Exon 2 of human cathepsin B derives from an Alu element

Isabelle M. Berquin[1], Mamoun Ahram, Bonnie F. Sloane*

Department of Pharmacology, Wayne State University, 540 E. Canfield Ave., Detroit, MI 48201, USA

Received 2 August 1997; revised version received 4 November 1997

Abstract Transcripts for the cysteine protease cathepsin B are alternatively spliced in the untranslated regions (UTRs). We show that a cathepsin B probe containing 5′-UTR sequences hybridized to an RNA of ~300 nt in addition to the typical 2.2 and 4.0 kbp mRNAs. Within this 5′-UTR, exon 2 was found to be homologous to Alu repetitive elements. Specifically, exon 2 was part of an Alu element interspersed with the cathepsin B gene. The ~300 nt band that hybridized to our cathepsin B probe likely corresponds to Alu transcripts, which are known to accumulate in human cells. Indeed, a similarly migrating band was detected with an authentic Alu probe. Thus, we suggest that primary transcripts for cathepsin B contain Alu sequences which are preserved as exon 2 in some fully spliced mRNAs.

Key words: Cathepsin B; Alu element; Transcript; RNA splicing

1. Introduction

The lysosomal cysteine protease cathepsin B has a general role in intracellular protein degradation, with specialized functions in certain cell types. Cathepsin B is frequently overexpressed and/or redistributed in malignant tumor cells; the various mechanisms responsible exploit every level of regulation of this enzyme's biosynthesis. One of the possible modes of regulation of human cathepsin B gene expression is alternative splicing, which occurs primarily in the 5′- and 3′-UTRs [1,2]. The 5′-UTR of human cathepsin B can be composed of at least four exons (exon 1, 2, 2a/2b and the first 15 bp of exon 3) which directly precede the translation initiation codon. In a previous study, we observed that exon 2 was present in most cDNAs isolated from cell lines, but absent in the dominant transcript variants freshly isolated from tissues [2]. The functional significance of alternative splicing of human cathepsin B transcripts is unknown.

While investigating the expression of cathepsin B in a normal human breast epithelial cell line and its transformed counterpart, we observed the presence of a small RNA species of approximately 300 nt that was detected with a specific cathepsin B cDNA probe. Here, we describe that the sequence responsible for this signal is exon 2, which cross-hybridizes to small transcripts derived from Alu repetitive elements. We show that a partial Alu element is present in the cathepsin B gene between exons 1 and 3. Our sequence analysis suggests that cathepsin B mRNAs containing exon 2 are formed by processing of primary transcripts at cryptic splice sites within this Alu element.

2. Materials and methods

2.1. cDNA and genomic constructs

Full-length cathepsin B cDNAs pLC34 and pLC43 were cloned from a gastric adenocarcinoma library as described [3]. pLC343, a variant cDNA clone that we used as a probe, was generated by ligating a 3.7 kbp *Eco*47III/*Nsi*I fragment from pLC43 and a 1.2 kbp *Eco*47III/*Nsi*I fragment from pLC34. Genomic cathepsin B clone pCBG1 was isolated from a human lymphocyte cosmid library as described previously [2]. The Alu cDNA probe Blur-2 [4] was a gift from Dr. Tom Mikkelsen, Henry Ford Hospital, Detroit, MI.

2.2. Cells

MCF-10A and MCF-10AneoT cells were maintained in Dulbecco's minimum essential medium/F12 medium supplemented with 5% horse serum, 20 ng/ml epidermal growth factor, 10 μg/ml insulin and 0.5 μg/ml hydrocortisone. Cells were grown in subconfluent monolayers under 5% CO_2.

2.3. Northern blotting

Total RNA was prepared according to the method of Chomczynski and Sacchi [5] or using Trizol reagent (Gibco-BRL) as described by the manufacturer. 20 μg RNA per lane was size-fractionated in 0.8% agarose gels containing 6.6% formaldehyde and transferred onto Nytran membranes (Schleicher and Schuell) by capillary blotting using Schleicher and Schuell's Turboblotter according to the manufacturer's instructions. cDNA probes were radiolabeled with [α-^{32}P]dCTP (3000 Ci/mmol) using Prime-It II random primer labeling kit (Stratagene). Hybridization was performed in 50% formamide/5×SSC/2×Denhardt's reagent/1% sodium dodecyl sulfate/250 mg/ml denatured salmon sperm DNA at 42°C for 16–20 h. Membranes were washed twice for 5 min at room temperature with 6×SSC/0.1% *N*-laurylsarcosine, twice for 5 min at 37°C with 1×SSC/0.5% sodium dodecyl sulfate, once for 1 h at 65°C with 0.1×SSC/1% sodium dodecyl sulfate, and then autoradiographed.

