Biochemistry (Moscow), Vol. 69, No. 6, 2004, pp. 697-701. Translated from Biokhimiya, Vol. 69, No. 6, 2004, pp. 856-861.

Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM03-227, March 28, 2004.

New Aspartic Proteinase of *Ulysses* Retrotransposon from *Drosophila virilis*

D. A. Volkov*, N. I. Dergousova, and L. D. Rumsh

Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10, Moscow 117871, Russia; fax: (7-095) 335-7103; E-mail: den@enzyme.siobc.ras.ru

Received September 4, 2003
Revision received December 3, 2003

Abstract—This work is focused on the investigation of a proteinase of *Ulysses* mobile genetic element from *Drosophila virilis*. The primary structure of this proteinase is suggested based on comparative analysis of amino acid sequences of aspartic proteinases from retroviruses and retrotransposons. The corresponding cDNA fragment has been cloned and expressed in *E. coli*. The protein accumulated in inclusion bodies. The recombinant protein (12 kD) was subjected to refolding and purified by affinity chromatography on pepstatin-agarose. Proteolytic activity of the protein was determined using oligopeptide substrates melittin and insulin B-chain. It was found that the maximum of the proteolytic activity is displayed at pH 5.5 as for the majority of aspartic proteinases. We observed that hydrolysis of B-chain of insulin was totally inhibited by pepstatin A in the micromolar concentration range. The molecular weight of the monomer of the *Ulysses* proteinase was determined by MALDI-TOF mass-spectrometry.

Key words: aspartic proteinases, retrotransposon, transposon *Ulysses*

Aspartic proteinases play a key role in various physiologic processes, such as food digestion, processing of hormones and neuropeptides, and protein utilization [1]. Aspartic proteinases include the proteinases of the majority of retroviruses; among those are human immunodeficiency virus (HIV) and analogous animal viruses. The active site of the aspartic proteinases is formed by two catalytically active residues of aspartic acid [2].

The study of transposons (mobile genetic elements that are widely found in bacterial, plant, animal, and human genomes) revealed that all of them contain the nucleotide sequences encoding proteins that are homologous to retroviral aspartic proteinases [3].

Apparently, the role of proteinases in this case is the processing of polyprotein predecessors, the products of translation of the corresponding mobile genetic elements (analogously to retroviral aspartic proteinases) [2], and release of functional proteins responsible for the transposition of the mobile element.

Retrotransposons are structurally similar to retroviral proviruses. Both retrotransposons and retroviruses are composed of an encoding part containing in different configurations one to three (or more) open frames, corresponding to gag, pol, and env proteins.

While the enzymes of a number of retroviruses (in particular HIV-1) are thoroughly investigated from both applied and fundamental points of view, the study of the enzymes of mobile elements has not yet started. The exception is two publications describing the isolation of the reverse transcriptase and integrase from MDG-4 mobile element and the determinations of their activity [4, 5], as well as investigation of autoprocessing of polyprotein of Ty3 mobile element from *Saccharomyces cerevisiae* [6].

In this connection, we studied the proteinase of *Ulysses* mobile element [7], which was found in the *Drosophila virilis* genome [5, 8].

To determine the conservative structural regions in the proposed sequence of the proteinase of *Ulysses* retrotransposon, we performed comparative analysis of amino acid sequences of known aspartic proteinases from exogenic and endogenic retroviruses and retrotransposons (Fig. 1), based on the use of "internal coordinate systems" for protein molecules [9]. The data provide information about the flanking sequences of the enzyme.

MATERIALS AND METHODS

The polymerase chain reaction was performed using a Techne apparatus (UK). Competent *E. coli* cells were

* To whom correspondence should be addressed.

