UDC 577.21.5

Phylogenetic study on structural elements of HIV-1 poly(A) region. 2. USE domain and TAR hairpin

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

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

m.i.zarudna@imbg.org.ua

The aim of this work was phylogenetic study on structural elements in the poly(A) region of human immunodeficiency virus type 1 (HIV-1), in particular the major upstream sequence element (USE), which stimulates polyadenylation of HIV-1 transcript, and the TAR (trans-activation response) hairpin, which juxtaposes spatially the AAUAAA and USE signals. **Methods**. The secondary structure of these elements has been predicted by UNA Fold program. **Results**. The structure of USE domain and TAR hairpin has been analysed in 1679 HIV-1 genomes and 17 genomes of simian immunodeficiency virus SIVcpzPtt. We found 376 and 588 different sequences for these elements, respectively, and revealed the most frequent base changes and subtype- and country-specific mutations. Only 43 % of HIV-1 isolates contain variants of the USE domain which occur with a frequency 5 % (the main variants) and 35 % of isolates contain main variants of the TAR hairpin. We found that the SIV USE domain and TAR hairpin most closely resemble those found in HIV-1 genomes of A/G-containing subtypes. **Conclusions**. The results of our large-scale phylogenetic study support a hypothesis on the interaction between $tRNA_3^{Lys}$ and the 3' end of HIV-1 genomic RNA and a controversial supposition of HIV-1 genome dimerization by the TAR-TAR kissing mechanism. Since the TAR hairpin is a target for developing antiviral drugs based on the inhibition of signal elements, the data on specific structural features of this hairpin may be useful for new antivirals design.

Keywords: HIV-1, SIVcpzPtt, poly(A) region, secondary structure, USE domain, TAR hairpin.

Introduction. The 5' and 3' untranslated regions (UTRs) of HIV-1 genome completely consist of structural elements directing multiple processes of viral replication, in particular, polyadenylation of HIV-1 pre-mRNA. In HIV-1, the identical sequences encompassing the AA UAAA hexamer and U/GU-rich downstream sequence element (DSE), which compose the core poly(A) site [1], are present at both the 5' and 3' ends of HIV-1 pre-mRNA and strict regulation is needed to repress premature polyadenylation at the 5' end of the transcript and stimulate the reaction at the 3' end. Repression is mediated in particular by the AuxDSE that is uniquely present at the 5' end of the transcript [2], whereas the use of the 3' poly(A) site is promoted by the major and minor USEs that are present exclusively at the 3' end [3, 4]. A partial occlusion of the AAUAAA hexamer by

base pairing in the upper part of the polyA hairpin allows the polyadenylation repression at the 5' end and activation at the 3' end [5].

The AAUAAA hexamer is a binding site for the cellular cleavage and polyadenylation specificity factor (CPSF). This factor can also bind HIV-1 major USE (CAGCUGCUUUUUGCCUGU) [6] probably via the interaction with its oligo(U) part [7]. The major USE is proposed to act as an initial binding site for CPSF which binds to the AAUAAA hexamer upon transient opening or «breathing» of the polyA hairpin [5]. The polyA hairpin and the major USE are juxtaposed spatially by the TAR hairpin [8]. The TAR and polyA hairpins are formed at both ends of HIV-1 pre-mRNA. In the 5' terminal region, the TAR hairpin binds the viral Tat protein to activate transcription elongation [9], and in the 3' terminal region this hairpin can be considered as a structural element of the poly(A) region.
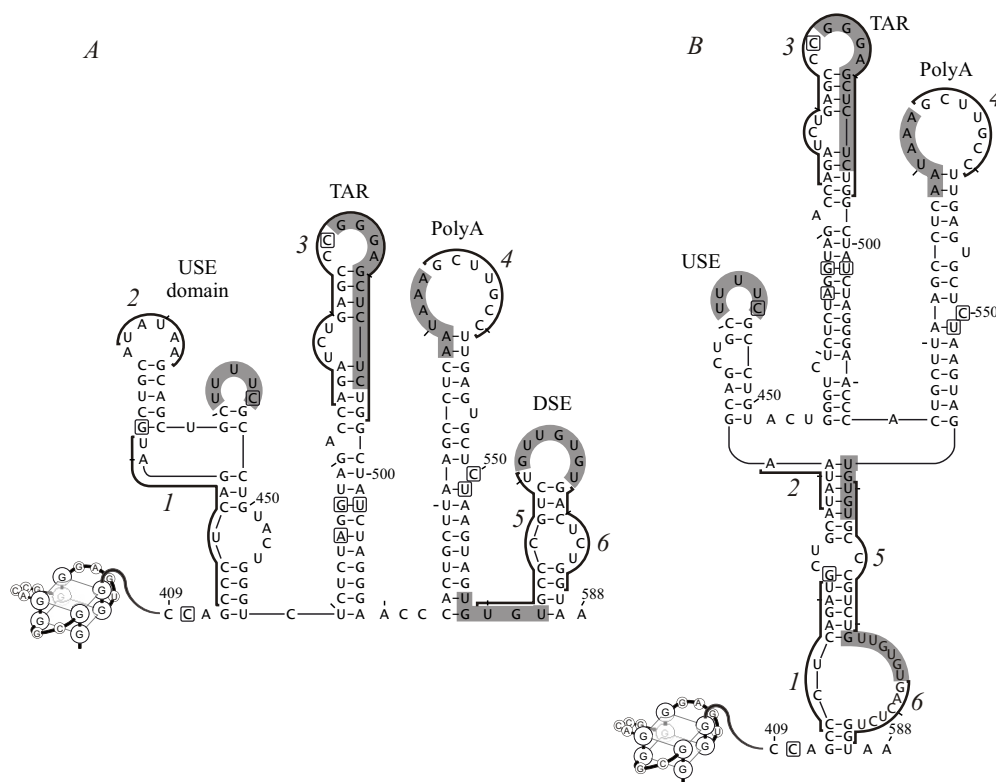
Fig. 1. The in-line (*A*) and domain (*B*) conformations of the complete 3' poly(A) region for HIV-1 isolate 99ZACM9 (GenBank accession number AF411967) of subtype C. The base changes are squared. The U-rich tract of USE, palindrome in the upper part of the TAR hairpin, AAUAAA hexamer and U/GU-rich DSEs are shadowed. Tract *1* is a binding site for the tRNA$_3^{Lys}$ at the 3' end of HIV-1 gRNA [11]; *2* corresponds to the TATAA box in HIV-1 proviral DNA; *3* is a binding site for Tat protein [9]; *4* is involved in the long distance interaction with the matrix coding region [12]; *5* – the 5' strand of U5-AUG duplex [13]; *6* – a primer activation signal (PAS) [14]. Tracts *2–5* are functional at the 5' end of HIV-1 gRNA

