Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 5214–5218, July 1986
Genetics

Analysis of the structure, transcripts, and protein products of *bcl-2*, the gene involved in human follicular lymphoma

(B-cell neoplasia/t(14;18) chromosome translocation/genetics of human lymphoma)

YOSHIHIDE TSUJIMOTO AND CARLO M. CROCE

The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104

Communicated by Hilary Koprowski, April 1, 1986

ABSTRACT **We have determined that the *bcl-2* (B-cell leukemia/lymphoma 2) gene is transcribed into three overlapping mRNAs, and we have cloned *bcl-2* cDNA sequences. Sequence analysis of the *bcl-2* cDNA clones and comparison of their sequences to their genomic counterparts indicate that the *bcl-2* gene contains at least two exons. The three *bcl-2* transcripts, which are 8.5, 5.5, and 3.5 kilobases (kb) long, overlap within the first exon, but only the 8.5-kb and 5.5-kb transcripts contain sequences of the second exon. The 8.5-kb and 5.5-kb transcripts seem to use different polyadenylylation sites. Sequence analysis of the cDNA clones corresponding to the 5.5-kb and 3.5-kb mRNAs indicates that the two *bcl-2* transcripts carry two overlapping open reading frames, one of which is 717 nucleotides long and codes for a protein (bcl-2α) of 239 amino acids and a molecular mass of 26 kDa, while the other codes for a protein of 205 amino acids (bcl-2β, molecular mass 22 kDa) that is identical to bcl-2α except at the carboxyl terminus. The *bcl-2* protein products in follicular lymphomas with or without *bcl-2* rearrangements are identical to the normal *bcl-2* products.**

Specific chromosome rearrangements, predominantly translocations and inversions, are observed in the great majority of human hematopoietic malignancies (1–3). In Burkitt lymphoma, the specific chromosomal translocations involve directly one of the human immunoglobulin loci and the c-*myc* oncogene (4–6). The consequence of the juxtaposition of the immunoglobulin locus and the c-*myc* oncogene is a deregulation of the transcription of the involved c-*myc* oncogene (5–7). Other specific chromosomal translocations involving band 14q32, where the immunoglobulin heavy-chain locus resides (8, 9), are observed in B-cell neoplasms. A t(11;14)(q13;q32) chromosome translocation has been observed in chronic lymphocytic leukemia of the B-cell type (10), in diffuse B-cell lymphoma (2, 3), and in multiple myeloma (11). This translocation juxtaposes the *bcl-1* (B-cell leukemia/lymphoma 1) locus to the heavy-chain locus (12, 13). In most cases of follicular lymphoma, one of the most common human hematopoietic malignancies, a t(14;18)(q32;q21) chromosome translocation has been observed (2, 3). By use of "chromosome walking" techniques, several chromosomal breakpoints involved in the t(14;18) chromosome translocation have been cloned and sequenced and the locus involved in follicular lymphoma has been identified (14–16). In this study, we have investigated the transcripts of the *bcl-2* gene and have identified the *bcl-2* open reading frames.

MATERIALS AND METHODS

RNA Extraction. Cytoplasmic RNA was extracted according to a published procedure (17). Poly(A)$^+$ RNA was selected by oligo(dT) column chromatography (18).

Construction of cDNA Library. Double-stranded cDNA was synthesized from mRNA by reverse transcriptase (Life Sciences, St. Petersburg, FL) using oligo(dT) as primer as described (19). After *Eco*RI linker ligation, the double-stranded cDNA was inserted into λgt10 or λgt11 phage vector (20). The primer-extension cDNA library was constructed by using synthetic oligonucleotide (15-mer) as primer for reverse transcriptase (19, 21).

RNA Blot Hybridization. RNA was glyoxalated, electrophoresed in 1% agarose gel, and blotted to nitrocellulose filter (22). Nitrocellulose filter-bound RNA was hybridized with ^{32}P-labeled probe in 50% (vol/vol) formamide/4× SSC/0.1% NaDodSO$_4$ at 37°C and finally washed with 2× SSC at 65°C. (1× SSC is 0.15 M NaCl/15 mM sodium citrate, pH 7.)

DNA Sequencing. Nucleotide sequences were determined by the chemical degradation method (23) or the chain-termination method (24). Both strands of DNA were sequenced.

