

Received 10 March 2000

Accepted 6 December 2000

THE DETECTION AND PURIFICATION OF A PROTEIN THAT RECOGNIZES A LINE 1 DNA REPETITIVE SEQUENCE

DMITRI LUKYANOV¹, JACEK ROGOLIŃSKI², OLGA PODGORNAYA¹
and JOANNA RZESZOWSKA-WOLNY²

¹ Institute of Cytology, Russian Academy of Sciences, Tikhoretsky pr. 4;
194064 St.Petersburg, Russia, ²Institute of Oncology, Department of
Experimental and Clinical Radiobiology, 44-100 Gliwice, Poland

Abstract: A protein which binds specifically to MspI8 (a 454 bp long repeated sequence highly homologous to the 5' untranslated region (5'UTR) of the LINE 1 sequence) was found and identified in nuclear extracts of rat liver cells. This protein was detected using the electrophoretic mobility shift assay (EMSA) and was purified by Q-Sepharose and DNA affinity chromatography. Its molecular mass was estimated by SDS electrophoresis as 29 kDa. The possibility that this protein (p29) is the rat analogue of human L1BPB-A, specific for the human LINE sequence LPE1, is discussed.

Key Words: Repetitive DNA Sequences, Untranslated Region (UTR), LINE 1 Binding Proteins, EMSA.

INTRODUCTION

Eukaryotic genomes contain differently organized repetitive sequences classified as tandem repeats (satellites) or dispersed repeats. Dispersed repetitive sequences are further divided into short and long interspersed elements (SINEs and LINEs respectively). LINEs are typically between 5 and 7 kbp long and are present in 10^4 to 10^5 copies per genome in all mammals [1,2]. Mammalian LINEs belong to the L1 family which makes up about 10% of the genome; all its members have a similar structure with at least two open reading frames (ORFs) and 5' and 3' untranslated regions (5'UTR and 3'UTR) [3-5]. The ORFs in human and mouse L1 sequences code for proteins which take part in the transposition process [3,4,6-8], allowing them to move within the genome

by a reverse transcription mechanism with an RNA intermediate; ORF1 encodes a highly basic protein able to form multimers and to interact specifically with single stranded RNA transcribed from ORF2, which encodes reverse transcriptase [6-8]. Full-length or fragmented LINEs (generally deleted at their 5'ends) are present in all human chromosomes and within many genes [9,10]; fragments are found in both the exons and introns of genes such as retinoblastoma, interleukin-2, PAX6 or factor VIII [10-12]. L1 sequences are transcribed in a few normal human tissues and in a variety of tumors including breast cancer [10, 13-15] but the mechanisms which regulate their expression are poorly understood. Therefore, the identification of proteins which take part in this process is important, and this paper is on proteins identified in rat liver nuclear extract. They specifically recognize a rat genomic sequence (MspI8) which is highly (68.9%) homologous to the 5' untranslated regions of some rat LINE 1 family members

MATERIALS AND METHODS

DNA

MspI8, part of a rat LINE repetitive sequence, was isolated and cloned into pUC19 as previously described [16]. It was excised from the plasmid with *HindIII* and *EcoRI*, separated from vector sequences by electrophoresis in a 0.5% agarose gel, and end-labeled with [α^{32} P]-dCTP using Klenow fragment polymerase [18].

Preparation of nuclei and nuclear extracts

Nuclei from three rat livers or pairs of testis were isolated as described earlier [19]; all the solutions were supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF; Sigma). Nuclei were extracted with a buffer containing 10 mM Tris-HCl, pH 8.0, 0.1mM MgCl₂, 0.2 mM PMSF, 0.5 mM dithiothreitol (DTT), 5% glycerol and 0.5 M NaCl for 1 h at 4°C, the suspension was centrifuged for 10 min. at 5000 x g, and the supernatant extract was saved. The protein concentration in the nuclear extracts was 2 mg/ml (liver) and 0.5 mg/ml (testis).