2.4. Sequence analysis

GenBank was searched for homology to the cathepsin B 5′-UTR using the BLAST algorithm. Optimal sequence alignment was determined using MacVector (Kodak, New Haven, CT).

3. Results and discussion

The normal human breast epithelial cell line MCF-10A and its ras-transfected variant MCF-10AneoT exhibit differences in the levels and subcellular distribution of the cysteine protease cathepsin B [6]. In order to determine whether cathepsin B is already altered at the level of the mRNA, we performed northern blot analyses using as a probe a cathepsin B cDNA clone (pLC343, Fig. 1A) containing the complete coding region plus additional 5′- and 3′-UTR sequences. Although MCF-10AneoT expresses higher levels of the protease than the parental line MCF-10A, their cathepsin B mRNA levels were similar (Fig. 1B). Interestingly, we detected a small RNA fragment of approximately 300 nt in addition to the 2.2 and

*Corresponding author. Fax: (1) (313) 577-6739.
E-mail: bsloane@med.wayne.edu

[1]Present address: Department of Radiation Oncology, University of Michigan, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA.

Abbreviations: SSC, sodium chloride/sodium citrate buffer; UTR, untranslated region

PII S0014-5793(97)01445-2

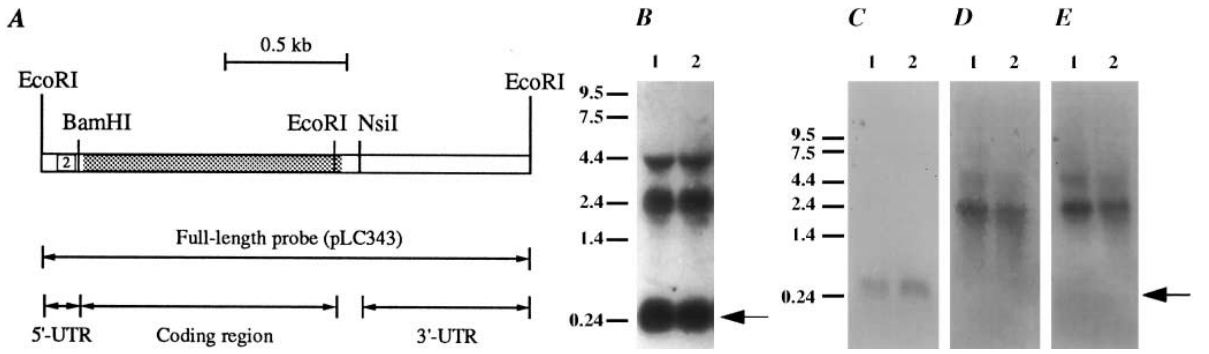

Fig. 1. A cDNA probe specific for the 5′-UTR of human cathepsin B detects a ~300 nt RNA species in MCF-10A (lane 1) and MCF-10AneoT (lane 2) cells. A: Partial restriction map of cathepsin B cDNA clone pLC343 with location of exon 2. The coding region is shaded. Fragments used as probes are indicated by arrows. B–E: Northern blot hybridization of total RNA with the following probes: B, full-length pLC343 probe; C, 5′-UTR probe; D, coding region probe; E, 3′-UTR probe. The position of molecular size markers is indicated to the left. Arrows indicate the ~300 nt RNA. Note that upon prolonged exposure the 2.2 and 4.0 kbp cathepsin B transcripts were also detected with the 5′-UTR probe (not shown).

4.0 kbp cathepsin B transcript species (Fig. 1B). The same band was consistently observed in other cells, including MCF-7 breast carcinoma, U87 glioblastoma, HD6-4V colon carcinoma and HT-1080 fibroblastoma cells (data not shown). Both the 2.2 and 4.0 kbp cathepsin B transcripts encode a 339 amino acid protein precursor that is later processed to a mature form of the enzyme [7,1]. Obviously, a 300 nt transcript would not be large enough to encode the entire preprocathepsin B. However, the strength of the signal observed under high stringency hybridization conditions implied that the small RNA fragment was homologous to part of the cathepsin B transcripts. To determine whether this RNA was related to the protein-encoding sequence or to untranslated regions of cathepsin B, we designed cDNA probes specific for the 5′-UTR, coding region and 3′-UTR (Fig. 1A). Only the 5′-UTR probe detected the ~300 nt RNA species (Fig. 1C,D,E). Therefore, this RNA cannot encode any polypeptide related to cathepsin B.

