

UDC 577.21.5

# Phylogenetic study on structural elements of HIV-1 poly(A) region. 1. PolyA and DSE hairpins

M. I. Zarudnaya, A. L. Potyahaylo, I. M. Kolomiets, D. M. Hovorun

Institute of Molecular Biology and Genetics, NAS of Ukraine  
150, Akademika Zabolotnoho Str., Kyiv, Ukraine, 03680

m.i.zarudna@imbg.org.ua

---

*Genome of human immunodeficiency virus type 1 (HIV-1) is highly heterogeneous. Aim. A phylogenetic study on structural elements of the HIV-1 poly(A) region, in particular polyA and DSE hairpins which compose a core poly(A) site. Methods. The secondary structure of the HIV-1 core poly(A) site has been predicted by the UNAFold program. Results. The structure of the polyA and DSE hairpins has been analysed in 1679 HIV-1 genomes of group M and 18 genomes of simian immunodeficiency virus SIVcpzPtt. We found 244 and 171 different sequences for the HIV-1 polyA and DSE hairpins, respectively. However 70 % of the HIV-1 isolates studied contain one of 7 variants of the polyA hairpin which occur with a frequency  $\geq 5$  % (main variants) and 79 % of the isolates contain one of 7 main variants of the DSE hairpin. We also revealed subtype and country specific mutations in these hairpins. We found that the SIV polyA hairpin most closely resembles that found in HIV-1 genomes of B/C subtypes. Conclusions. The results of our large-scale phylogenetic study support some structural models of the HIV-1 5' UTR, in particular the tertiary interaction between the polyA hairpin and the matrix region in HIV-1 gRNA. Possibly, the DSE hairpin appeared in the course of viral evolution of the HIV-1 group M. An exposure of the U/GU-rich element in the apical loop of DSE hairpin could significantly increase the efficiency of pre-mRNA polyadenylation in this HIV-1 group.*

*Keywords: HIV-1, SIVcpzPtt, poly(A) region, secondary structure, polyA hairpin, DSE hairpin.*

---

**Introduction.** Polyadenylation of the pre-mRNAs in mammals and their viruses depends on at least two sequence elements: the poly(A) signal (most often the AAUAAA hexamer) and the U/GU-rich downstream sequence element (DSE). These elements compose the core poly(A) site. Multiple protein factors assemble onto this site, in particular, the cleavage and polyadenylation specificity factor (CPSF) binds to the AAUAAA hexamer and the cleavage stimulation factor (CstF) binds to the U/GU-rich DSE [1, 2]. The efficiency of polyadenylation process can be modulated by upstream sequence elements (USEs) or/and auxiliary downstream sequence elements (AuxDSEs) [3]. In the HIV-1 retrovirus, the identical sequences encompassing AAUAAA and DSE are present at both the 5' and 3' ends of the HIV-1 pre-mRNA and a strict regulation is needed to re-

press the premature polyadenylation at the 5' end of the transcript and stimulate the reaction at the 3' end. In particular, usage of the 3' poly(A) site is promoted by the USEs that are present exclusively at the 3' end of the transcript [4].

In HIV-1 pre-mRNA, the AAUAAA hexamer is partly occluded by base pairing in the upper part of the polyA hairpin [5], the stability of which is delicately balanced to allow the regulation of polyadenylation reaction at the both ends [6]. Two putative DSEs, UGUGU and GUUGUGU, are located 6 and 19 nt downstream of the cleavage site, respectively. Recently we have first presented a structural model for the core poly(A) site at the 3' end of the HIV-1 pre-mRNA [7] (Fig 1). The tracts interacting with the polyadenylation factors as well as other functionally important elements are indicated in this figure. The other elements are functional at the 5' end of HIV-1 genomic RNA (gRNA), inasmuch as they



