FEBS 19717 FEBS Letters 421 (1998) 229–233

Long terminal repeats of human endogenous retrovirus K family (HERV-K) specifically bind host cell nuclear proteins

Sergey B. Akopov*, Lev G. Nikolaev, Pavel P. Khil, Yuri B. Lebedev, Eugene D. Sverdlov

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya 16/10, Moscow 117871, Russia

Received 24 November 1997; revised version received 2 December 1997

Abstract Solitary long terminal repeats (LTRs) of the human endogenous retroviruses, scattered in several thousand copies throughout the human genome, are potentially capable of affecting the expression of closely located genes. To assess their regulatory potential, the LTR sequences of one of the most abundant HERV families (HERV-K) were screened for the presence of binding sites for the host cell nuclear factors using mobility shift and UV-crosslinking assays. It was shown that the LTR sequences of two subfamilies harbor a specific binding site for a complex consisting of at least three proteins, ERF1, ERF2 and ERF3 of 98, 91 and 88 kDa apparent molecular mass, respectively. This binding site is located in the 5′ region of the LTR U3 element. The preservation of the specific protein binding site in different HERV-K LTR sequences suggests their possible role in regulation of nearby located genes.

Key words: Human endogenous retrovirus; HERV-K; Long terminal repeat; DNA binding protein; Transcription factor

1. Introduction

Human endogenous retroviral (HERV) sequences constitute an estimated 0.1–1% of the whole human genome (see [1–4] for review). To date, more than 10 distinct HERV families have been identified in the human genome. Of them, the HTDV/HERV-K [5,6] (or HERV-K (CUU), see [2]) family is considered to be the most active biologically and capable of forming virions in normal placenta and teratocarcinoma cells [7–9]. Although only about 50 full-length HERV-K copies have been identified so far in the human genome [6], several thousand solitary HERV-K-related long terminal repeats (LTRs) are scattered throughout most human chromosomes [10]. The detailed mapping of the HERV-K LTRs on human chromosome 19 [11,12] revealed frequent proximity of the LTRs and known genes suggesting the possible implication of the LTRs in the regulation of neighboring cellular genes, as demonstrated for other HERV families [13–19].

The HERV-K genes are transcribed in a variety of tissues and cell lines, suggesting that at least some of the HERV-K LTRs are active as promoters (see [3,7] and references therein). To be able to regulate cellular genes, the LTRs should contain some regulatory sequences with the capacity to bind regulatory proteins, e.g. transcription factors.

In this work we tested a near full-length HERV-K LTR for the ability to bind specifically cellular protein factors. We detected three proteins that form a specific complex with the LTR and can be the factors necessary for the LTR to perform its regulatory function.

2. Materials and methods

2.1. Cell culture

Jurkat and CHO (Chinese hamster ovary) cells were grown in RPMI 1640 medium supplemented with 10% fetal calf serum, penicillin (100 U/ml) and streptomycin (0.1 mg/ml), other cell lines were grown under the same conditions but in DMEM.

2.2. Oligonucleotides

Oligonucleotides and PCR primers (Table 1) were synthesized using an ASM-102U DNA synthesizer (Biosan, Russia).

2.3. Nuclear extracts

Nuclear extracts from cultured cells were prepared as described [20] with modifications [21]. For DNase I footprinting, Jurkat cell nuclear extract was additionally purified on a heparin-agarose column. 1 ml of the extract (10 mg/ml in protein) was diluted to 1 mg/ml with 20 mM HEPES-KOH, pH 7.9, 20% glycerol, 100 mM KCl, 0.5 mM EDTA, 1 mM dithiothreitol, 0.5 mM PMSF and then 0.66 volume of water was added to make the final KCl concentration 60 mM. The extract was then passed five times through a 2.5-ml heparin-agarose column equilibrated with nuclear extract buffer or NEB (12 mM HEPES-KOH, pH 7.9, 12% glycerol, 60 mM KCl, 0.3 mM EDTA, 0.6 mM dithiothreitol, 0.5 mM PMSF). The column was washed with 20 ml NEB and then with 3 ml NEB additionally supplied with KCl up to a final concentration of 360 mM. Proteins retained on the column were eluted with NEB containing 0.66 M KCl, dialyzed overnight against 100 volumes of NEB with 30 μM leupeptin and 0.5 mM PMSF at 4°C, roughly 5-fold concentrated using a Sartorius Centrisart 1 (10000) protein concentrator and stored frozen at −70°C.