Electrophoretic mobility shift assay (EMSA)

The specific binding of proteins to DNA was detected by the EMSA method [20]. *E.coli* genomic DNA sonicated to 0.5-2 kb in length and the cloned MspI8 insert fragment were used as nonspecific and specific competitors, respectively. The binding reactions (20-70 μ l) usually contained a 3 ng fragment of labeled MspI8, 2-50 μ l (0.1-10 μ g protein) of nuclear extract or chromatographic fractions, and 0.1-2 μ g of competitor DNA in a retardation buffer (A) containing 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM DTT and 5% glycerol. Column eluates were diluted so that the final salt concentration did not

exceed 250 mM. The binding mixture was incubated for 30 min. at room temperature and then electrophoresis was performed in a 4% polyacrylamide gel in 0.5 x TAE buffer (1 x TAE is 40 mM Tris-acetate, pH 8.0, 2 mM EDTA [18]). The gels were dried and exposed to X-ray film.

Ion exchange chromatography on Q-Sepharose and DEAE-Sepharose

Nuclear extracts were diluted 10-fold with buffer B (15 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, 0.5 mM DTT, 0.1 mM PMSF, 5% glycerol, 50 mM NaCl) and loaded onto a column of Q-Sepharose (1 cm x 20 cm) equilibrated in the same buffer. Proteins were eluted by a linear gradient of 0.05-1.5 M NaCl (40 ml) in buffer B and 1 ml fractions were collected and stored at -20^o C. The same conditions were used for DEAE-Sepharose columns (1 cm x 3 cm) except that proteins were eluted with a step gradient of buffer B containing 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.9, or 1 M NaCl (1 ml of each). All these procedures were carried out at room temperature.

Affinity purification

Proteins binding nonspecifically to DNA were first selected on an affinity matrix of salmon sperm DNA coupled to cellulose (Sigma) using 1-(3-dimethyl-aminopropyl)-3-ethylcarbodiimide hydrochloride (Sigma); 1 mg of sonicated salmon DNA (0.5-2 kbp) was attached to 1 ml of resin [21]. Q-Sepharose eluate fractions (200 µl) containing MspI8-binding activity were incubated with this matrix in 1 ml at room temperature for 30 min. After four 5-min washes of the matrix in 10 volumes of gel retardation buffer, bound proteins were eluted stepwise with 1 ml of the same buffer containing 0.2, 0.3, 0.4, 0.5, 0.7, 0.9 and 1 M NaCl and analysed by EMSA. In the second step, the fractions with maximal MspI8-binding activity were pooled and used for specific affinity selection on a matrix of linearized pMspI8 DNA (200 µg) covalently attached to 200 µl of cellulose. Pooled fractions diluted with buffer A (100 µl) were incubated with MspI8-cellulose and a varying amount of an excess of non-specific competitor DNA (*E.coli*) in 400 µl for 30 min. After four 5 min. washes with buffer A, elution was performed in one step using 200 µl of buffer A with 0.8 M NaCl.

SDS-polyacrylamide gel electrophoresis of proteins

Eluted proteins were analysed by SDS-PAGE in 12% polyacrylamide [22] and silver staining was performed by the standard method [23]. To determine molecular weights Prestained SDS Molecular Weight Markers were used (Sigma SDS-7B).

RESULTS

MspI8-binding proteins in nuclear extracts from rat liver and rat testis cells

Following the previous finding that the nuclear matrix proteins of rat testis show a high affinity for MspI8 [16, 17], the question was whether nuclear extracts contain proteins which specifically recognize and bind MspI8. The electrophoretic mobility shift assay (EMSA) was used to detect the complex formation of proteins with labeled MspI8 DNA incubated with nuclear extracts of liver or testis in the presence of increasing amounts of *E.coli* DNA as a nonspecific competitor. Specific bands formed by complexes of proteins with MspI8 were revealed in the presence of increasing amounts of competitor *E.coli* DNA (Fig. 1) using extracts from rat testis and liver nuclei; the latter was partially purified by DEAE-Sepharose column chromatography, eluting with a stepwise salt gradient, and the fraction containing maximal MspI8-binding activity (eluted with 0.3 M NaCl) was used. Proteins in crude nuclear extracts from rat testis also formed retarded bands with similar mobilities, and in this case two protein/DNA complexes were formed (Fig. 1B, arrows)