***In Vitro* Transcription and Translation.** The cDNA sequence was subcloned under the control of the promoter of *Escherichia coli* phage T7 RNA polymerase. Plasmid DNA was linearized by cutting with restriction enzyme. The capped RNA was synthesized from template DNA (1 μg) in 50 μl of 25 mM NaCl/40 mM Tris Cl, pH 7.5/8 mM $MgCl_2$/2 mM spermidine/5 mM dithiothreitol containing ATP, CTP, and UTP (each at 400 μM); GTP (100 μM); and m^{7}GpppG (400 μM). The size and integrity of the RNA synthesized was tested by electrophoresis followed by blot-hybridization analysis. The capped RNA was then translated in reticulocyte lysate in the presence of [^{35}S]methionine as described (25) or in micrococcal nuclease-treated reticulocyte lysate from Promega Biotec (Madison, WI).

RESULTS

Molecular Cloning of *bcl-2* cDNA. Chromosome 18 probes (probe A in Fig. 1) including the breakpoint hot spot of the t(14;18) chromosome translocation in follicular lymphoma detect mRNA in hematopoietic cell lines, including B and T cells (15). In order to analyze the gene for which the name *bcl-2* (B-cell leukemia/lymphoma 2) is proposed, a cDNA library was constructed from poly(A)$^+$ mRNA of the pre-B-cell leukemia cell line 380 (14). Double-stranded cDNAs were cloned in the λgt11 phage vector. By screening approximately 2×10^5 recombinants with chromosome 18 probe A (Fig. 1), we have obtained three independent cDNA clones (B3, -4, and -10) that overlap each other. As shown in Fig. 1, clone B3 contains 19 deoxyadenylate residues at the end, indicating that this clone represents the 3′ end of mRNA. The restriction maps of cDNA clones and genomic sequences are colinear from the 3′ end of the cDNA clone B3 until just before the *Bam*HI site of the cDNA sequence. The cDNA sequence beyond this point diverges from genomic sequences (data not

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: kb, kilobase(s).

