ISSN 1993-6842. Biopolymers and cell. 2008. vol. 24. N 6. Translated from Ukrainian

Distribution of hairpin-loop structures in plasmids of anthrax infectious agent

O. Yu. Limanskaya[1,2], A. P. Limanskii[1]

[1] I.I. Mechnikov Institute of Microbiology and Immunology of Academy of Medical Sciences of Ukraine
14 Pushkinska Str., Kharkiv, 61057 Ukraine

[2] National Research Centre "Institute for Experimental and Clinical Veterinary Medicine"
83 Pushkinska Str., Kharkiv, 61023 Ukraine

olga.limanskaya@mail.ru

An important biological function of hairpin-loop structures is the defense of RNA transcripts from degradation by different factors as well as the transcription regulation due to their formation in transcription terminators. The patterns of thermodynamically stable perfect and imperfect inverted repeats were determined for pXO1 and pXO2 plasmids of pathogenic Bacillus anthracis strains. A sequence analysis of these plasmids has shown the plasmid pXO1 contains 176 inverted repeats, the energy of which varies from –30.6 kcal/mol to –10.0 kcal/mol, and the plasmid pXO2 of B. anthracis contains 57 inverted sequences with energy from –27.2 kcal/mol to –10.0 kcal/mol. Physical maps of the pXO1 and pXO2 plasmids with located hairpins are presented. These hairpin-loop structures are shown to be localized in the sites of regulatory genes or the elements encoding proteins of unknown function.

Keywords: Bacillus anthracis, hairpin-loop structure, inverted repeat, cruciform structure.

Introduction. At present, the intensive efforts of scientists are aimed at the development of efficient approaches to analyze so called genetic texts, i.e. nucleotide sequences of genomes. A computer analysis is of special importance for studying DNA text due to the possibility to establish certain functions of different DNA fragments: search for structural genes, regulatory sites, etc. The accuracy of current computer methods in determ ining genes of a known nucleotide sequence is not higher than 70% [1].

Being a new branch of science, genomics is still descriptive and developing along with technical progress. The sequencing of genomes has deepened and extended our understanding of genetic information. At present, the most important parts of genome are considered to be exome (comprising only 1% of genome), introme, methylome, transcriptome (a set of all RNA transcripts in one or a population of cells), and variome (total genetic variations, characterizing the species, or a sum of single nucleotide polymorphisms) [2, 3].

In eukaryotes and prokaryotes the conservative fractions of genome, non-coding proteins, have very important functions, namely, they are a source of both sense and antisense non-coding RNAs of different variants of introns, comprising introme (sometimes introns may function as exons, and vice versa). The data on introme of microorganisms would allow to identify biologically significant properties, regulated by introns.

Functions of various genes are known to depend on many factors, however, the total number of elementary regulatory factors is thought to be considerably smaller than the total number of genes. An insignificant part of elementary factors (compared to the total number of genes) may be sufficient to regulate a considerable number of genes. Genome sequences of eukaryotes and prokaryotes contain a huge amount of information related to their molecular genetics. While scientists are developing approaches to obtain this information and solve various problems of genomics, much attention is given to search of repeats, since they comprise a large part of genome. In particular, human genome contains over 50% of repeats, some classes of which play a vital structural and functional role. Still, search for repeats is a highly challenging endeavor.

Bacillus anthracis is a large rod-shaped, Gram-positive, anaerobic bacterium, which is an etiological agent of anthrax, a dangerous and often fatal disease of both humans and animals. Along with *B. thuringiensis*, *B. cereus*, and *B. mycoides*, it belongs to *B. cereus* genus. These closely-related bacteria are animal (*B. anthracis* and *B. cereus*) and insect pathogens (*B. thuringiensis*). *B. anthracis* differs from the rest of the genus members by the presence of megaplasmids *pXO1* and *pXO2*, coding the synthesis of toxins and capsules, respectively, and providing for virulence of the bacterium. There are some known isolates of *B. anthracis* with one or two plasmids absent. Besides, there is a possibility of plasmid transfer among related species or natural loss of the *pXO*1 plasmid [4]. There are also data on the successful transfer of the *pXO1* plasmid into other bacteria and expression of the toxins genes, such as *lef* and *cya*, in heterologous systems [5]. The absence of one of plasmids results in the loss of pathogenic properties of *B. anthracis* strain [6].