Recently [10] we have first presented a structural model (in-line and domain conformations) for the complete 3' poly(A) region of HIV-1 pre-mRNA. Fig. 1 gives an example of this model for subtype C, the most prevalent HIV-1 strain in the world. The in-line structure (Fig. 1, *A*) includes a shortened TAR hairpin, polyA hairpin and new structural elements: the USE domain with the U-rich tract, exposed in the apical part of the truncated USE hairpin, and the DSE hairpin with the U/GU-rich tract exposed in its apical loop. The domain structure (Fig. 1, *B*) includes the USE, TAR and polyA hairpins in its upper part and the U/GU-rich tract exposed in the internal loop of the bottom duplex of this domain. The model also includes the minor USEs which according to our proposal [10] are structured into different G-quadruplexes and interact with heterogeneous ribonucleoprotein H (hnRNP H), that positively influences the polyadenylation process ([15] and refs therein).

The sequences interacting with polyadenylation factors and other functionally important elements are indicated in Fig. 1. Tracts *2–5* are functional at the 5' end of HIV-1 gRNA, inasmuch as they are located in the long terminal repeats of proviral DNA they are also duplicated at the 3' end of the transcript.

HIV-1 gRNA is highly heterogeneous in both the translated and untranslated regions. In particular, we found 376 and 588 different sequences for the USE domain and the TAR hairpin, respectively. These sequences contain up to seventeen base changes in comparison with the RefSec. In our previous article [16] we conducted phylogenetic study on the HIV-1 polyA and DSE hairpins. Here we investigated how mutations affect the secondary structure of the USE domain and the TAR hairpin and compared the structure of these elements in HIV-1 genomes and genomes of SIV of chimpanzee *Pan troglodytes troglodytes* (SIVcpz*Ptt*).

Materials and methods. The sequences encompassing the complete poly(A) region in HIV-1 and SIVcpz *Ptt* genomes have been extracted from the Entrez Nucleotide database of NCBI. We have examined all HIV-1 genomic sequences presented in this database by the end of 2010 and all corresponding genomic sequences from SIVcpz*Ptt* presented by the end of 2012. The secondary structure of the poly(A) region has been predicted by the UNAFold program [17]. The base changes in HIV-1 and SIVcpz*Ptt* pre-mRNA sequences were determined as compared to RefSec (GenBank accession number K03455). Nucleotide numbering starts

with 1 at the first nucleotide of each individual structural element.

Results and discussion. *USE domain*. We have shown earlier [10] that the USE domain is formed in foldings of the complete HIV-1 poly(A) region within 2 kcal/mol energy increment of the lowest free energy structure in 62–90 % of the isolates depending on subtype. The USE domains (dUSE) with combinations of base changes occurring with a frequency   5% (the main variants, dUSE1-dUSE7) are shown in Fig. 2. Their distribution by subtypes A, B, C and CRF01_AE, comprising large subpools, is given in Table 1. The total data on other subtypes and CRFs, comprising small subpools, are listed in the last column. As seen in Table 1, the main variants occur predominantly in HIV-1 isolates of certain subtypes, for example, dUSE1-dUSE3 – in subtype B isolates. The most frequent combinations of base changes in the USE domain of subtypes D and F isolates are G1A + C11G (48 %) and G1A + C11G + U27C + U34C (23 %), respectively. The variant of the USE domain dUSE8 found in 36 % of subtype G isolates is shown in Fig. 2.

