## Консервативные структурные мотивы в 3'-нетранслируемой области геномной PHK вируса SARS-CoV

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

Институт молекулярной биологии и генетики НАН Украины Ул. Академика Заболотного, 150, Киев, 03143, Украина E-mail: dhovorun@imbg.org.ua

Сигнальные последовательности в 3'-нетранслируемой области (UTR) «плюс»-РНК вирусов контролируют процессы репликации, транскрипции и трансляции РНК. В представленной работе методом компьютерного моделирования (Zuker, 2003) впервые изучены вторичная и третичная структуры 3' UTR геномной РНК вируса SARS-CoV. Показано, что эта область содержит последовательность, образующую псевдоузел, который по своей структуре наиболее близок к псевдоузлам, обнаруженным в РНК коронавирусов группы 1. Так же, как и в случае других коронавирусов, последовательность, соответствующая одной из цепей стебля 1 псевдоузла РНК вируса SARS-CoV, может участвовать в образовании альтернативного мотива — высокостабильной шпилькообразной структуры. Кроме этого, 3' UTR РНК вируса SARS-CoV содержит мотив 52т, найденный в 3' UTR РНК вируса IBV (коронавирус группы III) и отсутствующий в ряде РНК коронавирусов первой и второй групп. Обсуждаются также и другие структурные мотивы, образующиеся в 3' UTR РНК вируса SARS-CoV. Полученные нами результаты поддерживают существующую точку зрения о том, что вирус SARS-CoV является уникальным и не может быть отнесен ни к одной из трех известных групп коронавирусов.

Введение. В недавней публикации группа авторов [1] сообщила о том, что причиной вспышки тяжелого острого респираторного синдрома (severe acute respiratory syndrome, SARS) является новый коронавирус SARS-CoV. Филогенетический анализ и сравнение последовательностей показали, что SARS-CoV — уникальный коронавирус, имеющий лишь отдаленное родство с коронавирусами, охарактеризованными ранее [1—3]. Марра и соавт. [3] предположили, что он является представителем новой, четвертой, группы коронавирусов.

В настоящей работе с помощью компьютерного моделирования с использованием программы для предсказания вторичной структуры РНК, разработанной Цукером [4], мы впервые исследовали вторичную и третичную структуры 3'-нетранслируемой области геномной РНК SARS-CoV и сравнили ее со структурами этих же областей ранее охарактеризованных коронавирусов. Мы нашли, что 3'

UTR геномной PHK SARS-CoV содержит ряд консервативных структурных мотивов. Среди них имеется как мотив, характерный для коронавирусов группы I, так и мотив, найденный лишь в 3′ UTR PHK коронавируса группы III.

Вторичная структура 3' UTR PHK SARS-CoV. Геном коронавирусов (отряд Nidovirales, семейство Coronaviridae) состоит из одной молекулы позитивной одноцепочечной РНК длиной примерно 30000 нуклеотидов [2, 3]. Геномная РНК имеет на 5'конце кэп-структуру и поли(А)-последовательность на 3'-конце. Примерно две трети генома коронавирусов служат матрицей для синтеза репликаз 1А и 1Б, одна треть генома (3'-концевая) кодирует структурные и неструктурные белки. Репликазы синтезируются при трансляции геномной РНК, а остальные белки - при трансляции субгеномных РНК. Концевые последовательности всех субгеномных РНК одинаковы. 5'-Концы идентичны лидерной последовательности геномной РНК длиной примерно 70 нуклеотидов, а 3'-концы идентичны

