



УДК 577.21.5

© 2011

М. И. Зарудная, А. Л. Потягайло, В. В. Отенко, И. Н. Коломиец,
член-корреспондент НАН Украины Д. Н. Говорун

Вторичная структура основного поли(А)-сигнала про-мРНК вируса иммунодефицита человека

Вперше описано модель основного сигналу поліаденілювання про-мРНК ВІЛ-1. Показано, що “нижній” елемент (DSE) полі(А)-сигналу розташований у апікальній петлі шпильки завдовжки 26 нт, яку названо шпилькою DSE. Адекватність моделі підтверджено передбаченням вторинної структури області, яка включає шпильки TAR, polyA та DSE, для 414 ізолятів ВІЛ-1 різних субтипів та рекомбінантних форм. Вперше наведено також моделі вторинної структури 3'-кінця 3'-нетрансльованої ділянки про-мРНК ВІЛ-1 для ізолятів різних субтипів. Висловлено припущення, що U5(U)-шпилька домену PBS поряд з участю у зворотній транскрипції геномної РНК ВІЛ-1 може також брати участь у процесі поліаденілювання про-мРНК, забезпечуючи доступність DSE для білкового апарату поліаденілювання під час транскрипції провірусної ДНК.

Процесс полиаденилирования, существенный для регуляции экспрессии генов, направляется у млекопитающих и их вирусов двумя основными сигнальными элементами, расположенными по обеим сторонам от места расщепления транскрипта, к которому присоединяется 3'-концевая поли(А)-последовательность [1]. “Верхний” элемент сигнала полиаденилирования представляет собой гексамер (чаще всего AAUAAA), а “нижний” элемент сигнала (DSE), согласно нашему определению [2], состоит из U-богатых пентамеров (четыре U из пяти оснований) и/или GU/U-богатых пентамеров (два GU и один U). Пентамеры могут перекрываться, и в одном поли(А)-сигнале может быть несколько “нижних” элементов. Гексамер узнается белковым фактором специфичности расщепления/полиаденилирования CPSF, а “нижние” элементы — белковым фактором стимулирования расщепления CstF [1].

Вторичная структура сигналов полиаденилирования критична для их узнавания белковыми факторами, однако в настоящее время она определена лишь для отдельных элементов некоторых про-мРНК [2], включая про-мРНК вируса иммунодефицита человека (ВИЧ-1). Показано, что гексамер AAUAAA в про-мРНК ВИЧ-1 расположен в верхней

части шпильки длиной 45 нт (шпилька polyA) [3]. В связи с особенностями репликации ретровирусов идентичные поли(А)-сигналы расположены на обоих концах про-мРНК ВИЧ-1. Процесс полиаденилирования на 5'-конце должен подавляться (преждевременное полиаденилирование), а на 3'-конце — стимулироваться. Стимулирование происходит с помощью вспомогательного “верхнего” элемента USE поли(А)-сигнала про-мРНК ВИЧ-1, расположенного только на 3'-конце транскрипта, а ингибирование осуществляется с помощью нескольких механизмов (см. [4] и ссылки в этой работе). В частности, часть гексамера AAUAAA вовлечена в образование двойной спирали в шпильке polyA и малодоступна для CPSF как на 5'-конце, так и на 3'-конце про-мРНК ВИЧ-1.

В отличие от структуры шпильки polyA, представляемой одинаково в разных современных моделях вторичной структуры 5'-нетранслируемой области (5' UTR) про-мРНК ВИЧ-1, структура “нижних” элементов поли(А)-сигнала (UGUGU и GUUGUGU) в разных моделях различается (напр., [5, 6]) и зависит от структуры расположенных за шпилькой polyA домена PBS, направляющего обратную транскрипцию, и шпилек DIS, SD и Psi, необходимых для димеризации и упаковки геномной РНК. Область генома, расположенная за местом связывания праймера тРНК в домене PBS, присутствует только на 5'-конце транскрипта, поэтому на структуру “нижних” элементов истинного поли(А)-сигнала, находящегося на 3'-конце про-мРНК, влияет лишь структура участка, расположенного между “нижними” элементами и праймером. Насколько нам известно, модели вторичной структуры 3'-конца про-мРНК ВИЧ-1 в литературе не представлены и вторичная структура “нижних” элементов поли(А)-сигнала не определена.