2.4. Mobility shift assay and UV-crosslinking

LTR fragments for the mobility shift assay were radioactively labeled using PCR with a corresponding primer pair as described earlier [22]. Complementary oligonucleotides were annealed, labeled by filling-in the ends with Klenow enzyme and purified by polyacrylamide gel electrophoresis as described [23].

Mobility shift was performed as described [24] with modifications [23]. Two-dimensional mobility shift UV-crosslinking analysis was performed as follows. Approximately 200 000 cpm of labeled double-stranded oligonucleotide was mixed with 5 μg of poly-d(IC)-poly-d(IC) and 50 μg of nuclear extract proteins in a final volume of 60 μl NEB supplied with 3 μM leupeptin, and incubated for 20 min at room temperature. The mixture was applied to the first dimension 10% polyacrylamide gel prepared in 50 mM Tris-borate buffer, pH 8.3, 0.5 mM EDTA, and separated in the same buffer. The gel was irradiated with a UV lamp (254 nm, 6 μW/cm^2) for 3 min, an autoradiographically localized gel strip was cut out of the gel and immersed in 50 mM Tris-HCl, pH 6.8, 1% SDS, 2% 2-mercaptoethanol, 0.025% bromophenol blue for 15 min at 55°C. The gel strip was then placed on the top of a 12% SDS-polyacrylamide gel [25] and proteins crosslinked to the labeled oligonucleotide were separated in the second dimension according to their molecular masses. The oligonucleotides crosslinked to the proteins do not significantly change their electrophoretic mobility [26].

2.5. DNase I footprinting

To obtain either coding or non-coding end-labeled strands, an

*Corresponding author. Fax: (7) (95) 330-6538.
E-mail: akser@humgen.siobc.ras.ru

PII S0014-5793(97)01569-X

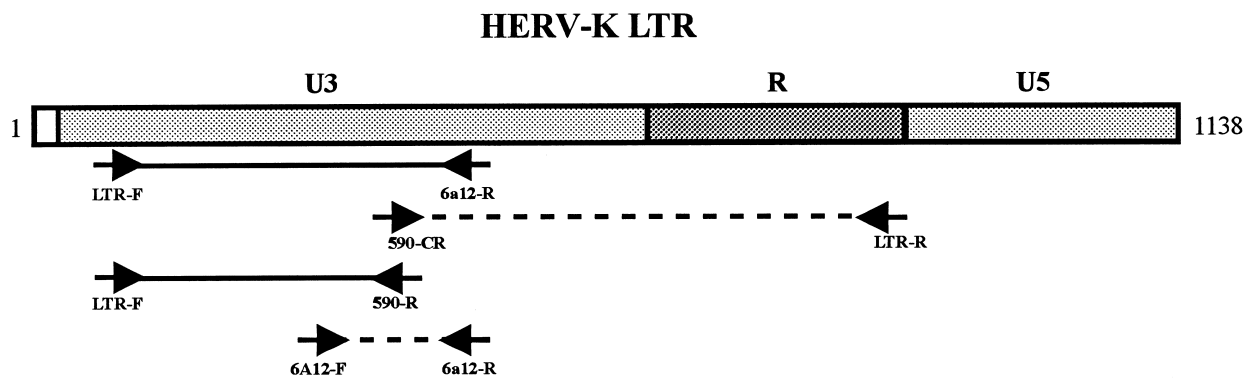

Fig. 1. Structure of the HERV-K LTR and positions of PCR primers and resulting fragments. Solid lines represent fragments capable of binding nuclear proteins (see text).