As seen from Fig. 2, all main USE domain variants possess a similar upper part with two small hairpins and a stem adopting two different conformations. The first short hairpin of 16 nt length has the AUAUAA apical loop exposing a sequence corresponding to the TATAA promoter element and the second short hairpin contains the U-rich tract in its apical loop. Of note, both hairpins are presented in the model of an entire HIV-1 genome based on SHAPE technology [18]. The whole USE domain has not been predicted in this model. The folding of the USE domain is determined by the formation of G-quadruplexes in the region of HIV-1 genomic RNA upstream of this domain and most likely occurs in HIV-1 pre-mRNA during polyadenylation in the presence of hnRNP H protein.

Base changes in HIV-1 USE domain occur mainly in the single-stranded stretches that do not alter an overall structure. The base changes located in the middle part of these stretches do not affect a free energy (dG) of the domain (compare dUSE1 and dUSE2, Fig. 2), while the base changes adjacent to double-stranded regions slightly affect dG (compare dUSE1 and dUSE4, Fig. 2).

The variant of USE domain with the double mutation C11G + G14A occurs with a frequency of 6.5 %

in HIV-1 isolates of subtype B (from USA, India and France). This mutation leads to break of the upper G-C base pair in the stem of the first short hairpin and to the formation of a USE domain-like structure exposing the U-rich tract with low stability (–9.5 kcal/mol in comparison with –14.8 kcal/mol for dUSE1 variant of the domain). It is likely that the USE signal in these isolates is functional predominantly in the domain conformation of the complete poly(A) region with the USE hairpin (Fig. 1, *B*).

Though the USE domain has the same length as the polyA hairpin (47 nt), it is significantly more heterogeneous. Only 43 % of the HIV-1 isolates studied contain one of the main variants of the USE domain, while about 70 % have one of the main variants of polyA hairpin as shown in our previous article [16]. In the rest of HIV-1 isolates, we considered diffrent sequences of the polyA hairpin as certain main variants of this hairpin containing rare and/or random mutations. This approach can be applied to all structural elements of the HIV-1 poly(A) region, including the USE domain and TAR hairpin. The structures of complete poly(A) region in the HIV-1 isolates studied in this work are presented in our database CESSHIV-1 currently available online at http://www.cesshiv1.org.

Base change frequency at each position of the HIV-1 USE domain is presented in Suplementary information, Table S1. Mutations in the region nt 413–421 (tract *1* in Fig. 1) corresponding to positions 2–10 of the USE domain (Fig. 2) occur with a frequency   1.5 %, except for positions 8 (4 %) and 10 (7 %). At these positions, the G8A and U10A/C are the most frequent base changes. The G8A (dUSE9, Fig. 2) and U10A (dUSE8, Fig. 2) were found in isolates of different subtypes, while the U10C (dUSE10, Fig. 2) is specific for isolates of CRF02_AG. These base changes do not alter the overall structure of the USE domain.

