Volume 319, number 1,2, 159–162 FEBS 12216 March 1993
© 1993 Federation of European Biochemical Societies 00145793/93/$6.00

Identification of CpG island at the 5′ end of murine leukemia inhibitory factor gene

Petr Kašpar, Michal Dvořák and Petr Bartůněk

Institute of Molecular Genetics, Czech Academy of Sciences, 166 37 Prague, Czech Republic

Received 26 January 1993

We identified a CpG island at the 5′ end of murine leukemia inhibitory factor gene (LIF). The CpG island is 0.6 kb long and covers most of the first exon and first intron. The region is non-methylated, its G+C content is 65% and relative frequency of CpG dinucleotide is 0.7. The block of 150 nucleotides, which is 72% conserved between murine, human, ovine and porcine genes, is a part of the CpG island. Two DNA fragments from this CpG island interact with nuclear proteins from NIH 3T3 cells. One fragment partially covers the block of conserved nucleotides. Human, ovine and porcine LIF genes also contain G+C- and CpG-rich sequences in the corresponding region.

Murine leukemia inhibitory factor; Methylation; CpG island; Gel retardation assay

1. INTRODUCTION

Leukemia inhibitory factor (LIF) is a secreted glycoprotein produced by a variety of cell types. LIF is known to have distinct in vitro activities. It induces macrophage differentiation in the M1 murine myeloid leukemia cells and suppresses their proliferation [1], it supports the proliferation of murine interleukin-3-dependent leukemic DA-1a cells [2] and inhibits differentiation of embryonic stem cells [3]. Furthermore, LIF promotes bone resorption in vitro [4]. It plays a role in stimulating the synthesis of several acute phase plasma proteins in hepatocytes [5], in inhibiting the activity of lipoprotein lipase in adipocytes [6] and in affecting the transmitter phenotype in sympathetic neurons [7].

Little is known about in vivo functions of LIF. Detectable levels of LIF expression were found in the extraembryonic part of the mouse embryo [8], in postnatal rat brain and footpad [9] and in the female reproductive tract of adult mouse [10].

Cross-species comparison of the sequence of murine, human, ovine and porcine LIF genes revealed a block of 150 nucleotides in the first intron, which is highly conserved between all four species [11].

We analyzed the base composition of the murine, human, ovine and porcine LIF genes [12] and found G+C-rich region which covers most of the first exon and the first intron of the LIF gene in all four species. The highly conserved block of nucleotides is a part of this G+C-rich region. The G+C-rich region is also rich in CpG dinucleotide. A high incidence of CpG dinucleotide concomitant with the absence of methylation are recognized features of CpG islands found at the 5′ end of many genes [13]. In order to obtain evidence, that the G+C-rich region is a CpG island, we analyzed methylation at the *Hpa*II restriction sites along the LIF gene in several murine tissues. We also examined binding of nuclear proteins to sites within this region.

Correspondence address: P. Kašpar, Institute of Molecular Genetics, Czech Academy of Sciences, Flemingovo nam. 2, 16637 Prague 6, Czech Republic. Fax: (42) (2) 311 9017.

2. MATERIALS AND METHODS

2.1. *Isolation of LIF gene*

Genomic clones were isolated from a λD-*Sau*3A genomic library from a CD 1 mouse (kindly provided by Ross Kinloch) using a murine LIF cDNA probe [14].

2.2. *Southern blot analysis*

Nuclear DNA (5 μg) was digested with an excess of *Hpa*II or *Msp*I enzymes and electrophoresed either in 3% Nu Sieve agarose gel, blotted onto nylon membrane (Gene Screen) and hybridized to the BP 1.1 probe spanning nucleotides 815–1,962 of murine LIF [12], or in 1% agarose gel, blotted and hybridized to the SH 4,9 probe spanning nucleotides 2,664–7,659. After washing at high stringency (0.1× SSC, 0.1× SDS, 65°C) membranes were autoradiographed.

2.3. *Multiple bandshift assay*

For multiple bandshift assay [15] the binding reaction contained from 5 to 50 μg of nuclear proteins, 1 ng of end-labelled *Hpa*II digest of the BP1.1 and 2–5 ng of poly(dI-dC). After 30 min on ice samples were loaded onto a 4% non-denaturing polyacrylamide gel.

2.4. *Gel retardation assay*

Gel retardation assay was carried out essentially as described for multiple bandshift assay except that the binding reaction contained 1 μg of poly(dI-dC), 0.25–0.5 ng of the end-labelled fragment and 5–10 μg of nuclear proteins.