The possibility of forming hairpin-loop structures attracted attention due to their capability to regulate a stability of microorganism's mRNAs [7, 8]. In [9] we determined the distribution of thermodynamically stable perfect inverted repeats for two isolates of slow growing Mycobacterium tuberculosis with complete genome (H37Rv and CDC1551). Regardless of a high level of homology (over 90%) of genomes of these mycobacteria isolates they differ in virulence level, namely, H37Rv is a laboratory strain, while clinical isolate CDC1551 is highly virulent. We proved both isolates to have eight long inverted repeats of 48-62 nucleotides, six of which coincide completely. At the same time in the CDC1551 genome (contrary to H37Rv) there is a highly stable hairpin of 58 nucleotides at 5'-end of DNA template chain [9]. It was supposed that localization of highly stable hairpin, G=-53.9 kcal/mol at 5'-end of DNA of the CDC1551 isolate may result in different degree of RNA transcripts stability or different level of transcription termination efficiency in the CDC1551 strain compared to the H37Rv isolate, which, in its turn, may be one of the reasons of different virulence of the strains, regardless of similarity in physical maps of their genomes.

The inverted repeats may also serve as recognition sites for recombinases. The presence of inverted repeats may testify to the probability of DNA sequences transfer among repeats due to the transcription or recombination [10]. An important biological function of hairpins is the protection of RNA transcripts of plasmids, determining virulence of *B. anthracis*, from degradation by different factors. We used computer analysis to characterize thermodynamically stable perfect and mismatched inverted repeats, forming hairpin-loop structures, or hairpins which may appear in the *pXO1* and *pXO2* plasmids of *B. anthracis* pathogenic strains, and presented the physical maps of plasmids with located hairpins.

Materials and Methods. Isolates of *pXO2* plasmid (number AE 011191 (NC 003981), 94829 bp) and *pXO1* plasmid (number AF 065504, 181654 bp) with complete genome were used. *Oligo* software (version 3.4) was used to search for perfect inverted repeats and determine their thermodynamic characteristics [11]. *RNA 2* software of GeneBee was used to search for mismatched repeats and determine their parameters [12].

Table 1
Thermodynamically stable hairpin-loop structures potentially formed by inverted repeats for pXO1 plasmid of pathogenic strain Bacillus anthracis

Stem length, b.p.	Loop length, b.p.	Free energy, kcal/mol	Position on plasmid
18	12	–21,5	24754–24801
20	8	–23,9	24990–25038
19	**4**	**–22,9**	**26110–26131**
18	**4**	**–28,3**	**26553–26592**
21	**10**	**–20,7**	**42014–42065**
25	12	–22,5	48010–48071
24	18	–30,6	55825–55890
19	**3**	**–21,9**	**60528–60568**
20	7	–28	63955–64001
13	**5**	**–22,9**	**71398–71428**
20	5	–30,5	72274–72318
21	**8**	**–29,8**	**100450–10049**
23	5	–25	105993–10604
16	**5**	**–21,7**	**109971–11000**
22	3	–21	114805–114851
20	**13**	**–29,7**	**120940–120991**
24	8	–21,2	121312–121367
24	**4**	**–24,3**	**136336–136387**
19	**4**	**–21,7**	**137579–137620**
22	**3**	**–25,3**	**146102–146148**
25	**3**	**–21,4**	**148426–148478**
18	4	–23	153484–153522
18	**3**	**–25**	**159742–159782**
20	3	–20,5	162053–162095
14	**4**	**–21**	**169071–169102**
17	17	–24,1	172043–172093
18	**9**	**–29,3**	**179421–179465**

Note. Perfect hairpins are shown in bold

Since software, used in this work, allows analyzing sequences, not exceeding 15000 bp, the complete sequences of *pXO1* and *pXO2* plasmids, obtained from GenBank database, were cut in fragments of 14000 bp.

