Structural Analysis of the Human BIN1 Gene

EVIDENCE FOR TISSUE-SPECIFIC TRANSCRIPTIONAL REGULATION AND ALTERNATE RNA SPLICING

(Received for publication, May 9, 1997, and in revised form, September 5, 1997)

Robert Wechsler-Reya‡, Daitoku Sakamura§, Jing Zhang, James Duhadaway, and George C. Prendergast¶

From The Wistar Institute, Philadelphia, Pennsylvania 19104

BIN1 is a putative tumor suppressor that was identified through its interaction with the MYC oncoprotein. To begin to identify elements of BIN1 whose alteration may contribute to malignancy, we cloned and characterized the human BIN1 gene and promoter. Nineteen exons were identified in a region of >54 kilobases, six of which were alternately spliced in a cell type-specific manner. One alternately spliced exon encodes part of the MYC-binding domain, suggesting that splicing controls the MYC-binding capacity of BIN1 polypeptides. Four other alternately spliced exons encode amphiphysin-related sequences that were included in brain-specific BIN1 species, also termed amphiphysin isoforms or amphiphysin-related sequences that were included in brain-specific BIN1 species, also termed amphiphysin isoforms or amphiphysin-related sequences that were included in brain-specific BIN1 species, also termed amphiphysin isoforms or

The identification of tumor suppressor genes in solid tumors is a major goal of cancer research. BIN1 is a novel MYC-interacting protein that has features of a tumor suppressor in certain carcinomas including those of the breast, liver, cervix, and prostate (1). BIN1 is related to amphiphysin, a neuronal protein that is a paraneoplastic autoimmune antigen associated with breast and lung cancer (2, 3), and to RVS167, a negative regulator of the cell cycle in yeast (4). Although widely expressed in normal cells, BIN1 is functionally deleted in ~50% of carcinoma cell lines and primary breast carcinomas examined (1). The human BIN1 gene is located at chromosome 2q14 (5), within a mid-2q region that is deleted in ~42% of metastatic prostate cancers (6). We have hypothesized that BIN1 is a tumor suppressor whose loss contributes to growth deregulation in cancer cells. As a prerequisite to examining this hypothesis, the exon organization, exon-intron boundaries, splice patterns, and promoter of the human BIN1 gene were defined.

MATERIALS AND METHODS

Cloning and Analysis—Genomic clones were isolated from a WI-38 diploid fibroblast AFIX phage library (a gift of L. Showe) using a BIN1 cDNA probe (1). The six phage inserts designated in Fig. 1 were subcloned in pBS† (Stratagene) and analyzed by extensive restriction mapping and Southern analysis with BIN1 cDNA probes (1). Large scale sequencing of genomic DNA that hybridized to BIN1 cDNA probes, comprising ~20 kb of the >54-kb locus, was determined using an automated DNA sequencer. The sequence data were assembled manually with assistance from MacVector and AssembLIGN software. Exons and other gene features were identified and/or confirmed by visual inspection or computer-aided comparison of BIN1 cDNAs and expressed sequence tags in GenBank‡, using SIM and ClustalW (pairwise or multiple sequence alignments), TBLASTN (DNA database comparisons), MatInspector, and TESS (promoter binding site identification), accessed through the Baylor College of Medicine Human Genome Center Home Page on the World Wide Web. The GenBank™ accession number for BIN1 cDNA is U86485 and for BIN1 genomic sequences are U83999–U84004.