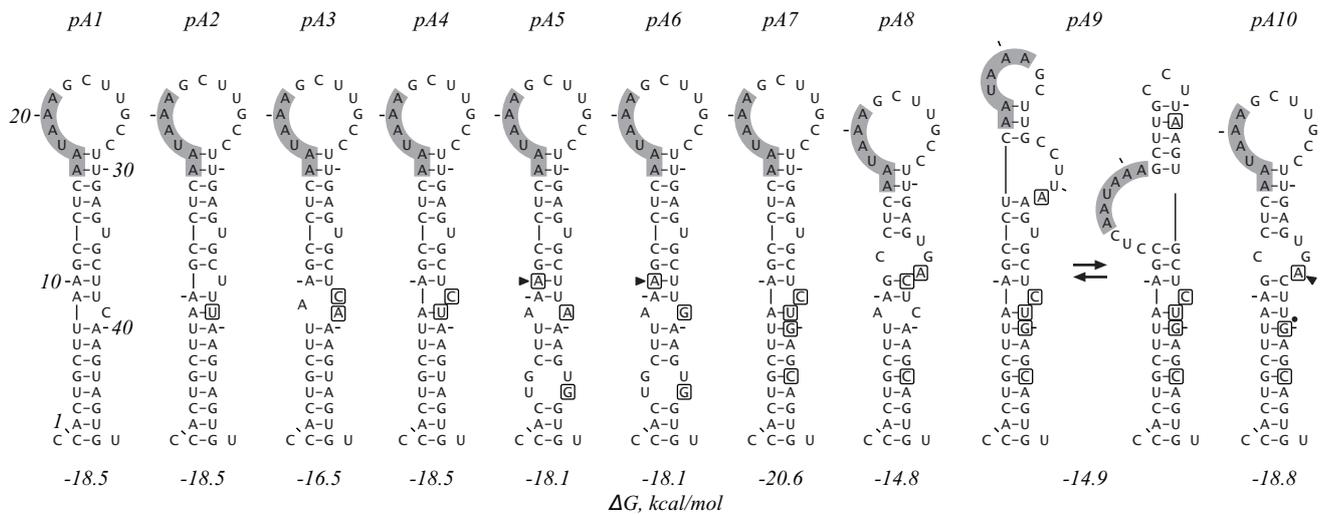


Fig. 2. Optimal structures of polyA hairpin variants in HIV-1 and SIVcpzPtt pre-mRNAs. The base changes as compared to RefSec are squared, insertions and deletions are indicated by triangle and spot, respectively. The AAUAAA hexamer is shadowed. pA1–pA7 – main variants of the polyA hairpin in HIV-1 isolates of group M; pA8 – rather frequent variant in subtype A isolates from Tanzania; pA9 – variant with the G31A base change; pA10 – polyA hairpin variant in SIVcpzPtt isolates

Table 1  
Occurrence of polyA hairpin variants in HIV-1 isolates of different subtypes (%)

Variant	Base changes	Subtype				
		A	B	C	CRF01_AE	Others
		Number of HIV-1 isolates (number of patients)				
		141 (122)	645 (258)	291 (234)	162 (61)	440 (322)
pA1	No	17	50	2	0	42
pA2	C39U	32	27	8	0	10
pA3	U38C + C39A	0	6	1	0	0.4
pA4	U38C + C39U	3	1	40	0	8
pA5	10_11ins A + C39A + A44G	0	0.1	0	46	1
pA6	10_11ins A + C39G + A44G	0	0	0	6	0
pA7	U38C + C39U + A40G + U43C	0	0.1	15	0	2
<b>Total</b>	–	<b>52</b>	<b>84</b>	<b>67</b>	<b>52</b>	<b>63</b>

in Table 2 have one of the main polyA hairpin variants such as pA4 (42 %), pA2 (8 %) and pA7 (4 %). The polyA hairpins in the rest of isolates can be considered as these three main variants with one or two base changes which occur with different frequencies in the HIV-1 isolates studied.

For example, A32G (GenBank acc. no. AM076852) occurs with a frequency below the error of sequencing and submission of HIV-1 genomic sequence into Gen Bank (0.5 % [12]), while A44G (AM076846) is very

frequent in HIV-1 isolates of subtype CRF01\_AE but is random in isolates of subtype C. Precisely rare and random base changes lead to a great variety in sequences of the polyA hairpin. The rare and random mutations presented in Table 2 did not affect greatly the overall structure of the polyA hairpin, except for G31A (AM076874). This base change occurs singly or in the combination with different mutations with a frequency of 2.1 % in the HIV-1 isolates studied. In most cases, it results in two alternative conformations of the hairpin with the

*Table 2*  
*Base changes in polyA hairpin of HIV-1 subtype C isolates from Mozambique*