3. RESULTS AND DISCUSSION

Our analysis of the G+C profile of murine LIF gene

revealed a region with a high G+C content (68% G+C) spanning nucleotides 930–1,530 [12] with the relative frequency of CpG dinucleotide 0.7 (Fig. 1). Upstream from the nucleotide 930 the G+C content is 52% and relative CpG frequency 0.25. The same low relative frequency was found downstream from the nucleotide 1,530 where the G+C content is 53%. Human, ovine and porcine LIF genes also contain G+C- and CpG-rich sequences in the corresponding region as murine LIF gene (data not shown).

Analysis of methylation at the *Hpa*II restriction sites along the murine LIF gene from various tissues is shown in Fig. 2. Schematic representation of methylation status is shown in Fig. 3. Within the G+C-rich region the DNA of neither tissue was methylated. The region downstream from the first intron is methylated (brain, thymus) or hypomethylated (embryonic brain, spleen, kidney, testis). Accordingly, we propose that murine LIF gene contains a CpG island at its 5′ end. The highly conserved block of nucleotides spanning nucleotides 1,313–1,458 is a part of this CpG island.

Next, we adopted two assays for studying interactions between potential regulatory proteins and the corresponding target sites in the CpG island of murine LIF gene. First, the multiple bandshift assay was performed [15]. The identification of DNA fragments interacting with proteins is based on decrease of the intensity of some bands. These bands, which represent DNA fragments of the restriction digest, fade or even disappear when the fragments interact with proteins. Three fragments, namely HP121 (spanning nucleotides 1,841–1,962), HH137 (1,221–1,358), HH174 (820–994) of *Hpa*II digest of the BP1.1 interact with nuclear extract from NIH 3T3 cells (Fig. 4). Gel retardation assay performed with these isolated DNA fragments confirmed results of the multiple bandshift assay (not shown); interaction of the fragment HH174 with nuclear extract was the weakest. Interestingly, fragment HH137 partially covers the highly conserved block of nucleotides.

We identified a G+C-rich region at the 5′ end of murine LIF gene. This element has all the attributes of a CpG island [16]: (a) high G+C content = 68%; (b) high relative frequency of CpG dinucleotide = 0.7; (c) length = 600 bp; and (d) non-methylated status. The corresponding regions in human, ovine and porcine LIF genes are likely to be CpG islands as well, because they meet demands for length, and high G+C and CpG contents. Unique base composition of CpG islands and pronounced lack of methylation make these structures distinctive points in the genome and predetermine them to interact with anticipated 'CG-binding proteins'. In all cases CpG islands were found to be structures important for transcriptional regulation of particular genes.

The expression of the LIF gene appears to be tightly regulated, as the levels of the constitutive expression in most tissues are extremely low. The only exception are the uterine endometrial glands of adult mice [10]. There are no conclusive data available about *cis*- and *trans*-regulatory elements of the LIF gene. The evolutionary conserved sequences upstream from transcription start sites – TATA-like elements, as well as the 150-nucleotide box within the first intron are candidates for this regulatory function. We have identified a conserved structure including sequences of the first exon and 5′ part of the first intron. This element is G+C- and CpG-rich.

Our methylation studies in murine tissues show that this element is CpG island. Our findings that the CpG island of murine LIF gene contains a block of 150 nucleotides which is highly conserved between species [11], and specifically interacts with nuclear proteins from NIH 3T3 fibroblasts, a potential source of LIF [17],

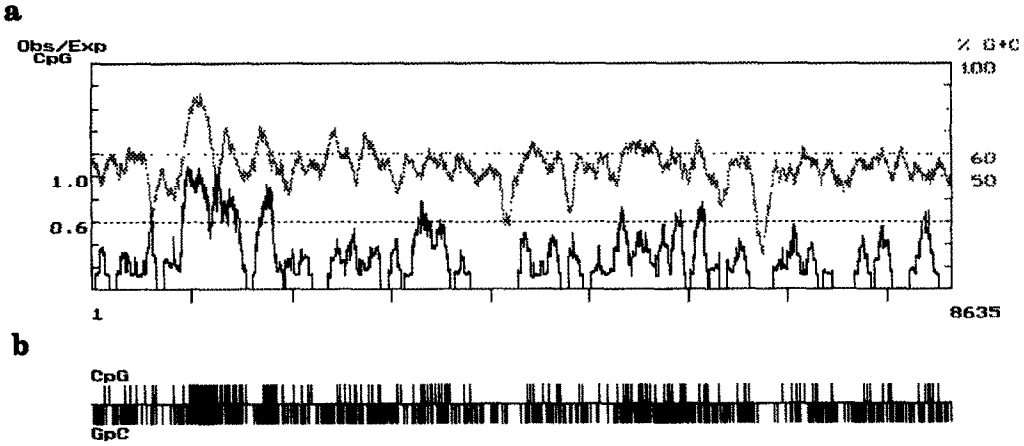

Fig. 1. Distribution of nucleotides G and C and of dinucleotides CpG and GpC along murine LIF gene (a). Percentage (%) G+C and distribution of CpG was calculated for 100 bp windows moving along sequence at 1 bp intervals. Distribution of CpG dinucleotide was measured as the ratio of observed frequency of CpG dinucleotide to frequency expected. (b) Distribution of CpG and GpC dinucleotides. Each vertical line represents CpG or GpC doublet.