RT-PCR—The substrate for RT-PCR was 2 μg of total cytoplasmic RNA isolated as described (7). Murine RNAs were a gift of L. Benjamin (Hadassah Hospital, Jerusalem). Human RNAs were isolated from WI-38 fibroblasts, HeLa cells, or Rh30 rhabdomyosarcoma cells that were cultured as described (1). RNA and 50 pmol of oligo(dT) (Pharmacia) were added to diethylypyrocarbonate-treated water (final volume 11 μl), heated 4 min at 70 °C, and quenched on ice. RT reactions (20 μl) were prepared by mixing 4 μl of 5 × buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM MgCl2), 1 μl of 25 mM dNTPs, 2 μl 0.1 μl dithiothreitol, 1 unit of RNase inhibitor, and 100 units of Moloney murine leukemia virus RT (Life Technologies, Inc.). This mixture was incubated 1 h at 42 °C, heated to 94 °C for 5 min, and diluted to 100 μl with diethylypyrocarbonate-treated water. 10 μl of the diluted reaction was used as substrate for 30 cycles of PCR (50 μl) (45 °C and 94 °C, 45 °C, 1 min 72 °C) with 0.5 unit of Taq polymerase (Pharmacia) in 1 × buffer, 0.2 mM dNTPs, 1.5 mM MgCl2. Separate PCR reactions were performed to generate 5′ (N terminus; exons 3–7), midsection (exons 6–11), and 3′ (C terminus; exons 11–16) segments of the BIN1 coding region. For amplifying products from human BIN1 message (1), the 5′- and 3′-primer pairs used were as follows: For the 5′ (N terminus) product, AAGGATCTCCGGACACTACT (C7/extra) and CACATGTACATGTCACAACC (nxt7α); for the midsection product, TGAAGCCCAAAATTGGCAAGGCC (dT3/ex) and TGGCTGAGATGGGGACTTG (5′-ATCG99); and for the 3′ (C terminus) product, GGGAGATCTCAGGATGTCGCTGAAAAGGGAAAACAGAG (99Fep) and GAGCTCGAGATGTCGACGGGCTTCTCTCAGTGAAGTTC (99SH3anti). For amplifying products from murine BIN1 message (8), the 5′- and 3′-primer pairs used were as follows: mNTsen1 (5′-CAGTGTCCTCCAGAATTTCC) and mNTanti1 (5′-AACACTTCTTGGGCTTTG); mNTsen1 (5′-AACCGCGAGGTGGTGCGGAG) and mNTanti1 (5′-ATCG99-5′TGTCCTGAGATGTCGACGGGCTTCTCTCAGTGAAGTTC) and mNTsen1 (5′-CAGTAGTACAGGACATTCC) and mNTanti1 (5′-CACCGCCCTCTGCTAAAAATTT). Products were fractionated on agarose gels, blotted and hybridized, or isolated and subcloned for DNA sequencing.

§ Recipient of the Christopher Davis Memorial Fellowship.
¶ Faculty Scholar of the American Cancer Society Junior Faculty Award and a Pew Scholar in the Biomedical Sciences. To whom correspondence should be addressed: The Wistar Institute, 3601 Spruce St., Philadelphia, PA 19104. Tel.: 215-898-3792; Fax: 215-898-2205; E-mail: prendergast@wista.wistar.upenn.edu.

* This work was supported by Grants DAMD17-96-1-6324 from the U. S. Army Breast Cancer Research Program and CN160 from the Canadian Cancer Society (to G.C.P.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) U86485, U83999–U84004.

† Medical Research Council of Canada Postdoctoral Fellow.
‡ Recipient of an American Cancer Society Junior Faculty Award and a Pew Scholar in the Biomedical Sciences. To whom correspondence should be addressed: The Wistar Institute, 3601 Spruce St., Philadelphia, PA 19104. Tel.: 215-898-3792; Fax: 215-898-2205; E-mail: prendergast@wista.wistar.upenn.edu.

The Journal of Biological Chemistry 1997, 272:31453–31458

© 1997 by The American Society for Biochemistry and Molecular Biology, Inc.
**RESULTS**

**BIN1 Gene Structure and Exon-Intron Organization**—A physical map of the human **BIN1** gene was constructed from a set of phage clones isolated from a WI-38 diploid fibroblast genomic library (see Fig. 1). To identify exons and exon-intron boundaries, the DNA sequence determined from six genomic segments (GenBank™ accession numbers U83998–U84004) was compared with **BIN1** cDNA sequences from several sources, including the original **BIN1** cDNA clone, RT-PCR products from human RNAs, and the DNA data base (1, 8, 10–12). With the exception of exon 1, all exons were located within a ~38-kb contig. An additional noncontiguous clone contained exon 1 and 5′-flanking sequences, with the latter extending ~3 kb upstream of the RNA cap site (see below). Given the structure of this clone, the size of intron 1 would be inferred to be at least 17 kb. Furthermore, it was concluded that the human **BIN1** gene spanned a minimum of 54 kb.