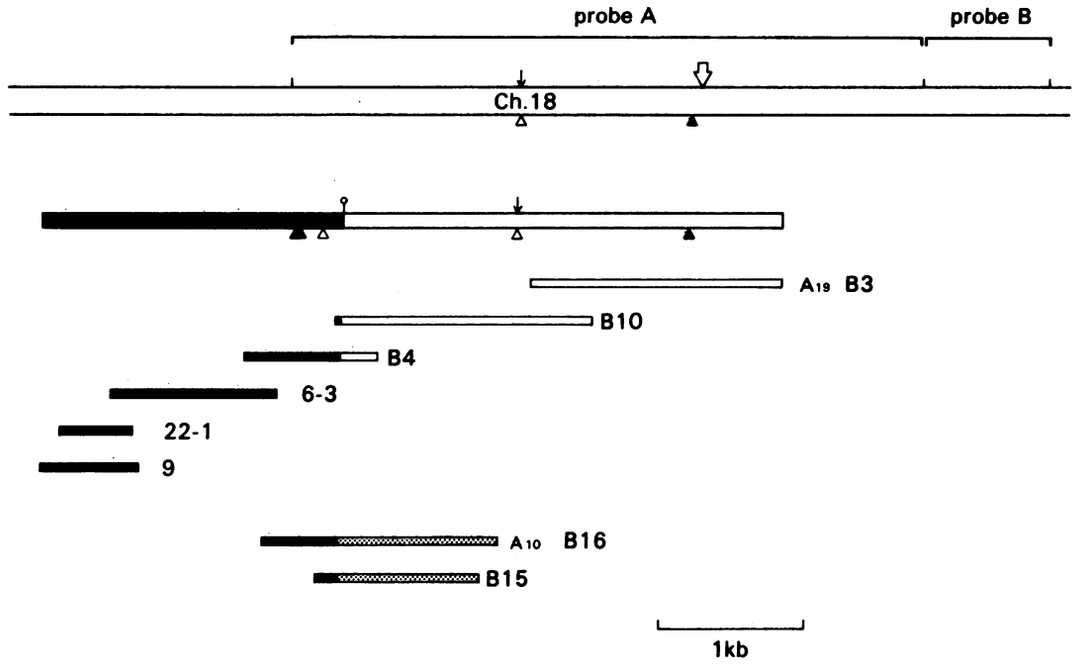

FIG. 1. *bcl-2* cDNA clones. Top double lines represent genomic (chromosome 18) restriction map corresponding to 3′ part of cDNA (second exon). The restriction map of genomic DNA was deduced by analysis of the several overlapping genomic clones, which have been described (14, 15). The hot spot of the breakpoint of the t(14;18) chromosome translocation is shown by the large open arrow. Structures of cDNA clones are shown below the genomic restriction map. Clones B3, B4, B10, B15, and B16 were obtained from a cell line 380 cDNA library made by using oligo(dT) as primer. The other clones, 22-1, 6-3, and 9, were obtained from the primer-extension library. For each clone, the open box represents the second exon and the filled or stippled dotted box represents the first exon. Lengths of oligo(dA) are indicated for clones B3 and B16. Restriction sites are shown for *Sst* I (open triangle), *Sst* II (solid triangle), *Hin*dIII (vertical line), *Bam*HI (circle with stem), and *Eco*RI (small arrow). kb, Kilobase.

shown). Thus, the cDNA sequences correspond to at least two genomic regions.

The 5′ part of cDNA clone B4 (5′ end to *Bam*HI site) detects another set of cDNA clones, B15 and B16 (Fig. 1). These two cDNA clones share the same sequences at the 5′ region with B4 but are totally different at the 3′ part from B3, B4, and B10 (see Fig. 1).

To obtain cDNA sequences further upstream, we also constructed a cDNA library by primer-extension and obtained a series of overlapping clones, 6-3, 22-1, and 9.

***bcl-2* mRNA.** As described above, cDNA cloning has revealed two different sets of clones, indicating that the *bcl-2* gene is transcribed into at least two mRNAs. In order to visualize these mRNA species, two different probes were used for blot hybridization. One is a genomic DNA (probe A, Fig. 1) that contains the 3′ exon. The other probe is a cDNA (clone 22-1) that corresponds to the 5′ exon.

The genomic DNA probe A detects 8.5-kb and 5.5-kb transcripts (Fig. 2*A*). On the other hand, the cDNA probe 22-1 detects an additional 3.5-kb transcript (Fig. 2*B*). The genomic DNA probe B, 3′ of the genomic probe A, hybridizes to 8.5-kb mRNA (data not shown), indicating that the 8.5-kb transcript uses a polyadenylylation site further downstream. Therefore, we conclude that the *bcl-2* gene is transcribed into three mRNAs of different sizes. The possibility that these mRNAs are derived from different but related genes is excluded by the fact that under the hybridization conditions used for RNA blot hybridization, the probes detect only one cellular gene.

It was shown previously (16) that the majority of follicular lymphomas carrying the t(14;18) translocation have breakpoints within a region about 100 base pairs in length on chromosome 18. The cDNA sequences indicate that the most common breakpoint hot spot for the t(14;18) translocation is within the 3′ exon (see Fig. 1). Follicular lymphoma cell line RS, with a t(14;18) translocation whose breakpoint is mapped in the hot-spot region (data not shown), shows aberrant-size *bcl-2* mRNAs, in addition to *bcl-2* transcript of the normal size (Fig. 2). This aberrant-size mRNA could be a chimeric RNA consisting of sequences derived from *bcl-2* and chromosome 14 sequences. That we observed 8.5-kb *bcl-2* transcript in RS follicular lymphoma cells suggests that both the