Tract *1* is a binding site for the $tRNA_3^{Lys}$ in HIV-1 3' UTR, this interaction promotes minus strand DNA transfer during reverse transcription [11]. Rare base changes G8A and U10A/C do not greatly affect duplex formation between tract *1* and $tRNA_3^{Lys}$. The U10A/C breaks an extreme base pair of this duplex and G8A breaks an internal base pair. Our data on conservation of tract *1* support the interaction between the $tRNA_3^{Lys}$ and 3' end of HIV-1 gRNA.
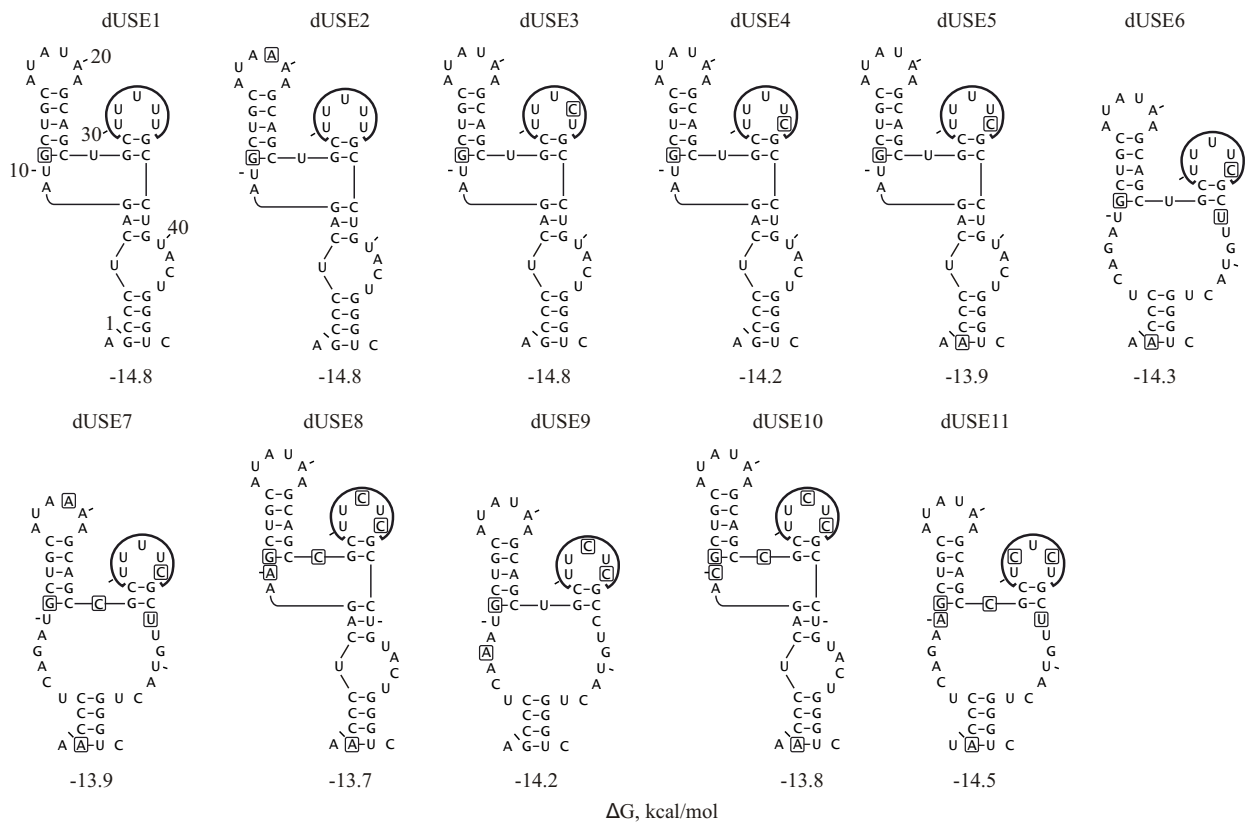
Fig. 2. Optimal and suboptimal structures of USE domain variants in HIV-1 and SIV pre-mRNAs. Free energy difference between optimal and suboptimal structures is –0.4, –1.6 and –0.1 kcal/mol for dUSE1, dUSE2 and dUSE7, respectively. The base changes as compared to RefSec are squared. The U-rich tract is indicated by line. dUSE1-dUSE7 are the main variants of the HIV-1 USE domain; dUSE8 and dUSE9-dUSE10 are variants of HIV-1 USE domain in subtype G and CRF02_AG isolates, respectively; dUSE11 is a variant of USE domain in SIVcpz*Ptt* isolates

Table 1
Occurrence of USE domain variants in HIV-1 isolates of different subtypes (%)

Variant	Base changes	Subtype				
		A	B	C	01_AE	Others
		Number of HIV-1 isolates (number of patients)				
		141 (122)	645 (258)	291 (234)	162 (61)	440 (322)
dUSE1	C11G	0	20	1	0	2
dUSE2	C11G + U19A	0	8	0	0	0
dUSE3	C11G + U33C	1	12	0	0	1
dUSE4	C11G + U34C	1	1	14	0	1
dUSE5	G1A + C11G+U34C	45	0	32	0	12
dUSE6	G1A + C11G + U34C + C37U	7	0	14	1	5
dUSE7	G1A + C11G + U19A + U27C + U34C + C37U	0	0	0	67	1
Total	–	**54**	**41**	**61**	**68**	**22**

The base change U19A (dUSE2 and dUSE7, Fig. 2) in the region of USE domain corresponding to the TA