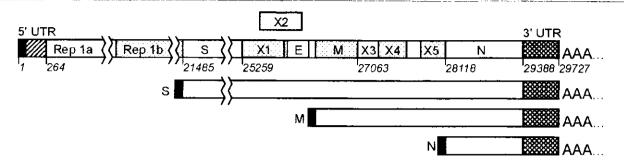


Рис. 1. Организация генома коронавируса SARS-CoV [2, 3]. Указаны области, транслируемые в следующие белки: Rep — репликаза; S (spike) — гликопротеин; E — малый белок оболочки; М — мембранный гликопротеин; N — нуклеокансидный белок; X1—X5 — неизвестные белки. Лидерная область 5' UTR выделена черным цветом. В нижней части рисунка показаны мРНК, кодирующие белки S, M и N

3'-концевой части генома. Организация генома вируса SARS-CoV, по данным работ [2, 3], показана на рис. 1. Субгеномные РНК коронавирусов образуются в результате прерываемой транскрипции, механизм которой в настоящее время неясен. Согласно модели, которой отдают предпочтение многие авторы, прерываемая транскрипция происходит при синтезе минус-цепи РНК [5].

Некоторыми авторами [6—9] установлено, что структурные элементы в 3′ UTR наиболее хорошо изученных коронавирусов, таких как вирус гепатита мыши (МНV) и коронавирус быка (ВСV), играют важную роль в репликации вирусов. Одним из этих элементов является псевдоузел, расположенный в 5′-концевой части 3′ UTR [6, 8]. Псевдоузлы с такой же локализацией содержатся не только в РНК МНV и ВСV, но и в РНК всех секвенированных коронавирусов [8].

Чтобы выяснить, могут ли псевдоузел и другие консервативные мотивы образовываться в 3' UTR PHK вируса SARS-CoV, мы исследовали вторичную структуру этой области, используя компьютерную программу для предсказания вторичной структуры РНК mfold (версия 3.1), разработанную Цукером [4]. Исследовали разные фрагменты геномной РНК SARS-CoV, расположенные в области -108/+339 нуклеотидов (нуклсотид +1 соответствует началу 3'-нетранслируемой области). Длина исследуемых фрагментов составляла от 80 до 250 нуклеотидов. Полученные результаты представлены на рис. 2. Следует отметить, что вторичная структура определенного участка РНК зависит от последовательности и длины фрагмента, в котором он содержится. В составе одних фрагментов этот участок сворачивается в локальную шпилькообразную структуру (шпильку), а в составе других -- образует пары оснований с отдаленными областями фрагмента.

На рис. 2 показаны лишь локальные шпилькообразные структуры. Образование некоторых из

них предсказывается независимо от фрагмента, выбранного для исследования (включающего данный участок). Например, шпилька SL 3 (stem-loop) (+98/+126 нуклеотидов) образуется как при сворачивании фрагмента -48/+132, так и при сворачивании последовательности +96/+242 нуклеотида. К неизменно предсказываемым шпилькам относятся также шпильки SL 1, SL 2 и SL 5. Шпильки SL 4 и SL 7 были предсказаны при сворачивании многих, но не всех исследованных фрагментов. При сворачивании фрагментов 3' UTR PHK SARS-CoV, содержащих участки, расположенные между шпильками SL 5 и SL 7, а также 3'-концевой участок 3' UTR, используемая программа предсказывает разные варианты. Шпильки SL 6 и SL 9 (рис. 2) будут обсуждаться ниже.

Можно предположить, что участки, сворачивающиеся одинаковым образом в составе разных фрагментов, скорее всего, будут сворачиваться таким же образом и в составе длинной нативной РНК, а участки, сворачивающиеся многими способами, будут участвовать во взаимодействии с отдаленными областями нативной РНК, приводя к образованию доменов. Если структура нативной РНК меняется на разных стадиях репликации вируса, как это, например, происходит при репликации вируса иммунодефицита человека [10, 11], тогда те или иные локальные шпильки могут образовываться лишь временно.

