

## SHORT COMMUNICATION

# Molecular Cloning, Physical Mapping, and Expression Analysis of a Novel Gene, BCL2L12, Encoding a Proline-Rich Protein with a Highly Conserved BH2 Domain of the Bcl-2 Family

Andreas Scorilas,<sup>\*†</sup> Lianna Kyriakopoulou,<sup>\*†</sup> George M. Yousef,<sup>\*†</sup> Linda K. Ashworth,<sup>‡</sup> Aku Kwamie,<sup>\*</sup> and Eleftherios P. Diamandis<sup>\*†,1</sup>

<sup>\*</sup>Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada; <sup>†</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5G 1X5, Canada; and <sup>‡</sup>Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California 94551

Received August 14, 2000; accepted November 20, 2000

**Members of the Bcl-2 family of apoptosis-regulating proteins contain at least one of the four evolutionarily conserved domains, termed BH1, BH2, BH3, or BH4. Here, we report the identification, cloning, physical mapping, and expression pattern of BCL2L12, a novel gene that encodes a BCL2-like proline-rich protein. Proline-rich sites have been shown to interact with Src homology region 3 (SH3) domains of several tyrosine kinases, mediating their oncogenic potential. This new gene maps to chromosome 19q13.3 and is located between the IRF3 and the PRMT1/HRMT1L2 genes, close to the RRAS gene. BCL2L12 is composed of seven coding exons and six intervening introns, spanning a genomic area of 8.8 kb. All of the exon-intron splice sites conform to the consensus sequence for eukaryotic splice sites. The BCL2L12 protein is composed of 334 amino acids, with a calculated molecular mass of 36.8 kDa and an isoelectric point of 9.45. The BCL2L12 protein contains one BH2 homology domain, one proline-rich region similar to the TC21 protein and, five consensus PXXP tetrapeptide sequences. BCL2L12 is expressed mainly in breast, thymus, prostate, fetal liver, colon, placenta, pancreas, small intestine, spinal cord, kidney, and bone marrow and to a lesser extent in many other tissues. We also identified one splice variant of BCL2L12 that is primarily expressed in skeletal muscle.** © 2001 Academic Press

Programmed cell death and apoptosis are highly regulated physiological processes (5, 18). Dysregulation of apoptosis results in several diseases including cancer,

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. AF289220.

<sup>1</sup>To whom correspondence and reprint requests should be addressed at Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. Telephone: (416) 586-8443. Fax: (416) 586-8628. E-mail: ediamandis@mtsinai.on.ca.

loss of tissue due to ischemia, and AIDS (19). Apoptotic events are regulated by a number of proteins that exert either a positive (pro-apoptotic) or a negative (anti-apoptotic) effect. Proteins participating in these events include members of the Bcl-2 family (1). Members of the Bcl-2 family are characterized by the presence of at least one of the BH1, BH2, BH3, or BH4 domains (14). The BH1 and BH2 domains are present in all anti-apoptotic proteins, while the BH3 domain is present in the pro-apoptotic members of the family. The BH4 domain appears to be present in the N-terminal domain of the anti-apoptotic Bcl-2, Bcl-X<sub>L</sub>, and Bcl-w proteins, and its function is not clear as yet.

In our search for genes involved in the process of apoptosis, we are using positional cloning techniques. A potential BH2 domain was identified in the (BAC)<sup>2</sup> clone BC42053, sequenced by the Department of Energy's Joint Genome Institute. Genomic sequences from this clone were in the form of 87 contigs of different lengths. The clone containing this gene was obtained from the Lawrence Livermore National Laboratory, and the genomic DNA was isolated. The chromosome 19 *EcoRI* restriction map (3) and long PCR strategies, using genomic DNA from this clone, were used to construct a contiguous area of the genomic area of interest. Bioinformatics approaches were used to predict the presence of new genes, and a putative new apoptosis-related, proline-rich protein was identified. Blast search and *EcoRI* restriction digestion analysis of the sequence of one adjacent (telomeric) cosmid clone (R31181) allowed us to identify the relative position of the new gene and other previously identified genes, RRAS, IRF3, and PRMT1, along the same chromosomal region (19q13.3). The sequence of the putative new gene was then verified by

<sup>2</sup> Abbreviations used: BCL2L12, BCL2-like 12 (proline-rich); IRF3, interferon regulatory factor 3; EST, expressed sequence tag; RT-PCR, reverse transcription-polymerase chain reaction; BAC, bacterial artificial chromosome; PAC, P1-derived artificial chromosome.