Accession number	Patient	Base changes in polyA hairpin											PolyA hairpin			
		Position														
		7	10	31	32	36	37	38	39	40	43	44				
AM076840	PG174		A10G					U38A	C39U							
AM076842	PG174		A10G					U38A	C39U							
AM076844	PG81							U38C	C39U							pA4
AM076846	PG56							U38C	C39U				A44G			
AM076848	PG5							U38C	C39U							pA4
AM076850	PG5							U38C	C39U							pA4
AM076852	PG6				A32G				C39U							
AM076854	PG184							U38C	C39U							pA4
AM076856	PG196							U38C	C39U							pA4
AM076858	PG182							U38C	C39U	A40G						
AM076860	PG182							U38C	C39U	A40G						
AM076862	PG188							U38C	C39U							pA4
AM076864	PG189					C36U		U38C	C39U	A40G	U43C					
AM076866	PG203							U38C	C39U	A40G						
AM076868	PG215								C39U							pA2
AM076870	PG224							U38C	C39U							pA4
AM076872	PG227							U38C	C39U	A40G	U43C					pA7
AM076874	PG227			G31A				U38C	C39U	A40G	U43C					
AM076876	PG227	U7C				C36U		U38C	C39U	A40G	U43C					
AM076878	PG230					C36A	U37C		C39U							
AM076880	PG231							U38C	C39U							pA4
AM076882	PG177								C39U							pA2
AM076884	PG178							U38C	C39U							pA4
AM076886	PG217							U38C	C39U							pA4
AM076888	PG218					C36U		U38C	C39U							
AM076891	PG36							U38C	C39U							pA4

same free energy and complete exposure of the AAUA AA hexamer in one of these conformations (for example, see pA9 in Fig. 2).

As seen in Table 2, the double mutation U38C + C39U (both as the pA4 variant and in combination with other base changes) is prevailing among the HIV-

1 isolates from Mozambique (77%). The availability of a certain motif containing one or several frequent base changes in combination with rare and/or random mutations is also characteristic of the polyA hairpins in HIV-1 isolates of other subtypes. It is also a peculiarity of other structural elements in the HIV-1 polyA region.