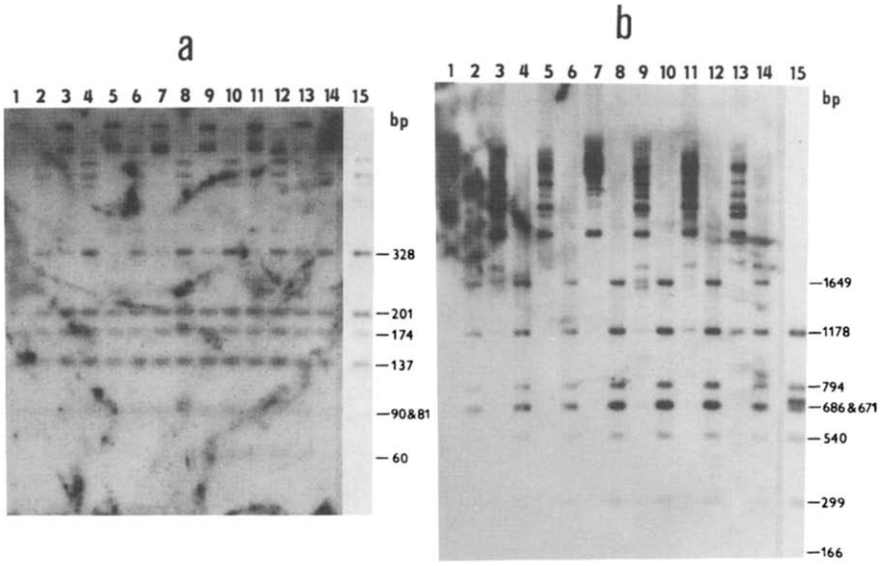

Fig. 2. Southern blot analysis of the methylation pattern in the first exon and first intron (a), and in the region downstream from the first intron (b) of the murine LIF gene. DNA from NIH 3T3 cells (lanes 1,2), or genomic DNA from brain of 16-day-old embryos (lanes 3,4), adult brain (lanes 5,6), thymus (lanes 7,8), spleen (lanes 9,10), kidney (lanes 11,12) and testis (lanes 13,14) was digested with *Hpa*II (lanes 1,3,5,7,9,11,13) or *Msp*I (lanes 2,4,6,8,10,12,14) blotted and hybridized to the ^{32}P-labelled BP1.1 (a) or to the SH4.9 (b) probes. Lanes 15 represent BP1.1 fragment or SH4.9 fragment digested with *Hpa*II.

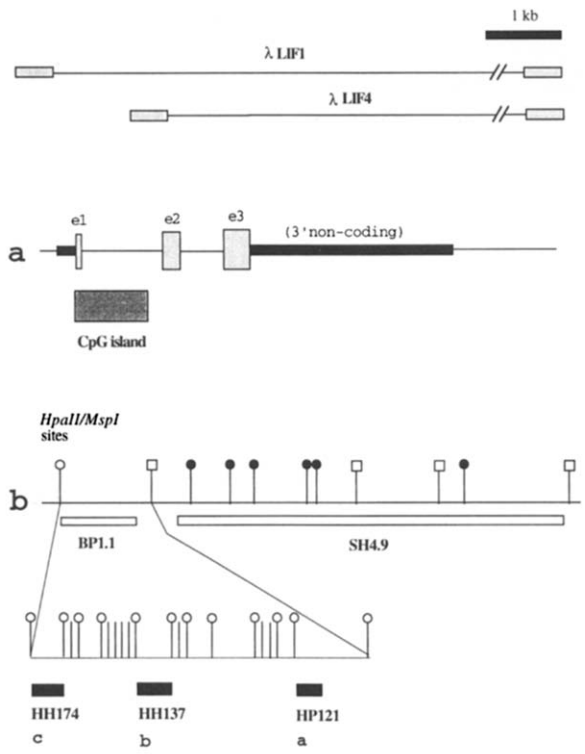

Fig. 3. Structure and methylation pattern of the LIF gene. (a) Exon–intron structure of the murine LIF gene and CpG island. Exons are depicted as stippled boxes and the position of introns as thin lines, thick lines are non-coding regions presented in mRNA. (b) *Hpa*II/*Msp*I restriction map of the LIF gene. Opened boxes = probes used for the analysis of the methylation; closed boxes = probes used and for the study of DNA–protein interaction. Vertical bars represent *Hpa*II/*Msp*I sites with indication of their methylation status. Opened circles = non-methylated sites; closed circles = methylated sites; squares = tissue-specifically hypomethylated sites.