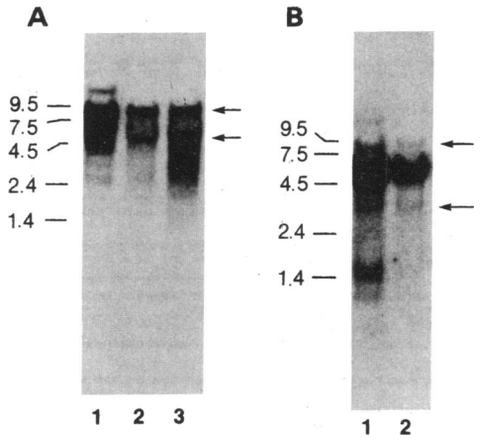

FIG. 2. RNA blot hybridization. Poly(A)$^+$ mRNA (2.5 μg) isolated from cell lines was electrophoresed, transferred to nitrocellulose, and hybridized with ^{32}P-labeled DNA probes. (*A*) Lanes: 1, RS follicular lymphoma RNA; 2, GM697 pre-B-cell leukemia; 3, 380 pre-B-cell leukemia (14). The RNA was hybridized with the genomic DNA probe A (see Fig. 1). (*B*) Lanes: 1, RS follicular lymphoma; 2, GM697 (pre-B-cell leukemia). The RNA was hybridized with cDNA clone 22-1 (see Fig. 1). Arrows indicate 8.5-, 5.5-, and 3.5-kb transcripts. RNA ladder (Bethesda Research Laboratories) was used as size (kb) markers. Since probe 22-1 cross-hybridizes with 28S rRNA very strongly, the hybridization of 5.5-kb transcript seems to be hidden behind rRNA cross-hybridization.

normal and the rearranged *bcl-2* genes are expressed in the lymphoma cells. A different situation occurs in Burkitt lymphoma, where the uninvolved c-*myc* gene is transcriptionally silent (7).

Nucleotide Sequence of the *bcl-2* Gene. Nucleotide sequence derived from 5.5-kb *bcl-2* mRNA is shown in Fig. 3*A*. The DNA sequence of 5105 nucleotides reveals one possible open reading frame, encoding 239 amino acid residues (bcl-2α

