Molecular Cloning of a Novel Human Leukemia-associated Gene

EVIDENCE OF CONSERVATION IN ANIMAL SPECIES*

(Received for publication, March 23, 1989)

Xiao-xiang Zhu, a,b Karen Kozarsky,* John R. Strahler,* Christoph Eckerskorn,d
Friedrich Lotspeich,d Randa Melhem,* John Lowe,* David A. Fox,* Sam M. Hanash,*
and George F. Atweh a,c

From the Veterans Administration Medical Center and the Departments of Internal Medicine, Pediatrics, and Pathology,
University of Michigan, Ann Arbor, Michigan 48109 and the Max-Planck-Institut für Biochemie, D-8033 Martinsried,
Federal Republic of Germany

We have recently described an 18-kilodalton polypeptide (p18) that is present in much greater abundance in acute leukemia blast cells (myeloid and lymphoid) than in resting or proliferating nonleukemic lymphoid cells or chronic lymphoid and myeloid leukemia cells. In this report we describe the cloning of two different sized full-length cDNAs that code for p18. The two cDNAs differ in their 3' noncoding regions as a result of alternative polyadenylation. Analysis of the complete nucleotide sequence and the corresponding amino acid sequence did not reveal significant homology to any previously described sequences. We show evidence that this gene is highly conserved in several animal species and low stringency hybridization studies suggest that the p18 gene may be a member of a family of partially homologous genes in the human genome.

It is generally believed that more than one event is required for the transformation of a normal cell to a malignant cell. The application of molecular biology to the study of the pathogenesis of cancer has led to the discovery of many oncogenes which may be involved in malignant transformation. In spite of the identification of more than 50 oncogenes in several animal species and low stringency hybridization studies suggesting that the p18 gene may be a member of a family of partially homologous genes in the human genome, however, no evidence of oncogene activation has been found.

We have used a different approach for the investigation of the molecular alterations in acute leukemia in humans. The approach is based on direct analysis of the polypeptide constituents of leukemic cells using two-dimensional polyacrylamide gel electrophoresis (PAGE) to identify polypeptide alterations in leukemic cells relative to non-leukemic cells. This approach led to the identification of an 18-kDa polypeptide (p18) that is present in much greater abundance in cells from patients with acute leukemia of different subtypes than in normal peripheral blood lymphocytes, nonleukemic proliferating lymphoid cells, bone marrow cells, or cells from patients with chronic lymphoid or myeloid leukemia (7). In a previous study, p18 was isolated from two-dimensional polyacrylamide gels, and the sequence of a tryptic decapetide was derived by gas-phase microsequencing (7). In this paper, we report the successful cloning of the cDNA that codes for p18. The availability of this cDNA allowed us to deduce the primary structure of the protein and to perform a study of its conservation in different species.

MATERIALS AND METHODS

Two-dimensional Electrophoresis and Amino Acid Sequence Analysis—A human T cell leukemia cell line (HSB-2) was grown in tissue culture. The cell pellet was solubilized and analyzed by two-dimensional PAGE as previously described (7). The protein of interest was digested with trypsin within the gel matrix, and the sequence of eluted tryptic fragments was determined as described (8).

cDNA Cloning—Poly(A)+ RNA was isolated from the leukemic T cell line HSB-2 and used as template for the synthesis of double-stranded cDNA as described by Gubler and Hoffman (9). The CDM7 plasmid vector that was used in the construction of the cDNA library was the generous gift of Brian Seed (Massachusetts General Hospital) (10, 11). CDM7 was digested with BstXI and ligated to double-stranded cDNA to which BstXI adaptors had been added (10). The resulting recombinant DNA was used to transform competent MC1061/p32 cells to generate the cDNA library.

DNA Sequence Analysis—After generating a limited restriction map of the two different cDNA clones, DNA sequence analysis was performed by the dye deoxy chain termination method of Sanger et al. (14) according to the strategy illustrated in Fig. 1. All sequence analysis was performed using single-stranded M13 clones, some primed with the universal primer and others with synthetic oligonucleotides.

RNA Analysis—RNA was extracted from cells of chronic myelogenous leukemia in acute blastic transformation (K562 cells) (15).
and used to screen the HSB-2 cDNA library. Two independent clones which hybridized strongly to this probe mixture were identified and characterized as described below.