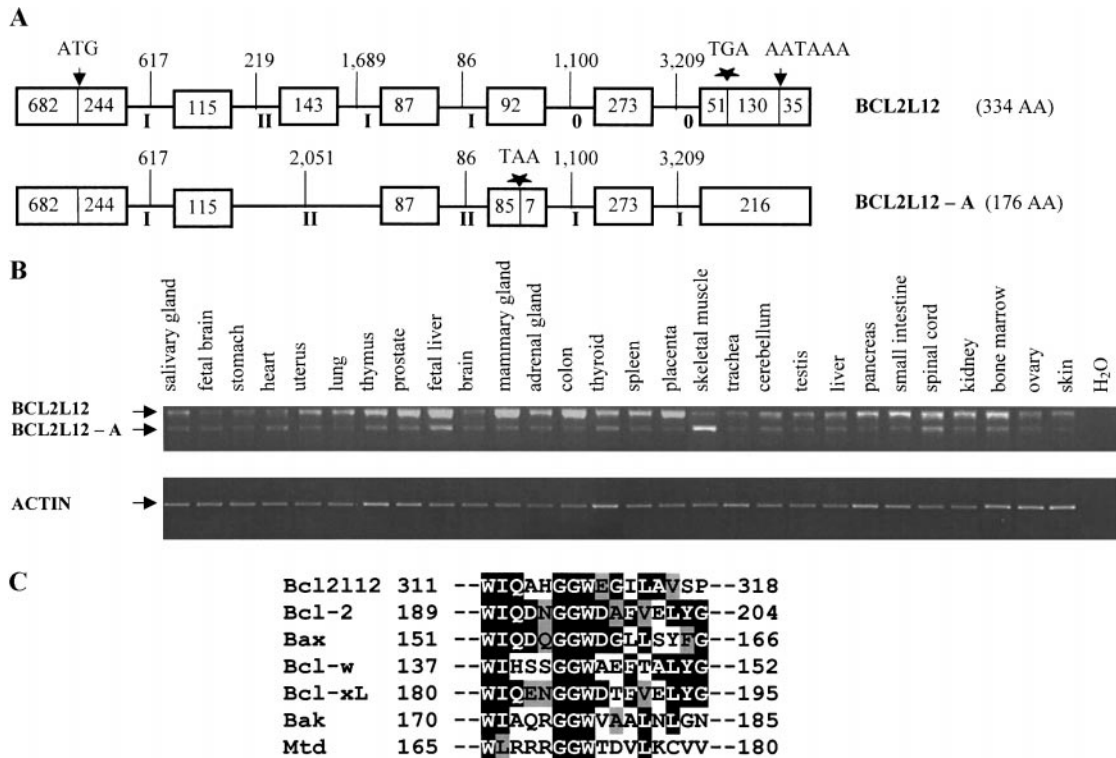
gggggtacaaacaggaaacctctcttcccacatcctcccacatcctcccagagaggtctcttaa  
 tcagctagggttccctcccacatccatggcacatcctcccagcgtcttcagtcgacagc  
 cctccaactccctcaattccggccagttatagacctctccctcccacacatcagtaagaagt  
 AAGTATCCCCACTTTTCAGGGTAGCATCCTCTTTCCCTAGTTATTTTCAGTGCAGAATCGATTACG  
 TGTAAGCCTCTACCTCTGCACCTACCCCAATGTAGACCCCTTCCACAGTCCCCCTACAGAAGA  
 TCTCCCAAAACATTTCCAACACAGAAGTTCATTCATTCCTCCCGACTGCCCGCTCCCTCCT  
 CCAGGACAGAGCACGCCCCAGGTTTCGAGACTGAACTACCCCATCTACGGAGTCCACCCCTCCC  
 TCTCTAGAACCCCTCAGTGTAGACGCCTCCTCTTTCTCTCTTCCCTCCCTACTTTTCTAGTTT  
 TTCTGCAGCCGACCTCCAGCGTCGGCCACTGTAGCTCCTTCCCTGCTCCACTTTCCCGAGATAC  
 ACAGCCTGCCTTTCCACTCTCGTGCAGACCCATCCTCTTTCCCGCTCCTCGCTCTCGGACGCCA  
 CCAACGCTCTCCCGGGCTCTTCCCGGTTACCTACGATGGAAGGTCGGGGCGTGGGGCAGCTGGA  
 ACCACCCCTGTCTTGGAGCTCCGGGTAGCTCTCAACTCGAGGCTGCCACCCCTTCCCGTC  
 AGCTGAGCTCTAGAGCGCTGGGGCTTTCTTTTGGATCGACTCTTGAATAAAACCACTTTATC  
 ATTCTTTGGGTAACAGACCCAAAAGCCG **(ATG)** GGA CGG CCC GCT GGG CTG TTC CCG  
 CCC CTA TGC CCT TTT TTG GGT TTC CGG CCA GAG GCA TGC TGG GAG CGT  
 P L C P F L G F R P E A C W E R  
 CAC ATG CAA ATT GAG CGT GCA CCC AGC GTT CCG CCC TTT CTA CGC TGG  
 H M Q I E R A P S V P P F L R W  
 GCC GGT TAT CGA CCC GGC CCA GTG CGC AGG CGC GGG AAA GTT GAA CTA  
 A G Y R P G P V R R R G K V E L  
 ATA AAG TTT GTA CGA GTT CAG TGG AGG AGA CCG CAA GTT GAG TGG AGG  
 I K F V R V Q W R R P Q V E W R  
 AGG CGG CGG TGG GGC CCC GGA CCA **gtc**agcggg...intron-1...tctcac  
**R R R W G P G**  
**ag**GT GCC TCC ATG GCA GGC TCT GAA GAG CTG GGG CTC CGG GAA GAC ACG  
 A S M A G S E E L G L R E D T  
 CTG AGG GTC CTA GCT GCC TTC CTT AGG CGT GGT GAG GCT GCC GGG TCT  
 L R V L A A F L R R G E A A G S  
 CCT GTT CCA ACT CCA CCT **ag**taagagg...intron-2...cctctcc**aga** AGC  
 P V P T P P R S  
 CCT GCC CAA GAA GAG CCA ACA GAC TTC CTG AGC CGC CTT CGA AGA TGT  