Results and Discussion. Repeats may be direct and inverted, perfect (complete coincidence of their sequences) or imperfect (containing mismatches). Inverted repeats in RNA and DNA molecules may be in two different conformational states – either as single- or double-stranded helix, or in the form of hairpin-loop structure, consisting of a double-stranded stem and single-stranded loop. Earlier it was shown that under physiological conditions the superhelical DNA with inverted repeats (palindromes) may form hairpins as fragments of cruciform structure with the stem length not less 7 bp and the loop not exceeding 4-5 bp in [13-15]. We have used these parameters to determine thermodynamically stable perfect inverted repeats in the sequences of *pXO1* (Table 1) and *pXO2* plasmids of the pathogenic strain A2012 *B. anthracis* (Table 2).

The computer analysis revealed that the *pXO1* plasmid contains 67 hairpin-loop structures with the loop not exceeding 5 nucleotides (Fig.1). The free energy (- *G*) of 11 hairpins is over 20 kcal/mol (Table 1). It is noteworthy that the *pXO1* plasmid has four perfect inverted repeats with high values of - *G* (over 20 kcal/mol) and the loop of 8-13 nucleotides (positions 5, 11, 12, 27 in Table 1). We believe that this fact is in favor of the *in vivo* existence of hairpins with loops of 8-13 nucleotides.

For convenience, the palindromes, potentially capable of forming hairpins due to interstranded complementary pairing of nucleotides in single-stranded DNA and RNA, may be divided into long and short repeats (over 45 nucleotides and less than 45 nucleotides, respectively). The *pXO1* plasmid may contain 17 long palindromes of 45-66 nucleotides with *G* ranging from -30.6 to -20.7 kcal/mol, and 10 short palindromes of 31-43 nucleotides with *G* from -20.5 to -30.6 kcal/mol (Table 1). It should be noted that the majority of hairpin structures on the physical map of the *pXO1* plasmid is located either in the area of regulatory genes or in the elements of unknown function (Fig.1, *b*). This fact proves the well known literature data on the presence of cruciform structures in the regulatory fragments, in many transcription terminators, in particular [16].

Table 2
Thermodynamically stable hairpin-loop structures, potentially formed by inverted repeats, for pXO2 plasmid of pathogenic strain Bacillus anthracis

Stem length, b.p.	Loop length, b.p.	Free energy, kcal/mol	Position on plasmid
20	9	–20,0	13585–13634
17	4	–24,3	29190–29227
22	7	–27,2	37712–37762
20	**6**	**–21,2**	**44958–45003**
20	**11**	**–27,1**	**48478–48528**
20	8	–22,4	59914–59961

Note. Perfect hairpins are shown in bold.

The *pXO2* plasmid may contain 25 hairpin structures with the loop up to 5 nucleotides (Fig.2). Besides, there are two long perfect inverted repeats with the loop exceeding 5 nucleotides and - *G* over 20 kcal/mol (positions 4, 5 in Table 2).

Contrary to the *pXO1* plasmid, the *pXO2* plasmid contains neither long palindromes with the loop up to 5 nucleotides, nor a single perfect short palindrome with - *G* over 20 kcal/mol (Table 2). Fig.2, *b* shows a physical map of the *pXO2* plasmid with positions of hairpin structures found. It should be noted that functions of the majority of genes of the *pXO2* plasmid are yet to be determined, and the data obtained can be analyzed thereafter. According to the results, the perfect and mismatched inverted repeats are highly stable genetic elements for both plasmids. The analysis of potentially probable secondary structures allows the supposition on the biological function of mentioned hairpins being the protection of RNA transcripts of plasmids, causing virulence of bacteria, from degrading by different factors.