LTR-F/590-R fragment (333 bp in length) was PCR-amplified using 5′-end labeled LTR-F or 590-R primers (see Fig. 1 and Table 1). To this end, 10 pmol of each primer was incubated with 5 μCi of [γ-^{32}P]ATP (Institute for Physics and Power Engineering, Obninsk, Russia) and 10 U of polynucleotide kinase (Promega) in 15 μl of 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 0.5 mM dithiothreitol, 50 μg/ml bovine serum albumin at 37°C for 45 min. 10 pmol of the labeled primer was added to the PCR mixture containing 10 pmol of the corresponding unlabeled primer, 5 ng of template plasmid DNA, 250 μM each dNTP, 1.5 mM $MgCl_2$, 1×PCR reaction buffer (Perkin-Elmer Cetus) and 2 U of the AmpliTaq DNA polymerase in a final volume of 50 μl. PCR amplification was performed for 18 cycles of the following profile: 94°C for 48 s, 60°C for 48 s, 72°C for 60 s. The whole reaction mixture was then applied to a 10% non-denaturing polyacrylamide gel, the labeled fragments localized by autoradiography, cut out from the gel, and eluted. The footprinting was done as described [27].

2.6. Database search and alignment of sequences

Full-length LTRs of the HERV-K family were identified in GenBank using the BLAST WWW server at NCBI (available at http://www.ncbi.nlm.gov/BLAST). The subfamily-specific consensus sequences of HERV-K LTR were used as search probes. LTRs were aligned using ClustalW ver. 1.6 [28] and the alignment was then manually refined in GDE ver. 2.0. [29]. For transcription factor search the TRANSFAC (ver. 3.1) database [30] was used with the help of the TESS Internet search tool (http://agave.humgen.upenn.edu/tess/index.html).

3. Results and discussion

Fragments of the specific LTR representing the 'old' HERV-K subfamily and mapped to the chromosome 19p12 locus [31] were amplified by PCR using several primer pairs. Positions of primers and PCR fragments within the LTR sequences are shown in Fig. 1, and sequences of the primers are presented in Table 1. Overlapping fragments of the whole LTR were labeled in the course of the PCR reaction using one labeled nucleotide [22], gel-purified, incubated with Jurkat cell nuclear extract and applied to a non-denaturing polyacrylamide gel (Fig. 2). Of four overlapping fragments tested, a strong protein binding band (indicated by an arrow in Fig. 2) was detected for two fragments amplified with the primer pairs LTR-F/6A12-R and LTR-F/590-R. These results located an apparently unique binding site of the LTR between the LTR-F and 6A12-F primer binding sites in a fragment of 230 bp within the LTR U3 region.

To further characterize the protein binding site, DNase I protection experiments were performed (Fig. 3). The protected sequence of about 28 bp was located in the 5′ part of the U3 region. It is flanked by DNAse I hypersensitive sites and spans both previously identified putative enhancer core and glucocorticoid receptor binding sites [6,10]. The most efficiently protected sequence was situated immediately 5′ to these two sites.

To assess general conservation and stability of this region in evolution, the alignment of the full-length sequences of the HERV-K LTRs found in GenBank was done as described in Section 2. The region of interest represented by six characteristic sequences of the alignment is shown in Fig. 4A. The main feature of the region is the presence of a 6-bp site deleted from the first three LTRs of the younger subfamily but not from the other three LTRs of the older subfamily (see [31] for

Table 1
Structures and description of oligonucleotides

Oligonucleotide	Description	Structure 5′-3′
LTR-F	LTR 5′-end primer	TGTAGAAAGAAGTAGACATAGGAGACT
LTR-R	LTR 3′-end primer	AAAGACACAGAGACAAAGTATAGA
590-R	LTR internal primer	GCCATATTTCAGACTATCACATGG
590-CR	Complementary to 590-R	CCATGTGATAGTCTGAAATATGGC
6A12-F	LTR internal primer	ATCACCATTCTCCCAGTCTCA
6A12-R	LTR internal primer	TTCACTAATCCTCCTCAGCCACA
LTR1	LTR protein binding site, coding	AATAGGAGACTCCATTTTGTTCTGTACTAAGA
LTR1-C	LTR protein binding site, non-coding	TTTCTTAGTACAGAACAAAATGGAGTCTCCTA
LTR2	LTR protein binding site, coding	AATAAGAGACTCCATTTTGAAAAAGACCTGTACTTTAA
LTR2-C	LTR protein binding site, non-coding	TTTTAAAGTACAGGTCTTTTTCAAAATGGAGTCTCTTA
HIV	HIV-1 enhancer core, coding	CTACAAGGGACTTTCCGCTGGGGACTTTCCAGGG
HIV-C	HIV-1 enhancer core, non-coding	AGCCCTGGAAAGTCCCCAGCGGAAAGTCCCTTGT
HIVM	Mutant HIV-1 enhancer, coding	CTACAACTCACTTTCCGCTGCTCACTTTCCAGGG
HIVM-C	Mutant HIV-1 enhancer, non-coding	AGCCCTGGAAAGTGAGCAGCGGAAAGTGAGTTGT