 P A Q E E P T D F L S R L R R C  
 CTT CCC TGC TCC CTG GGG CGA GGA GCA GCC CCC TCT GAG TCC CCT CGG  
 L P C S L G R G A A P S E S P R  
 CCT TGC TCT CTG CCC ATC CGC CCC TGC TAT GGT TTA GAG CCT **gtc**aaagag  
 P C S L P I R P C Y G L E P G  
 ...intron-3...tcaccc**ag**GC CCA GCT ACT CCA GAC TTC TAT GCT TTG GTG  
 P A T P D F Y A L V  
 GCC CAG CGG CTG GAA CAG CTG GTC CAA GAG CAG CTG AAA TCT CCG CCC  
 A Q R L E Q L V Q E Q L K S P P  
 AGC CCA **gtg**agagca...intron-4...ctgcctcc**ag**AA TTA CAG GGT CCC CCA  
 S P E L Q G P P  
 TCG ACA GAG AAG GAA GCC ATA CTG CGG AGG CTG GTG GCC CTG CTG GAG  
 S T E K E A I L R R L V A L L E  
 GAG GAG GCA GAA GTC ATT AAC CAG AAG**gtg**gatggg...intron-5...ccctac**ag**  
 E E A E V I N Q K  
 CTG GCC TCG GAC CCC GCC CTG CGC AGC AAG CTG GTC CGC CTG TCC TCC  
 L A S D P A L R S K L V R L S S  
 GAC TCT TTC GCC CGC CTG GTG GAG CTG TTC TGT AGC CGG GAT GAC AGC  
 D S F A R L V E L F C S R D C D S  
 TCT CGC CCA AGC CGA GCA TGC CCC GGG CCC CCG CCT CCT TCC CCG GAG  
 S R P S R A C P G P P P P S P E  
CCC CTG GCC CGC CTG GCC CTA GCC ATG GAG CTG AGC CGG CGC GTG GCC  
 P L A R L A L A M E L S R R V A  
 GGG CTG GGG GGC ACC CTG GCC GGA CTC AGC GTG GAG CAC GTG CAC AGC  
 G L G G T L A G L S V E H V H S  
 TTC ACG CC TGG ATC CAG GCC CAC GGG GGC TGG**gtg**agccgct...intron-6  
 F T P **W I Q A H G G W**  
 ...cccttcc**ag**GAG GGC ATC CTG GCT GTT TCA CCC GTG GAC TTG AAC TTG  
 E G I L A V S P V D L N L  
 CCA TTG GAC **(TGA)** GCTCTTTCTCAGAAGCTGCTACAAGATGACACCTCATGTCCTGCCCTC  
 P L D \*  
 TTCGTGTGCTTTTCCAAAGTCTTCTATTCCACTCAGGGCTGTGGGGTGGTGGTTGCCCTACCTG  
 TTTTGGCCAAAATAAATTGTTTAAAACCTTTTCTTATTA AAAACGTTACAAAGtattcaattg  
 tctgttcttacacttgggaactaaagttaaggagaggagatttggttcagagatttc  
 tgtggcttacgcaaatgaggaaagaaccagttcattcattctagaatattttatttgc