0006-2979/04/6906-0697 ©2004 MAIK "Nauka / Interperiodica"

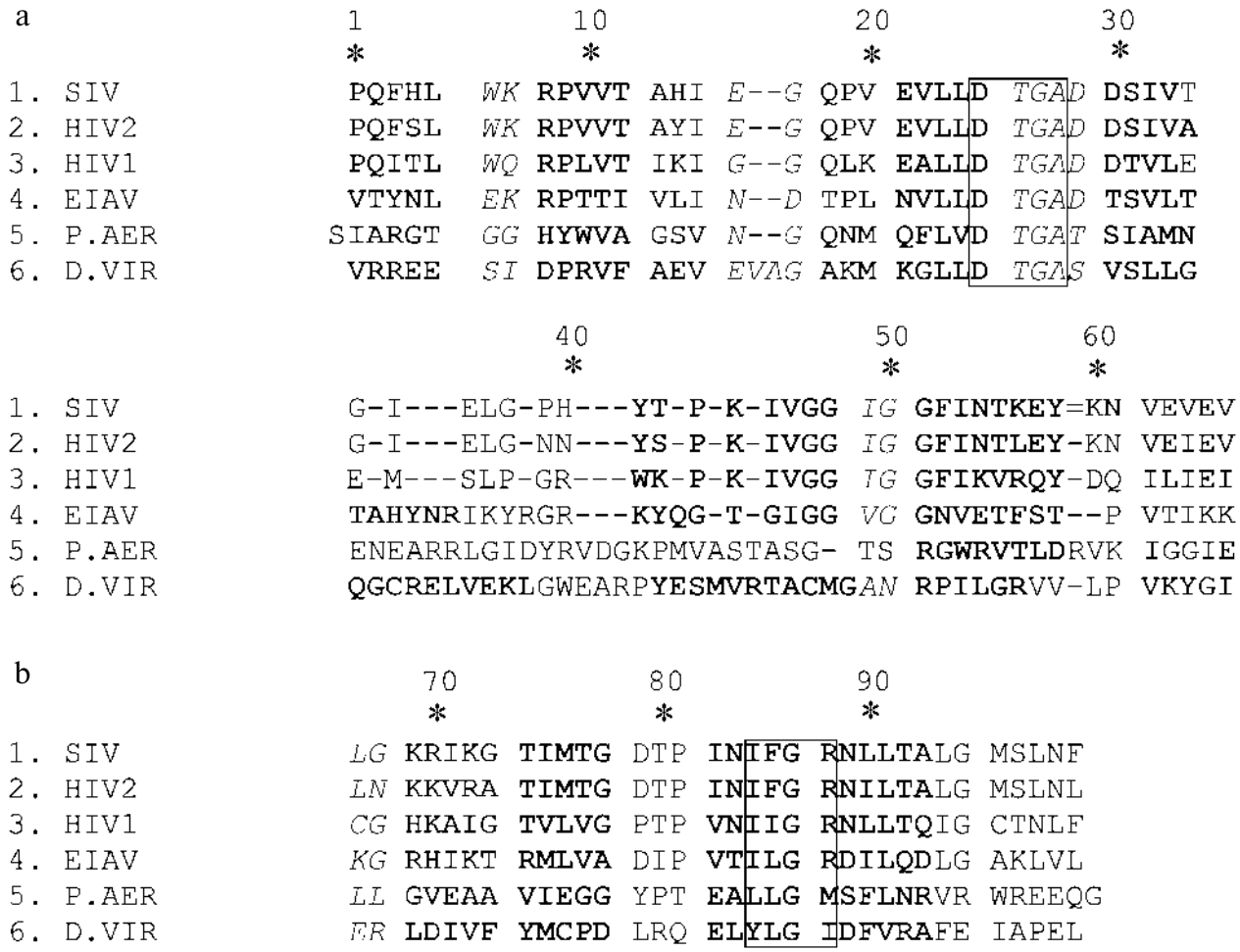

Fig. 1. Comparison of amino acid sequences of aspartic proteinases from retroviruses and retrotransposons. Conservative sequences in the active site and C-terminal regions are in the frame. a) Active site region; b) C-terminal template sequence. HIV-1 and HIV-2 are proteinases from human immunodeficiency virus type 1 and 2; SIV, proteinase from simian immunodeficiency virus; D.Vir, proteinase from *Ulysses* transposon of *Drosophila virilis*; EIAV, proteinase from equine infectious anemia virus; P.AER, aspartic proteinase from *Pseudomonas aeruginosa*.