The DNA sequence of each exon and proximal introns are shown in Fig. 2. Based on the characteristics they encode, the **BIN1** exons can be grouped into four sets, termed the BAR (**BIN1/amphiphysin/RVS167**-related), unique, brain-specific, and protein-protein interaction sets, respectively. Exons 1–8 encode the BAR domain of **BIN1** (1). In this group, exon 1 included a different 5′-UTR and N-terminal coding sequence (MAEMGSKG) compared with the original **BIN1** cDNA (ML-WNV) (1). The genomic sequence was judged to accurately represent the 5′-end of the **BIN1** mRNA, because (i) expressed sequence tag and cDNA sequences identical to the genomic but not the 5′ **BIN1** cDNA sequence were present in the DNA data base; (ii) cDNAs whose structure matched the original cDNA clone could not be identified by RT-PCR in any tissue; and (iii) the 5′-end of the cDNA was found to contain an inversion of 64 bp derived from the middle of the cDNA (previously missed because the inversion fortuitously contained a translation initiation site). Exons 9–11 encode a unique region of **BIN1** that is functionally undefined and unrelated to amphiphysin and RVS167. The unique-1 (U1) and unique-2 (U2) regions are encoded by exons 9 and 11, respectively, and separated by a nuclear localization-like motif encoded by exon 10. Exons 12A–12D encode amphiphysin-related sequences that were not found in the original **BIN1** cDNA (see Fig. 2B). These exons are spliced into larger isoforms of **BIN1** message detected in brain and muscle, alternately termed amphiphysin isoforms or amphiphysin II (1, 10–12) (see below). Exons 13–16 encode the C-terminal region of **BIN1** implicated in protein-protein interactions. Exons 13–15 and 15–16 encode the Myc-binding domain and the Src homology 3 (SH3) domain, respectively, the latter of which is also a feature of amphiphysin and RVS167 (1).

**Alternate Splicing of **BIN1** RNA**—To examine patterns of **BIN1** splicing, RT-PCR was performed using RNAs isolated from three human cell lines, WI-38 diploid fibroblasts, HeLa cervical carcinoma, and Rh30 rhabdomyosarcoma (a muscle tumor line). The observations in human cells were extended using RNAs isolated from a set of normal murine tissues. All cells and tissues examined were previously shown to express **BIN1** RNA by Northern analysis (1). Oligonucleotide primers derived from **BIN1** sequences were used for RT-PCR of fragments spanning exons 3–7 (5′-end), 7–11 (midsection), and 11–16 (3′-end). For analysis of murine RNAs, oligonucleotide primer sequences were derived from the sequence of **SH3P9**, a murine **BIN1** cDNA (GenBank™ accession number U60884; and Ref. 8). Products from these reactions were fractionated on agarose gels, blotted, and hybridized to a **BIN1** cDNA probe or blunted and sequenced.

Amplification of the 5′-end of **BIN1** yielded a single product in all cell lines examined, indicating that this region was not subjected to alternate splicing in these cells. In contrast, amplification of the central and 3′ regions revealed several alternate splicing events. In the central region, two products of similar abundance were observed in WI-38 fibroblasts that
FIG. 2. Exon-intron structure. A, exon and proximal intron sequences. The figure is read left to right with complete exons shown in the left panel and introns following, each shown in the right panel. Register is 50 bp per line. Sizes of intron gaps noted are exact, if sequenced; otherwise estimated from restriction mapping.