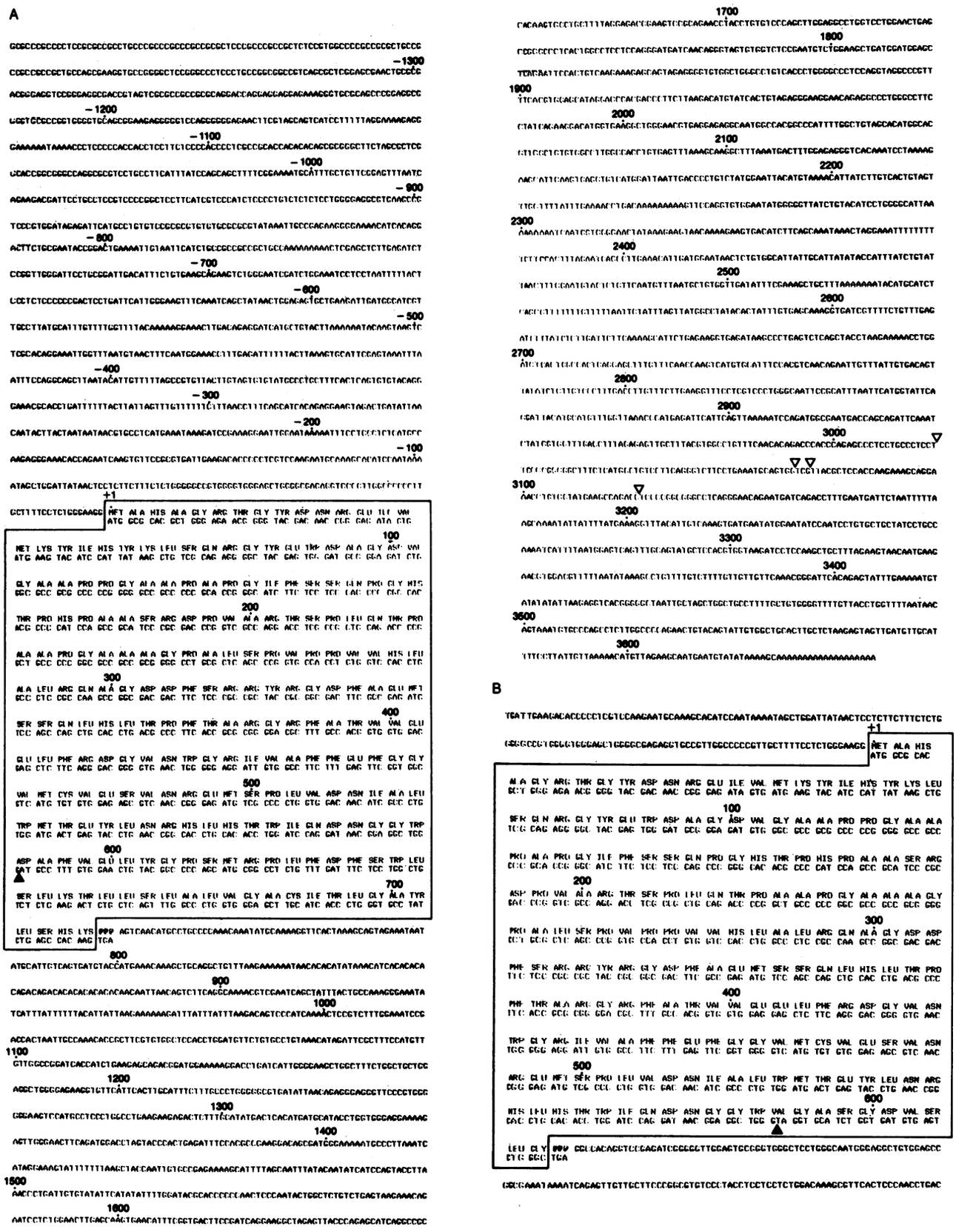

FIG. 3. (*Legend appears at the bottom of the opposite page.*)

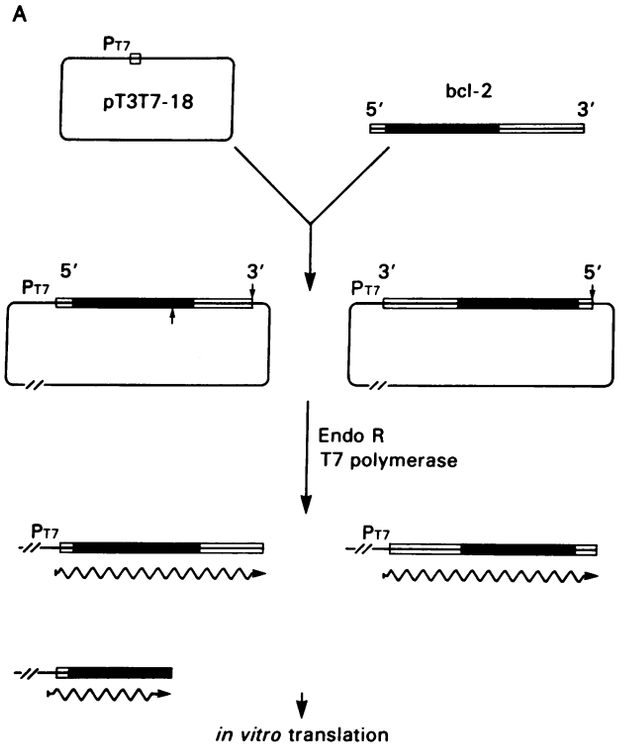

FIG. 4. *In vitro* translation of *bcl-2* mRNA. (*A*) Strategy for preparation of *bcl-2* mRNA *in vitro*. cDNA sequence (−93 to +1348) corresponding to 5.5-kb transcript was subcloned in vector pT3T7-18 (Bethesda Research Laboratories), which contains the promoter for T7 RNA polymerase (P_{T7}). Open reading frame is shown by the solid box. After linearization of template DNAs by cutting with restriction endonuclease (Endo R) at the end of cDNA insert and/or within the open reading frame (at +563), *bcl-2* mRNA was produced as described in *Materials and Methods*. About 0.5 μg of RNA (shown by wavy lines) was assayed by translation *in vitro*. (*B*) Analysis of *in vitro* translation products by NaDodSO$_4$/12.5% polyacrylamide gel electrophoresis. Lanes: 1, no RNA; 2, antisense *bcl-2* RNA; 3, sense RNA; 4, sense RNA truncated within the open reading frame (at +563 in Fig. 3). Positions and sizes (kDa) of marker proteins are given at left.

protein). The nucleotide sequence corresponding to the 3.5-kb *bcl-2* mRNA (Fig. 3*B*) codes for another protein, consisting of 205 amino acid residues (bcl-2β protein), which is identical to bcl-2α except at the carboxyl terminus. Thus, the *bcl-2* gene seems to code for two different protein products.