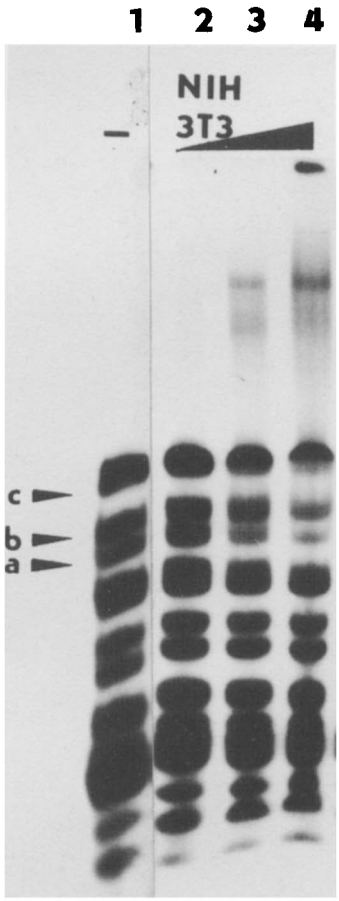

Fig. 4. DNA–protein interaction in the CpG island of the murine LIF gene. (a) Multiple bandshift assay with *Hpa*II-digested fragment BP1.1. Fading bands represent fragments HP121 (a), HH137 (b) and HH174 (c). Lanes 1, 2, 3, 4 = 0, 1, 3, 10 μg of NIH 3T3 nuclear proteins added, respectively.

suggest the role of this region in transcriptional regulation of the LIF gene.

Acknowledgements: The authors are thankful to Dr. J. Pačes for help with analyzing DNA sequences.

REFERENCES

[1] Tomida, M., Yamamoto-Yamaguchi, Y. and Hozumi, M. (1984) J. Biol. Chem. 259, 10978–10982.
[2] Moreau, J.F., Donaldson, D.D., Bennett, F., Witeck, K., Giannoti, J., Clark, S.C. and Wong, G.G. (1988) Nature 366, 690–692.
[3] Williams, R.L., Hilton, D.J., Pease, S., Wilson, T.A., Stewart, C., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A. and Gough, N.M. (1988) Nature 336, 684–687.
[4] Abe, E., Tanaka, H., Ishimi, Y., Miyawa, C., Hayashi, T., Nagasawa, H., Tomida, M., Yamaguchi, Y., Hozumi, M. and Suda, T. (1986) Proc. Natl. Acad. Sci. USA 83, 5958–5962.
[5] Baumann, H. and Wong, G. (1989) J. Immunol. 143, 1163–1167.
[6] Mori, M., Yamaguchi, K. and Abe, K. (1989) Biochem. Biophys. Res. Commun. 160, 1085–1092.
[7] Yamamori, T., Fukada, K., Aebersold, R., Korsching, S., Fann, M.J. and Paterson, P. (1989) Science 246, 1412–1416.
[8] Conquet, F. and Brulet, P. (1990) Mol. Cell. Biol. 10, 3801–3805.
[9] Yamamori, T. (1991) Proc. Natl. Acad. Sci. USA 88, 7298–7302.
[10] Bhatt, H., Brunet, L.J. and Stewart, C.L. (1991) Proc. Natl. Acad. Sci. USA 88, 11408–11412.
[11] Wilson, T.A., Metcalf, D. and Gough, N.M. (1992) Eur. J. Biochem. 204, 21–30.
[12] Stahl, J., Gearing, P.D., Wilson, T.A., Brown, M.A., King, J.A. and Gough, M. (1990) J. Biol. Chem. 265, 8833–8841.
[13] Bird, A.P. (1986) Nature 321, 209–213.
[14] Stewart, C.L., Kašpar, P., Brunet, L.J., Bhatt, H., Gadi, I., Kotgen, F. and Abbondanzo, S.J. (1992) Nature 359, 76–79.
[15] Kozmik, Z. and Pačes, V. (1990) Gene 90, 287–291.
[16] Gardiner-Garden, M. and Frommer, M. (1987) J. Mol. Biol. 196, 261–282.
[17] Rathjen, P.D., Toth, S., Willis, A., Heath, J.K. and Smith, A.G. (1990) Cell 62, 1105–1114.