As shown in [18-20], the inverted repeats form cruciform structures in negatively superhelical DNA both *in vivo* and *in vitro*. There are specific biochemical [21] and biophysical methods to detect hairpin structures *in vitro*. While molecular biological methods, based on cutting a hairpin loop with a nuclease (specific for single-stranded DNA), allow determining the sequence

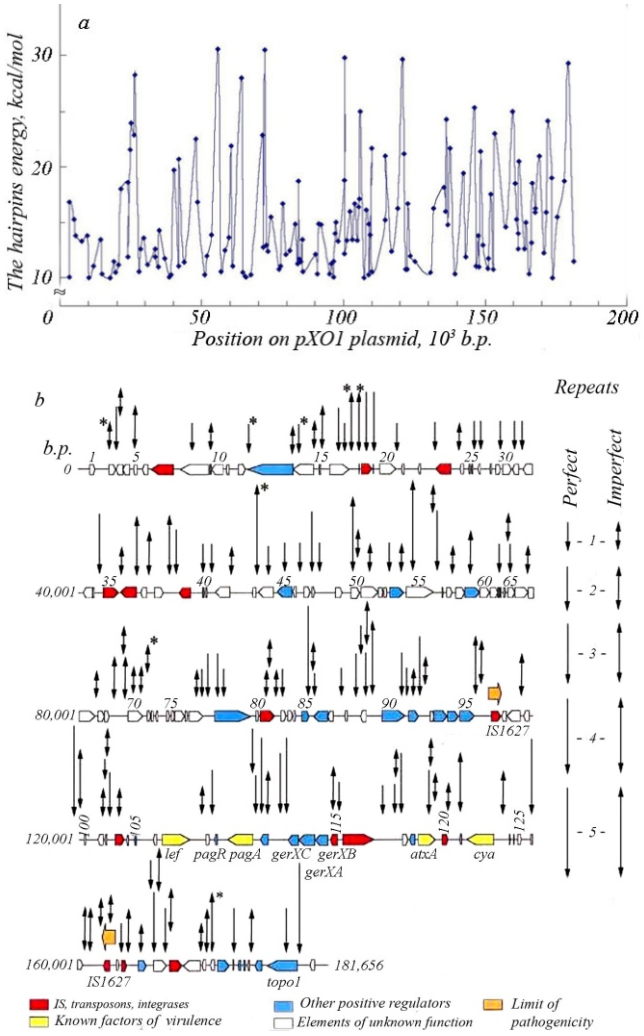

Fig.1 Histogram of distribution of possible hairpin-loop structures on *pXO1* plasmid (181654 bp) of anthrax agent (*a*) and physical map of *pXO1* plasmid of *Bacillus anthracis* [17] (*b*). Figures show positions of known genes of toxins, elements IS1623, and expected positions of 143 open reading frames. Encoding genes: *lef* - endopeptidase of lethal factor, *cya* - calmodulin-sensitive adenylate cyclase; *pagA* - protective antigen; *topo I* - topoisomerase I; *gerXA, ger XB* - development of spores; *atxA* - positive transregulator of expression of antrax toxin gene; *pagR* - transcriptional repressor. Arrows indicate positions of thermodynamically stable perfect and mismatched hairpin structures; asterisks show hairpin structures, where the loop exceeds 12 nucleotides; hairpins with the free energy over: 1 - 10; 2 - 15; 3 - 20: 4 - 25; 5 - 30 kcal/mol

and locating the inverted repeat on the genome, modern methods of nanobiotechnology, first of all scanning probe microscopy, enable direct visualization of a hairpin structure [22].