Отличительной особенностью ВИЧ-1 является чрезвычайно высокая гетерогенность генома, что, с одной стороны, значительно осложняет диагностику ВИЧ-инфекции, а с другой — позволяет использовать филогенетический подход для исследования механизмов репликации вируса.

В настоящей работе приведены результаты исследования структуры “нижних” элементов поли(А)-сигнала и 3' нетранслируемой области про-мРНК ВИЧ-1 с помощью предсказания вторичной структуры и филогенетического анализа.

Первичные последовательности про-мРНК ВИЧ-1, включающие поли(А)-сигнал, были выбраны из ГенБанка. Вторичную структуру исследуемой области определяли с помощью программы mfold Дзукера [7] со следующими параметрами: 5% субоптимальности и window 0. Для автоматизации процесса предсказания вторичной структуры и поиска мутаций использовалось разработанное нами программное обеспечение.

Для определения вторичной структуры “нижних” элементов поли(А)-сигнала про-мРНК ВИЧ-1 для изолята НХВ2 с реферативной последовательностью генома (субтип В) мы свернули несколько фрагментов транскрипта, которые включали шпильку polyA, оба GU/U-богатых тракта, а также расположенные дальше участки разной длины вплоть до конца транскрипта, т.е. до начала места связывания тРНК в домене PBS. Обнаружено, что область транскрипта, включающая только шпильку polyA и оба GU/U-богатых тракта, сворачивается с образованием шпильки длиной 26 нт (рис. 1, а), которую мы назвали шпилькой DSE. В апикальной петле этой шпильки находится второй “нижний” элемент GUUGUGU, включающий два перекрывающихся GU/U-богатых пентамера GUUGU и UGUGU, оба пентамера доступны для связывания с CstF. Первый “нижний” элемент UGUGU должен функционировать менее эффективно из-за частичного вовлечения в спа-

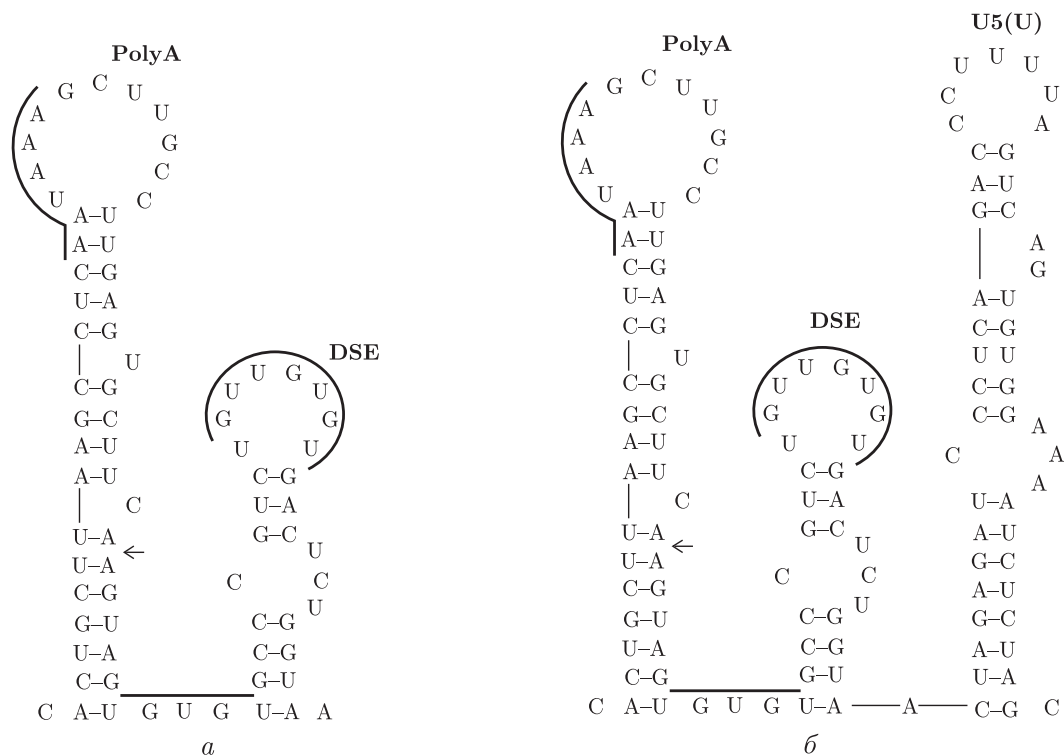


Рис. 1. Вторичная структура основного поли(А)-сигнала (а) и 3'-конца 3' UTR (б) про-мРНК ВИЧ-1 для изолята НХВ2 (RefSec) субтипа В. Стрелкой указано место расщепления транскрипта. Гексамер ААУААА и “нижние” элементы поли(А)-сигнала обозначены линией

ривание с основаниями соседних шпильек (см. рис. 1, а) и из-за близости к месту расщепления [8].