One of the recombinant clones contained a 1.5-kb cDNA insert and the other contained a 1-kb insert. Both inserts were sequenced in their entirety using the strategy outlined in Fig. 1. The obtained nucleotide sequence is shown in Fig. 3. The two cDNAs were identical in their 5'-untranslated regions and their coding regions. They differed, however, in their 3'-untranslated regions. Both cDNAs code for the same 149-amino acid polypeptide that starts with an ATG at position 103 of both transcripts and ends with TAA terminator at position 551. The translated amino acid sequence contained a perfect match for three tryptic peptides whose sequence was determined by gas-phase microsequencing (the original decapeptide described earlier (7) and two additional peptides that have been sequenced since the initial publication: 1) Ala-Ile-Glu-Glu-Asn-Asn-Phe-Ser-Lys, (2) Lys-Leu-Glu-Ala-Ala-Glu-Glu-Arg, (3) Asp-Leu-Ser-Leu-Glu-Glu-Ile-Gln-Lys). The calculated molecular weight of the translated polypeptide is 17,302 with a predicted isoelectric point of 5.64. These are in close agreement with a measured molecular weight of 18,000 and a pI of 5.7 (7).

Northern blot analysis (Fig. 4) was performed using RNA isolated from HeLa cells (lane 1), monkey kidney (COS) cells (lane 2), and K562 cells (lane 3). The filter was hybridized to a probe from the common 5' region of the two p18 cDNA clones. A 1.5- and 1-kb band were seen in RNA from human cells (lanes 1 and 3) while a single 1-kb band was seen in RNA from monkey cells (lane 2). In order to map the site of transcription initiation more accurately and to establish that the cDNA clones are full-length copies of p18 mRNA, we performed primer extension analysis. A DNA fragment which extends from position 103 to position 204 was 5'-end-labeled and used to prime cDNA synthesis using total RNA from monkey cells as template. An extension product of 205 nucleotides was noted (data not shown). This corresponds very closely to the 5'-end of the p18 cDNA and suggests that both clones are full-length cDNAs.

To study the organization of the p18 gene in the human genome, we performed the Southern analysis shown in Fig. 5. Human genomic DNA was digested with seven different restriction enzymes and probed with a p18 cDNA probe. All lanes show either one or two hybridizing fragments which suggests a gene of limited size and complexity. We then performed Southern blot analysis using DNA isolated from human, monkey, chimpanzee, dog, cow, pig, duck, hamster, and mouse. When the filter was probed with the human p18 cDNA and washed at low stringency to detect partial homologies, we were able to detect hybridizing bands in every animal DNA tested (Fig. 6A).

When we compared the pattern of the hybridizing fragments in EcoRI-digested human DNA at high stringency (Fig. 5, lane 3) and at low stringency (Fig. 6A, lane 1), we noted two additional bands in the low stringency blot that were not seen at higher stringency. This suggested to us the possibility that another gene may exist in the human genome which has partial homology to the p18 gene. To confirm that the additional fragments represent partially homologous sequences rather than partial digestion of human DNA with EcoRI, we rehybridized the same filter shown in Fig. 6A to p18 cDNA and washed it at high stringency. The autoradiograph in Fig. 6B shows the disappearance of the two additional bands noted above. This suggests the existence of sequences in the human genome that have partial homology to the p18 gene.
**Molecular Cloning of a Leukemia-associated Gene**

We have recently described the use of high resolution protein separation techniques (two-dimensional PAGE) for the identification of polypeptides that are aberrantly expressed in leukemic cells (7,24). This led to the identification of an 18-kDa polypeptide that is present in significantly greater amounts in leukemic cells than in non-leukemic white blood cells (7).

Our preliminary studies suggested that the increased amount of p18 in leukemic cells is not related to specific cell lineage, differentiation stage, or cell proliferation (7). In this report, we describe the cloning of the gene which codes for this polypeptide and describe the complete structure of the cDNA and its translated protein product.

Several aspects of the cloned gene deserve further comment. The cDNA inserts are derivatives of alternatively polyadenylated mRNAs transcribed from the same gene. The significance of the utilization of different polyadenylation signals in the p18 gene is not clear at present. Since the resulting mRNAs differ only in their 3'-untranslated regions, the use of the different splice sites does not alter the translated amino acid sequence.

**DISCUSSION**

We have recently described the use of high resolution protein separation techniques (two-dimensional PAGE) for the identification of polypeptides that are aberrantly expressed in leukemic cells (7,24). This led to the identification of an 18-kDa polypeptide that is present in significantly greater amounts in leukemic cells than in non-leukemic white blood cells (7). Our preliminary studies suggested that the increased amount of p18 in leukemic cells is not related to specific cell lineage, differentiation stage, or cell proliferation (7). In this report, we describe the cloning of the gene which codes for this polypeptide and describe the complete structure of the cDNA and its translated protein product.