transformed on a Pulser electroporator (BioRad, USA). The cells were cultivated using a thermostatted Incubator Shaker Series 25 (New Brunswick Scientific Co. Inc., USA). Optical density measurements were performed using an UltroSpec 4050 spectrophotometer (LKB, Sweden). The cells were ultrasonicated on a UZDN-A unit (Russia). HPLC was carried out using a Beckman system (USA). Centrifugation was carried out using GR 412 centrifuge (Jouan, France) and 5415 microcentrifuge (Eppendorf, Germany).

Cloning of *Ulysses* proteinase gene. Plasmid vector pET-23a(+) (Invitrogen, USA) was used for cloning. The proteinase gene was obtained by polymerase chain reaction (PCR) technique; pHXL vector containing cDNA of *Ulysses* transposon was used as a template (kindly provided by Prof. M. B. Evgen'ev). Two oligonucleotide primers were employed for amplification: 1) 5′-AA AAA CAT ATG GTG AGA CGG GAG GAG AG-3′, containing a nucleotide sequence corresponding to 5′-terminus of the proteinase gene. The primer contained the recognition site for *Nde*I restrictase at the 5′-terminus (underlined). An antisense primer 2, 5′-AA AAA AAG CTT TTA CAG CTC TGG TGC GAT CTC A-3′, contained a sequence that is complementary to the 3′-terminal region of the proteinase gene; *Hind*III site (underlined) is introduced for the subsequent cloning. A DNA fragment, obtained as a result of PCR, was analyzed in 1.5% agarose gel, then treated with *Nde*I and *Hind*III restrictases and cloned into pET-23a(+) vector, initially treated by the same restriction enzymes.

Expression and purification of the proteinase. Cloning was performed in *E. coli* cells (strain HB 101). For expression of the corresponding gene by a recombinant plasmid the competent *E. coli*/BL21(DE3) were

transformed. *E. coli* cells which were transformed by pET-*Ulysses* plasmid were grown overnight at 37°C in 10 ml of LB medium in the presence of 100 μg/ml ampicillin. Overnight cell culture was diluted 100 times with LB medium, which contained ampicillin at a concentration of 50 μg/ml, and grown for 3 h at 37°C. When the optical density of the culture reached 0.6-0.8 at 600 nm, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.4 mM, and the mixture was incubated for 3 h under the same conditions. The cell culture was cooled at 4°C, precipitated by centrifugation at 5000 rpm, washed with STE buffer, and stored at −20°C. The inclusion bodies were extracted according to the earlier described technique [10]. The precipitate, which mainly contained the expressed proteins (as inclusion bodies), was dissolved in buffer A (50 mM Tris-HCl, pH 8.0, 25 mM NaCl, 5 mM EDTA, 8 M urea, 0.2% β-mercaptoethanol) at the concentration of 1 mg/ml, and then centrifuged for 30 min at 20,000 rpm. Then the inclusion bodies were subjected to refolding (renaturation). The refolding of denatured target proteins was performed by dialysis against buffer B (25 mM phosphate buffer, pH 7.0, 25 mM NaCl, 5 mM EDTA, 0.2% β-mercaptoethanol) at 4°C [11]. After that, the protein was additionally purified by affinity chromatography on pepstatin-agarose. The protein was applied on the column (40 × 15 mm) containing pepstatin-agarose equilibrated with buffer B. The protein was eluted with buffer C (100 mM Tris-HCl buffer, pH 9.5, 25 mM NaCl, 5 mM EDTA, 0.2% β-mercaptoethanol). Fractions were pooled and stored at 4°C. Protein concentration was determined as described by Bradford [12].