characterization of these two subfamilies). As seen in Fig. 3, this deletion lies in the very center of the region protected by protein(s). To study the role of this deletion in the protein binding, two different double-stranded oligonucleotides representing the consensus of these two groups of sequences were designed (Table 1) and then used for the electrophoretic mobility shift assay (Fig. 4B). Lane 1 shows the result of the mobility shift using LTR1/LTR1-C double-stranded oligonucleotide (see Table 1 for sequence) as a labeled probe and nuclear extract from Jurkat cells. This oligonucleotide was designed based on the consensus sequence of the younger LTR subfamily characterized by the presence of the 6-bp AAAAGA deletion (Fig. 4A). Apart from two major visible shifted **a** and **b** bands (Fig. 4B), several faint bands could also be detected. Other lanes show the results of similar mobility shift experiments but in the presence of 10- and 20-fold molar excess of various non-labeled competitors which either contained (LTR1/LTR1-C oligonucleotide, identical to the probe) or did not contain (LTR2/LTR2-C oligonucleotide) the 6-bp deletion (lanes 2, 3 and 3, 4, respectively). Lanes 6 and 7, and lanes 8 and 9 illustrate competitor properties of HIV/HIV-C and HIVM/HIVM-C oligonucleotides representing the native and mutated enhancer regions of human immunodeficiency virus type 1 [23]. Fig. 4B demonstrates that the protein(s) in the major shifted band **a** is readily displaced from its complexes with labeled oligonucleotides by relatively small molar excesses of both oligonucleotide competitors irrespective of the presence of the 6-bp deletion. At the same time, a functional lentiviral enhancer sequence known to bind several transcription factors, including NF-κB [23], does not compete for the protein factor(s) binding the LTR sequence. The less intense shifted band **b** behaved differently. Its intensity was changed insignificantly upon addition of both HERV-K and HIV-1 oligonucleotides, indicating low binding specificity of the protein in this complex. Therefore, only protein(s) responsible for the appearance of the shifted band **a** can be considered to specifically bind the HERV-K LTR.

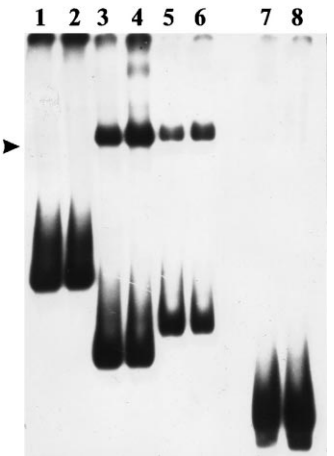

Fig. 2. Electrophoretic mobility shift assay using the HERV-K LTR fragments as probes and nuclear extracts from Jurkat cells. The lanes with uneven and even numbers contained 1 and 3 μg of the extract protein, respectively. Lanes 1, 2: the probe produced by primers 590-CR and LTR-R, length 549 bp. Lanes 3, 4: the probe produced by primers LTR-F and 6A12-R, length 294 bp. Lanes 5, 6: the probe produced by primers LTR-F and 590-R, length 333 bp. Lanes 7, 8: the probe produced by primers 6A12-F and 6A12-R, length 201 bp. The arrow indicates a single major protein-DNA complex.