B, similarity between exon 12A–12D and amphiphysin sequences. Alignment of the predicted open reading frames (ORF) of BIN1 exons with amino acids 297–457 of human amphiphysin (18).
differed in the presence or absence of exon 10 sequences. Messages including exon 10 sequences were not detected in BIN1 messages from any of the other tissues or cell lines examined, suggesting that this splice form was relatively uncommon and thus regulated. Amplification of the 3' end revealed additional splice forms. Two products of similar abundance were detected in all cell types that differed in the presence of exon 13 sequences, which encodes part of the Myc-binding domain (1). The coordinate appearance of each species suggested that exon 13 was alternately spliced but in an unregulated fashion. Additional species that included sequences derived from exons 12A–12D were detected in murine brain (E9.5 RNA also included one of these species). Interestingly, whereas exon 12A–12D sequences were not detected in any other normal murine tissues, exon 12A was included in RNA species in each of the established human cell lines. It was unclear whether the difference in exon 12A splicing reflected tissue-specific regulation, cell line establishment, or neoplastic transformation. Nevertheless, taken together with the brain-specific events, this observation suggested that splicing of exons 12A–12D was uncommon in most tissues and thus may be regulated.

To determine which combinations of exons appeared in various BIN1 RNAs, we performed RT-PCR using exon 9 and 16 primers (which span all the alternately spliced exons) and subcloned and sequenced the products. We confined this analysis to RNAs isolated from WI-38 and HeLa cells, where exons 10, 12A, and 13 are alternately spliced, to focus on the events in proliferating rather than postmitotic cells. The results, which are summarized in Fig. 3B, showed that seven of the eight RNA species theoretically possible in these cell lines were in fact generated (the one that was not detected was the +12A–13 species). We concluded that exons 10, 12A–12D, and 13 of BIN1 were alternately spliced and that splicing of exons 10 and 12A–12D splicing was likely to be regulated.

**Definition of 5'-flanking Sequences Sufficient for Basal Transcription and MyoD Activation**—Definition of the BIN1 promoter was of interest for two reasons. First, previous work suggested that epigenetic mechanisms might underlie the loss of BIN1 expression in breast tumor cells (1). Therefore, characterization of the BIN1 promoter would permit an examination of tumor DNA for alterations in DNA methylation or transcription factor interactions that might account for loss of expression. Second, we have observed previously that BIN1 is expressed at high levels in skeletal muscle and murine C2C12

---

2 K.Elliott, D. Sakamuro, W. Du, and G. C. Prendergast, submitted for publication.
myoblasts (1, 14). For this reason, we predicted that the BIN1 promoter might be activated by MyoD, a master regulator of muscle cell differentiation (15).

To identify the BIN1 promoter, it was first necessary to pinpoint the site(s) of transcription initiation. To this end, primer extension analysis was performed on RNA from WI-38 diploid fibroblasts. By comparing the genomic sequence to that of a murine BIN1 cDNA (8), which has a long 5'-UTR, a primer was able to map the 5'-end of the BIN1 RNA in WI-38 cells to the guanine residue designated +1 in Fig. 5.

Determination of the genomic sequence upstream of the RNA cap site indicated that the 5'-flanking region was GC-rich and lacked a TATA box but contained a consensus binding site for TBP at −79 identified by MatInspector (13), and a single E box consensus site for MyoD at −237 identified by visual inspection.

The transcriptional potential of the 5'-flanking region was tested in a transient transfection assay. An 886-bp BglII-flanking region was cloned into the luciferase reporter plasmid pGL2-basic, allowing transcription to be initiated at the endogenous BIN1 RNA (14). As shown in Fig. 5C, within two days after transfection, pGL2-Bgl exhibited ~100-fold greater activity that the control plasmid pGL2-basic (see Fig. 6A).

To determine whether pGL2-Bgl included sequences that were sufficient for regulated expression, the plasmid was introduced with or without a MyoD expression vector into 10T1/2 fibroblasts (which do not express MyoD but differentiate into myoblasts in response to it). As a positive control for MyoD responsiveness, a second set of transfections used a luciferase reporter driven by a mutated ornithine carboxylase promoter that contains the E box response element. 3 CPACΔmut-luc is a positive control reporter for MyoD responsiveness (see text). The data represent the average of two trials.

**FIG. 6.** Promoter activity of 5'-flanking sequences. A, basal transcription. C2C12 cells were transfected with 5 μg of the luciferase reporters indicated after transfection, pGL2-Bgl exhibited 7-fold of the basal level by MyoD cotransfection (see Fig. 6B).