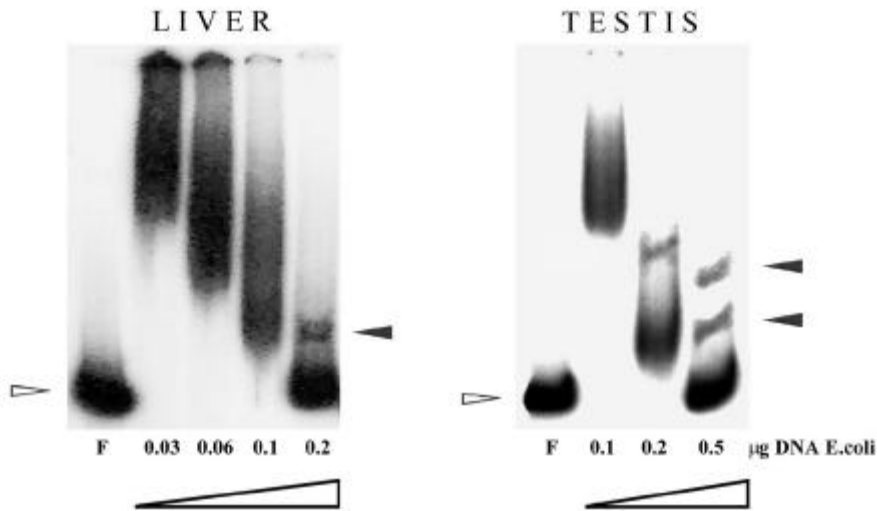


Fig. 1. Electrophoretic mobility shift assay for complexes formed by proteins in nuclear extracts of rat liver and testis with the LINE 1 sequence MspI8.

Labeled MspI8 (3 ng) was incubated with proteins (2 μ g) from a rat liver nuclear extract which had been eluted from a DEAE-Sepharose column in 0.3M NaCl (left) or with a crude rat testis nuclear extract (1.5 μ g protein) (right) and increasing amounts of nonspecific competitor DNA (*E.coli*). In all the figures, lane F contained only MspI8, the open arrows denote MspI8, and the solid arrows protein-DNA complexes.

To test the specificity of the complexes formed, an examination of *E. coli* DNA and non-labeled MspI8 DNA as competitors was performed. At low concentrations of a non-specific competitor, high molecular weight complexes were formed, visible as a smear during electrophoresis (Fig. 2). A stepwise increase of the amount of the competitor revealed a specific complex, clearly seen when 600 ng of *E. coli* DNA was used as a competitor and stable up to an approximately 600-fold molar excess (1800 ng) of nonspecific competitor (Fig. 2).

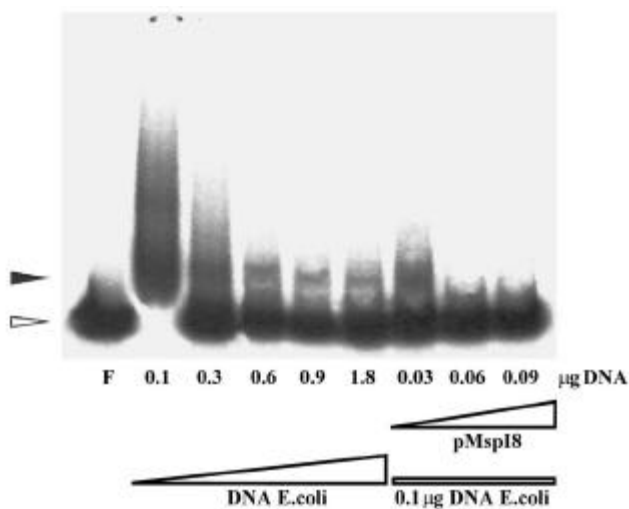


Fig. 2. An electrophoretic mobility shift assay of complexes with MspI8 formed by proteins from rat liver nuclear extract, eluted from a DEAE-Sephacel column in 0.3M NaCl, in the presence of different amounts of nonspecific (*E. coli*) or specific (unlabeled MspI8) competitor DNA. Each assay contained 3 ng of labeled MspI8 and 2 µg of eluted proteins.