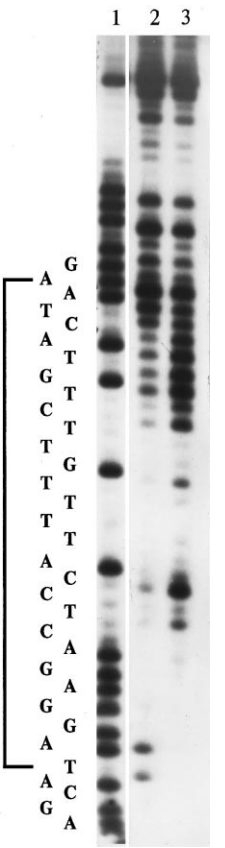

Fig. 3. DNase I protection of the HERV-K LTR fragment LTR-F/590-R. Lane 1, A+G sequence marker. Lane 2, 50 μg of the purified nuclear extract protein added. Lane 3, no protein added.

The mobility shift using a longer double-stranded oligonucleotide LTR2/LTR2-C as a labeled probe and a shorter one (LTR1/LTR1-C) as a competitor showed that the latter is similarly capable of displacing the protein(s) from the complex formed by the former (data not shown). Therefore, despite significant differences between the binding region primary structures, both subfamilies of the HERV-K LTRs are capable of forming a specific complex with the same protein(s) when incubated with the nuclear extract from Jurkat cells. Similar electrophoretic mobility shift experiments with nuclear extracts from HeLa and CHO cell lines apparently produced the same two major shifted bands, with slightly less mobility in the latter case (data not shown). Therefore, the proteins that form complexes **a** and **b** with the HERV-K LTR sequence seem to be present in different human and mammalian cell types.

To characterize the protein-DNA complexes in more detail, the UV-crosslinking technique was employed [21–23]. The protein-DNA complexes were resolved by a mobility shift gel (as in Fig. 4B), the corresponding gel strip was cut out, irradiated with UV light, and the crosslinked complexes of proteins with labeled oligonucleotides were separated in the second dimension according to their molecular masses (Fig. 4C). The non-specific complex **b** (Fig. 4B) produced a single protein spot (**b** in Fig. 4C) with an apparent molecular mass of 56 kDa, whereas the major mobility shift band **a** produced three protein spots corresponding to 98, 91 and 81 kDa. All

