

Біоінформатичний аналіз інвертованих повторів геному коронавірусів

О. Ю. Лиманська^{1,2}

¹ДУ «Інститут мікробіології та імунології ім. І. І. Мечнікова АМН України»
Вул. Пушкінська, 14, Харків, Україна, 61057

²Національний науковий центр «Інститут експериментальної і клінічної ветеринарної медицини» УААН
Вул. Пушкінська, 83, Харків, Україна, 61023

olga.limanskaya@mail.ru

Мета. Створення карт локалізації досконалих і недосконалих потенційних шпилькових структур у геномі коронавірусів людини і тварин. **Методи.** Біоінформатичний аналіз нуклеотидних послідовностей коронавірусів, атомно-силова мікроскопія. **Результати.** Визначено термодинамічно стабільні досконалі та недосконалі інвертовані повтори, які утворюють шпилькові структури, що можуть виникати у геномній РНК коронавірусів людини і тварин – вірусів тяжкого гострого респіраторного синдрому, гепатиту миші, епідемічної діареї свині, трансмісивного гастроентериту та бичачого коронавірусу. Створено карти локалізації шпильок (які є одним із ланцюгів сигнальних механізмів функціонування геному) на геномі коронавірусів. **Висновки.** Основними сайтами локалізації потенційно можливих консервативних структурних мотивів є гени реплікази та глікопротеїнів шпильок коронавірусів. Шпилькові структури є консервативними елементами всередині набору ізолятів одного виду коронавірусів.

Ключові слова: вірус тяжкого гострого респіраторного синдрому, коронавірус, шпилькова структура, інвертований повтор.

Вступ. Хрестоподібні структури, які можуть формуватися в нуклеїнових кислотах інвертованими повторами, поряд з іншими неканонічними утвореннями є важливими елементами геному та відіграють певну біологічну роль. Вважають, що вони залучені до регуляції реплікації ДНК і транскрипції [1, 2].

Методом флуоресцентної проточної цитометрії визначено, що в ядрі знаходиться $\sim 10^5$ хрестоподібних структур [3].

Досконалі та недосконалі інвертовані повтори, окрім того, що відіграють певну роль у мутагенезі, асоційовані з низкою генетичних захворювань лю-

дини (спадковий ангіоневроз, дефіцит антитромбіну, дефіцит сироваткової холінестерази) [4].

Незважаючи на досить інтенсивне вивчення локалізації паліндромів у геномі різних організмів, роль і розподіл шпилькових структур у геномі вірусів і бактерій залишаються нез'ясованими. Виходячи з цього, нами здійснено пошук потенційних шпилькових структур у геномі коронавірусів. До родини коронавірусів (CoV) належать вірус епідемічної діареї свині, вірус інфекційного бронхіту, вірус гепатиту миші, вірус трансмісивного гастроентериту свині. Існують також коронавіруси людини, великої рогатої худоби, коня, кішки. За даними секвенування, вірус тяжкого гострого респіраторного синдрому людини (ТГРС-

CoV) також можна віднести до ряду *Nidovirales* родини *Coronaviridae* роду *Coronavirus*.

Коронавіруси підрозділяються на три серогрупи, кожна з яких має перехресну серологічну реактивність та схожу організацію геному. Усі знані на цей час коронавіруси людини належить до груп I і II. ТГРС-CoV утворює нову групу IV, оскільки проведені генетичні та антигенні дослідження продемонстрували його віддаленість від усіх відомих груп коронавірусів [5].

У даній роботі знайдено розподіл досконалих і недосконалих інвертованих повторів у геномі деяких коронавірусів, до яких відносять і вкрай небезпечний для людини вірус тяжкого гострого респіраторного синдрому. На основі аналізу отриманих карт локалізації потенційних шпилькових структур показано, що розподіл інвертованих повторів збігається для ізолятів всередині одного виду коронавірусів.

Отже, порівняння розподілу шпилькових структур може слугувати ще одним інструментом (поряд з філогенетичним аналізом) дослідження еволюційних взаємовідносин та геномної організації не тільки коронавірусів, а й представників інших видів.