The same result was obtained with 100 ng of *E. coli* DNA supplemented with 30 ng of the specific competitor (unlabeled MspI8), showing that the specific competitor acts efficiently at much lower concentrations. The fact that the higher molecular weight complexes were competed more efficiently by MspI8 DNA than by *E. coli* DNA suggests that these complexes are also formed by proteins with affinity for MspI8. They may contain aggregates of the proteins in the lower molecular weight complex.

Purification of MspI8-binding proteins

Proteins in rat liver nuclear extracts were fractionated by three successive steps: ion exchange chromatography on Q-Sepharose, followed by binding and elution from non-specific (salmon sperm) and specific (MspI8) DNA affinity matrices.

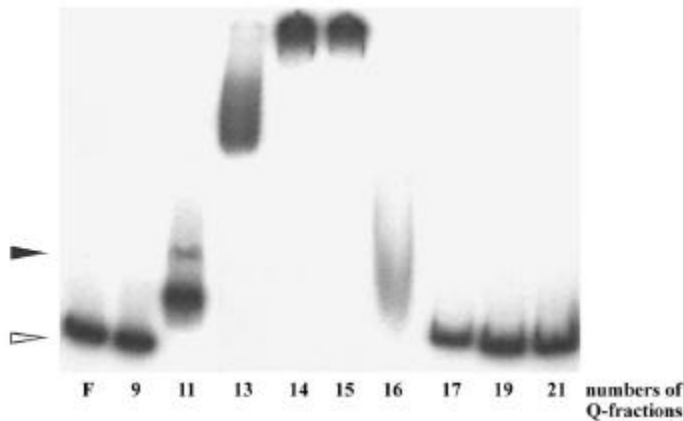


Fig. 3. Purification of MspI8 binding proteins on a Q-Sepharose column. Eluted fractions were assayed by EMSA using unlabeled *E.coli* competitor DNA. Part of the linear elution gradient is shown. Samples contained 3 ng of labeled MspI8, 5 μ l of eluted fractions, and 1500 ng of *E.coli* DNA.

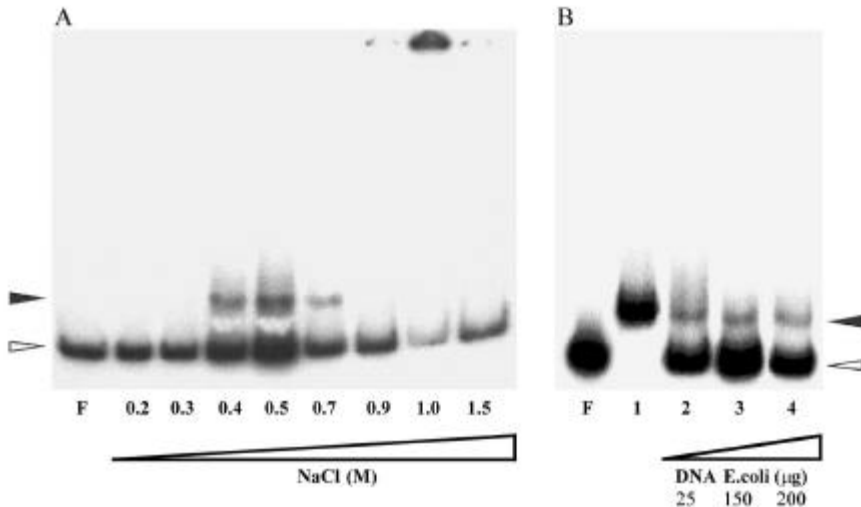


Fig. 4. EMSA assays of proteins eluted from an affinity chromatography resin.

A. Proteins eluted from the nonspecific DNA affinity matrix. Q-Sepharose eluted fraction 13 (Fig. 3) was diluted 5-fold with buffer A to reduce the salt concentration to less than 150 mM and bound to salmon sperm DNA-cellulose. Each sample contained 10 μ l of eluted fraction, 2 ng of labeled MspI8, and 50 ng of *E.coli* DNA. The salt concentration (M) at which fractions were eluted is shown below.

B. Proteins eluted from the specific DNA affinity matrix MspI8-cellulose. Samples contained 3 ng of labeled MspI8, and in lane 1: 2 μ l of Sepharose Q fraction 13 and 2 μ g of *E.coli* DNA; in lanes 2-4: 10 μ l of fractions eluted from pMspI8-cellulose in 0.8 M NaCl; incubation was done in the presence of 25, 150 or 200 μ g of *E.coli* DNA.