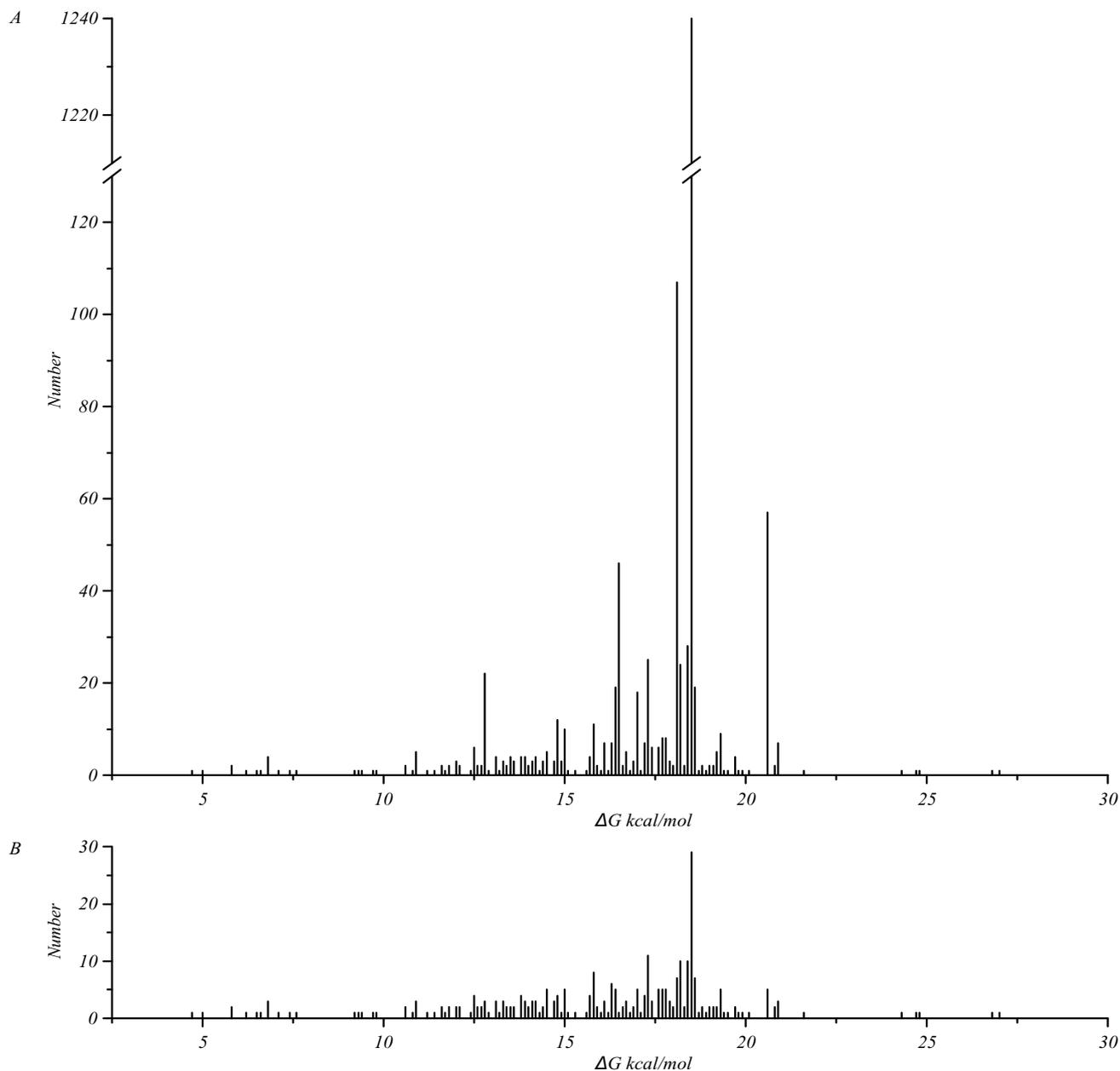


Fig. 3. Free energy distribution of HIV-1 polyA hairpin variants: *A* – all 1863 polyA hairpins; *B* – 286 HIV-1 polyA hairpins with different sequences (see Section PolyA hairpin)

In general the base changes which led to significant alterations in the polyA hairpin structure occurred only in 4 % of all HIV-1 genomes studied. The structures of the polyA region in the HIV-1 isolates studied are presented in our database CESSHIV-1 which is currently available online at <http://www.cesshiv1.org>. Base change frequency at each position of the polyA hairpin is presented in Suppl. information (Table S1). Rather frequent mu-

tations in this element are an insertion between positions 10, 11 and base changes at positions 37–40, 43, and 44.

HIV-1 gRNA contains several strong binding sites for multifunctional virion infectivity factor (Vif) including the polyA hairpin [13]. In particular, Vif-RNA interaction is important for viral particle assembly. Vif specifically binds to the stem of polyA hairpin, however structural and sequence determinants of this binding

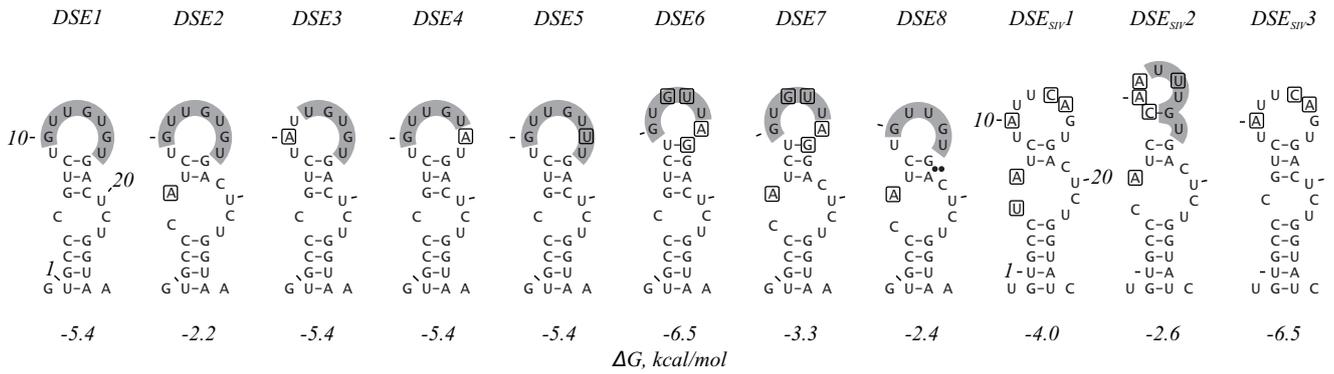


Fig. 4. Optimal structures of DSE hairpin variants in HIV-1 and SIVcpzPtt pre-mRNAs. The base changes as compared to RefSec are squared, deletions are indicated by spot. The U/GU-rich DSE signal is shadowed. DSE1–DSE7 – main variants of the DSE hairpin in HIV-1 isolates of group M (for DSE2 a suboptimal structure with free energy difference of –0.9 kcal/mol is shown); DSE8 – frequent variant in subtype D isolates; DSE<sub>SIV1</sub>–DSE<sub>SIV3</sub> – DSE-like hairpin variants in SIVcpzPtt isolates

Table 3  
Occurrence of DSE hairpin variants in HIV-1 isolates of different subtypes (%)

Variant	Base changes	Subtype				
		A	B	C	CRF01_AE	Others
		Number of HIV-1 isolates (number of patients)				
		141 (122)	645 (258)	291 (234)	162 (61)	440 (322)
DSE1	RS	30.5	76.0	65.3	0	58.6
DSE2	G6A	3.5	7.3	6.5	0	7.7
DSE3	G10A	0	2.2	5.5	0	1.8
DSE4	G15A	36.9	0.5	1.0	0	3.9
DSE5	G15U	7.1	0.1	0.7	0	3.6
DSE6	U12G + G13U + G15A + U16G	0	0	0	43.8	0.9
DSE7	G6A + U12G + G13U + G15A + U16G	0	0	0	13.6	0
<b>Total</b>	–	<b>78</b>	<b>86</b>	<b>79</b>	<b>57</b>	<b>76</b>

are not defined. Various defects in the stem of polyA hairpin (Fig. 2) can modulate its binding to Vif.