Матеріали і методи. У роботі використано повністю секвеновані послідовності ізолятів вірусу тяжкого гострого респіраторного синдрому (ТГРС, SARS-CoV) (номери AY27848, AY279354, AY268070 для бази даних GenBank), бичачого коронавірусу (AF220295), вірусу трансмісивного гастроентериту (NC_002306), вірусу інфекційного бронхіту (AY251817, AY251816, AF391157, AF391156, AF391154, AY237817, AY223860, AF470630, AF470629, AF470628, AF467921), вірусу епідемічної діареї свині (NC_003436), коронавірусу миші (NC_001846), коронавірусу кішки (AY204523, AY204524, AY204525), а також плазмиди *pGEMEX* (X65317).

Для пошуку досконалих і недосконалих інвертованих повторів та визначення їхніх термодинамічних характеристик застосовували відповідно програми Oligo (версія 3.4) [6] та RNA2 пакета GeneBee [7].

Використано атомно-силовий мікроскоп (АСМ) Nanoscope III з D-сканером (Veeco Instruments Inc.,

США). АСМ-зображення зразка суперспіральної ДНК плазмиди *pUC8* (довжина 2665 п. н.) після нанесення на стандартну амінослоду записували у повітрі в режимі «висота» із застосуванням віброуючого варіанта АСМ та незагострених зондів фірми КТЕК International (Російська Федерація) з резонансною частотою 300–360 кГц. Зразок готували відповідно до раніше наведеної методики [8].

Результати і обговорення. Відомо, що хресто-подібні структури можуть бути включені до промоторів та областей термінації транскрипції – наявність хрестів є сигналом для зупинки РНК-полімерази, термінації синтезу РНК-транскриптів з наступною дисоціацією комплексу РНК-полімераза–ДНК–РНК-транскрипт. Одним із зазначених термінаторів транскрипції для T7 РНК-полімерази є область термінації транскрипції плазмиди *pGEMEX* – внутрішній термінатор транскрипції довжиною ~ 90 п. н., ефективність якого становить 70–80 % [9]. Проаналізувавши область термінації транскрипції ДНК *pGEMEX* на наявність термодинамічно стабільних інвертованих повторів, ми знайшли недосконалий інвертований повтор довжиною 28 п. н., вільна енергія якого – G становить 11,2 ккал/моль. Існування термінації транскрипції T7 РНК-полімерази при елонгації транскрипції на матриці ДНК *pGEMEX*, що містить даний інвертований повтор в області термінатора, раніше продемонстровано *in vitro* [10]. Тому, виходячи з параметрів шпильки ДНК *pGEMEX* та літературних даних щодо параметрів шпильок хрестоподібних структур, які спостерігали в експериментах *in vivo* [11] та *in vitro* [12], для подальшого аналізу обрано шпильку, довжина петлі яких не перевищувала, як правило, 5 нуклеотидів, а мінімальна енергія – G становила ~ 9 ккал/моль.

На основі визначених потенційних (тобто термодинамічно стабільних) шпильок у геномі вірусу ТГРС (таблиця) побудовано діаграму їхнього розподілу на фізичній карті геному ТГРС (рис. 1). Зазначимо, що виявлені шпильки є консервативними структурними мотивами для вірусу ТГРС. Порівняння їхньої локалізації на геномі декількох ізолятів ТГРС показало, що положення їх для переважної більшості шпильок збігається. Це, на наш погляд, може вказувати на певну роль шпилькових

Досконалі та недосконалі термодинамічно стабільні шпилькоподібні структури, що можуть бути потенційно утворені інвертованими повторами, в геномній РНК вірусу тяжкого гострого респіраторного синдрому (номер AY291451 для бази даних GenBank)