**FIG. 1.** Genomic organization and partial genomic sequence of the BCL2L12 gene. Intronic sequences are not shown except for the splice junction areas. Introns are shown with lowercase letters and exons with uppercase letters. The coding nucleotides are shown in triplets. For full sequence, see GenBank Accession No. AF289220. The start and stop codons are encircled, and the exon-intron junctions are in black square boxes. The BH2 homology domain is highlighted in black. The possible PXXP motifs are underlined, and the TC21 identical motif is doubly overlined. The putative TATA box and polyadenylation signal are in boldface type.

various experimental approaches, including EST database searching, RT-PCR screening of 28 tissues, and PCR product sequencing from both directions for each tissue. The 3' and 5' termini of the gene were verified using RACE technology.

As shown in Fig. 1, the BCL2L12 (from Bcl-2-like 12)

gene is formed of seven coding exons and six intervening introns, spanning an area of 8772 bp of genomic sequence on chromosome 19q13.3. All of the exon-intron splice sites (mGT...AGm) conform to the consensus sequence for eukaryotic splice sites (8). Lengths of the coding exons are 926, 115, 143, 87, 92, 273, and



**FIG. 2.** (A) Schematic representation of the BCL2L12 gene. Exons are represented by boxes, and introns are represented by connecting lines. Numbers inside the boxes or above the lines refer to basepairs, while the amino acid numbers (AA) are given in parentheses at the end of the line. Roman numerals indicate intron phases. (B) Tissue expression of the BCL2L12 gene, as determined by RT-PCR. The splice variant (lower PCR band) is highly expressed in skeletal muscle. (C) Alignment of the BH2 deduced amino acid sequence of BCL2L12 with members of the Bcl-2 multigene family. Identical amino acids are highlighted in black and similar residues in gray.

216 bp, respectively. The predicted protein-coding region of the gene is formed of 1005 bp, encoding a deduced 334-amino-acid polypeptide with a predicted molecular mass of 36.8 kDa and an isoelectric point of 9.45. An (AATAAA) sequence is identical with a consensus polyadenylation signal (13) and is followed, after 35 nucleotides, by the poly(A) tail not found in the genomic sequence. Another (ATATAA) sequence was reported to occur as a natural polyadenylation variant in 12% of vertebrate mRNA sequences (15).

The highlighted region in Fig. 1 indicates a 16-amino-acid sequence, found in the BH2 domain of several members of the Bcl-2 family of proteins (14). The amino acid loop (WIQXXGGW) at positions 311–318 was also found in the Bcl-2, Bax, and Bcl-X<sub>L</sub> proteins (Fig. 2C). No BH1, BH3, and BH4 domains were found. Bcl-2 family genes containing only one BH domain have been reported (9). Although most Bcl-2 family members contain a combination of the BH1, BH2, BH3, or BH4 domains, a truncated form of the Bax protein (11) and the mammalian activator of apoptosis, Hara-kiri (9), contain only one domain. Thus, it appears that a full complement of BH domains may not be necessary to confer apoptotic or anti-apoptotic activity. Through the use of sequence analysis tools, we were able to identify various putative posttranslational modification sites. There are numerous potential sites for O-glycosylation. Furthermore, several possible sites of phosphorylation have been identified for

cAMP-dependent protein kinase, protein kinase C, and casein kinase 2. In addition, several N-myristoylation sites have been predicted (Table 1).