Several aspects of the cloned gene deserve further comment. It appears that the cDNAs that were isolated are derivatives of alternatively polyadenylated mRNAs transcribed from the same gene. The significance of the utilization of different polyadenylation signals in the p18 gene is not clear at present. Since the resulting mRNAs differ only in their 3’-untranslated regions, the use of the different splice sites does...
Molecular Cloning of a Leukemia-associated Gene

1.5-
1.0-
0.5-
0.0-

FIG. 4. Northern blot analysis of p18 mRNA. Total cellular RNA from the HeLa cells (lane 1), COS cells (lane 2), and K562 cells (lane 3) was probed with a p18 cDNA probe. 1.5- and 1-kb bands are seen in HeLa cells and K562 cells (human cell lines) and a 1-kb band is seen in COS cells (monkey cell line).

FIG. 5. Southern blot analysis of the human p18 gene. Human genomic DNA was digested with BamHI (lane 1), BglII (lane 2), EcoRI (lane 3), HindIII (lane 4), KpnI (lane 5), PstI (lane 6), and Sall (lane 7). Hybridization of the Southern filter to the p18 cDNA probe at high stringency led to the detection of one or two bands in every lane. The position of the size markers is shown on the left.

not generate diversity at the protein level, which is the case in several alternatively polyadenylated genes (25, 26). It is conceivable that alternative polyadenylation may serve a regulatory function by generating mRNAs of different stabilities. In HeLa cells and K562 cells, the majority of the p18 mRNA is a product of proximal polyadenylation while in monkey cells (COS), all the p18 mRNA is a product of proximal polyadenylation. Further studies are needed to investigate the significance of these observations and explore a possible regulatory role for the use of different polyadenylation sites.

When the derived DNA and protein sequences were compared with sequences present in Genebank, EMBL, and PIR data bases, no significant homology to known sequences was detected. Analysis of the translated amino acid sequence did not reveal a signal peptide which suggests that p18 is not a secreted protein. Our previous studies have suggested cytosolic localization of p18, based on crude separation of nuclear and cytoplasmic fractions (7). These studies, however, do not exclude some minor nuclear localization. Analysis of the amino acid sequence of p18 did not show any of the well known features of transcriptional regulators (leucine zipper (27), zinc fingers (28), or homeo box sequences (29)). These findings, along with the small size of the protein argue against a regulatory function mediated by DNA binding to promoter or enhancer sequences.

In previous studies, we observed a moderate increase in p18 expression in stimulated lymphocytes relative to resting lymphocytes (7). We recently examined the effect of lymphoid stimulation on the phosphorylation of the p18 gene product by 32P-labeling of lymphoid cells. Our data suggest that at least two phosphorylated forms of p18 increase in amount following lymphoid activation. Interestingly, Feuerstein et al. (30) identified an abundant cytosolic phosphoprotein (pp17) (Mr = 17,000, pI 5.5) in HL-60 promyelocytic leukemia cells whose phosphorylation is increased after exposure to phorbol esters in vitro. They suggested that this protein may play a role in intracellular propagation of growth regulatory signals and proposed to call it "prosolin" (31). Pasmanter et al. (32) described a group of phosphoproteins which they called p19 (Mr = 19,000, pI 5.9, 5.7, and 5.4) whose phosphorylation is stimulated in endocrine tumor cell lines by a variety of secretagogues. Later on, the same group identified and purified a similar protein from bovine brain (33). An antibody that they raised to the purified protein reacted with similar polypeptides in many different species from man to mouse and was present at high levels in HL-60 promyelocytic leukemia cells (34).

2 S. M. Hanash and N. Hailat, manuscript in preparation.
Finally, Sobel et al. (35) described a group of ubiquitous phosphoproteins which they called "stathmin" (Mr, 18,000, pl 5.8–6) that are very abundant in rat brain. They also noted that the phosphorylation of these proteins is regulated by a variety of extracellular effectors which induce different target cellular responses. They pointed out the similarities between the proteins they described and those described by Feuerstein et al. (30) and Pasmanter et al. (32) and suggested that these proteins may provide a relay between extracellular signals and intracellular substrates. They speculated that these proteins may be involved in the regulation of the proliferation, differentiation, and/or functions of the many different cell types in which they were discovered (35).

The protein we describe here shares many of the properties of the proteins described above. It has similar migration properties by two-dimensional PAGE (Mr, 18,000, pl 5.6) and is phosphorylated upon stimulation of lymphoid cells. It is a major cytosolic protein that is highly conserved in evolution (Fig. 6A) and may be a member of a family of related genes (Fig. 6B). Northern blot analysis showed a very high level of expression of p18 in human brain. All these data suggest that p18, p19, prosolin, and stathmin may be different names for the same protein. More studies need to be performed to explore possible regulatory roles such proteins may play in the different cellular processes in which they are involved.