The 5′-UTR of human cathepsin B mRNAs is variable, due to the existence of several possible splicing variants [1,2]. The 5′-UTR-specific cDNA probe used in this study included the last 59 bp of exon 1, all of exon 2 (88 bp) and the first 15 bp of exon 3. We searched the GenBank database for sequences with homology to this 5′-UTR region, and found that exon 2 shared 79% nucleotide identity with an Alu repetitive element (Fig. 2A). Extending the homology search to a genomic DNA segment containing exon 2 and flanking sequences, we identified a partial Alu element spanning 155 bp of the cathepsin B gene from the end of intron 1 to the first few base pairs of intron 2 (Fig. 2B). This partial element roughly corresponds to a right Alu monomer [8] and is located in the reverse orientation relative to the cathepsin B gene. Near-consensus splice donor and acceptor sites are present at the cathepsin B intron/exon junctions within the Alu element (Fig. 2B). This suggests that during transcription of the cathepsin B gene, the Alu element is incorporated in primary transcripts, and that a portion of this element is retained as exon 2 in some fully spliced mRNAs. In other mRNAs, alternative splicing schemes eliminate exon 2 and thus Alu sequences as well.

Alu elements contain an internal RNA polymerase III promoter and can be transcribed in vitro [9–11]. A few active Alu elements are thought to spread within the genome through retrotransposition [9,10]. Alu transcript species of ~300 nt and ~120 nt have been detected in human cells [12,13].

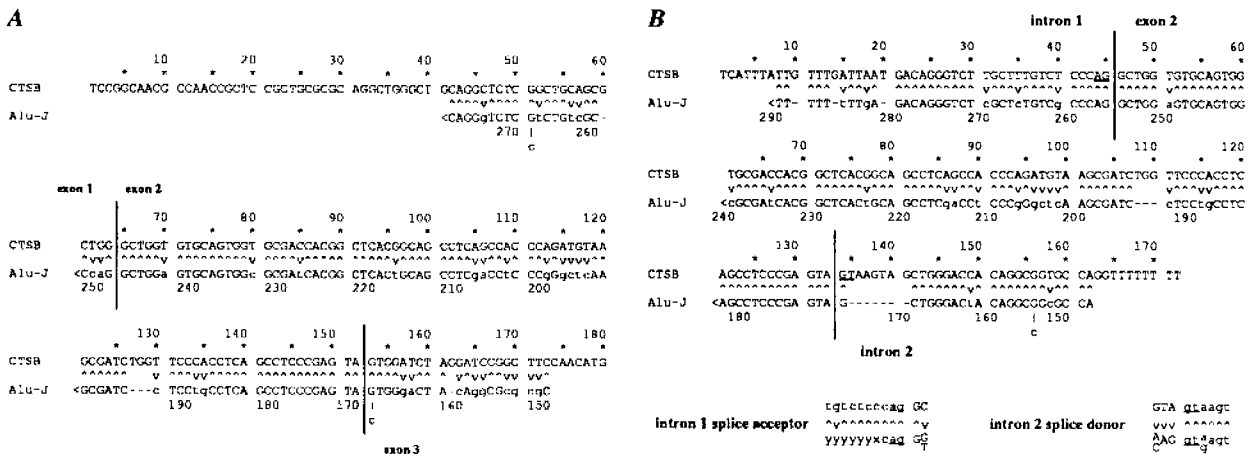

Fig. 2. Exon 2 of human cathepsin B derives from an Alu element. Sequence alignment of cathepsin B cDNA (A) and genomic DNA (B) with an Alu consensus site. Highest homology was found with consensus for the Alu-J subfamily. The symbol ^ indicates a sequence match, v a mismatch, - a deletion, | an insertion. The genomic cathepsin B intron/exon boundaries (determined in [1]) are marked by a vertical line. Comparison of intron/exon boundaries with consensus splice donor and acceptor sites is also indicated.

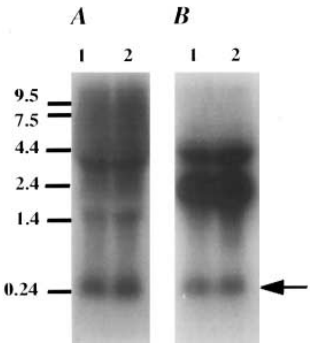

Fig. 3. The ~300 nt RNA detected in MCF-10A (lane 1) and MCF-10AneoT (lane 2) cells with the cathepsin B probe co-migrates with Alu transcripts. A: Northern blot hybridization of total RNA with the Blur-2 Alu probe. B: Northern blot hybridization with the full-length cathepsin B probe. The position of molecular size markers is indicated to the left. Arrow indicates the ~300 nt RNA.