After affinity chromatography, the proteins were concentrated by ultra-filtration using a YM10 membrane (Amicon, Holland) to the concentration of 0.5 mg/ml, and stored at −70°C. Purity of the samples was monitored by SDS-PAGE in 15% gels according to Laemmli [13].

Determination of *Ulysses* proteinase activity. Melittin and B-chain of insulin were hydrolyzed by the proteinase at pH 5.5 for 1 h at 37°C in buffer D (50 mM acetate buffer, 100 mM NaCl, 5 mM EDTA, 0.2% β-mercaptoethanol). The enzyme concentration was 0.6 μM, substrate concentration 43 μM. The hydrolysis products were separated by HPLC on an Ultrasphere C_{18} column (Beckman Ultrasphere, 2 × 150 mm, 5 μm). The sample was eluted by a step gradient of acetonitrile in water in the presence of 0.1% trifluoroacetic acid: 0 to 40% in 40 min, 40 to 50% in 10 min at elution rate of 0.5 ml/min. The products were detected at 220 nm. Amino acid sequences of the resulting peptides were obtained by N-terminal sequencing.

Determination of pH optimum of activity. Melittin and B-chain of insulin were used as substrates. The pH dependence was studied in the pH range between 2 and 11. The following buffer solutions were used: Gly-HCl (pH 2-4), acetate (pH 5-6), phosphate (pH 7), Tris-HCl (pH 8-9), Gly-NaOH (pH 10-11).

Pepstatin inhibition of hydrolysis of B-chain of insulin by *Ulysses* proteinase. Enzyme solution with final concentration of 0.6 μM was incubated with pepstatin for 20 min. Then the substrate was added to the final concentration of 43 μM. The reaction was performed in buffer D at 37°C for 1 h. The inhibitor concentration was varied of the range 0.219-0.0073 μM.

Determination of molecular weight of *Ulysses* proteinase. The molecular weight of the proteinase was determined by MALDI-TOF mass-spectrometry technique using a VISION 2000 apparatus (Thermo BioAnalysis, USA). The sample was irradiated by a UV laser with wavelength of 337 nm and maximum energy of 250 μJ in pulse mode with a pulse duration of 3 nsec. Deionized protein samples were mixed with a matrix containing 2,5-dihydroxybenzoic acid dissolved in 25% methanol with 0.1% trifluoroacetic acid. The N-terminal amino acid sequences of the protein were analyzed by automatic degradation according to Edman using a 470A gas-phase sequencer (Applied Biosystems, USA). C-Terminal sequences were analyzed using carboxypeptidase A.

RESULTS AND DISCUSSION

Choice of the cloning fragment. DNA of *Ulysses* retrotransposon contains 10,633 bp. This retrotransposon is flanked by two unusually large straight repetitions, 2136 bp long. *Ulysses* exhibited the typical organization of LTR-containing retrotransposons with matrix and capsid protein domains, which are encoded in the first open reading frame. Proteinase, reversed transcriptase, RNase H, and integrase domains are encoded in the second reading frame. However, *Ulysses* lacks the third reading frame that is present in the certain retrotransposons, which may encode an evn-like protein [8].

The choice of cDNA fragment for cloning encountered certain difficulties. As mentioned earlier, the alignment of primary structures of *Ulysses* and HIV-1 proteinases presented by Scheinker et al. [8] did not appear convincing for the C-terminal part of the molecule, despite an overall good match. The template sequence hydrophobic amino acid–hydrophobic amino acid–glycine–arginine [9] (Fig. 1), which is standard for all retroviral proteinases and characteristic for the molecule segment passing through the active loop of the active site close to the C-terminal segment, was located in a completely different region of the molecule. This contradiction could be resolved if one supposes that the segments linking the loop with β-layer of the molecule are shortened, and the entire molecule is shorter than the HIV-1 proteinase molecule. On the other hand, it was suggested that the *Ulysses* proteinase molecule can be sufficiently

longer than HIV-1 proteinase molecule due to its C-terminal region. Hence, the decision was to clone a fragment slightly enlarged at both ends compared to the alignment data, assuming that during the autoprocessing the proteinase itself will provide the required length of the molecule. The cloned fragment of cDNA corresponded to amino acid residues numbered 119-231 for the second reading frame.