Acknowledgments—The expert technical assistance of Howard Brickner and secretarial help of Ann Sellitti are greatly appreciated.

REFERENCES

1. Bishop, J. M. (1987) Science 235, 305–311
2. Eva, A., Tronick, S. R., Gol, R. A., Pierce, J. H., and Aaronson, S. A. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 4926
3. Dalla-Favera, R., Brezini, M., Erickson, J., Patterson, D., Gallo, R., and Croce, C. M. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 7824
4. de Kleijn, A., van Kessel, A. G., Groenewold, G., Bartram, C. R., Hagemeijer, A., Boetsma, D., Spurr, N. K., Heisterkamp, N., Groffen, J., and Stephenson, J. R. (1982) Nature 300, 765–767
5. Collins, S., and Groudine, M. (1982) Nature 298, 679–681
6. Hayward, W. S., Neel, B. G., and Astrin, S. M. (1981) Nature 290, 475–480
7. Hanash, S. M., Strahler, J. R., Kuick, R., Chu, E. H. Y., and Nichols, D. (1988) J. Biol. Chem. 263, 12813–12815
8. Eckerskorn, C., Strahler, J., Hanash, S., and Lotspeich, F. (1989) in Two-dimensional Electrophoresis (Endler, L., and Hanash, S., eds) VCH Publisher, Weinheim, in press
9. Gubler, U., and Hoffman, B. J. (1983) Gene (Amst.) 25, 263–269
10. Aruffo, A., and Seed, B. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 8750–8757
11. Seed, B. (1987) Nature 329, 840–842
12. Benton, W. D., and Davis, R. W. (1977) Science 196, 180–182
13. Wood, W. I., Gitschier, J., Lasky, L. A., and Lawn, R. M. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 1585–1588
14. Sanger, F., and Coulson, A. R. (1975) J. Mol. Biol. 94, 441–448
15. Lozzio, C. B., and Lozzio, B. B. (1975) Blood 45, 221–234
16. Gluzman, Y. (1981) Cell 23, 175–182
17. Favaloro, J., Treisman, R., and Kamen, R. (1980) Methods Enzymol. 65, 718–749
18. Maniatis, T., Fritch, E. F., and Sambrook, J. (1982) Molecular Cloning, pp. 200–201, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
19. Atweh, G. F., Anagnou, N. P., Shearin, J., Forget, B. G., and Kaufman, R. E. (1986) Nucleic Acids Res. 13, 777–790
20. Atweh, G. F., Wong, C., Reed, R., Antonarakis, S. E., Zhu, D., Ghosh, P. K., Maniatis, T., Forget, B. G., and Kazazian, H. H., Jr. (1987) Blood 70, 147–151
21. Blin, N., and Stafford, D. W. (1976) Nucleic Acids Res. 3, 2303–2308
22. Southern, E. M. (1975) J. Mol. Biol. 98, 503–517
23. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499–560
24. Hanash, S. M., Baier, L. J., McCurry, L., and Schwartz, S. A. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 807–811
25. Rosenfeld, M. G., Mermod, J. J., Amasa, S. G., Swanson, L. W., Sawchenki, P. E., Rivier, J., Vale, W. W., and Evans, R. M. (1980) Nature 294, 129–135
26. Darnell, J., Lodish, H., and Baltimore, D. (1986) in Molecular Cell Biology, pp. 485–487, Scientific American Books, New York
27. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) Science 240, 1769–1764
28. Miller, J., McClachlan, A. D., and Klug, A. (1986) EMBO J. 4, 1639–1644
29. Levine, M., and Hoey, T. (1988) Cell 55, 537–540
30. Feuerstein, N., and Cooper, H. L. (1983) J. Biol. Chem. 258, 10766–10773
31. Braverman, R., Bhattacharya, B., Feuerstein, N., and Cooper, H. L. (1986) J. Biol. Chem. 261, 14342–14348
32. Pasmanter, R., Danoff, A., Fleischer, N., and Schubart, U. K. (1986) Endocrinology 119, 1229–1238
33. Schubart, U. K., Alago, W., and Danoff, A. (1987) J. Biol. Chem. 262, 18071–18077
34. Schubart, U. K. (1989) J. Biol. Chem. 264, 12156–12160
35. Sobel, A., Bouterin, M. C., Beretta, L., Chneiweiss, H., Doze, V., and Peyro-Saint-Paul, H. (1989) J. Biol. Chem. 264, 3765–3772