TAA box in HIV-1 proviral DNA (tract *2* in Fig. 1) occurs in almost all HIV-1 isolates of CRF01_AE and in
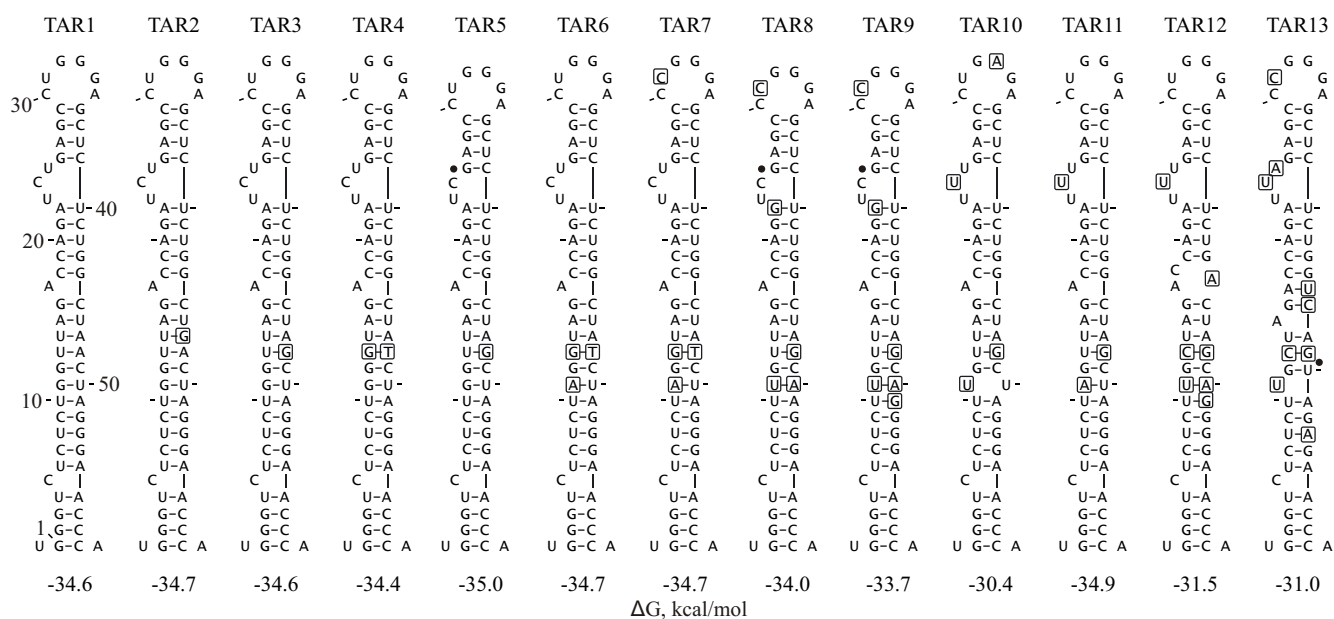
Fig. 3. Optimal structures of TAR hairpin variants in HIV-1 and SIV pre-mRNAs. The base changes as compared to RefSec are squared. Deletions are indicated by spot. TAR1-TAR9 – main variants; TAR10, TAR11 and TAR12 – frequent variants in HIV-1 isolates of subtypes D, F and G, respectively; TAR13 – TAR hairpin variant in SIVcpz*Ptt* isolates

Table 2
Occurrence of TAR hairpin variants in HIV-1 isolates of different subtypes (%)

Variant	Base changes	Subtype				
		A	B	C	01_AE	Others
		Number of HIV-1 isolates (number of patients)				
		141 (122)	645 (258)	291 (234)	162 (61)	440 (322)
TAR1	No	0	33	0	0	0.2
TAR2	A47G	0	7	0	0	0.2
TAR3	A48G	0	17	0	0	2
TAR4	U13G + A48U	1	0	6	0	2
TAR5	U25del + A48G	23	0	0	0	0.2
TAR6	G11A + U13G + A48U	0	0	10	0	2
TAR7	G11A + U13G + U31C + A48U	0	0	5	0	0.4
TAR8	G11U + A22G + U25del + U31C + A48G + U50A	0	0	0	32	0
TAR9	G11U + A22G + U25del + U31C + A48G + U50A + A51G	0	0	0	31	0
Total	–	**24**	**57**	**21**	**63**	**7.0**

27 % of isolates of subtype B, mainly from South Korea (79 %). It leads to the change of TATAA box to TAAAA tract, which was shown to be a functional sequence in HIV-1 isolates of CRF01_AE and subtype J [19].

The most frequent mutations in the U-rich tract of the USE domain, which is a binding site for a polyadenyla-tion factor CPSF, are U-to-C base changes. Besides, U-to-G base changes and deletions of U residue are rather often. The base change U32C occurs very frequently in HIV-1 isolates of CRF02_AG (90 %) and subtype G (79 %), while U33C occurs most frequently in HIV-1 isolates of subtype B (25 %). The base change U34C

33

occurs with a high frequency (85–100 %) in the HIV-1 isolates of all subtypes except for subtypes B and D.

The USE domain in all 17 SIV genomes (dUSE11, Fig. 2) is similar to that in HIV-1 isolates. Since almost all SIV isolates contain the base change C37U in the USE domain (Supplementary material, Table S2), the structure of its middle part is similar to that of HIV-1 dUSE6 and dUSE7 (Fig. 2). The most frequent base changes in SIV-1 USE domain are G1A, U10del/U10A, C11G, U27C, C37U and U31C + U33C or U32C + U34C. The combinations of base changes in the SIV USE domain most closely resemble those found in the HIV-1 USE domain of A/G-containing subtypes.

TAR hairpin. The variants of HIV-1 TAR hairpin with the combinations of base changes occurring with a frequency ⩾ 5% (the main variants TAR1-TAR9) are shown in Fig. 3. Their distribution by subtype is given in Table 2. Frequent variants of the TAR hairpin in HIV-1 isolates of subtypes D (20 %), F (38 %) and G (21 %) forming small subpools are also shown in Fig. 3 (TAR10-TAR12). Like the USE domain, the TAR hairpin sequence is much more heterogeneous than that of the polyA and DSE hairpins. The main variants of the TAR hairpin occur only in 35 % of the isolates studied.