Although their function is not known, Alu transcripts have been suggested to play a role in translation [14]. The ~300 nt Alu RNA is polyadenylated and represents a dimeric Alu repeat, whereas the poly(A)$^-$ ~120 nt RNA corresponds to a left Alu monomer [12,13]. This is consistent with our 5′-UTR probe detecting the ~300 nt RNA, but not the ~120 nt RNA. To confirm that the 300 nt RNA fragment is Alu-related, we probed an RNA blot from MCF-10A and MCF-10AneoT cells successively with an Alu probe and with the full-length cathepsin B probe. As expected, the Alu probe detected an RNA fragment that co-migrated with the band detected with the cathepsin B cDNA probe (Fig. 3).

In the human genome, there are an estimated 900 000 copies of Alu repetitive DNA, accounting for approximately 5% of genomic DNA [15]. Interspersed Alu sequences are also present in numerous transcripts for protein-encoding genes. In a survey of a subset of the GenBank database, 5% of fully spliced human cDNAs were found to contain Alu sequences, primarily in the 3′- and 5′-UTRs [16]. Alu elements in the coding region are much less frequent, and have been suggested to play a role in the recent evolution of some proteins [17], although Alu regions can be mistakenly interpreted as belonging to an open reading frame [18]. The presence of Alu elements in the 3′- and 5′-UTR of human transcripts may influence their regulation, since UTRs are involved in the control of RNA stability and translation. Structural or sequence elements within Alu repeats may interact with RNA-binding proteins or modify the topology of other regulatory regions in mRNAs. Interestingly, computer modeling of possible RNA secondary structure elements indicated that cathepsin B mRNAs with exon 2 contained extended regions of potential base-pairing (B.F. Sloane and B.A. Frosch, data not shown). Interspersed Alu elements may also increase the diversity of transcripts by allowing for additional alternative splicing events. Interestingly, the Alu-derived exon 2 is skipped in the majority of cathepsin B transcripts expressed in tissues, whereas it is incorporated in transcripts from cells grown in vitro [2]. The significance of this alternative splicing scheme for the regulation of expression of cathepsin B transcripts is currently under investigation.

Acknowledgements: This study was supported by U.S. Public Health Service Grant CA36481. I.M.B. was supported in part by an institutional training grant (T32 CA09531) from the National Cancer Institute.

References

[1] Gong, Q., Chan, S.J., Bajkowski, A.S., Steiner, D.F. and Frankfater, A. (1993) DNA Cell Biol. 12, 299–309.
[2] Berquin, I.M., Cao, L., Fong, D. and Sloane, B.F. (1995) Gene 159, 143–149.
[3] Cao, L., Taggart, R.T., Berquin, I.M., Moin, K., Fong, D. and Sloane, B.F. (1994) Gene 139, 163–169.
[4] Rubin, C.M., Houck, C.M., Deininger, P.L., Friedmann, T. and Schmid, C.W. (1980) Nature 284, 372–374.
[5] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 161, 156–159.
[6] Sloane, B.F., Moin, K., Sameni, M., Tait, L.R., Rozhin, J. and Ziegler, G. (1994) J. Cell Sci. 107, 373–384.
[7] Chan, S.J., San Segundo, B., McCormick, M.B. and Steiner, D.F. (1986) Proc. Natl. Acad. Sci. USA 83, 7721–7725.
[8] Daniels, G.R. and Deininger, P.L. (1983) Nucleic Acids Res. 11, 7595–7610.
[9] Deininger, P.L., Batzer, M.A., Hutchison III, C.A. and Edgell, M.H. (1992) Trends Genet. 8, 307–311.
[10] Schmid, C. and Maraia, R. (1992) Curr. Opin. Genet. Dev. 2, 874–882.
[11] Liu, W.M., Leeflang, E.P. and Schmid, C.W. (1992) Biochim. Biophys. Acta 1132, 306–308.
[12] Matera, A.G., Hellmann, U. and Schmid, C.W. (1990) Mol. Cell. Biol. 10, 5424–5432.
[13] Chang, D.Y. and Maraia, R.J. (1993) J. Biol. Chem. 268, 6423–6428.
[14] Liu, W.M., Maraia, R.J., Rubin, C.M. and Schmid, C.W. (1994) Nucleic Acids Res. 22, 1087–1095.
[15] Britten, R.J., Baron, W.F., Stout, D.B. and Davidson, E.H. (1988) Proc. Natl. Acad. Sci. USA 85, 4770–4774.
[16] Yulug, I.G., Yulug, A. and Fisher, E.M.C. (1995) Genomics 27, 544–548.
[17] Makalowski, W., Mitchell, G.A. and Labuda, D. (1994) Trends Genet. 10, 188–193.
[18] Claverie, J.M. (1992) Genomics 12, 838–841.