На рис. З показана схема сворачивания фрагмента РНК SARS-CoV -43/+353 нуклеотида, включающего всю З' UTR и часть З'-концевой поли(А)-последовательности. Представлена одна из двух наиболее энергетически выгодных структур. Вторая структура имеет отличие от первой лишь в области +190/+267 нуклеотидов. Как видно из этого рисунка, шпильки SL 1—3, а также SL 5 сохраняются и при сворачивании полной З'-нетранслируемой области. Шпилька SL 6 сохранилась только в струк-

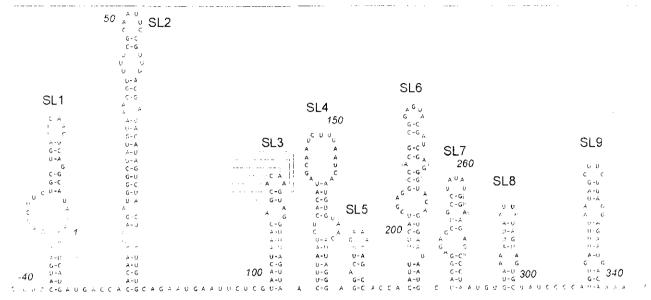


Рис. 2. Вторичная структура 3′ UTR геномной РНК вируса SARS-CoV. Первичная последовательность вируса была взята из электронной базы данных нуклеотидных последовательностей Национального центра биотехнологии и информации США (NCBI, http://www.ncbi.nlm.nih.gov/Database/index.html). Шпилькообразные структуры предсказаны методом компьютерного моделирования по программе mfold [4]. SL 1—9 — шпилькообразные структуры. Стоп-кодон трансляции белка N взят в рамку. Относительно линий, соединяющих шпильки SL 2 и SL 3, а также об отмеченных последовательностях — см. текст. Свободные энергии шпилек SL 1—9 равны: -15,0; -22,5; -8,3; -6,2; -5,4; -9,4; -10,3; -1,2 и -5,2 ккал/моль соответственно

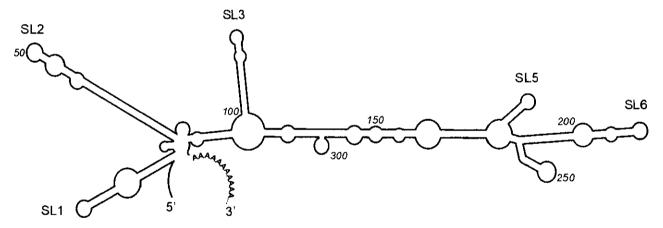


Рис. 3. Схема сворачивания фрагмента РНК вируса SARS-CoV, состоящего из 43 концевых нуклеотидов области, кодирующей белок N, полной 3'-нетранслируемой области и части 3'-концевой поли(A)-последовательности

туре, показанной на рис. 3. Остальные шпильки, представленные на рис. 2, при сворачивании полной 3′ UTR не образуются; соответствующие им последовательности принимают участие во взаимодействии с отдаленными областями фрагмента.

Консервативные структурные элементы в 3'нетранслируемой области РНК вируса SARS-CoV. На рис. 4 а, б, представлены схемы псевдоузлов, которые, по данным Вильямса и соавт. [8], могут быть образованы участками, расположенными в 3' UTR коронавирусов. Псевдоузлы, образуемые РНК разных коронавирусов, похожи по форме и примерно одинаково локализованы, но различаются первичной последовательностью. На рис. 4, а, показана схема предложенного Вильямсом консенсусного псевдоузла, в который сворачиваются РНК короновирусов групп I и II, а на рис. 4, б, — схема псевдоузла, образуемого РНК вируса инфекционного бронхита птиц (IBV) (вирус группы III). Как видно, псевдоузлы, в которые сворачиваются РНК коронавирусов группы III, значительно отличаются по размерам петель и стеблей от псевдоузлов, образуемых РНК коронавирусов групп I и II. Предлагаемая нами схема псевдоузла, который может