№	Довжина стебла, п. н.	Довжина петлі, н.	-G, ккал/моль	Положення на геномі	Білок	Тип повтору
1	15	4	16,9	2570–2603	Репліказа А	Недосконалий-2
2	14	4	12,6	2959–2990	Репліказа А	Досконалий-1
3	17	3	14,5	2894–2929	Репліказа А	Недосконалий-1
4	15	4	11,5	3261–3294	Репліказа А	Недосконалий-1
5	10	3	12,7	3481–3503	Репліказа А	Досконалий-1
6	16	4	12,3	4051–4086	Репліказа А	Досконалий-1
7	15	4	14,6	5668–5701	Репліказа А	Недосконалий-1
8	12	6	15,1	5338–5367	Репліказа А	Недосконалий-2
9	20	5	17,2	6983–7028	Репліказа А	Недосконалий-2
10	11	5	11,5	7140–7166	Репліказа А	Досконалий-1
11	11	4	10,3	8634–8659	Репліказа А	Досконалий-1
12	14	4	12,3	10921–10952	Репліказа А	Досконалий-1
13	12	7	10,3	12314–12344	Репліказа А	Недосконалий-1
14	11	3	10,7	13363–13387	Репліказа А	Недосконалий-1
15	13	5	13,8	13404–13434	Репліказа А	Недосконалий-1
16	14	4	16,3	13888–13919	Репліказа Б	Недосконалий-2
17	16	3	15,7	13945–13979	Репліказа Б	Недосконалий-2
18	10	3	12,7	16460–16482	Репліказа Б	Недосконалий-1
19	11	4	11,4	17166–17191	Репліказа Б	Недосконалий-1
20	17	4	16,1	17521–17558	Репліказа Б	Недосконалий-2
21	8	3	11,5	20230–20248	Репліказа Б	Досконалий-1
22	15	6	11,3	21376–21411	Репліказа Б	Недосконалий-1
23	11	4	11,8	23447–23472	Глікопротеїн S	Недосконалий-1
24	17	5	16,9	23636–23674	Глікопротеїн S	Недосконалий-2
25	19	4	16,6	24075–24116	Глікопротеїн S	Недосконалий-2
26	10	6	10,4	26102–26127	Ген E	Досконалий-1

Примітка. Типи повторів недосконалий-1 (досконалий-1), недосконалий-2 відповідають шпилькам із значенням енергії (-G) понад 10 і 15 ккал/моль відповідно. Виділено позиції шпильок, послідовності та вторинні структури яких наведено на рис. 2.

структур у ланцюгу сигнальних механізмів функціонування вірусу ТГРС.

Усі проаналізовані повтори розподілили на два види – досконалі і недосконалі (стебло яких містить

некомплементарні нуклеотиди або делеції нуклеотидів в одному з ланцюгів стебла шпильки). Окрім того, повтори диференціювали на три групи за рівнем енергії. Першу групу склали повтори з енер-