При сворачивании самого протяженного фрагмента длиной 122 нт образуются три шпильки (см. рис. 1, б) — polyA, DSE и U5(U)-шпилька, которая является составной частью домена PBS в изолятах ВИЧ-1 ряда субтипов [5]. Можно предположить, что U5(U)-шпилька не только принимает участие в инициации обратной транскрипции геномной РНК ВИЧ-1, но и обеспечивает доступность второго GU/U-богатого “нижнего” элемента для белкового аппарата полиаденилирования как в начале транскрипции поли(А)-сигнала, так и в конце транскрипции всей провирусной ДНК. Следует отметить, что второй GU/U-богатый “нижний” элемент экспонирован и при сворачивании фрагментов промежуточной длины либо в линкере (рис. 2, а), либо во внутренней петле шпильки длиной 39 нт (см. рис. 2, б). В последнем случае экспонируется пентамер UGUGU.

Поскольку область про-мРНК ВИЧ-1, соответствующая шпильке DSE, состоит из довольно консервативных участков [5], мы предположили, что структура основного поли(А)-сигнала, определенная нами для изолята НХВ2 (см. рис. 1, а), является универсальной. Для проверки этого предположения мы предсказали вторичную структуру области, включающей последовательно расположенные шпильки TAR, polyA и DSE, для 414 изолятов ВИЧ-1 разных субтипов и рекомбинантных форм.

Несмотря на то, что шпилька TAR, необходимая для стимулирования элонгации транскрипции провирусной ДНК, непосредственно не участвует в процессе полиаденилирова-