Proteins were eluted with a linear salt concentration gradient and tested by EMSA. Proteins showing an affinity for MspI8 were detected in fractions eluted from Q-Sepharose at 0.3 – 0.5 M NaCl (Fig. 3, fractions 11-16). Similar results were obtained using DEAE-Sepharose column chromatography (data not shown). The amount of these proteins in these fractions was high enough to bind all the labeled DNA.

In the presence of higher amounts of *E.coli* DNA, one complex with the same molecular weight as that seen in lane 11 of figure 3 was formed (data not shown). The protein fraction with the maximum MspI8-binding activity from Q-Sepharose (fraction 13, Fig. 3) was then incubated with a non-specific DNA affinity matrix (salmon sperm DNA-cellulose) and proteins were eluted stepwise with increasing salt concentrations from 0.2 to 1.5 M NaCl and tested by EMSA (Fig. 4 A). Proteins which were eluted in 0.4, 0.5 and 0.7 M NaCl formed complexes with MspI8, which had mobilities similar to those seen in figure 2. The electrophoretic pattern of the proteins in these fractions was simpler than that of the nuclear extracts, but too complex to allow the proteins responsible for binding MspI8 to be recognized (Fig. 5).

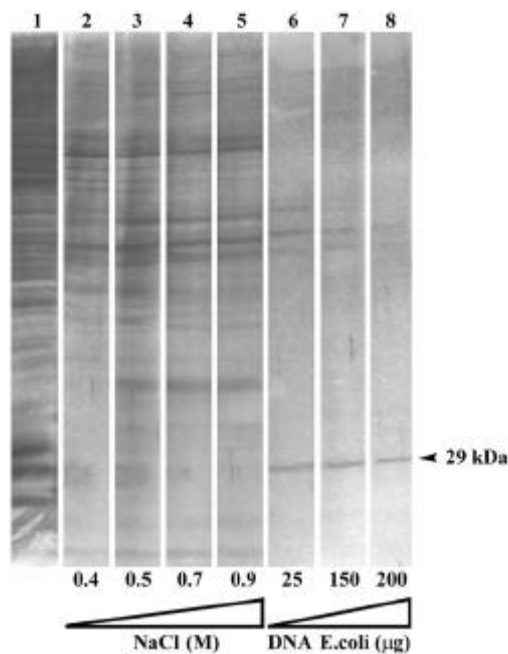


Fig. 5. SDS-PAGE of proteins binding to MspI8.

Lane 1: fraction 13 from Q-Sepharose (18 µg protein); lanes 2-5: 20 µl of proteins eluted from salmon sperm DNA-cellulose with 0.4, 0.5, 0.7, and 0.9 M NaCl, respectively; lanes 6-8: proteins eluted from MspI8-cellulose after binding in the presence of 25, 150 and 200 µg of *E.coli* DNA, respectively.

The fractions eluted from salmon DNA-cellulose with 0.4 and 0.5 M NaCl were pooled and further purified on a specific affinity matrix of linearized pMspI8-cellulose. Binding was performed in the presence of three different concentrations (25, 150 and 200 μ g) of *E.coli* DNA as a nonspecific competitor. After washing, the bound proteins were eluted with a buffer containing 0.8 M NaCl and assayed by EMSA (Fig. 4B).

A protein of approximately 29 kDa was the most prominent in the eluate from MspI8-cellulose (Fig. 5). This protein, which we term p29, may be responsible for the formation of the specific complex with MspI8 seen in figure 4.

DISCUSSION

In humans, LINE transcripts are frequently found in various tumours, but at very low levels, if at all, in normal tissues. In mice, L1 transcripts were detected in spermatocytes [10,14, 28]. Expression is essential for the retrotransposition of LINE 1 sequences, but it is a relatively rare phenomenon in spite of the high repetitiveness of L1 in the mammalian genome. Unlike retroviruses and retrotransposons, L1s have no long terminal repeats (LTRs) which serve as promoters [24]. Their 5'UTRs are necessary for transcription [25] but in contrast to the conserved ORFs, there is no sequence similarity among the 5'UTRs of different species [26]. However, they have similar features: they are GC-rich, they undergo 5' truncation and may consist of small repetitive sequences [15, 25, 27]. The truncation and methylation of CpG dinucleotides probably takes part in the repression of LINE transcription [25, 27], whereas the increased copy numbers of small repetitive sequences could lead to more efficient transcription of L1 [27].