The BCL2L12 protein was found to have proline-rich sites. One PPPP site as well as five PP amino acid sites is present in this protein. Eight putative PXXP motifs were also identified (Fig. 1). Proline-rich motifs are characterized by the presence of the consensus PXXP tetrapeptide, found in all proline-rich proteins identified to date (2). It is known that SH3 domains recognize proline-rich sequences and that all known SH3-bind-

**TABLE 1**

**Putative Posttranslational Modification Sites in the Novel BCL2L12 Gene**

Modification	Residue	BCL2L12 <sup>a</sup>	BCL2L12-A <sup>a</sup>
O-glycosylation	Thr	117	117, 151
	Ser	113, 121, 263, 273	113, 145
Protein kinase C phosphorylation	Thr	97, 204	97
	Ser	151, 259, 287	—
Casein kinase II phosphorylation	Thr	204	—
	Ser	255, 325	153
N-myristoylation	Gly	82, 145, 292,	82, 130, 159,
		294, 295, 316	166, 171

<sup>a</sup> The residues are numbered according to the sequence shown in Fig. 1 and our GenBank submission (Accession No. AF289220).



ing proteins contain proline-rich regions with at least one PXXP motif (2, 6, 10, 16). Proline-rich domains have been identified in a number of diverse proteins such as epidermal growth factors, phosphatidylinositol 3-kinase, and, more recently, the small GTPase RRAS protein and members of the RRAS superfamily such as the TC21 protein (10, 17). Moreover, the amino acid loop (PPSPEP) at positions 271–276 of the BCL2L12 protein is identical with the PXXP motif present in the RRAS and TC21 oncogenes (17). This motif is required for integrin activation.

PCR screening for BCL2L12 transcripts using gene-specific primers revealed the presence of two bands in most of the tissue cDNAs examined. The two bands were gel-purified, cloned, and sequenced. The upper band represents the classical form of the gene, and the lower band is a splice variant. As shown in Fig. 2A, this variant (BCL2L12-A) is missing exon 3 (143 bp). This splice variant is expected to encode a truncated protein of 176 amino acids with five PP proline sites, two putative PXXP motifs, and no BH2 homology domain.

By identifying the position of the RRAS, IRF3, PRMT1/HRMT1L2, and BCL2L12 genes along these clones, we were able to define precisely the relative location and the direction of transcription of these four genes. PRMT1 is the most telomeric, and its direction of transcription is from centromere to telomere, followed by BCL2L12, which is more centromeric and transcribes in the same direction. The distance between the two genes is 3.4 kb. RRAS and IRF3 genes are more centromeric, located at a distance of 25.1 and 0.3 kb from BCL2L12, respectively, and are transcribed in the opposite direction.

RT-PCR analysis, with primers GGA GAC CGC AAG TTG AGT GG and GTC ATC CCG GCT ACA GAA CA from different human tissues, was used to identify the expression pattern of BCL2L12. Actin was used as a control gene. In each of the 26 adult and 2 fetal tissues tested, a BCL2L12 transcript was identified. As shown in Fig. 2B, the classic form of the BCL2L12 gene is highly expressed in the thymus, prostate, fetal liver, mammary, colon, placenta, small intestine, kidney, and bone marrow. Lower levels of expression are also seen in all other tissues tested. The splice variant is highly expressed in skeletal muscle, fetal liver, and spinal cord. In skeletal muscle, the levels of BCL2L12-A transcripts are higher than the classic form, in contrast to other tissues. To verify the RT-PCR specificity, representative PCR products were cloned and sequenced. These results were reproducible with different sets of primers and different reaction conditions.

In summary, we have cloned a novel human gene, BCL2L12, that maps to chromosome 19q.13.3–q13.4 between the RRAS and the PRMT1 genes. The predicted 36.8-kDa BCL2L12 protein contained a BH2 domain and several small proline-rich motifs (PXXP). Both the BH2 and the proline-rich domains play important roles in the productive interaction of proteins

during events such as apoptosis or its repression, epidermal factor signaling, cellular localization of cytoplasmic proteins, and activation of the phosphatidylinositol 3-kinase in response to IgM cross-linking (4, 7, 10, 12). The presence of the BH2 domain, present in most anti-apoptotic proteins, together with the proline-rich motifs suggests that the BCL2L12 protein may participate in protein–protein interactions. To our knowledge, this is the first gene identified encoding a protein that contains both a proline-rich domain and a BH2 domain. The presence of both domains, if functional, may represent the crossroads between proteins involved in the regulation of apoptotic events and proteins containing SH3 domains.

#### ACKNOWLEDGMENTS

A portion of this work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48.