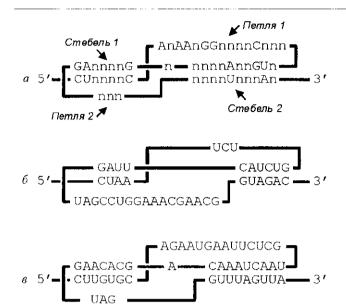


Рис. 4. Схемы псевдоузлов, образуемых последовательностями 3' UTR геномных РНК коронавирусов: a — консенсус для псевдоузлов, образуемых РНК коронавирусов первой и второй групп [8];  $\delta$  — псевдоузел РНК IBV [8];  $\delta$  — псевдоузел РНК SARS-CoV

образоваться в 3′ UTR вируса SARS-CoV, показана на рис. 4, в, а на рис. 2 нуклеотиды, образующие пары оснований в стебле 1 псевдоузла, соединены линиями. Как видно из рис. 4, этот псевдоузлы, образуемые РНК коронавирусов первой и второй групп. Более детальное сравнение предсказанного нами псевдоузла со структурой восьми псевдоузлов, приведенных в работе [8], показало, что по первичной последовательности стебля 1 псевдоузел, образуемый РНК SARS-CoV, наиболее близок к псевдоузлам, образуемым РНК коронавирусов первой группы.

Авторы работ [6, 7] сообщили о том, что большая шпилькообразная структура с внутренними петлями длиной 68 нуклеотидов образуется сразу же за стоп-кодоном трансляции белка N в 3' UTR коронавируса MHV. Нуклеотиды, расположенные в 3'-концевой части этой шпильки, участвуют в образовании двух альтернативных структур. Они являются либо частью шпильки длиной 68 нуклеотидов, либо участвуют в образовании псевдоузла вместе с прилегающей шпилькой, стебель которой образует стебель 2 в псевдоузле. Обе структуры имеются также и в 3' UTR коронавируса быка [7, 8]. И большая шпилька, и псевдоузел существенны для репликации вируса МНV [6, 8], однако механизмы их функционирования неизвестны. Как видно из рис. 2, большая шпилька, расположенная в 5'-концевой части 3' UTR, и прилегающая к ней шпилька, стебель которой образует стебель 2 псевдоузла, содержатся также и в 3' UTR SARS-CoV. Это шпильки SL 2 и SL 3. Интересно отметить, что они образуются и при сворачивании полной 3' UTR (рис. 3).

Кроме шпилек SL 2 и SL 3 в 3'-нетранслируемой области вируса SARS-CoV содержится еще один консервативный мотив. Джонассен и соавт. [12] показали, что 3' UTR отдельных представителей коронавирусов и пикорнавирусов содержит мотив s2m (stem-loop II-like motif), подобный по первичной последовательности и вторичной структуре высококонсервативной шпильке II, локализованной в 3'-концевой части астровирусов человека и животных. Авторы предположили, что присутствие этого мотива в трех семействах вирусов является результатом процессов рекомбинации между неродственными вирусами.

Марра и соавт. [3] обнаружили мотив s2m в 3' UTR SARS-CoV (участок +202/+233 нуклеотида). Этот фрагмент отличается от соответствующего фрагмента РНК IBV только по одной позиции (+204 нуклеотида). В РНК IBV фрагмент, соответствующий фрагменту +202/+232 нуклеотида РНК SARS-CoV, образует верхнюю, наиболее консервативную часть шпильки s2m, в которой имеется несколько одноцепочечных участков [12]. Нижняя часть этой шпильки представляет собой двухцепочечный участок из пяти пар оснований без внутренних петель. Первичная последовательность нижнего участка неконсервативна, но в большинстве случаев в ней заменены не отдельные нуклеотиды, а пары нуклеотидов и сохранена вторичная структура.