The TAR hairpin without base changes and all other main variants have two 1 nt bulges and a 2–3 nt bulge in the stem. In general, about 60 % of the HIV-1 isolates studied contain the TAR hairpin with a similar structure. All main variants of the TAR hairpin, except for TAR4, are subtype specific (Table 2). We also observed country specific combinations of base changes in the TAR hairpin.

For example, the double mutation U25del + A48G (TAR5) occurs only in HIV-1 isolates of subtype A from Russia (64 %) and Ukraine (25 %), while most of subtype A isolates from Tanzania contain the TAR hairpin with additional base changes G11U and/or U50G or others.

In 35 % of HIV-1 isolates, the TAR hairpin stem is moderately altered mainly due to the formation of small internal loops, for example, TAR10 and TAR12 (Fig. 3). In 5 % of HIV-1 isolates the TAR hairpin structure is severely altered (truncated stem, greatly disturbed upper part or complete absence of the TAR hairpin structure). The defective TAR hairpin can impact overall folding of the 5' and 3' UTRs in HIV-1 gRNA and inhibit cer-

tain processes of viral replication [20, 21]. Moreover, the TAR hairpin is one of microRNAs produced in HIV-1 [22] and defects in the TAR hairpin may impact their functioning.

Base change frequency at each position of the TAR hairpin is presented in Suplementary information, Table S3. Tract *3* (Fig. 1) which corresponds to positions 20–43 of the TAR hairpin is sufficient and necessary for binding Tat protein [9]. This tract is not highly conserved. The most frequent base changes occur at the second and third positions of the 3 nt bulge (positions 23–25) in the TAR hairpin stem, the first position being highly conserved. In particular, C24U was found in HIV-1 isolates of all subtypes, except for CRF01_AE, and deletion at position 25 was found mainly in isolates of subtypes A and CRF01_AE (Suplementary information, Table S4).

Other frequent base changes in the upper part of the TAR hairpin are A22G, U31C, G33A and G44A. The A22G resulting in the A-U to G-U substitution (TAR8 and TAR9, Fig. 3) was found in almost all CRF01_AE isolates. The U31C and G33A occur in the apical loop of the TAR hairpin, the U31C was found in almost all CRF01_AE isolates (TAR8 and TAR9, Fig. 3) and in 32% of the subtype C isolates (TAR7, Fig. 3), while the G33A – in 57 % of the subtype D isolates (TAR10, Fig. 3). The G44A resulting in internal loop ACxA instead of an A bulge was found in 72 % of CRF02_AG isolates and 36 % of subtype G isolates (TAR12, Fig. 3).

Tract *3* also encompasses the GGGAGCUCUC palindrome (shadowed in Fig. 1) that is supposed to participate in the HIV-1 gRNA dimerization [23–25]. This palindrome corresponds to positions 32–41 of the TAR hairpin. The results of our phylogenetic study showed that the GGGAGCUCUC palindrome is well conserved. Rare base changes (G32A and G33A) occur only at its first and second positions with a frequency of 3 and 5 %, respectively. The G32A breaks two extreme base pairs of the TAR palindrome duplex, while the G33A (specific for subtype D isolates) causes the G-U to A-U substitution in this duplex.

At present a role of the GGGAGCUCUC palindrome in the HIV-1 gRNA dimerization is not entirely clear. Jalalirad *et al.* emphasize a dominant role of the TAR 3 nt bulge in the HIV-1 genome dimerization without involving TAR-TAR kissing: as an RNA binding

site, as a flexible hinge between the TAR upper and lower stems, and as a protein binding site [23].

Our data on palindrome conservation support a hypothesis on the HIV-1 gRNA dimerization through the TAR-TAR kissing mechanism. We propose that one of numerous destabilizing proteins, for the binding of which the TAR bulge is essential, for example Vif (viral infectivity factor) [26], promotes the TAR palindrome exposure and thus facilitates TAR-TAR kissing. The bulge deletion is supposed to prevent specific interaction of destabilizing protein(s) with the TAR hairpin and inhibit formation of the palindrome duplex. By analogy with a partial occlusion of the AAUAAA hexamer in the polyA hairpin, that is essential for polyadenylation regulation, the occlusion of the palindrome by the TAR hairpin structure may be also functionally important. In particular, it may assure the involvement of the TAR hairpin in HIV-1 genome dimerization only at certain stage of the replication cycle.

By conducting a limited survey of HIV-1 and SIV TAR hairpin in HIV-1 and SIV genomes, Berkhout *et al*. first demonstrated its similarity in both genomes [27]. We found 14 different sequences of the SIV TAR hairpin which contain from 3 up to 13 base changes as compared to HIV-1 RefSec (Supplementary information, Table S5). In most SIVcpz*Ptt* isolates, the bottom part of the TAR hairpin stem (below the 3 nt bulge) contains 1–2 internal loops, three 1nt bulges (TAR13, Fig. 3) or 2 nt bulges. In two SIV isolates, the TAR hairpin has a moderately altered upper part with 4 nt bulge and 5 nt apical loop. The most frequent base changes in the SIV TAR hairpin are G11U and A48G. Among rather frequent are U13C, C24U, U25A, U46C and C49del. Like the USE domain, combinations of base changes in the SIV TAR hairpin most closely resemble those found in HIV-1 USE domain of A/G-containing subtypes.