The interaction of nuclear proteins with the 5'UTRs of LINES is likely to be crucial for control of LINE expression and transposition, and we have detected proteins which bind specifically to the 5' fragment of MspI8, a rat L1 LINE element [16,17], in nuclear extracts from rat liver and testis cells, and have partially purified one such protein from rat liver cells. This protein could take part in the regulation of LINE expression. The relationship of this protein to those which protect the A, B and C sites of the human 5'UTR LINE in HeLa cell nuclear extract [24] remains to be studied. MspI8 shows homology to the A-site (positions 360-367, 2 mismatches) and to the C-site (positions 58-46 of the noncoding strand, 3 mismatches). Two proteins, L1PBP-A and L1PBP-B, bind to the same region of the human sequence LPE1 [29]. There is no identical region in MspI8 but one (position 209-187, noncoding strand) is similar, i.e. C-rich (12 Cs in 22 bases of LPE1, 13 Cs in 22 bases of MspI8) and with the same ends [16,17]. L1PBP-A may be a transcription activator and L1PBP-B a suppressor [29]. The molecular mass of protein p29 described here is similar to

that of L1PBP-A. We plan to use synthetic oligonucleotides [29] to examine if its binding site is the same as that of L1PBP-A and L1PBP-B.

The most important region for regulation of LINE expression, the 5'UTR, is the most divergent in closely related species such as rat and human, and the proteins which bind to it are of special interest. Further investigation of p29 as the predominant protein which binds to the 5'UTR region of rat L1 may lead to a better understanding of the significance of LINE sequences in the mammalian genome.

Acknowledgements. The authors thank Joanna Lanuszewska for her help with the illustrations and Ronald Hancock for catalysing this collaboration. This research was supported partly by the Committee for Scientific Research /KBN/ - grant no 4P05A 015 19, and grants from the Human Genome DOE (USA) and Wellcome Trust (UK).

REFERENCES

1. Singer, M.F. and Skowronski, J. Making sense out of LINES: long interspersed repeat sequences in mammalian genomes. **Trends Biochem. Sci.** 10 (1985) 119-122.
2. Fanning, T.G. and Singer, M.F. LINE 1: a mammalian transposable element. **Biochim. Biophys. Acta** 910 (1987) 203-212.
3. Hattori, M., Kuhara, S., Takenaka, O. and Sakaki, Y. L1 family of repetitive sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. **Nature**, 321 (1986) 625-628.
4. Mathias, S.L., Scott, A.F., Kazazian, H.H.Jr., Boeke, J.D. and Gabriel, A. Reverse transcriptase encoded by a human transposable element. **Science** 254 (1991) 1808-1810.
5. Charlesworth, B., Sniegowski, P. and Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature** 371 (1989) 215-220.
6. Kolosha, V.O. and Martin, S.L. In vitro properties of first ORF protein in mouse LINE 1 support its role in ribonucleoprotein particle formation during retrotransposition. **Proc. Natl. Acad. Sci. USA** 94 (1997) 10155-10160.
7. Hohjoh, H. and Singer, M.F. Cytoplasmic ribonucleoprotein complexes containing human LINE 1 protein and RNA. **EMBO J.** 15 (1996) 630-639.
8. Hohjoh, H. and Singer, M.F. Sequence-specific single-strand RNA binding protein encoded by the human LINE 1 retrotransposon. **EMBO J.** 16 (1997) 6034-6043.
9. Drechsler, M. and Royer-Pokora, B. A LINE element is present at the site of 300 kb deletion starting in intron 10 of the PAX6 gene in a case of familial aniridia. **Hum. Genet.** 98 (1996) 297-303.