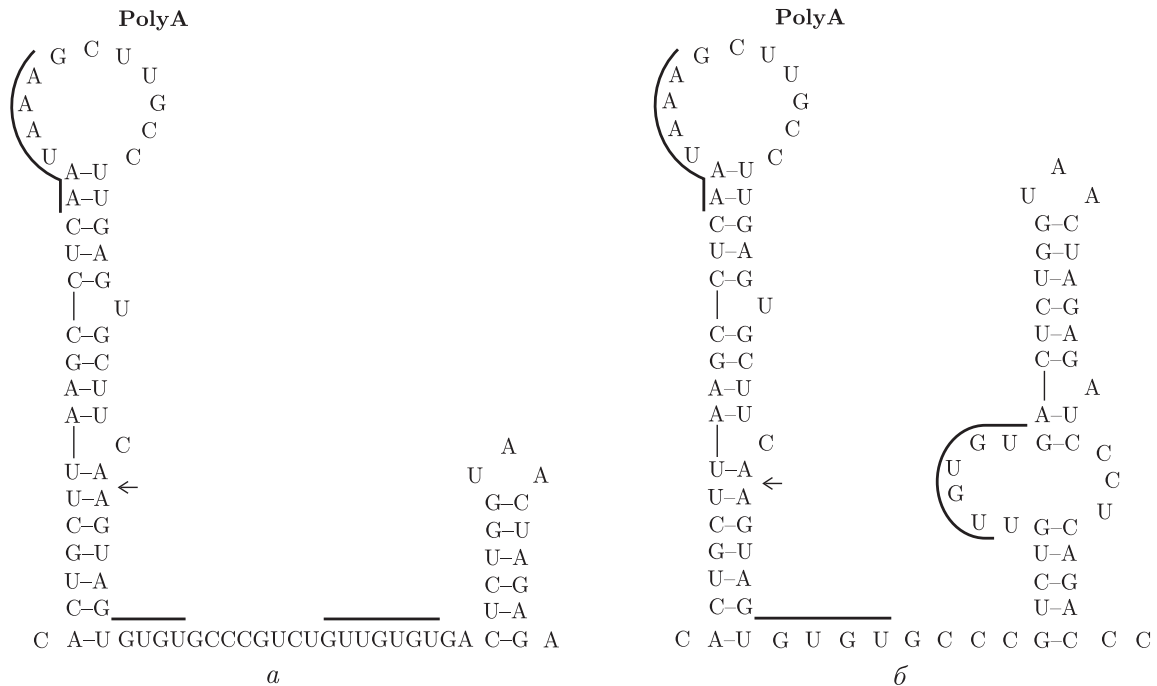


Рис. 2. Вторичная структура фрагментов 3'-конца 3' UTR про-мРНК ВИЧ-1 для изолята HXB2 (RefSec) субтипа В. Длина фрагмента — 83 нт (*a*) и 95 нт (*б*). Стрелкой указано место расщепления транскрипта. Гексамер AAUAAA и “нижние” элементы поли(А)-сигнала обозначены линией

ния, она является важной структурной частью поли(А)-сигнала [9, 10]. Эта шпилька, находясь между вспомогательным “верхним” элементом поли(А)-сигнала и шпилькой polyA, пространственно сближает эти элементы, тем самым обеспечивая эффективное функционирование 3'-концевого поли(А)-сигнала [9]. Кроме того, мутации, дестабилизирующие шпильку TAR, могут ингибировать процесс полиаденилирования вследствие удлинения соседней шпильки polyA за счет нуклеотидов, высвобождающихся из TAR [10], а стабильность шпильки polyA критична для стимуляции полиаденилирования на 3'-конце транскрипта и подавления на 5'-конце [4].

Мы обнаружили, что шпильки, подобные TAR, polyA и новой шпильке DSE, образуются на 3'-конце почти всех исследованных про-мРНК ВИЧ-1 (92%), несмотря на гетерогенность первичной последовательности исследуемой области. Замены оснований в геномах ВИЧ-1 разных субтипов приводят в основном к небольшим изменениям в структуре отдельных шпилек. Для примера на рис. 3 приведены вторичные структуры основного поли(А)-сигнала для изолятов ВИЧ-1 субтипов А и С. Интересно, что замена G → U в шпильке DSE приводит к образованию тракта GUUGUUU (см. рис. 3, *a*), который включает три перекрывающихся пентамера GUUGU, UUGUU и UGUUU, два из них являются U-богатыми пентамерами.

Таким образом, филогенетические исследования свидетельствуют в пользу того, что “нижний” элемент основного поли(А)-сигнала про-мРНК ВИЧ-1 экспонирован в апикальной петле шпильки DSE.

Для определения вторичной структуры 3'-конца 3' UTR про-мРНК ВИЧ-1 в изолятах разных субтипов мы свернули соответствующую область для 400 изолятов. Обнару-