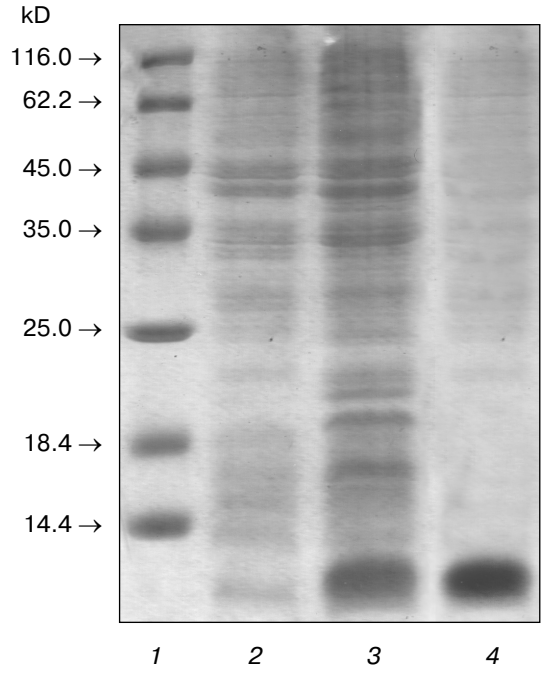

Fig. 2. Expression of *Ulysses* proteinase in *E. coli*/BL21(DE3) and purification of the recombinant protein. Data of SDS-PAGE in 15% gels: *1*) protein standards; *2*) total protein content in *E. coli* transformed with pET-*Ulysses* without induction; *3*) total protein content in *E. coli* transformed with pET-*Ulysses* after induction by IPTG; *4*) protein sample after affinity chromatography.

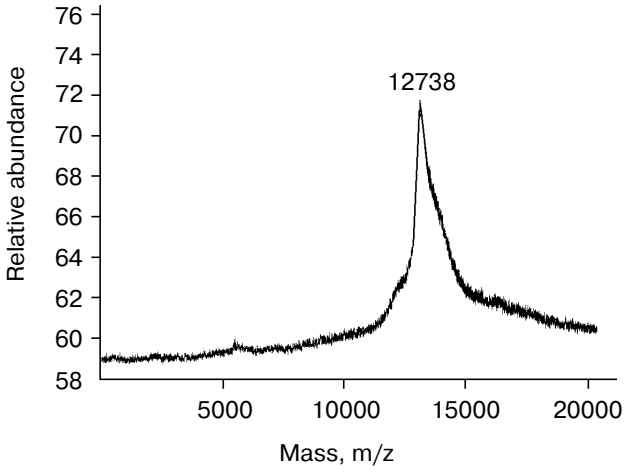

Fig. 3. MALDI MS spectrum of *Ulysses* proteinase.

Cloning of *Ulysses* proteinase gene. Plasmid vector pET-23a(+) under the control of T7 bacteriophage promoter was used for cloning. The cloning was performed in *E. coli* cells (HB 101 strain), which did not contain a copy of the RNA-polymerase T7 gene. The obtained pET-*Ulysses* vector was later used for the production of recombinant *Ulysses* proteinase. The recombinant plasmid was used for the transformation of competent *E. coli*/BL21(DE3) strain, which contained a chromosomal copy of RNA-polymerase gene of T7 under the control of lac UV5 promotor, inducible by isopropyl-β-D-thiogalactopyranoside (IPTG).