10. Kuo, K.W., Sheu, H.M., Huang, Y.S. and Leung, W.C. Expression of transposon LINE 1 is relatively human specific and function of the transcripts may be proliferation essential. **Biochem. Biophys. Res. Commun.** 253 (1998) 566-570.
11. Tauchi, H., Matsuura, S., Isomura, M., Kinjo, T. Nakamura, A., Sakamoto, S., Kondo, N., Endo, S., Komatsu, K. and Nakamura, Y. Sequence analysis of 800-kb genomic DNA region on chromosome 8q21 that contains the Nijmegen breakage syndrome gene, NBS1. **Genomics** 55 (1999) 242-247.
12. Van de Water, N., Williams, R., Ockelford, P. and Browett, P. A 30.7 kb deletion within factor VIII gene associated with LINE 1 element insertion. **Thromb. Haemost.** 79 (1998) 938-942.
13. Swergold, G.D. Identification, characterization, and cell specificity of a human LINE 1 promoter. **Mol. Cell. Biol.** 10 (1990) 6718-6729.
14. Asch, H.L., Eliacin, E., Fanning, T.G., Connolly, J.L., Brattbauer, G. and Asch, B.B. Comparative expression of the LINE 1 p 40 protein in human breast carcinomas and normal breast tissues. **Oncol. Res.** 8 (1996) 239-247.
15. Hata, K. and Sakaki, Y. Identification of critical CpG sites for repression of L1 transcription by DNA methylation. **Gene** 189 (1997) 227-234.
16. Rzeszowska-Wolny, J. and Rogoliński, J. MspI8, the repetitive sequence specifically interacting with nuclear matrix of rat testis cells. **Acta Biochim. Polon.** 41 (1994) 459-471.
17. Rogoliński, J., Widzak, P. and Rzeszowska-Wolny, J. Interactions of rat repetitive sequence MspI8 with nuclear matrix proteins during spermatogenesis. **Acta Biochim. Polon.** 43 (1996) 319-324.
18. Sambrook, J., Fritsch, E.F. and Maniatis, T. **Molecular cloning. A laboratory manual.** 2nd edition, Cold Spring Harbor Laboratory Press, 1989.
19. Berezney, R. and Coffey, D.S. Identification of nuclear protein matrix. **Biochem. Biophys. Res. Commun.** 60 (1974) 1410-1417.
20. Strauss, F. and Varshavsky, A. A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. **Cell** 37 (1984) 889-901.
21. Dean, P.D.G., Johnson, W.S. and Middle, F.A. **Affinity Chromatography.** IRL Press, Oxford, Washington DC, 1985.
22. Laemmli, U.K. Cleavage of the structural proteins during the assembly of the head of bacteriophage T4. **Nature** 227 (1970) 680-682.
23. Dunn, M.J. **Gel Electrophoresis: Proteins.** BIOS Scientific Publishers, Oxford NY, 1993.
24. Minakami, R., Kurose, K., Etoh, K., Furuhashi, Y., Hattori, M. and Sakaki, Y. Identification of an internal cis-element for the human L1 transcription and a nuclear factor(s) binding to the element. **Nucl. Acids Res.** 20 (1992) 3139-3145.

25. DeBerardinis, R.J. and Kazazian, H.H. Jr. Analysis of the promoter from an expanding mouse retrotransposon subfamily. **Genomics** 56 (1999) 317-323.
26. Furano, A.V., Robb, S.M. and Robb, F.T. The structure of the regulatory region of the rat L1 (L1Rn, long interspersed repeated) DNA family of transposable elements. **Nucl. Acids Res.** 16 (1988) 9215-9231.
27. Woodcock, D.M., Lawler, C.B., Linsenmeyer, M.E., Doherty, J.P. and Warren, W.D. Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. **J. Biol. Chem.** 272 (1997) 7810-7816.
28. Branciforte, D. and Martin, S.L. Developmental and cell type specificity of LINE 1 expression in mouse testis: implication for transposition. **Mol. Cell Biol.** 14 (1994) 2584-2592.
29. Mathias, S.L and Scott, A.F. Promoter binding proteins of an active human L1 retrotransposon. **Biochem. Biophys. Res. Commun.** 191 (1993) 625-632.
30. Erukashvily, N. I., Lobov, I.B., Kukalev, A. and Podgornaya, O.I. A nuclear matrix protein related to intermediate filaments proteins is a member of the complex binding aliphoid DNA in vitro. **Cell Biol. Int.** 24 (2000), 483-492.