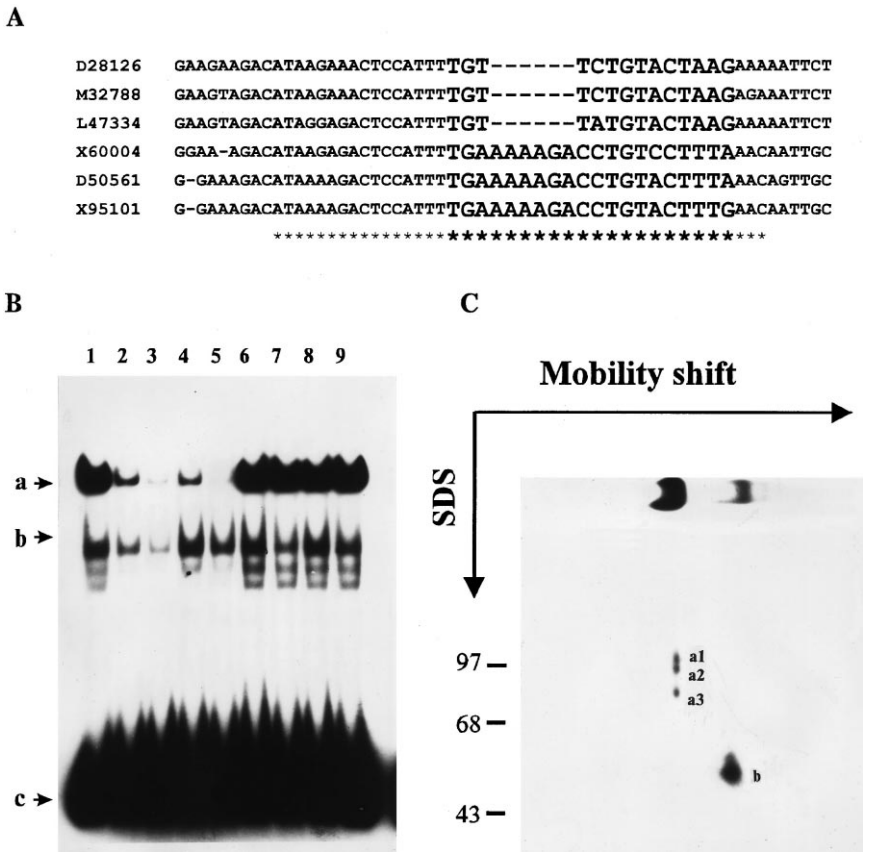

Fig. 4. Identification of the LTR binding proteins. A: Alignment of the sequences of the selected LTRs in the protein binding region (marked by asterisks). GRE and enhancer core elements previously identified by sequence homology [5,6] are enlarged. Left column: GenBank accession numbers of the corresponding sequences. Full-length LTRs of HERV-K were identified in GenBank based on homology with consensus sequences (see Section 2). B: Electrophoretic mobility shift and competition assays using the double-stranded labeled oligonucleotide LTR1/LTR1-C (see Table 1) as template and the nuclear extract from Jurkat cells. **a, b**: major protein-DNA complexes; **c**: free oligonucleotide probe. Lane 1, no competitor added. Lanes 2–9, electrophoretic mobility shift gels with 10- and 20-fold molar excess of different unlabeled double-stranded oligonucleotide competitors: lanes 2, 3: LTR1/LTR1-C; lanes 4, 5: LTR2/LTR2-C; lanes 6, 7: HIV/HIV-C; lanes 8, 9: HIVM/HIVM-C. C: Two-dimensional electrophoresis of the HERV-K LTR-protein complexes: first dimension, non-denaturing 10% polyacrylamide gel; second dimension, 12% SDS polyacrylamide gel after UV-crosslinking of the complexes. **a1** (ERF1), **a2** (ERF2), **a3** (ERF3) and **b**, proteins bound to the LTR. For details see text.

spots were located strictly on the same vertical line, indicating that they were part of the same DNA-protein complex. All three proteins could be crosslinked to DNA by UV light, which means that they are in close contact with it. This can explain why the region protected from the DNase I cleavage by these proteins is rather long (about 30 bp). We designated the proteins ERF1, ERF2 and ERF3 (endogenous retrovirus factors) in order of decreasing molecular masses. However, the complex might also include other protein(s) participating in the binding through protein-protein interactions.

An extensive search of the TRANSFAC database [30] did not reveal proteins both capable of binding our oligonucleotides and having molecular masses close to those of ERF1, ERF2 and ERF3 proteins.

Although the characterization of the identified proteins is a matter for further experiments, the results presented here clearly demonstrate that different kinds of HERV-K LTRs preserved the potential for specific interaction with cellular proteins that is a prerequisite for their involvement in functional interplay in the genome.

Acknowledgements: We are grateful to Drs. V.K. Potapov and N.V. Skaptsova for oligonucleotide synthesis and to Dr. B.O. Glotov for critical reading of the manuscript. This work was supported by grants from the Russian National Human Genome Project and the Russian Foundation for Basic Research.

References

[1] Leib-Mosch, C., Brack-Werner, R., Werner, T., Bachmann, M., Faff, O., Erfle, V. and Hehlmann, R. (1990) Cancer Res. 50, 5636–5642.
[2] Larsson, E., Kato, N. and Cohen, M. (1989) Curr. Top. Microbiol. Immunol. 148, 115–132.
[3] Lower, R., Lower, J. and Kurth, R. (1996) Proc. Natl. Acad. Sci. USA 93, 5177–5184.
[4] Patience, C., Wilkinson, D.A. and Weiss, R.A. (1997) Trends Genet. 13, 116–120.
[5] Ono, M. (1986) J. Virol. 58, 937–944.
[6] Ono, M., Yasunaga, T., Miyata, T. and Ushikubo, H. (1986) J. Virol. 60, 589–598.
[7] Tonjes, R.R., Lower, R., Boller, K., Denner, J., Hasenmaier, B., Kirsch, H., Konig, H., Korbmacher, C., Limbach, C., Lugert, R., Phelps, R.C., Scherer, J., Thelen, K., Lower, J. and Kurth, R. (1996) J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 13, 261–267.
[8] Simpson, G.R., Patience, C., Lower, R., Tonjes, R.R., Moore, H.D., Weiss, R.A. and Boyd, M.T. (1996) Virology 222, 451–456.
[9] Patience, C., Simpson, G.R., Colletta, A.A., Welch, H.M., Weiss, R.A. and Boyd, M.T. (1996) J. Virol. 70, 2654–2657.