Expression and purification. SDS-PAGE was used to test the expression of the recombinant gene in *E. coli* cells. Figure 2 shows an electrophoregram of the original proteins after expression of the corresponding gene, and also an electrophoregram of the obtained target protein at different stages of proteinase extraction and purification. The target protein (~12 kD) was accumulated in cells in the form of inclusion bodies with high yield. The extracted inclusion bodies were dissolved in a buffer containing 8 M urea, and after that subjected to refolding. Refolding of the denatured target proteins was performed by dialysis against 25 mM phosphate buffer, pH 7.0, at 4°C. The protein was further purified by affinity chromatography on pepstatin-agarose. Bound protein was eluted with 20 mM Tris-HCl buffer, pH 9.5. The purity of the resulting samples was at least 90%.

The ability of recombinant protein to bind with pepstatin-agarose suggests that the substrate-binding center corresponding to aspartic proteinases is formed in the course of refolding. Therefore, it could be expected that the protein would have proteolytic activity.

Determination of molecular weight. Precise determination of molecular weight of the proteinase was carried out by mass-spectrometry. The molecular weight of the enzyme is 12,738 daltons (Fig. 3). N- and C-terminal analyses of amino acid sequences were also performed. The six N-terminal amino acid residues were determined: Val-Arg-Arg-Glu-Glu-Ser. Protein treatment by carboxypeptidase A identified the first C-terminal amino acid of the protein as Leu. These data are in good correlation with the estimated mass of *Ulysses* proteinase monomer and imply the absence of autoprocessing of the expressed protein.

Investigation of enzymatic properties. The enzymatic properties of the proteinase were studied. As we demonstrated earlier [7], *Ulysses* proteinase efficiently hydrolyzes melittin. To determine the substrate specificity of the proteinase, we also investigated another polyprotein substrate, B-chain of insulin, which is widely used for the assessment of the substrate specificity of aspartic proteinases. *Ulysses* proteinase efficiently hydrolyzed B-

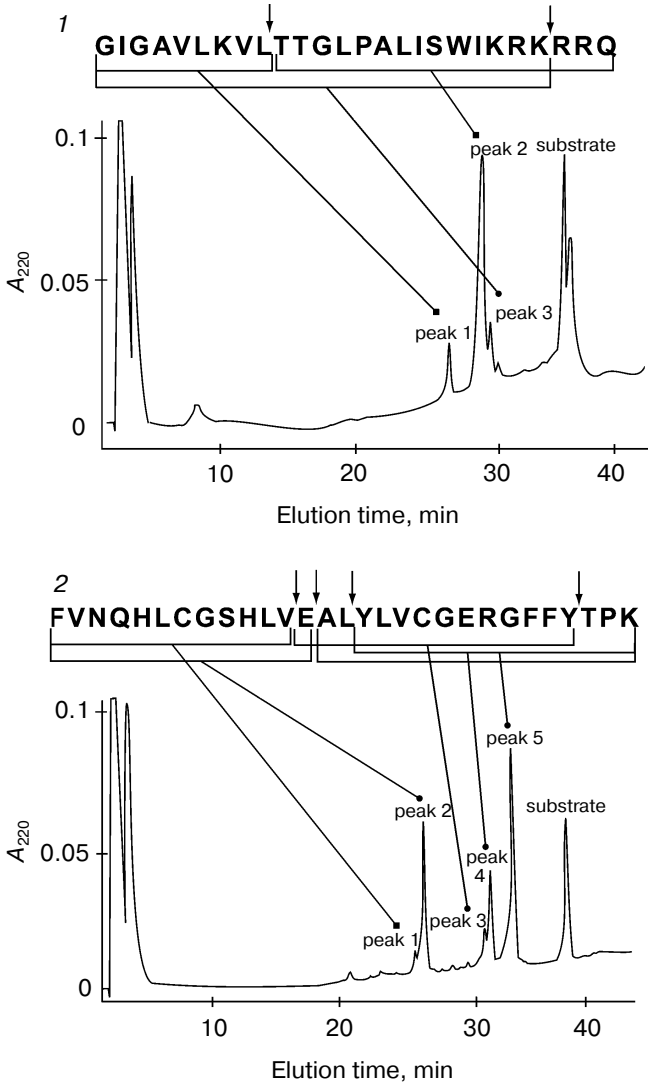

Fig. 4. HPLC of the enzymatic hydrolysis (pH 5.5) products of melittin (*1*) and B-chain of insulin (*2*) by *Ulysses* proteinase. An Ultrasphere C18 column was equilibrated with aqueous 0.1% TFA solution. Hydrolyzed bonds are indicated by arrows.