To determine whether the open reading frame observed on the basis of the cDNA sequences is biologically active, we used the strategy shown in Fig. 4*A*. The cDNA sequence with the open reading frame for bcl-2α protein was cloned so that it was under control of the promoter of the *E. coli* phage T7 and was transcribed by T7 RNA polymerase *in vitro*. The RNA transcripts made *in vitro* were assayed in an *in vitro* translation system, and the protein products, labeled with [^{35}S]methionine, were analyzed by NaDodSO$_4$/PAGE (Fig. 4*B*). Antisense RNA, transcribed from DNA constructs in which the cDNA sequence was inserted in the wrong orientation with respect to the T7 promoter, did not produce any protein. On the other hand, sense RNA produced two major protein products of approximately 28–30 kDa, close to the size expected for the *bcl-2* protein products from the DNA sequence analysis. The two proteins are presumably initiated from the two ATG codons at nucleotide positions +1 and +46. The small size differences observed on NaDodSO$_4$/PAGE could be due to the high proline content of the *bcl-2* gene products. To confirm that these 28- to 30-kDa proteins are derived from the open reading frame of *bcl-2*, sense RNA was made from a DNA construct in which the open reading frame was truncated by cutting with restriction enzyme *Bam*HI (at +563 in Fig. 3*A*). This truncated sense RNA produced proteins smaller than 28–30 kDa (Fig. 4*B*), confirming that the 28- to 30-kDa proteins are derived from the *bcl-2* open reading frame shown in Fig. 3*A*. Computer search did not show any homology of bcl-2α and bcl-2β to any protein whose sequence is in the Dayhoff protein database (National Biomedical Research Foundation, Washington, DC).

Genomic Organization of *bcl-2*. As described above, cDNA sequences corresponding to the 5.5-kb transcript consist of at least two exons. To identify the 5′ exon(s), a genomic clone, λ18-27, was obtained that hybridizes with the 5′ region of the B4 clone (5′ end to *Bam*HI site). Restriction mapping and DNA sequencing of part of this genomic clone indicated that 18-27 DNA has all cDNA sequences 5′ of the splice site and that the genomic DNA sequences are colinear to B16 and B15 cDNA sequences without any intervening sequence (data not shown). Thus, the 3.5-kb mRNA is transcribed from this 5′ exon, without any intron. On the other hand, the 5.5-kb transcript is produced by splicing within the first exon and joining to the second (3′) exon.

DISCUSSION

It was reasoned that it should be possible to identify and characterize human genes involved in the pathogenesis of B-cell leukemias and lymphomas by using overlapping clones to "walk" upstream from the involved immunoglobulin heavy-chain locus (12–16). By use of this approach, the t(11;14)(q13;q32) and t(14;18)(q32;q21) chromosome breakpoints in human B-cell malignancies were cloned by us and our coworkers (12–16). At chromosome band 18q21, a gene, *bcl-2*, that is activated by its proximity to the heavy-chain locus was identified (14–16). Subsequently, two other laboratories have cloned the t(14;18) chromosome breakpoints (26, 27). In most follicular lymphomas, the chromosome 18 breakpoints involve the 3′ end of the *bcl-2* gene or are 3′ of the involved *bcl-2* gene. In this study, we have investigated the structure of the *bcl-2* gene and its transcripts by cloning

FIG. 3. (*on opposite page*). Nucleotide sequence of *bcl-2* cDNA determined from overlapping cDNA clones shown in Fig. 1. Open reading frames are boxed. (*A*). cDNA sequence corresponding to 5.5-kb *bcl-2* transcript. (*B*) cDNA sequence corresponding to 3.5-kb transcript. Only DNA sequence within or immediately flanking the open reading frame is shown. The splicing site is shown by a solid triangle in *A* and in *B*. The breakpoints of the t(14;18) translocation in four cases of follicular lymphoma that were previously reported (15, 16) are shown by open triangles in *A*.

bcl-2 cDNA and comparing the cDNA clones to the *bcl-2* genomic DNA. The results of this analysis indicate that the *bcl-2* gene consists of at least two exons. There *bcl-2* transcripts, 8.5 kb, 5.5 kb, and 3.5 kb long, have been identified, the smaller transcripts being derived only from the first exon. The 8.5-kb transcript has more sequence at the 3′ end than the 5.5-kb transcript. This extra 3′ noncoding sequence might derive from another exon(s) or a larger second exon.