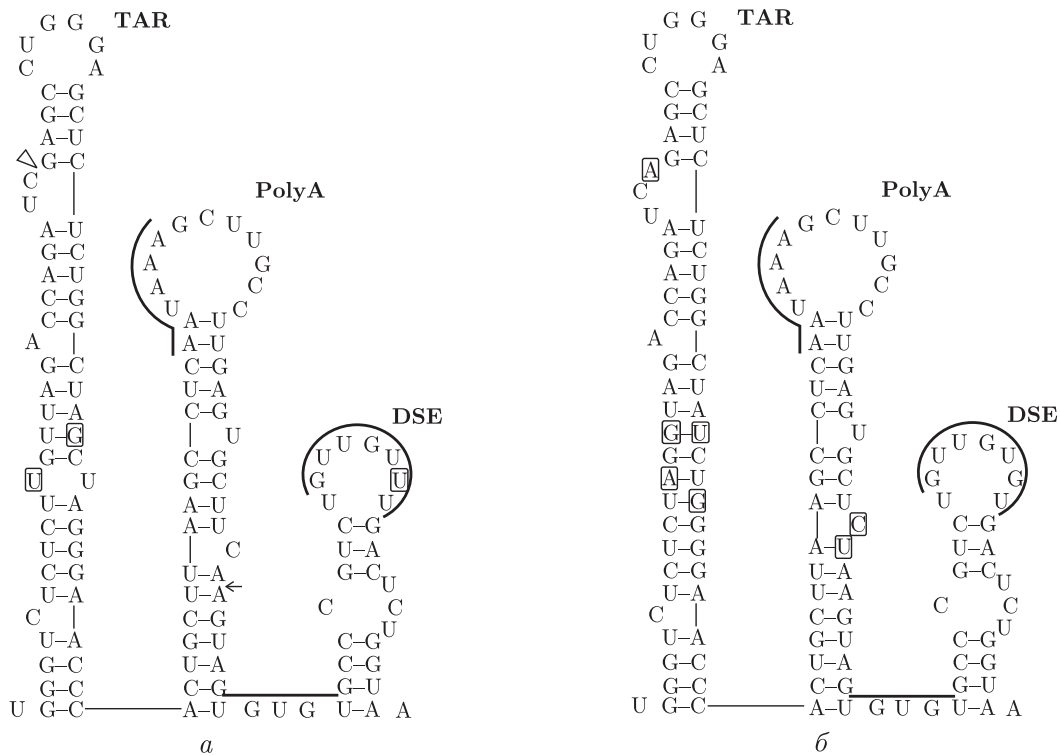


Рис. 3. Вторичная структура поли(А)-сигнала про-мРНК ВИЧ-1 для изолята pUG031-A1 (номер АВ098330) субтипа А (а) и изолята p02ZMJMC18 (номер АВ254156) субтипа С (б). Гексамер AAUAAA и “нижние” элементы поли(А)-сигнала обозначены линией. Замены оснований по сравнению с реферативной последовательностью генома HXB2 (RefSec) отмечены квадратами

жено, что структура, подобная той, которая показана на рис. 1, б для изолята HXB2, образуется также на 3'-конце про-мРНК всех других изолятов субтипа В, практически всех изолятов субтипа С, а также многих изолятов других субтипов и рекомбинантных форм.

В большинстве изолятов ВИЧ-1 субтипов Е (А/Е) и G, домен PBS которых содержит вместо U5(U)-шпильки U5(AA)-шпильку, а также в некоторых изолятах других субтипов и рекомбинантных форм 3'-конец 3' UTR сворачивается в альтернативную структуру с образованием шпильки polyA, длинной шпильки, содержащей второй “нижний” элемент (гексамер UGUGU) во внутренней петле, а также укороченной U5(AA)-шпильки (рис. 4, а). Шпилька с DSE во внутренней петле является удлиненной формой шпильки длиной 39 нт, образующейся при сворачивании фрагмента про-мРНК изолята HXB2 длиной 95 нт (см. рис. 2, б). Следует отметить, что в изолятах с альтернативной структурой 3'-конца про-мРНК шпильки DSE и U5(U) могут образовываться в субоптимальных структурах (см. рис. 4, б).

Таким образом, наши филогенетические исследования вторичной структуры поли(А)-сигнала про-мРНК ВИЧ-1 и 3'-конца 3' нетранслируемой области этой про-мРНК показали, что “нижний” элемент функционального поли(А)-сигнала расположен в апикальной петле шпильки DSE и полностью доступен для белкового фактора CstF как в начале транскрипции поли(А)-сигнала, так и в конце транскрипции провирусной ДНК.