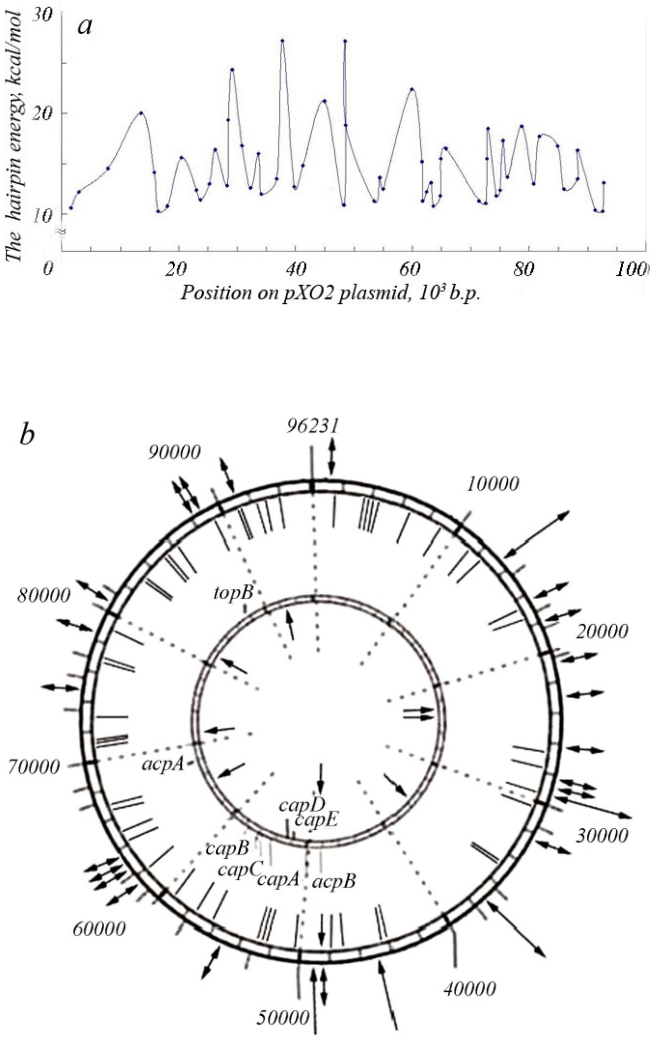

Fig.2 Histogram of distribution of possible hairpin-loop structures on *pXO2* plasmid (94829 bp) of anthrax agent (a) and physical map of *pXO2* plasmid of *B. anthracis* (number AF188935, 96231 nucleotides long) (b). Location of known virulent genes (*capA, capB, capC, capD, cape, topB*) and positive trans-regulators (*acpA, acpB*) is presented in the inner circle.
Arrows indicate positions of potentially perfect and mismatched hairpin structures, the loop of which does not exceed 12 nucleotides. Other symbols: - mismatched and perfect hairpins, free energy of about 10 kcal/mol, the loop up to 8 nucleotides; - perfect hairpins, free energy over 10 kcal/mol; - perfect hairpins, free energy over 20 kcal/mol; - mismatched hairpins, free energy over 10 kcal/mol; - mismatched hairpins, free energy over 20 kcal/mol

RNA molecules are known as the most labile macromolecules present in cells. The level of mRNA is regulated at the stages of synthesis and degradation. The stability of mRNA is determined by the combination of *trans*- and *cis*-factors, the former including exo- and endoribonucleases, and the latter being divided into two classes – stabilizers and destabilizers. Destabilizing elements provide binding of nucleases and initiation of degradation processes, while stabilizers prevent degradation of mRNA, blocking the action of different nucleases. One of the possible mechanisms of mRNA stabilization is related to the formation of non-canonical (hairpin) structure at 3'- or 5'-end of mRNA. For instance, the formation of such hairpin at 5'-end stabilizes mRNA of *Escherichia coli* through prevention of interaction between 5'-end of mRNA and RNAse E [23].

Besides, hairpin structures may often appear inside the inner transcription terminators *i.e.* specific fragments of DNA template, where an elongation complex of RNA polymerase – DNA template – RNA trascript stops and usually dissociates. Though it is true that many bacteria (anthrax agent among them) seldom contain classic transcription terminators, it may be supposed that hairpin-loop structures found are formed to provide regulation at the level of transcription.

Thus, the analysis of sequences of the *pXO1* and *pXO2* plasmids of *B. anthracis* allowed to propose the physical maps of plasmids with located perfect and mismatched inverted repeats, potentially capable of forming thermodynamically stable hairpin-loop structures. The length of highly stable hairpins ranges from 66 to 19 nucleotides, their free energy being -30.6 to -10.3 kcal/mol.

The majority of hairpins, defined at the physical maps of *pXO1* and *pXO2* plasmids, are located either in the area of positive trans-regulators or in the elements of unknown function. We suppose that the location of hairpin structures in the genome of *pXO1* and *pXO2* plasmids is not random; similar to long homopurine tracts, potentially capable of forming triplexes, they may be situated within promoters and terminators of transcription as well as near "hot" spots of recombination [24].

The work was supported by grants of AMS 47/2002 and AMS 72/2007, the Academy of Medical Sciences of Ukraine.