Base changes in the AAUAAA hexamer and the neighbouring GCUUGCC tract occur with a frequency below the error level, except for a position 28 in the polyA hairpin at which mutations occur with a frequency of 0.7 %. We have also shown high conservation of the sequence GGCAAGC in the matrix region of HIV-1 gRNA for a large pool of HIV-1 isolates. Thus, our data support a statement about the long distance interaction between the GCUUGCC and GGCAAGC sequences important for the tertiary structure of the 5' untranslated region of HIV-1 gRNA [8].

As mentioned in Introduction, a balanced stability of the polyA hairpin is crucial for HIV-1 replication [6, 14]. Both destabilization and stabilization of the wild type hairpin inhibits polyadenylation of HIV-1 pre-mRNA at the 3' end of the transcript. Stabilization of this hairpin by 10.4 kcal/mol blocked the access of polyadenylation factors to the hexamer and led to complete loss of polyadenylation at the 3' end of HIV-1 pre-mRNA [6, 14]. Destabilization by 8.5 kcal/mol increased the efficiency of premature polyadenylation at the 5' end of HIV-1 transcript from 5–10 % to 30–40 % [14], that affects less severely polyadenylation at the 3' end than stabilization of the polyA hairpin. It was

shown by the method of virus evolution that HIV-1 mutants with stabilized or destabilized polyA hairpin have improved their replication capacity via drifting towards a hairpin of thermodynamic stability that is close to the value of the wild type hairpin [15].

Free energy distribution of the HIV-1 polyA hairpins is presented in Fig. 3, A. Here we have analysed the HIV-1 genomes which contain both the complete poly(A) region (see Materials and methods) and incomplete regions encompassing a polyA hairpin sequence (see CESSHIV-1 database). In sum we have analyzed 1863 genomes (from 1072 patients) which contain 286 different sequences of the polyA hairpin. For two patients with multiple identical sequences of the polyA hairpin, we considered for analysis only one inpatient sequence.

The major peak in Fig. 3, A, corresponds to the HIV-1 isolates containing the polyA hairpins with free energy (dG) of  $-18.5$  kcal/mol. The main variants of the polyA hairpin pA1, pA2 and pA4 make a major contribution to this peak (92 %), while other main variants contribute to minor peaks at  $-16.5$ ,  $-18.1$  and  $-20.6$  kcal/mol. The minor peak at  $-12.8$  kcal/mol corresponds mostly to the polyA hairpins with the base change G31A and double mutation G31A + C39U.

Free energy distribution of the 286 HIV-1 polyA hairpins with different sequences is shown in Fig. 3, B. In general both distributions are similar. The main peak in Fig. 3, B, also corresponds to  $-18.5$  kcal/mol. A maximal number of different sequences of the polyA hairpin (29 ones) contribute to this peak. Both distributions are skewed, in particular the number of HIV-1 isolates with the polyA hairpins with  $dG > -18.5$  kcal/mol (26.9 %) significantly exceeds that with  $dG < -18.5$  kcal/mol (6.5 %).

Free energy distributions of the HIV-1 polyA hairpins for different subtypes (A, B, C, D, G, CRF01\_AE and CRF02\_AG) are similar to those presented in Fig. 3, A. All distributions are skewed and have the maximum at  $-18.5$  kcal/mol, except that the plot for CRF01\_AE has a maximum at  $-18.1$  kcal/mol. Thus, a free energy of the polyA hairpin lies within the range of  $-10.9 \div -20.9$  kcal/mol in most HIV-1 isolates (98.5 %), and the polyA hairpin with dG of  $-18.5$  kcal/mol is found in 66.6 % isolates.