В РНК SARS-CoV шпилька SL 6 (рис. 2) подобна по структуре шпильке s2m. Шпилька SL 6 также может образовываться и при сворачивании полной 3' UTR (рис. 3). Следовательно, коронавирус SARS, наряду с коронавирусом IBV, относится к тем немногим представителям семейства коронавирусов, которые содержат шпильку s2m в 3'-нетранслируемой области. Ряд коронавирусов группы I, по данным Джонассена и соавт. [12], и группы II, по нашим данным, этот мотив не содержат.

Таким образом, в 3' UTR вируса SARS-CoV имеются три консервативных структурных мотива (шпильки SL 2, SL 3 и SL 6). Информации о других вторичных структурах, показанных на рис. 2, в настоящее время нет, однако имеются сведения о некоторых первичных последовательностях, локализованных в 3' UTR коронавирусов.

Согласно литературным данным, для репликации дефектных интерферирующих РНК, представ-

ляющих собой геном с большими делециями, они должны содержать 3'-концевую область родительской геномной РНК длиной 417—463 нуклеотида (в случае коронавируса МНV), включающую всю 3' UTR ([6] и ссылки в в этой работе). Однако для синтеза «минус»-цепи РНК достаточно присутствия лишь 55 3'-концевых нуклеотидов и части 3'-концевой поли(А)-последовательности [13]. Предполагают, что концевые участки подобной длины нужны и для репликации недефектных вирусов.

Лиу и Джонсон ([9] и ссылки в этой работе) показали, что в 3′ UTR PHK MHV имеются два элемента, с которыми связываются клеточные белки: один расположен на 3′-конце области, а другой — примерно в ее середине. В настоящее время белки полностью не охарактеризованы. З′-Концевой элемент, связывающий клеточные белки, представляет собой 42 З′-концевых нуклеотида, т. е. находится в области геномной РНК, необходимой для синтеза «минус»-цепи [14].

На рис. 2 мы показали единственную локальную вторичную структуру (шпилька SL 9), образующуюся при сворачивании 3'-концевых фрагментов РНК SARS-CoV. Наряду с этим вариантом сворачивания программа предсказывает и другие варианты. В них 42 3'-концевых нуклеотида участвуют во взаимодействии с отдаленными участками 3' UTR, как это, например, видно на рис 3. Однако роль шпильки SL 9 неясна, поскольку в настоящее время неизвестно, важна ли вторичная структура для связывания клеточных белков с 3'-концом геномных РНК коронавирусов.

По данным авторов работы [15], место связывания клеточных белков, расположенное примерно в середине 3' UTR PHK MHV, содержит последовательность UGAGAGAGUU, существенную для связывания этих белков и репликации вируса. 3'-Концевое место связывания также содержит подобную последовательность — UGAAUGAAGUU [14, 16]. В случае РНК ВСУ 3'-концевая последовательность идентична последовательности в РНК MHV, а последовательность, расположенная в середине 3' UTR, несколько отличается (UUGGAGA-AAGU) [17]. Мы полагаем, что в случае вируса SARS-CoV одно из мест связывания клеточных белков, возможно, находится в шпильке SL 5 (UUGAAAGAGCC), учитывая, что 3' UTR этого вируса содержит s2m мотив, отсутствующий в РНК MHV и BCV.

В последовательности UGAGAGAAGUU РНК МНУ содержится фрагмент AGAAG, необходимый, по данным авторов работы [18], для связывания клеточного белка гяРНП A1. Авторы [18] полагают, что взаимодействие этого белка, связывающе-

гося с 3' UTR PHK MHV, с белком PTB (polypyrimidine tract-binding protein), связывающимся с 5' UTR этой PHK, обеспечивает контакт между двумя концами геномной PHK, который может быть существенным для репликации и транскрипции PHK.

На рис. 2 в шпильке SL 7 отмечена последовательность GGAAGAGC. По литературным данным [9], это единственный элемент, присутствующий в 3' UTR PHK коронавирусов всех трех групп. Его функция неизвестна.