О. Ю. Лиманська, О. П. Лиманський

Розподіл шпилькових структур у плазмідах збудника сибірської виразки

Резюме

Однією з важливих біологічних функцій шпилькових структур є захист РНК-транскриптів від деградувальної дії різних факторів, а також регуляція транскрипції за рахунок їхнього формування у термінаторах транскрипції. Проведено пошук та визначено розподіли термодинамічно стабільних досконалих і недосконалих інвертованих повторів у плазмідах pXO1 і pXO2 патогенних штамів Bacillus anthracis. Аналіз послідовностей плазмід pXO1 і pXO2 B. anthracis виявив, що перша містить 176 інвертованих послідовностей з енергією від –30,6 до –10,0 ккал/моль, а друга – 57 шпильок з енергією від –27,2 до –10,0 ккал/моль. Представлено фізичні карти плазмід pXO1 і pXO2 з локалізованими шпильковими структурами. Показано, що останні на фізичних картах плазмід pXO1 і pXO2 розташовані в ділянці регуляторних генів або в елементах з невизначеною функцією.

Ключові слова: Bacillus anthracis, шпилькова структура, інвертований повтор, хрестоподібна структура.

О. Ю. Лиманская, А. П. Лиманский

Распределение шпилечных структур в плазмидах возбудителя сибирской язвы

Резюме

Одной из важных биологических функций шпилечных структур является защита РНК-транскриптов от деградирующего действия разных факторов, а также регуляция транскрипции за счет их формирования в терминаторах транскрипции. Проведен поиск и определены распределения термодинамически стабильных совершенных и несовершенных инвертированных повторов в плазмидах pXO1 и pXO2 патогенных штаммов Bacillus anthracis. Анализ последовательностей плазмид pXO1 и pXO2 B. anthracis выявил, что первая содержит 176 инвертированных последовательностей с энергией от –30,6 до –10,0 ккал/моль, а вторая – 57 шпилек с энергией от –27,2 до –10,0 ккал/моль. Представлены физические карты плазмид pXO1 и pXO2 с локализованными шпилечными структурами. Показано, что последние на физических картах плазмид pXO1 и pXO2 локализованы в области регуляторных генов или в элементах с неопределенной функцией.

Ключевые слова: Bacillus anthracis, шпилечная структура, инвертированный повтор, крестообразная структура.