#### REFERENCES

- Adams, J. M., and Cory, S. (1998). The Bcl-2 protein family: Arbiters of cell survival. *Science* **281**: 1322–1326.
- Alexandropoulos, K., Cheng, G., and Baltimore, D. (1995). Proline-rich sequences that bind to Src homology 3 domains with individual specificities. *Proc. Natl. Acad. Sci. USA* **92**: 3110–3114.
- Ashworth, L. K., Batzer, M. A., Brandriff, B., Branscomb, E., de Jong, P., Garcia, E., Garnes, J. A., Gordon, L. A., Lamerdin, J. E., Lennon, G., *et al.* (1995). An integrated metric physical map of human chromosome 19. *Nat. Genet.* **11**: 422–427.
- Bar-Sagi, D., Rotin, D., Batzer, A., Mandiyan, V., and Schlessinger, J. (1993). SH3 domains direct cellular localization of signaling molecules. *Cell* **74**: 83–91.
- Choi, W. S., Yoon, S. Y., Chang, H., Choi, E. J., Rhim, H., Jin, B. K., Oh, T. H., Krajewski, S., Reed, J. C., and Oh, Y. J. (2000). Correlation between structure of Bcl-2 and its inhibitory function of JNK and caspase activity in dopaminergic neuronal apoptosis. *J. Neurochem.* **74**: 1621–1626.
- Cohen, G. B., Ren, R., and Baltimore, D. (1995). Modular binding domains in signal transduction proteins. *Cell* **80**: 237–248.
- Gout, I., Dhand, R., Hiles, I. D., Fry, M. J., Panayotou, G., Das, P., Truong, O., Totty, N. F., Hsuan, J., Booker, G. W., *et al.* (1993). The GTPase dynamin binds to and is activated by a subset of SH3 domains. *Cell* **75**: 25–36.
- Iida, Y. (1990). Quantification analysis of 5'-splice signal sequences in mRNA precursors. Mutations in 5'-splice signal sequence of human beta-globin gene and beta-thalassemia. *J. Theor. Biol.* **145**: 523–533.
- Inohara, N., Ding, L., Chen, S., and Nunez, G. (1997). Harakiri, a novel regulator of cell death, encodes a protein that activates apoptosis and interacts selectively with survival-promoting proteins Bcl-2 and Bcl-X(L). *EMBO J.* **16**: 1686–1694.
- Lowenstein, E. J., Daly, R. J., Batzer, A. G., Li, W., Margolis, B., Lammers, R., Ullrich, A., Skolnik, E. Y., Bar-Sagi, D., and Schlessinger, J. (1992). The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. *Cell* **70**: 431–442.
- Ottillie, S., Diaz, J. L., Chang, J., Wilson, G., Tuffo, K. M., Weeks, S., McConnell, M., Wang, Y., Oltersdorf, T., and Fritz, L. C. (1997). Structural and functional complementation of an inactive Bcl-2 mutant by Bax truncation. *J. Biol. Chem.* **272**: 16955–16961.

12. Pleiman, C. M., Hertz, W. M., and Cambier, J. C. (1994). Activation of phosphatidylinositol-3' kinase by Src-family kinase SH3 binding to the p85 subunit. *Science* **263**: 1609–1612.
13. Proudfoot, N. J., Cheng, C. C., and Brownlee, G. G. (1976). Sequence analysis of eukaryotic mRNA. *Prog. Nucleic Acid Res. Mol. Biol.* **19**: 123–134.
14. Reed, J. C., Zha, H., Aime-Sempe, C., Takayama, S., and Wang, H. G. (1996). Structure–function analysis of Bcl-2 family proteins. Regulators of programmed cell death. *Adv. Exp. Med. Biol.* **406**: 99–112.
15. Sheets, M. D., Ogg, S. C., and Wickens, M. P. (1990). Point mutations in AAUAAA and the poly(A) addition site: Effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res.* **18**: 5799–5805.
16. Sparks, A. B., Rider, J. E., Hoffman, N. G., Fowlkes, D. M., Quillam, L. A., and Kay, B. K. (1996). Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCgamma, Crk, and Grb2. *Proc. Natl. Acad. Sci. USA* **93**: 1540–1544.
17. Wang, B., Zou, J. X., Ek-Rylander, B., and Ruoslahti, E. (2000). R-Ras contains a proline-rich site that binds to SH3 domains and is required for integrin activation by R-Ras. *J. Biol. Chem.* **275**: 5222–5227.
18. Wyllie, A. H. (1993). Apoptosis (The 1992 Frank Rose Memorial Lecture). *Br. J. Cancer* **67**: 205–208.
19. Zha, H., Aime-Sempe, C., Sato, T., and Reed, J. C. (1996). Proapoptotic protein Bax heterodimerizes with Bcl-2 and homodimerizes with Bax via a novel domain (BH3) distinct from BH1 and BH2. *J. Biol. Chem.* **271**: 7440–7444.