Conclusions. Our large-scale phylogenetic analysis data on the structural elements of the HIV-1 complete poly(A) region presented here and in our previous article [16] support a number of functional RNA-RNA interactions in HIV-1 gRNA. Our findings are believed to be useful in diagnostics of AIDS, monitoring the ways of HIV-1 spread and for design of new therapeutics to inhibit viral replication. In particular, the TAR hairpin is a target for developing antiviral therapy based on signal element inhibition [28]. Considering a significant degree of heterogeneity of this element, our data on its structural variability can be helpful in drugs design.

Філогенетичний аналіз структурних елементів в області полі(A) ВІЛ-1. 2. Домен USE і шпилька TAR

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

Резюме

Мета. Провести філогенетичний аналіз структурних елементів області полі(A) у про-мРНК ВІЛ-1, зокрема, основного «верхнього» елемента (USE), який стимулює поліаденілювання транскрипту ВІЛ-1, і шпильки TAR, що просторово зближує сигнали AAUAAA і USE. *Методи*. Передбачення вторинної структури цих елементів здійснювали за допомогою програми UNAfold. *Результати*. Структуру домену USE і шпильки TAR проаналізовано для 1679 геномів ВІЛ-1 групи M та 17 геномів вірусу мавп SIVcpzPtt. У геномі ВІЛ-1 ми виявили 376 і 588 різних послідовностей для домену USE і шпильки TAR відповідно, а також найчастіші мутації і мутації, специфічні для певного субтипу або країни. Лише 43 % усіх досліджених геномів ВІЛ-1 містять варіанти домену USE, які зустрічаються з частотою 5 % (основні варіанти), і 35 % геномів включають основні варіанти TAR. Встановлено також, що домен USE і шпилька TAR з геному SIVcpzPtt дуже схожі на відповідні елементи з геному ВІЛ-1 A/G-вмісних субтипів. *Висновки*. Результати наших широкомасштабних досліджень підтримують гіпотезу щодо взаємодії між $tRNA_3^{Lys}$ та 3'-кінцем геномної РНК ВІЛ-1, а також спірну гіпотезу димеризації шпильки TAR через взаємодію паліндромів. Оскільки шпилька TAR є об'єктом антивірусної терапії, спрямованої на пригнічення сигнальних елементів, дані стосовно специфічних особливостей цієї шпильки можуть бути корисними для створення нових лікарських препаратів.

Ключові слова: ВІЛ-1, SIVcpzPtt, область полі(A), вторинна структура, USE, TAR.

Филогенетический анализ структурных элементов в области поли(A) ВИЧ-1. 2. Домен USE и шпилька TAR

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

Резюме

Цель. Провести филогенетический анализ структурных элементов области поли(A) ВИЧ-1, в частности, основного «верхнего» элемента (USE), стимулирующего полиаденилирование транскрипта ВИЧ-1, и шпильки TAR, пространственно совмещающей сигналы AAUAAA и USE. *Методы*. Вторичная структура этих элементов предсказана программой UNAfold. *Результаты*. Структуры домена USE и шпильки TAR проанализированы в 1679 геномах ВИЧ-1 и 17 геномах вируса обезьян SIVcpzPtt. В геноме ВИЛ-1 мы выявили соответственно 376 и 588 различных последовательностей для домена USE и шпильки TAR, а также наиболее частые мутации и мутации, специфические для определенного субтипа или страны. Только 43 % всех исследованных геномов ВИЧ-1 содержат варианты домена USE, встречающиеся с частотой 5 % (основные варианты), и 35 % геномов включают основные варианты TAR. Мы обнаружили, что домен USE и шпилька TAR в геноме SIVcpz Ptt наиболее близки по структуре к соответствующим элементам в геномах ВИЧ-1 A/G-содержащих субтипов. *Выводы*. Результаты наших широкомасштабных исследований свидетельствуют в пользу гипотезы взаимодействия между $tRNA_3^{Lys}$ и 3'-концом генома ВИЧ-1, а также спорной гипотезы димериза-

ции шпильки TAR через взаимодействие палиндромов. Поскольку шпилька TAR является объектом антивирусной терапии, направленной на подавление сигнальных элементов, данные о специфических особенностях этой шпильки могут быть полезными для создания новых лекарственных препаратов.

Ключевые слова: ВИЧ-1, SIVcpzPtt, область поли(А), вторичная структура, USE, TAR.

REFERENCES

1. *Chan S, Choi EA, Shi Y.* Pre-mRNA 3'-end processing complex assembly and function. Wiley Interdiscip. *Rev RNA.* 2011; **2**(3): 321–35.

2. *Ashe MP, Furger A, Proudfoot NJ.* Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. *RNA.* 2000; **6**(2):170–7.

3. *DeZazzo JD, Kilpatrick JE, Imperiale MJ.* Involvement of long terminal repeat U3 sequences overlapping the transcription control region in human immunodeficiency virus type 1 mRNA 3' end formation. *Mol Cell Biol.* 1991; **11**(3):1624–30.