**DISCUSSION**

We have characterized the structure and some of the regulatory features of the human BIN1 gene. Nineteen exons were identified by visual inspection.

We observed that the activity of both reporters was increased up to 7-fold of the basal level by MyoD cotransfection (see Fig. 6B). The effect was dose-dependent, because higher ratios of MyoD:reporter plasmids increased reporter activity. We constructed a promoter sufficient for directing transcription in myoblasts. 3 J. Cleveland, unpublished results.
identified within a \( \geq 54 \)-kb region of DNA previously mapped to chromosome 2q14 (5). The primary BIN1 transcript was found to be extensively spliced, resulting in at least seven different species in proliferating cells and an even larger number in postmitotic cells of the brain. Characterization of the BIN1 promoter defined a region sufficient to direct inducible transcription in muscle cells where BIN1 is highly expressed. Thus, BIN1 is subjected to tissue-specific regulation at the levels of transcription and splicing.

Exons 10 and 13 were two of the three exons found to be alternately spliced in proliferating cells. Exon 10 splicing was relatively uncommon, because it appeared only in messages from WI-38 in addition to skeletal muscle (the source of the original BIN1 cDNA). Exon 10 encodes a basic amino acid-rich region that closely resembles a nuclear localization signal. However, this region may not act as a nuclear localization signal, because we have observed recently that its presence is neither necessary nor sufficient for nuclear localization (14, 16). Therefore, exon 10 splicing probably has other implications. Exon 13 splicing was ubiquitous but apparently unregulated, because an approximately similar quantity of +13 and –13 RNA species were detected in all cells examined. Since this exon encodes a significant part of the Myc-binding domain, it is likely that its alternate splicing affects the MYC-interacting potential of BIN1. An interesting possibility implied by our results is that, taken together, there may be two classes of BIN1 polypeptides that exist in cells, one that can interact with MYC and one that cannot.

Four exons identified in this study, 12A–12D, were not included in the original BIN1 cDNA but were detected by RT-PCR in a subset of messages in the brain. The existence of brain-specific exons was suggested previously by Northern analysis, which revealed a larger message(s) in the brain in addition to the ubiquitously expressed smaller species (1). Our findings confirmed those of others who have recently identified exon 12A–12D sequences in brain and muscle cDNA species, alternately termed amphiphysin isoforms or amphiphysin-II (10–12). Another cDNA species identified by these workers implies the presence of an additional 93-bp brain-specific exon in intron 6 (which is unrelated to amphiphysin or RVS167); however for unknown reasons we were unable to confirm its presence in the expected location either by RT-PCR or direct DNA sequencing. Exons 12A–12D encode sequences related to amphiphysin, so their introduction would be expected to increase the amphiphysin-like character of BIN1. Alternate splicing of exon 12A–12D in a subset of brain messages may therefore provide a mechanism to augment or vary certain amphiphysin functions in neurons while retaining BIN1 functions in the same cell.

Interestingly, we found that exon 12A was spliced into a subset of messages in the human cell lines WI-38, HeLa, and Rh-30, but not into messages in normal nonneuronal tissues. The significance of exon 12A splicing in these cells is unclear. However, since we did not detect the +12A–13 isoform in cells, an interesting possibility is that +12A and –13 isoforms are functionally redundant (that is, they each lack the ability to interact with and inhibit the oncopgenic properties of MYC). If so, the appearance of +12 isoforms in WI-38 and HeLa cells may reflect an aberrant splicing event that relieves MYC inhibition by BIN1 (1) thereby promoting immortalization or establishment. In general, the extensive splicing we have documented in BIN1 opens the possibility that splice site mutations or altered splicing via epigenetic mechanisms may be germane to tumorigenesis. Two important goals of future work will be to (i) assess the activities of different splice forms of BIN1 for MYC interaction, cell localization, and inhibition of neoplastic cell growth, and (ii) determine whether there are altered splice patterns in tumor cells that could compromise the growth inhibitory activity of BIN1.

