatosis type 1 gene and IRA1 and IRA2 negative regulators of the *RAS* cyclic AMP pathway. *Nature* **347:** 291–294.
5. Campagnoni, A. T., Verdi, J. M., Verity, A. N., and Amur-Umarjee, S. (1990). Posttranscriptional events in the expression of myelin basic protein genes. *Ann. N.Y. Acad. Sci.* **605:** 270–279.
6. Cawthon, R. M., Weiss, R., Xu, G., Viskochil, D., Culver, M., Stevens, J., Robertson, M., Dunn, D., Gesteland, R., O'Connell, P., and White, R. (1990). A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. *Cell* **62:** 193–201.
7. Colwell, G., Li, B., Forrest, D., and Brackenbury, R. (1992). Conserved regulatory elements in the promoter region of the N-CAM gene. *Genomics* **14:** 875–882.
8. Daston, M. M., Scrable, H., Nordlund, M., Sturbaum, A. K., Nissen, L. M., and Ratner, N. (1992). The protein product of the neurofibromatosis type 1 gene product is expressed in highest abundance in neurons, Schwann cells, and oligodendrocytes. *Neuron* **8:** 415–428.
9. Dynan, W. S. (1986). Promoters for house-keeping genes. *Trends Genet.* **2:** 196–197.
10. Gumucio, D. L., Heillstedt-Williamson, H., Gray, T. A., Tarlé, S. A., Shelton, D. A., Tagle, D. A., Gragowski, L., Slightom, J. L., Goodman, M., and Collins, F. S. (1992). Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human γ and ϵ globin gene. *Mol. Cell. Biol.* **12:** 4919–4929.
11. Gutmann, D. H., Tennekoon, G. I., Cole, J. L., Collins, F. S., and Rutkowski, J. L. (1993). Modulation of neurofibromatosis type 1 gene product, neurofibromin, during Schwann cell differentiation. *J. Neurosci. Res.* **36:** 216–223.
12. Locker, J., and Buzard, G. (1990). A dictionary of transcription control sequences. *DNA Seq.* **1:** 3–11.
13. Marchuk, D. A., Saulino, A. M., Tavakkol, R., Swaroop, M., Wallace, M. R., Andersen, L. B., Mitchell, A. L., Gutmann, D. H., Boguski, M., and Collins, F. S. (1991). cDNA cloning of the type 1 neurofibromatosis gene: Complete sequence of the *NF1* gene product. *Genomics* **11:** 931–940.
14. Martin, G. A., Viskochil, D., Bollag, G., McCabe, P. C., Crosier, W. J., Haubruck, H., Conroy, L., Clark, R., O'Connell, P., Cawthon, R. M., Innis, M. A., and McCormick, F. (1990). The GAP-related domain of the neurofibromatosis type 1 gene product interacts with *ras* p21. *Cell* **63:** 843–849.
15. Martin-Gallardo, A., Marchuk, D. A., Gocayne, J., Kerlavage, A. R., McCombie, W. R., Venter, J. C., Collins, F. S., and Wallace, M. R. (1992). Sequencing and analysis of genomic fragments from the *NF1* locus. *DNA Seq.* **3:** 237–243.
16. McCombie, W. R., Kirkness, E., Fleming, J. T., Kerlavage, A. R., Iovannisci, D. M., and Martin-Gallardo, A. (1991). The use of Exonuclease III deletions in automated DNA sequencing. *Methods* **3:** 33–40.
17. Riccardi, V. M., and Lewis, R. A. (1986). "Neurofibromatosis: Phenotype, Natural History, and Pathogenesis," Johns Hopkins Univ. Press, Baltimore, MD.
18. Viskochil, D., Buchberg, A. M., Xu, G., Cawthon, R. M., Stevens, J., Wolff, R. K., Culver, M., Carey, J. C., Copeland, N. G., Jenkins, N. A., White, R., and O'Connell, P. (1990). Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. *Cell* **62:** 187–192.
19. Wallace, M. R., Marchuk, D. A., Andersen, L. B., Letcher, R., Oder, H., Saulino, A. M., Fountain, J. W., Bereton, A., Nicholson, J., Mitchell, A. L., Brownstein, B., and Collins, F. S. (1990). Type 1 neurofibromatosis gene: Identification of a large transcript disrupted in three NF1 patients. *Science* **249:** 181–186.
20. Xu, G., Lin, B., Tanaka, K., Dunn, D., Wood, D., Gesteland, R., White, R., Weiss, R., and Tamanoi, F. (1990). The catalytic domain of the neurofibromatosis type 1 gene product stimulates ras GTPase and complements ira mutants of *S. cerevisiae*. *Cell* **63:** 835–841.
21. Zacksenhaus, E., Gill, R. M., Phillips, R. A., and Gallie, B. L. (1993). Molecular cloning and characterization of the mouse *RB1* promoter. *Oncogene* **8:** 2343–2351.