В заключение отметим, что, с одной стороны, 3′ UTR PHK вируса SARS-CoV содержит мотив s2m, присутствующий в PHK вируса группы III (IBV) и отсутствующий в ряде PHK вирусов первой и второй групп, а, с другой, — эта область содержит псевдоузел, значительно отличающийся по своим характеристикам от псевдоузла, образуемого PHK IBV, и близкий к псевдоузлам, образующимся в PHK вирусов группы I. В результате, предложенная нами вторичная структура З′ UTR PHK вируса SARS-CoV поддерживает существующую точку зрения [1-3] о том, что этот вирус является уникальным и не может быть отнесен ни к одной из трех существующих групп коронавирусов.

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

Conservative structural motifs in the 3' untranslated region of SARS coronavirus

Summary

Signal elements in the 3' untranslated region (UTR) of «plus»-RNA viruses function as control elements in RNA replication, transcription and translation. Here we performed computer-assisted (Zuker, 2003) secondary structure analysis and prediction of tertiary structure of 3' UTR of SARS coronavirus. We found that this region contains a sequence potentially able to form a pseudoknot which was earlier observed in 3' UTR of every coronavirus genome that was sequenced. The SARS-CoV pseudoknot structure is similar to the pseudoknot structure of the group 1 coronaviruses. As in the case of other coronaviruses the formation of the SARS-CoV pseudoknot interferes with the formation of a bulged stem-loop structure with high negative free energy value located in the most upstream part of 3' UTR. Besides the SARS-CoV 3' UTR, like the same region of IBV coronavirus, contains the s2m motif which is absent in 3' UTR of a number of group 1 and 2 coronaviruses. Other structural motifs in 3' UTR of SARS-CoV has been discussed as well. As a result, the SARS-CoV 3' UTR structure established supports the present opinion on a unique character of this virus which cannot be assigned to any of three known groups of coronaviruses.

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

Консервативні структурні мотиви в 3'-нетрансльованій області геномної PHK вірусу SARS-CoV

Резюме

Сигнальні послідовності в 3'-нетрансльованій області (UTR)

«плюс»-РНК вірусів контролюють процеси реплікації, транскрипції та трансляції РНК. У цій роботі методом комп'ютерного моделювания (Zuker, 2003) вперше вивчено вторинну і третинну структуру З' UTR геномної РНК вірусу SARS-CoV. Показано, що ця область містить послідовність, яка утворює псевдовузол, найближчий за свою структурою до псевдовузлів, виявлених в РНК коронавірусів групи І. Як і у випадку інших коронавірусів, послідовність, що відповідає одному з ланцюгів стебла І псевдовузла РНК вірусу SARS-CoV, може брати участь в утворенні альтернативного мотивувисокостибільної шпилькоподібної структури. Окрім того, 3' UTR PHK вірусу SARS-CoV містить мотив s2m, виявлений в аналогічній області вірусу IBV (коронавірус групи III), але відсутній в низці РНК коронавірусів груп І і ІІ. Обговорюються також інші структурні мотиви, які утворюються в 3' UTR PHK вірусу SARS-CoV. Отримані нами результати підтримують існуючу точку зору про те, що вірус SARS-CoV є унікальним і не може бути віднесений до жодної з трьох відомих груп коронавірусів.