chain of insulin. Hydrolysis of both polypeptide substrates by the proteinase was performed at pH 5.0. The reaction mixture was analyzed by HPLC using a C_{18} column (Fig. 4). The data on the substrate specificity of *Ulysses* proteinase are illustrated in the figure. It can be seen that *Ulysses* proteinase hydrolyses the bonds formed both by hydrophobic and charged amino acids, showing the dual nature of the enzyme specificity.

We have studied the pH dependence of melittin hydrolysis by a proteinase from *Drosophila virilis* mobile element. The reaction was performed in the pH range between 2 and 10. The pH optimum for the proteinase is 5.5. This fact relates our proteinase with known aspartic proteinases, which display maxima of activity at acidic pH values.

The majority of aspartic proteinases including HIV-1 proteinase are well inhibited by pepstatin A [14]. We studied the pepstatin A inhibition of hydrolysis of B-chain of insulin by *Ulysses* proteinase. It was found that the hydrolysis of the polypeptide is nearly completely suppressed in the presence of 10-μM inhibitor in the reaction mixture.

This work was supported by the Russian Foundation for Basic Research (project No. 02-04-49108 and No. HSh-1781.2003.4).

The authors are grateful to M. B. Evgen'ev for providing the vector containing the cDNA sequence of the *Ulysses* transposon. The authors would like to thank N. S. Andreeva for her kind attention to this work. The authors also express their gratitude to K. B. Ignatov (Institute of Bioorganic Chemistry, Russian Academy of Sciences) for providing DNA-ligase and DNA-polymerase, and to N. S. Bystrov (Institute of Bioorganic Chemistry, Russian Academy of Sciences) for synthesis of oligonucleotide primers.

REFERENCES

1. Davies, D. R. (1990) *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 189-215.
2. Dunn, B. M., Goodenow, M. M., Gustchina, A., and Wlodawer, A. (2002) *Genome Biol.*, **3(4),** 3006.1-3006.7.
3. Archipova, I. R., Lubomirskaja, N. V., and Ilyin, Yu. V. (1995) *Drosophila Retrotransposon*s, R. J. Landes Press Comp. Osttames, USA.
4. Ivanov, A. V., Ivanov, V. A., Melnikov, A. A., Siunov, A. V., Fodor, A. A., and Ilyin, Yu. V. (1991) *EMBO J.*, **10**, 2489-2495.
5. Glukhov, I. L., Ivanov, V. A., and Ilyin, Yu. V. (2000) *Mol. Biol. (Moscow)*, **34**, 277.
6. Kirchner, J., and Sandmeyer, S. (1993) *J. Virol.*, **67**, 19-28.
7. Volkov, D. A., Savvateeva, L. V., Dergousova, N. I., and Rumsh, L. D. (2002) *Vopr. Med. Khim.*, **48**, 553-560.
8. Scheinker, V. S., Lozovskaja, E. R., Bishop, J. G., Corces, V. G., and Evgen'ev, M. B. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 9615-9619.
9. Andreeva, N. S. (1991) *Adv. Exp. Med. Biol.*, **306**, 559-572.
10. Lin, K., and Chang, S. (1991) *BioTechniques*, **11**, 748-752.
11. Dergousova, N. I., Amerik, A. Yu., Volynskaya, A. M., and Rumsh, L. D. (1996) *Appl. Biochem. Biotechnol.*, **61**, 97-107.
12. Bradford, M. M. (1976) *Analyt. Biochem.*, **72**, 248-254.
13. Laemmli, U. K. (1970*) Nature*, **227**, 680-685.
14. Tyagi, S. C. (1992) *Biochem. Cell. Biol.*, **70**, 309-315.