Sequence analysis of the *bcl-2* cDNA indicates that the *bcl-2* gene codes for two protein products, bcl-2α and bcl-2β, the first of 239 amino acids and the second of 205 amino acids, that differ at the carboxyl terminus. Since the predicted amino acid sequences of the *bcl-2*-encoded proteins do not show transmembrane domain or leader peptide, the *bcl-2* proteins do not seem to be either membrane or secreted proteins. Since most of the chromosome 18 breakpoints occur either within the second exon or 3′ of the second exon, we conclude that the *bcl-2* protein products in follicular lymphomas are identical to the normal *bcl-2* protein products. Since the *bcl-2* gene involved in the t(14;18) translocation rearranges within a region 3′ of the *bcl-2* coding sequences, we conclude that the transcription of the involved *bcl-2* gene and of the heavy-chain locus occurs in the same direction and that the 3′ end of the *bcl-2* gene is more proximal than its 5′ end at band 18q21.

We gratefully acknowledge Drs. R. Gemmill and R. Morgan for providing us with the follicular lymphoma cell line RS and Dr. R. Ricciardi for helping us with the *in vitro* translation system. We thank M. Bashir and R. Moore for excellent technical assistance and Mrs. K. A. Reinersmann for preparation of this manuscript. This research was supported by Grant CA39860 from the National Cancer Institute and by Grant CD-185 from The American Cancer Society.

1. van den Berghe, H., Parloir, C., David, G., Michaux, J. L. & Sokal, G. (1979) *Cancer* **44**, 188–195.
2. Yunis, J. J., Oken, M. M., Theologides, A., Howe, B. B. & Kaplan, M. E. (1984) *Cancer Genet. Cytogenet.* **13**, 17–28.
3. Yunis, J. J., Oken, M. M., Kaplan, M. E., Ensrud, K. M., Howe, R. B. & Theologides, A. (1982) *N. Engl. J. Med.* **307**, 1231–1236.
4. Dalla Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R. C. & Croce, C. M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7824–7827.
5. Croce, C. M., Thierfelder, W., Erikson, J., Nishikura, K., Finan, J., Lenoir, G. & Nowell, P. C. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6922–6926.
6. Erikson, J., Nishikura, K., ar-Rushdi, A., Finan, J., Emanuel, B. S., Lenoir, G., Nowell, P. C. & Croce, C. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7581–7585.
7. Nishikura, K., ar-Rushdi, A., Erikson, J., Watt, R., Rovera, G. & Croce, C. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4822–4826.
8. Croce, C. M., Shander, M., Martinis, J., Cicurel, L., D'Ancona, G. G., Dolby, T. W. & Koprowski, H. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3416–3419.
9. Erikson, J., Finan, J., Nowell, P. C. & Croce, C. M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5611–5615.
10. Nowell, P. C., Shankey, T. V., Finan, J., Guerry, D. & Besa, E. (1981) *Blood* **57**, 444–451.
11. van den Berghe, H., Vermaelen, K., Louwagie, A., Criel, A., Mecucci, C. & Vaerman, J. P. (1984) *Cancer Genet. Cytogenet.* **11**, 381–387.
12. Tsujimoto, Y., Yunis, J., Onorato-Showe, L., Nowell, P. C. & Croce, C. M. (1984) *Science* **224**, 399–402.
13. Tsujimoto, Y., Jaffe, E., Cossman, J., Gorham, J., Nowell, P. C. & Croce, C. M. (1985) *Nature (London)* **315**, 340–343.
14. Tsujimoto, Y., Finger, L. R., Yunis, J., Nowell, P. C. & Croce, C. M. (1985) *Science* **226**, 1097–1099.
15. Tsujimoto, Y., Cossman, J., Jaffe, E. & Croce, C. M. (1985) *Science* **228**, 1440–1443.
16. Tsujimoto, Y., Gorham, J., Cossman, J., Jaffe, E. & Croce, C. M. (1985) *Science* **229**, 1390–1393.
17. ar-Rushdi, A., Tan, K. B. & Croce, C. M. (1982) *Somatic Cell Genet.* **8**, 151–161.
18. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
20. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
21. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
22. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
23. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
24. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
25. Ricciardi, R. P., Miller, J. S. & Roberts, B. E. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4927–4931.
26. Bakhshi, A., Jensen, J. P., Goldman, P., Wright, J. J., McBride, O. W., Epstein, A. L. & Korsmeyer, S. J. (1985) *Cell* **41**, 899–906.
27. Cleary, M. L. & Sklar, J. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7439–7443.