## СПИСОК ЛИТЕРАТУРЫ

- 1. Ksiazek T. G., Erdman D., Goldsmith C. S., Zaki S. R., Peret T., Emery S., Tong S., Urbani C., Comer J. A., Lim W. A novel coronavirus associated with severe acute respiratory syndrome // N. Engl. J. Med.—2003.—348, N 20.—P. 1953—1966.
- Rota P. A., Oberste M. S., Monroe S. S., Nix W. A., Campagnoli R., Icenogle J. P., Peñaranda S., Bankamp B., Maher K., Chen M.-H. Characterization of a novel coronavirus associated with severe acute respiratory syndrome // Science.—2003.—In press (Publ. online 1 May 2003; 10.1126/science.1085952.)
- 3. Marra M. A., Jones S. J. M., Astell C. R., Holt R. A., Brooks-Wilson A., Butterfield Y. S. N., Khattra J., Asano J. K., Barber S. A., Chan S. Y. The genome sequence of the SARS-associated coronavirus // Science.—2003.—In press (Publ. online 1 May 2003; 10.1126/science.1085953.)
- Zuker M. Mfold web server for nucleic acid folding and hybridization prediction // Nucl. Acids Res.—2003.—31, N 13.—P. 1—10.
- Sawicki S. G., Sawicki D. L. A new model for coronavirus transcription // Adv. Exp. Med. Biol.—1998.—440.— P. 215—219.
- Hsue B., Hartshorne T., Masters P. S. Characterization of an essential RNA secondary structure in the 3' untranslated region of the murine coronavirus genome // J. Virol.—2000.—74, N. 15.—P. 6911—6921.
- Hsue B., Masters P. S. A bulged stem-loop structure in the 3' untranslated region of the genome of the coronavirus mouse

- hepatitis virus is essential for replication // J. Virol.—1997.—71, N 10.—P. 7567—7578.
- Williams G. D., Chang R.-Y., Brian D. A. A phylogenetically conserved hairpin-type 3' untranslated region pseudoknot functions in coronavirus RNA replication // J. Virol.—1999.—73, N 10.—P. 8349—8355.
- Liu Q., Johnson R. F., Leibowitz J. L. Secondary structural elements within the 3' untranslated region of mouse hepatitis virus strain JHM genomic RNA // J. Virol.—2001.—75, N 24.—P. 12105—12113.
- Berkhout B., Ooms M., Beerens N., Huthoff H., Southern E., Verhoef K. In vitro evidence that the untranslated leader of the HIV-1 genome is an RNA checkpoint that regulates multiple functions through conformational changes // J. Biol. Chem.— 2002.—277. N 22.—P. 19967—19975.
- Huthoff H., Berkhout B. Multiple secondary structure rearrangements during HIV-1 RNA dimerization // Biochemistry.— 2002.—41, N 33.—P. 10439—10445.
- 12. Jonassen C. M., Jonassen T. Ø., Grinde B. A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus // J. Gen. Virol.—1998.—79, pt 4.—P. 715—718.
- Lin Y.-J., Liao C.-L., Lai M. M. C. Identification of the cis-acting signal for minus-strand RNA synthesis of a murine coronavirus: implications for the role of minus-strand RNA in RNA replication and transcription // J. Virol.—1994.—68, N.12.—P. 8131.—8140
- 14. Yu W., Leibowitz J. L. Specific binding of host cellular proteins to multiple sites within the 3' end of mouse hepatitis virus genomic RNA // J. Virol.—1995.—69, N 4.—P. 2016—2023.
- Liu Q., Yu W., Leibowitz J. L. A specific host cellular protein binding element near the 3' end of mouse hepatitis virus genomic RNA // Virology.—1997.—232, N 1.—P. 74—85.
- 16. Yu W., Leibowitz J. L. A conserved motif at the 3' end of mouse hepatitis virus genomic RNA required for host protein binding and viral RNA replication // Virology.—1995.—214, N 1.—P. 128—138.
- Spagnolo J. F., Hogue B. G. Host protein interactions with the 3' end of bovine coronavirus RNA and the requirement of the poly (A) tail for coronavirus defective genome replication // J. Virol. —2000. —74. N 11.—P. 5053—5065.
- Virol. —2000. —74, N 11. —P. 5053—5065.
  18. Huang P., Lai M. M. C. Heterogeneous nuclear ribonucleo-protein A1 binds to the 3' untranslated region and mediates potential 5'-3'-end cross talks of mouse hepatitis virus RNA // J. Virol. —2001. —75, N 11. —P. 5009—5017.

УДК 577.21

Надійшла до редакції 08.05.03