To gain insight into the evolution of frequent mutations in HIV-1 polyA hairpin we studied a structure of the core poly(A) site in 18 isolates of SIVcpzPtt that is

an ancestor of HIV-1 group M [16]. Similarity of the polyA hairpin in SIVcpzPtt and HIV-1 was first reported by Berkhout *et al.* [5]. The SIV polyA hairpin presented in their article possesses the upper part identical to that in HIV-1 HXB2 genome and a stem with two bulges  $0 \times 2$  and  $0 \times 1$ . In our study we found 14 different sequences of the SIV polyA hairpin, including the sequences identical to the HIV-1 polyA hairpin variants pA1, pA2 and pA4. All SIV isolates have the polyA hairpin with the common upper part, except for three isolates in which the apical loop is elongated by 2 nucleotides (nt) thus exposing 5 nt instead of 4 nt of the AAUAAA hexamer.

The combination of base changes 35\_36insA + C39del + A40G + U43C can be considered as a frequent motif of the SIV polyA hairpin (Supplementary information, Table S2). Three SIV isolates have the polyA hairpin containing exclusively this motif (pA10, Fig. 2). The double mutation 35\_36insA + C39del was found in three HIV-1 polyA hairpins (two isolates of subtype C and one of subtype B) and A40G + U43C is a constituent of the combination of base changes in the HIV-1 polyA variant pA7 specific for subtype C isolates.

*DSE hairpin.* The variants of HIV-1 DSE hairpin with combinations of base changes occurring with a frequency  $\geq 5$  % (the main variants, DSE1-DSE7) are shown in Fig. 4. Their distribution by subtype is given in Table 3. The variants DSE1-DSE3 are found in HIV-1 isolates of different subtypes, while the variants DSE4-DSE7 occur predominantly in HIV-1 isolates of certain subtype. HIV-1 isolates of subtypes D, F and G commonly have the DSE hairpin without mutations (DSE1), except that the variant with the base changes G6A + U16del + G17del (DSE8, Fig. 4) occurs rather frequently in HIV-1 isolates of subtype D (19 %). About 80 % of the HIV-1 isolates studied have one of the main DSE hairpin variants.

The apical loop is the most variable region of the DSE hairpin (Supplementary information, Table S3). In the U/GU-rich tract, frequent base changes occur at positions 12, 13, 15 and 16. The combination of these four base changes, which is specific for CRF01\_AE, does not impair DSE signal (DSE6, Fig. 4). According to our description of the mammalian DSE region [17], it is composed of certain U/GU-rich pentamers (in particular the GUUGU, UGUGU or GUGUU tracts) which are located at different distances from each other. All the-

se pentamers are found in the apical loop of the main variants of the DSE hairpin as single or overlapping tracts. In DSE5, the GUUGU pentamer overlaps the UGUUU tract which is a U-rich (URE) DSE signal of the type «a four out of five base URE» [18]. In some HIV-1 isolates of CRF01\_AE, the mutations at positions 12 and/or 13 impair DSE signal by preventing a complete exposure of the U/GU-rich tract, which can decrease an efficiency of the polyadenylation process. It concerns mainly the group of 46 HIV-1 isolates from 3 patients, which possess the DSE hairpin with a combination of base changes G6A + C8U + U12del + G15C + U16A + 17G18.

The sequences of both strands of the DSE hairpin stem are well conserved. The rather frequent base change G6A destabilizes the hairpin and leads to partial occlusion of the U/GU-rich tract in the optimal structure; however this tract is completely exposed in a the suboptimal structure with close free energy (DSE2, Fig. 4). Rare mutations (4 %) occur at position 17 (deletion and G17A base change) and between positions 17 and 18 (insertion). The deletion occurs in 39 % of HIV-1 isolates of subtype D and does not affect DSE signal exposure (DSE8, Fig. 4). The insertion 17G18 in combinations with some base changes impairs DSE signal exposure, mainly in the above mentioned group of 46 isolates of CRF01\_AE.

The 5' strand of the DSE hairpin together with two neighboring residues (tract 2 in Fig. 1) correspond to the 5' strand of the U5-AUG duplex formed at the 5' end of HIV-1 gRNA. The mutation G6A in the DSE hairpin corresponds to G9651A in HXB2 genome resulting in a G-U to A-U base pair substitution in the duplex. We have shown that the 3' strand of this duplex is also well conserved in a large pool of HIV-1 isolates, which supports the formation of the U5-AUG duplex. The conservation of tract 3 encompassing PAS signal supports its functional importance, including tRNA<sub>3</sub><sup>Lys</sup>-PAS interaction.

In the SIV isolates studied, we have not found a hairpin similar to the HIV-1 DSE hairpin as distinct from the polyA hairpin occurred in both HIV-1 and SIV. However a hairpin with the bottom duplex like in the HIV-1 DSE hairpin (frequently with an additional base pair G-U) was found in almost all SIV isolates (DSE<sub>SIV</sub>1-DSE<sub>SIV</sub>3, Fig. 4). The main difference between the HIV-1 DSE hairpin and its SIV analogue is the absence of the U/GU-rich tract in the DSE<sub>SIV</sub> apical loop. Only

one SIV hairpin has such tract but not completely exposed (DSE<sub>SIV</sub>2, Fig. 4). Probably, SIV isolates use UGUGU signal just downstream of the polyA hairpin (Fig. 1) that is partly occluded.