REFERENCES

1. *Проблемы* и перспективы молекулярной генетики.–М.: Наука, 2004.–Т. 2.–334 с.
2. *Subramanian A., Tamayo P., Mootha V. K., Mukherjee S., Ebert B. L., Gillette M. A., Paulovich A., Pomeroy S. L., Golub T. R., Lander E. S., Mesirov J. P.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles // Proc. Nat. Acad. Sci. USA.–2005.–**102**.–P. 15545–15550.
3. *Katayama S., Tomaru Y., Kasukawa T., Waki K., Nakanishi M., Nakamura M., Nishida H., Yap C. C., Suzuki M., Kawai J., Suzuki H., Carninci P., Hayashizaki Y., Wells C., Frith M., Ravasi T., Pang K. C., Hallinan J., Mattick J., Hume D. A., Lipovich L., Batalov S., Engstrom P. G., Mizuno Y., Faghihi M. A., Sandelin A., Chalk A. M., Mottagui-Tabar S., Liang Z., Lenhard B., Wahlestedt C.* Antisense transcription in the mammalian transcriptome // Science.–2005.–**309**.–P. 1564–1566.
4. *Chunsun R., Kyunghee L., Cheonkwon Y., Won K. S., Hee-Bok O.* Sensitive and rapid quantitative detection of anthrax spores isolated from soil samples by real-time PCR // Microbiol. Immunol.–2003.–**47**.–P. 693–699.
5. *Makino S. I., Cheun H. I., Watarai M., Uchida I., Takeshi K.* Detection of anthrax spores from the air by real-time PCR // Lett. Appl. Microbiol.–2001.–**33**.–P. 237–240.
6. *Ellerbrok H., Nattermann H., Ozel M., Beutin L., Appel B., Pauli G.* Rapid and sensitive identification of pathogenic and apathogenic *Bacillus anthracis* by real-time PCR // FEMS Microbiol. Lett.–2002.–**214**.–P. 51–59.
7. *Deutscher M., Li Z.* Exoribonucleases and their multiple role in RNA metabolizm // Progr. Nucl. Acids Res. Mol. Biol.–2001.–**66**.–P. 67–105.
8. *Grunberg-Manago M.* Messenger RNA stability and its role in control of gene expression in bacteria and phages // Annu. Rev. Genet.–1999.–**33**.–P. 193–227.
9. *Лиманский А. П., Лиманская О. Ю., Волянский Ю. Л.* Компьютерный анализ инвертированных повторов в геноме микобактерий туберкулеза // Журн. микробиологии, эпидемиологии и иммунобиологии.–2004.–№ 5.–С. 48–52.
10. *Hacker J., Kaper J.* Pathogenicity islands and the evolution of microbes // Annu. Rev. Microbiol.–2000.–**54**.–P. 641–679.
11. *Rychlik W., Spencer W., Rhoads R.* Optimization of the annealing temperature for DNA amplification *in vitro* // Nucl. Acids Res.–1990.–**18**.–P. 6409–6417.
12. *Бродский Л. И., Драчев А. Л., Татузов Р. Л., Чумаков К. М.* Пакет прикладных программ для анализа последовательностей биополимеров: GeneBee // Биополимеры и клетка.–1991.–**7**.–С. 10–14.
13. *Lilley D.* Hairpin-loop formation by inverted repeats in supercoiled DNA molecules // Proc. Nat. Acad. Sci. USA.–1980.–**77**.–P. 6468–6472.
14. *Lyamichev V., Panyutin I., Mirkin S.* The absence of cruciform structures from *pAO3* plasmid DNA *in vivo* // J. Biomol. Struct. and Dyn.–1984.–**2**.–P. 291–301.
15. *Sinden R., Pettijohn D.* Cruciform transitions in DNA // J. Biol. Chem.– 1984.–**259**.–P. 6593–6600.
16. *Kashlev M., Komissarova N.* Transcription termination: primary intermediates and secondary adducts // J. Biol. Chem.–2002.–**277**.–P. 14501–14508.
17. *Okinaka R. T., Cloud K., Hampton O., Hoffmaster A. R., Hill K. K., Keim P., Koehler T. M., Lamke G., Kumano S., Mahillon J., Manter D., Martinez Y., Ricke D., Svensson R., Jackson P. J.* Sequence and organization of *pXO1*, the large *Bacillus anthracis* plasmid harborning the anthrax toxin genes // J. Bacteriol.–1999.–**181**.–P. 6509–6515.

18. *Panyutin I., Lyamichev V., Lyubchenko Y.* A sharp structural transition in *pAO3* plasmid DNA caused by increased superhelix density // FEBS Lett.–1982.–**148**.–P. 297–301.
19. *Panyutin I., Klishko V., Lyamichev V*. Kinetics of cruciform formation and stability of cruciform structure in superhelical DNA // J. Biomol. Struct. and Dyn.–1984.–**1**.–P. 1311–1324.
20. *Vologodskii A*. Formation of unusual structures in the supercoiled DNA. Influence of transitions // Mol. Biol.–1988.–**22**.–P. 687–692.
21. *Nasar F., Jankowski C., Nag D.* Long palindromic sequences induce double-stranded breaks during meiosis in yeast // Mol. and Cell. Biol.–2000.–**20**.–P. 3449–3458.
22. *Лиманський О. П., Лиманська О. Ю*. Вивчення геномної ДНК мікроорганізмів методом атомно-силової мікроскопії // Цитология и генетика.–2002 –**36**.–С. 30–36.
23. *Unniraman S., Chatterji M., Nagaraja V.* A hairpin near the 5' end stabilises the DNA gyrase mRNA in *Mycobacterium smegmatis* // Nucl. Acids Res.–2002.–**30**.–P. 5376–5381.
24. *Katayama T., Inoue N., Torigoe H*. Location of the triplex DNA-binding domain of *Saccaromyces cerevisiae* Stm1 protein // Nucl. Acids Res.–2007.–**35**.–P. 123–124.

UDC 577.2:573.6:616-0227
Received 30.05.08