4. *Valsamakis A, Zeichner S, Carswell S, Alwine JC.* The human immunodeficiency virus type 1 polyadenylylation signal: a 3' long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylylation. *Proc Natl Acad Sci USA.* 1991; **88**(6):2108–12.

5. *Klasens BI, 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**(2): 446–54.

6. *Gilmartin GM, Fleming ES, Oetjen J, Graveley BR.* CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. *Genes Dev.* 1995; **9**(1):72–83.

7. *Kaufmann I, Martin G, Friedlein A, Langen H, Keller W.* Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J.* 2004; **23**(3):616–26.

8. *Gilmartin GM, Fleming ES, Oetjen J.* Activation of HIV-1 pre-mRNA 3' processing *in vitro* requires both an upstream element and TAR. *EMBO J.* 1992; **11**(12):4419–28.

9. *Ott M, Geyer M, Zhou Q.* The control of HIV transcription: keeping RNA polymerase II on track. *Cell Host Microbe.* 2011; **10** (5):426–35.

10. *Zarudnaya MI, Potyahaylo AL, Kolomiets IM, Hovorun DM.* Structural model of the complete poly(A) region of HIV-1 pre-mRNA. *J Biomol Struct Dyn.* 2013; **31**(10):1044–56.

11. *Piekna-Przybylska D, Dykes C, Demeter LM, Bambara RA.* Sequences in the U3 region of human immunodeficiency virus 1 improve efficiency of minus strand transfer in infected cells. *Virology.* 2011; **410**(2):368–74.

12. *Paillart JC, 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**(8):5995–6004.

13. *Abbink TE, 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**(13):11601–11.

14. *Ooms M, Cupac D, Abbink TE, 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**(5):1649–59.

15. *Decorsiere A, Cayrel A, Vagner S, Millevoi S.* Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage. *Genes Dev*. 2011; **25**(3):220–5.

16. *Zarudnaya MI, Potyahaylo AL, Kolomiets IM, Hovorun DM.* Phylogenetic study on structural elements of HIV-1 poly(A) region. 1. PolyA and DSE hairpins. *Biopolym Cell.* 2013; **29**(6): 454–62.

17. *Markham NR, Zuker M.* UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol.* 2008; **453**:3–31.

18. *Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, Burch CL, Weeks KM.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature.* 2009; **460**(7256): 711–6.

19. *Ramirez de Arellano E, Soriano V, Alcamil J, Holguin A.* New findings on transcription regulation across different HIV-1 subtypes. *AIDS Rev.* 2006; **8**(1):9–16.

20. *Vrolijk MM, Harwig A, Berkhout B, Das AT.* Destabilization of the TAR hairpin leads to extension of the polyA hairpin and inhibition of HIV-1 polyadenylation. *Retrovirology.* 2009; **6**:13.

21. *Vrolijk MM, Harwig A, Berkhout B.* Opening of the TAR hairpin in the HIV-1 genome causes aberrant RNA dimerization and packaging. *Retrovirology.* 2012; **9**:59.

22. *Schopman NC, Willemsen M, Liu YP, Bradley T, van Kampen A, Baas F, Berkhout B, Haasnoot J.* Deep sequencing of virus-infected cells reveals HIV-encoded small RNAs. *Nucleic Acids Res.* 2012; **40**(1):414–27.

23. *Jalalirad M, Saadatmand J, Laughrea M.* Dominant role of the 5' TAR bulge in dimerization of HIV-1 genomic RNA, but no evidence of TAR-TAR kissing during *in vivo* virus assembly. *Biochemistry.* 2012; **51**(18):3744–58.

24. *Andersen ES, Contera SA, Knudsen B, Damgaard CK, Besenbacher F, Kjems J.* Role of the trans-activation response element in dimerization of HIV-1 RNA. *J Biol Chem.* 2004; **279**(21): 22243–9.

25. *Pallesen J.* Structure of the HIV-1 5' untranslated region dimer alone and in complex with gold nanocolloids: support of a TAR-TAR-containing 5' dimer linkage site (DLS) and a 3' DIS-DIS-containing DLS. *Biochemistry.* 2011; **50**(28):6170–7.

26. *Freisz S, Mezher J, Hafirassou L, Wolff P, Nomine Y, Romier C, Dumas P, Ennifar E.* Sequence and structure requirements for specific recognition of HIV-1 TAR and DIS RNA by the HIV-1 Vif protein. *RNA Biol.* 2012; **9**(7):966–77.

27. *Berkhout B.* Structural features in TAR RNA of human and simian immunodeficiency viruses: a phylogenetic analysis. *Nucleic Acids Res.* 1992; **20**(1):27–31.

28. *Stelzer AC, Frank AT, Kratz JD, Swanson MD, Gonzalez-Hernandez MJ, Lee J, Andricioaei I, Markovitz DM, AlHashimi HM.* Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat Chem Biol.* 2011; **7**(8):553–9.