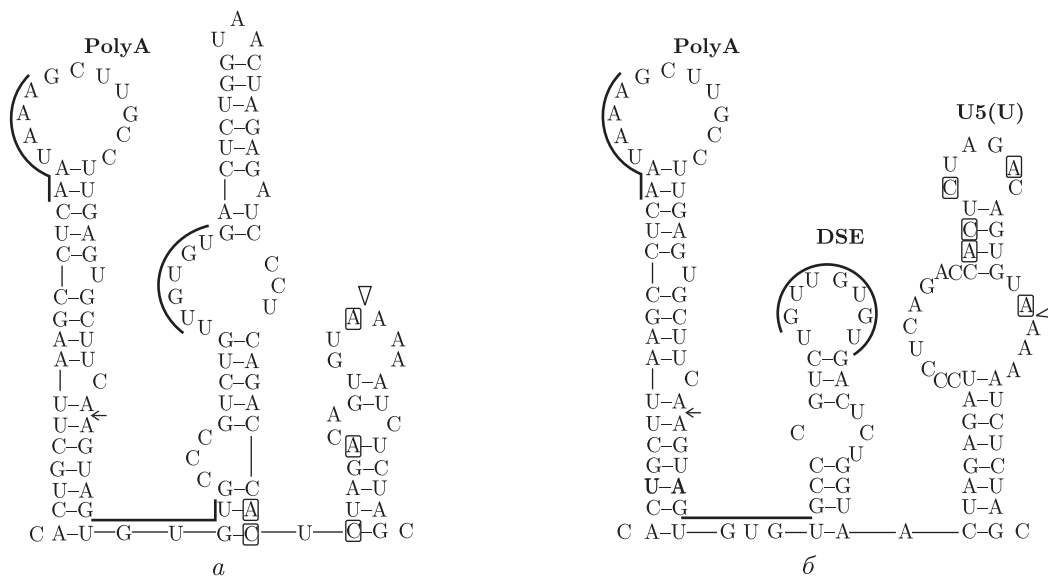


Рис. 4. Вторичная структура 3'-конца 3' UTR про-мРНК ВИЧ-1 для изолята GHNJ175 (номер AB231 893) субтипа G. *a* — оптимальная структура, *б* — субоптимальная структура, $\Delta\Delta G = 0,6$ ккал/моль. Стрелкой указано место расщепления транскрипта. Гексамер AAUAAA и “нижние” элементы поли(А)-сигнала обозначены линией. Замены оснований по сравнению с реферативной последовательностью генома HXB2 (RefSec) отмечены квадратами

Исследование выполнено при частичной поддержке гранта Украинского научно-технологического центра (проект 4302).

1. Zhao J., Hyman L., Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis // *Microbiol. Mol. Biol. Rev.* – 1999. – **63**, No 2. – P. 405–445.
2. Zarudnaya M. I., Kolomiets I. M., Potyahaiko A. L., Hovorun D. M. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures // *Nucl. Acids Res.* – 2003. – **31**, No 5. – P. 1375–1386.
3. 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**, No 1. – P. 276–281.
4. 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 // *Nucl. Acids Res.* – 1999. – **27**, № 2. – P. 446–454.
5. Abbink T. E. M., 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**, No 13. – P. 11601–11611.
6. Wilkinson K. A., Gorelick R. J., Vasa S. M. et al. High-throughput SHAPE analysis reveals structures in HIV – 1 genomic RNA strongly conserved across distinct biological states // *PLoS Biol.* – 2008. – **6**, No 4. – e96.
7. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction // *Nucl. Acids Res.* – 2003. – **31**, No 13. – P. 3406–3415.
8. Chou Z. F., Chen F., Wilusz J. Sequence and position requirements for uridylylate-rich downstream elements of polyadenylation signals // *Ibid.* – 1994. – **22**, No 13. – P. 2525–2531.
9. Gilmartin G. M., Fleming E. S., Oetjen J. Activation of HIV-1 pre-mRNA 3' processing in vitro requires both an upstream element and TAR // *EMBO J.* – 1992. – **11**, No 12. – P. 4419–4428.
10. Vrolijk M. M., Harwig A., Berkhout B., Das A. T. Destabilization of the TAR hairpin leads to extension of the polyA hairpin and inhibition of HIV-1 polyadenylation // *Retrovirology.* – 2009. – **6**. – P. 13.

*Институт молекулярной биологии
и генетики НАН Украины, Киев*

Поступило в редакцию 21.04.2010

M. I. Zarudnaya, A. L. Potyahaylo, V. V. Otenko, I. N. Kolomiets,
Corresponding Member of the NAS of Ukraine **D. M. Hovorun**

The secondary structure of core poly(A) signal of human immunodeficiency virus pre-mRNA

A model of core polyadenylation signal of HIV-1 pre-mRNA has been first presented. Downstream sequence element (DSE) of poly(A) signal has been shown to be located in the apical loop of 26 nt hairpin which we named DSE hairpin. The proposed model is supported by the secondary structure prediction of a genomic RNA region encompassing TAR, polyA and DSE hairpins for 414 HIV-1 isolates of different subtypes and recombinant forms. Models of the secondary structure of 3' end of HIV-1 pre-mRNA 3' untranslated region for isolates of different subtypes have been also first presented. U5(U) hairpin of PBS domain has been supposed to take part in the polyadenylation process alongside with participation in reverse transcription. This hairpin provides the DSE exposure to the protein polyadenylation machinery during the proviral DNA transcription.