The strains of HIV-1 are classified into four groups: M, N, O and P which are of chimpanzee or gorilla origin [16]. We have not found any DSE hairpin exposing the U/GU-rich tract in either HIV-1 groups N, O and P (36 isolates) or SIV from gorilla (5 isolates). We hypothesize that the U/GU-rich tract exposed in the apical loop of DSE hairpin has been acquired in HIV-1 group M in the course of evolution. At present it is poorly understood why only group M, but not other HIV-1 groups, resulted in a global pandemic [16]. An effective DSE may be one of the features that makes group M much more prevalent than groups N, O and P.

*М. І. Зарудна, А. Л. Потягайло, І. М. Коломієць, Д. М. Говорун*

Філогенетичний аналіз структурних елементів в області полі(А) ВІЛ-1. 1. Шпильки поліА та DSE

Резюме

*Геном вірусу імунodefіциту людини (ВІЛ-1) надзвичайно гетерогенний. Мета.* Провести філогенетичний аналіз структурних елементів області полі(А) ВІЛ-1, зокрема, шпильок поліА і DSE, які складають основний сайт полі(А). *Методи.* Передбачення вторинної структури основного полі(А)-сайта здійснювали за допомогою програми UNAFold. *Результати.* Структуру шпильок поліА і DSE проаналізовано для 1679 геномів ВІЛ-1 групи М та 18 геномів вірусу імунodefіциту мавп SIVcpzPtt: у геномах ВІЛ-1 виявлено 244 і 171 різні послідовності шпильок поліА і DSE відповідно. Однак 70 % вивчених геномів містили один із семи варіантів шпильки поліА, які зустрічалися з частотою ≥5 % (основні варіанти), і 79 % ізолятів містять один із семи основних варіантів шпильки DSE. Виявлено також специфічні для певного субтипу або країни мутації зазначених шпильок. Встановлено, що шпилька поліА з геному SIVcpzPtt дуже схожа на таку з геному ВІЛ-1 субтипу В і С. *Висновки.* Результати наших широкомасштабних досліджень підтримують деякі структурні моделі 5' нетрансльованої ділянки геному ВІЛ-1, зокрема, третинну взаємодію між шпилькою поліА і областю гена матричного білка. Шпилька DSE, можливо, утворилася у процесі еволюції ВІЛ-1 групи М. Експонування U/GU-збагаченого елемента в апікальній петлі цієї шпильки могло значно підвищити ефективність поліаденілювання про-мРНК у ВІЛ-1 цієї групи.

*Ключові слова:* ВІЛ-1, SIVcpzPtt, область полі(А), вторинна структура, шпилька поліА, шпилька DSE.

*М. И. Зарудная, А. Л. Потягайло, И. Н. Коломиец, Д. Н. Говорун*

Филогенетический анализ структурных элементов в области поли(А) ВИЧ-1. 1. Шпильки полиА и DSE

Резюме

*Геном ВИЧ-1 чрезвычайно гетерогенен. Цель.* Провести филогенетический анализ структурных элементов области (поли)А ВИЧ-

1, в частности, шпильки полиА и DSE, составляющих основной поли(А)-сайт. **Методы.** Вторичная структура основного поли(А)-сайта предсказана с помощью программы UNA Fold. **Результаты.** Структура шпильки полиА и DSE проанализирована в 1679 геномах ВИЧ-1 группы М и 18 геномах вируса иммунодефицита обезьян SIVcpzPtt: в геномах ВИЧ-1 выявлены соответственно 244 и 171 различная последовательность шпильки полиА и DSE. Однако 70 % изученных геномов содержат один из семи вариантов шпильки полиА, встречающихся с частотой  $\geq 5$  % (основные варианты) и 79 % изолятов содержат один из семи основных вариантов шпильки DSE. Обнаружены также варианты шпильки, специфические для определенного субтипа или страны. Выявлено, что шпилька полиА в геноме SIVcpzPtt наиболее близка по структуре таковой в геноме ВИЧ-1 субтипов В и С. **Выводы.** Результаты наших широкомасштабных исследований свидетельствуют в пользу некоторых структурных моделей 5' нетранслируемой области генома ВИЧ-1, в частности, наличия третичного взаимодействия между шпилькой полиА и участком в гене матричного белка. Шпилька DSE, возможно, образовалась в процессе эволюции ВИЧ-1 группы М. Экспонирование U/GU- богатого элемента в апикальной петле шпильки DSE могло значить повысить эффективность полиаденилирования про-мРНК в ВИЧ-1 данной группы.