[10] Leib-Mosch, C., Haltmeier, M., Werner, T., Geigl, E.M., Brack-Werner, R., Francke, U., Erfle, V. and Hehlmann, R. (1993) Genomics 18, 261–269.
[11] Lebedev, Y.B., Volik, S.V., Obradovich, D., Ermolaeva, O., Ashworth, L.K., Lennon, G.G. and Sverdlov, E.D. (1995) Mol. Gen. Genet. 247, 742–748.
[12] Vinogradova, T., Volik, S., Lebedev, Yu., Shevchenko, Yu., Lavrentyeva, I., Khil, P., Grzeschik, K.-H., Ashworth, L. and Sverdlov, E. (1997) Gene (in press).
[13] Liu, A.Y. and Abraham, B.A. (1991) Cancer Res. 51, 4107–4110.
[14] Feuchter, A.E., Freeman, J.D. and Mager, D.L. (1992) Genomics 13, 1237–1246.
[15] Ting, C.N., Rosenberg, M.P., Snow, C.M., Samuelson, L.C. and Meisler, M.H. (1992) Genes Dev. 6, 1457–1465.
[16] Samuelson, L.C., Phillips, R.S. and Swanberg, L.J. (1996) Mol. Biol. Evol. 13, 767–779.
[17] Kowalski, P.E., Freeman, J.D., Nelson, D.T. and Mager, D.L. (1997) Genomics 39, 38–46.
[18] Di Cristofano, A., Strazullo, M., Longo, L. and La Mantia, G. (1995) Nucleic Acids Res. 23, 2823–2830.
[19] Goodchild, N.L., Wilkinson, D.A. and Mager, D.I. (1992) Gene 121, 287–294.
[20] Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) Nucleic Acids Res. 11, 1475–1489.
[21] Bogomolova, A.E., Braga, E.A. and Nikolaev, L.G. (1989) Mol. Biol. (Moscow) 23, 1057–1063.
[22] Nikolaev, L.G., Tsevegiyn, T., Akopov, S.B., Ashworth, L.K. and Sverdlov, E.D. (1996) Nucleic Acids Res. 24, 1330–1336.
[23] Nikolaev, L.G. (1996) Mol. Biol. (Moscow) 30, 714–720. English translation: Mol. Biol. (1996), 30, 425–428.
[24] Fried, M. and Crothers, D.M. (1981) Nucleic Acids Res. 9, 6505–6525.
[25] Laemmli, U.K. and Favre, M. (1973) J. Mol. Biol. 80, 601–613.
[26] Nikolaev, L.G., Glotov, B.O., Belyavsky, A.V., Grachev, S.A. and Levin, A.V. (1988) Nucleic Acids Res. 16, 519–535.
[27] Galas, D. and Schmitz, A. (1978) Nucleic Acids Res. 5, 3157–3170.
[28] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[29] Smith, S. (1995) Genetic Data Environment (GDE) version 2.0.
[30] Wingender, E., Kel, A.E., Kel, O.V., Karas, H., Heinemeyer, T., Dietze, P., Knueppel, R., Romashenko, A.G. and Kolchanov, N.A. (1997) Nucleic Acids Res. 25, 265–268.
[31] Khil, P.P., Kostina, M.B., Azhikina, T.L., Kolesnik, T.B., Lebedev, Yu.B. and Sverdlov, E.D. (1997) Bioorgan. Khim. (Moscow) 23, 434–440. English translation: Russ. J. Bioorgan. Chem. 23, 406–411.