The BIN1 promoter is characterized by a high CpG content but otherwise exhibits the features of a housekeeping promoter. We showed that the muscle determination factor MyoD can up-regulate the BIN1 promoter and identified an E box site that might mediate MyoD-induced activation. Transcriptional activation by MyoD and/or other helix-loop-helix proteins may contribute to the strong up-regulation of BIN1 levels in differentiated neurons and muscle cells (1, 10). Because the BIN1 promoter is rich in CpG residues, it is highly susceptible to the alterations in CpG methylation status, which are common in cancer cells and form the basis for loss of some tumor suppressors such as p16INK4 in lung cancers (17). Other mechanisms by which promoter activity might be altered in cancer cells include genetic mutations or epigenetic changes in the activity of transcriptional regulatory factors. The stage is now set to determine the basis for the frequent loss of BIN1 expression in certain solid tumors such as breast carcinoma (1).

Acknowledgments—We thank R. Bucanefusca and X. Gong for technical assistance. For providing reagents, we are grateful to L. Benjamin for murine RNAs, J. Cleveland for the MyoD-responsive ODCminiLuc reporter plasmid, D. Goldhamer for murine C2C12 and 10T<sub>1/2</sub> cells, and a β-actin promoter-driven MyoD expression vector, and L. Shove for the WI-38 AP1x genomic library. We acknowledge the Wistar Institute Core Facilities for oligonucleotide synthesis and DNA sequencing.

REFERENCES
1. Sakamuro, D., Elliott, K., Wechler-Reya, R., and Prendergast, G. C. (1996) *Nat. Genet.* 14, 69–77
2. David, C., Solimena, M., and De Camilli, P. (1994) *FEBS Lett.* 351, 73–79
3. Droppko, K. J. (1996) *Ann. Neurol.* 39, 659–667
4. Bauer, F., Urdaci, M., Aigle, M., and Crouzet, M. (1993) *Mol. Cell. Biol.* 13, 5070–5084
5. Negorev, D., Reithman, H., Wechler-Reya, R., Sakamuro, D., Prendergast, G. C., and Simon, D. (1996) *Genomics* 33, 329–331
6. Chen, C., Bova, G. S., Moore, D. H., Small, E. J., Carroll, P. R., Pin, S. S., Epstein, J. I., Isaacs, W. B., and Jensen, R. H. (1996) *Cancer Res.* 56, 3091–3102
7. Prendergast, G. C., and Cole, M. D. (1989) *Mol. Cell. Biol.* 9, 124–134
8. Sparke, A. B., Hoffman, N. G., McConnell, S. J., Fowlkes, D. M., and Kay, B. K. (1996) *Nat. Biotechnol.* 14, 741–744
9. Butler, M. H., David, C., Ochoa, G.-C., Freyberg, Z., Daniell, L., Grabs, D., Cremona, O., and De Camilli, P. (1997) *J. Cell Biol.* 137, 1355–1367
10. Ramjaun, A. R., Micheva, K. D., Bouchelet, I., and McPherson, P. S. (1997) *J. Biol. Chem.* 272, 16700–16706
11. Tsumori, K., Maeda, Y., Tatsui, K., Seki, S., and Tokunaga, A. (1997) *Biochem. Biophys. Res. Commun.* 236, 178–183
12. Quandt, K., Frech, K., Haras, H., Wingender, E., and Werner, T. (1995) *Nucleic Acids Res.* 23, 4878–4884
13. Wechler-Reya, R., Elliott, K., and Prendergast, G. C. (1998) *Mol. Cell. Biol.*, in press
14. Tapp, P., Davis, R. L., Thayer, M. J., Cheng, P.-F., Weintraub, H., and Lassar, A. B. (1988) *Science* 242, 405–411
15. Wechler-Reya, R., Elliott, K., Herlyn, M., and Prendergast, G. C. (1997) *Cancer Res.* 57, 3258–3263
16. Merlo, A., Herman, J. G., Mao, L., Lee, D. J., Gabrielson, E., Burger, P. C., Baylin, S. B., and Sidransky, D. (1995) *Nat. Med.* 1, 686–692
17. Yamamoto, R., Li, X., Winter, S., Francke, U., and Kililman, M. W. (1995) *Hum. Mol. Genet.* 4, 265–268