**Ключевые слова:** ВИЧ-1, SIVcpzPtt, область поли(А), вторичная структура, шпилька полиА, шпилька DSE.

## REFERENCES

- Zarudnaya M. I. mRNA polyadenylation. 1. 3'-end formation of vertebrates' mRNAs // *Biopolym. Cell.*—2001.—17, N 2.—P. 93–108.
- Chan S., Choi E. A., Shi Y. Pre-mRNA 3'-end processing complex assembly and function // *Wiley Interdiscip. Rev. RNA.*—2011.—2, N 3.—P. 321–335.
- Zarudnaya M. I., Potyahaylo A. L., Kolomiets I. M., Hovorun D. M. Auxiliary elements of mammalian pre-mRNAs polyadenylation signals // *Biopolym. Cell.*—2002.—18, N 6.—P. 500–517.
- Valsamakis A., Zeichner S., Carswell S., Alwine J. C. The human immunodeficiency virus type 1 polyadenylation signal: a 3' long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylation // *Proc. Natl Acad. Sci. USA.*—1991.—88, N 6.—P. 2108–2112.
- Berkhout B., Klaver B., Das A. T. A conserved hairpin structure predicted for the poly(A) signal of human and simian immunodeficiency viruses // *Virology.*—1995.—207, N 1.—P. 276–281.
- Klasens B. I., Thiesen M., Virtanen A., Berkhout B. The ability of the HIV-1 AAUAAA signal to bind polyadenylation factors is controlled by local RNA structure // *Nucleic Acids Res.*—1999.—27, N 2.—P. 446–454.
- Zarudnaya M. I., Potyahaylo A. L., Otenko V. V., Kolomiets I. N., Hovorun D. M. The secondary structure of core poly(A) signal of human immunodeficiency virus pre-mRNA // *Reports of the Nat. Acad. Sci. of Ukraine.*—2011.—N 4.—P. 170–176.
- Paillart J. C., Skripkin E., Ehresmann B., Ehresmann C., Marquet R. *In vitro* evidence for a long range pseudoknot in the 5'-untranslated and matrix coding regions of HIV-1 genomic RNA // *J. Biol. Chem.*—2002.—277, N 8.—P. 5995–6004.
- Abbink T. E., Berkhout B. A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon // *J. Biol. Chem.*—2003.—278, N 13.—P. 11601–11611.
- Ooms M., Cupac D., Abbink T. E., Huthoff H., Berkhout B. The availability of the primer activation signal (PAS) affects the efficiency of HIV-1 reverse transcription initiation // *Nucleic Acids Res.*—2007.—35, N 5.—P. 1649–1659.
- Markham N. R., Zuker M. UNAFold: software for nucleic acid folding and hybridization // *Methods Mol. Biol.*—2008.—453.—P. 3–31.
- Yuan Y., Kerwood D. J., Paoletti A. C., Shubsda M. F., Borer P. N. Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1 x 3 internal loop // *Biochemistry.*—2003.—42, N 18.—P. 5259–5269.
- Henriet S., Richer D., Bernacchi S., Decroly E., Vigne R., Ehresmann B., Ehresmann C., Paillart J. C., Marquet R. Cooperative and specific binding of Vif to the 5' region of HIV-1 genomic RNA // *J. Mol. Biol.*—2005.—354, N 1.—P. 55–72.
- Das A. T., Klaver B., Berkhout B. A hairpin structure in the R region of the human immunodeficiency virus type 1 RNA genome is instrumental in polyadenylation site selection // *J. Virol.*—1999.—73, N 1.—P. 81–91.
- Berkhout B. HIV-1 as RNA evolution machine // *RNA Biol.*—2011.—8, N 2.—P. 225–229.
- Sharp P. M., Hahn B. H. Origins of HIV and the AIDS pandemic // *Cold Spring Harb. Perspect. Med.*—2011.—1, N 1.—a006841.
- Zarudnaya M. I., Kolomiets I. M., Potyahaylo A. L., Hovorun D. M. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures // *Nucleic Acids Res.*—2003.—31, N 5.—P. 1375–1386.
- Chou Z.-F., Chen F., Wilusz J. Sequence and position requirements for uridylylate-rich downstream elements of polyadenylation signals // *Nucleic Acids Res.*—1994.—22, N 13.—P. 2525–2531.

Received 18.06.13