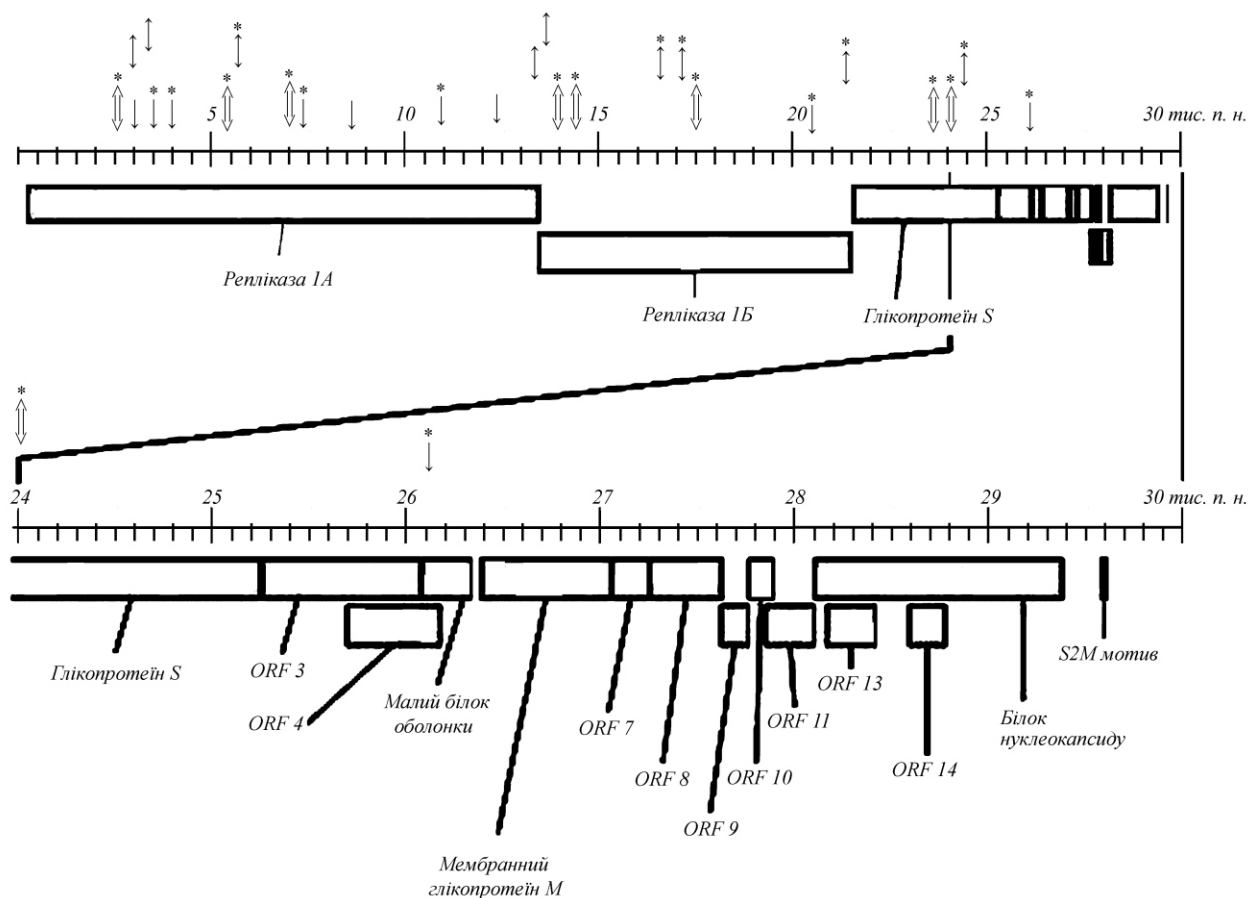


Рис. 1. Фізична карта вірусу тяжкого гострого респіраторного синдрому (номер AY291451) з наведеними позиціями відомих генів. Стрілки вказують на позиції визначених термодинамічно стабільних досконалих і недосконалих шпилькових структур; зірочками позначено шпилькові структури, положення яких збігається з положенням останніх для іншого ізоляту вірусу ТГРС (номер AY278488); \downarrow , \updownarrow – недосконалі шпильки з енергією – G понад 10 і 15 ккал/моль відповідно; \dashv – досконалі шпильки з енергією понад 10 ккал/моль

гією ($-G$) 10–15, другу – 15–20, третю – 20 ккал/моль.

Послідовність і вторинну структуру двох типів недосконалих інвертованих повторів, енергія яких становить понад -15 ккал/моль, наведено на рис. 2. Аналогічним чином нами отримано діаграми розподілу та параметри шпилькоподібних структур для бичачого коронавірусу (рис. 3), коронавірусу миші (рис. 4, *a*), вірусу епідемічної діареї свині (рис. 4, *б*) та вірусу трансмісивного гастроентериту свині (рис. 4, *в*).

Варто відмітити, що приблизно дві третини геному коронавірусів є матрицею для синтезу репліказ 1А та 1Б, одна третина геному кодує структурні (білок нуклеопротеїну, зовнішні глікопротеїни *S*, *M* і *E*), а також низку неструктурних білків (рис. 1).

Серед проаналізованих нами послідовностей ізолятів коронавірусів тварин і вірусу ТГРС найбільшу «схильність» до утворення шпилькоподібних структур виявлено для вірусу ТГРС – показано можливість утворення до 26 для одного ізоляту (таблиця). При цьому переважна більшість шпилькоподібних структур (20 з 26) може утворюватися в генах, що кодують репліказу (рис. 1).

У роботі [13] за допомогою комп'ютерного моделювання (програма для передбачення вторинної структури РНК) досліджено вторинну та третинну структури 3'-нетрансльованої області (UTR, untranslated region) геномної РНК ТГРС-CoV, структурні елементи якої відіграють важливу роль у реплікації вірусів, та порівняно зі структурами цих же областей раніше охарактеризованих коронавірусів.

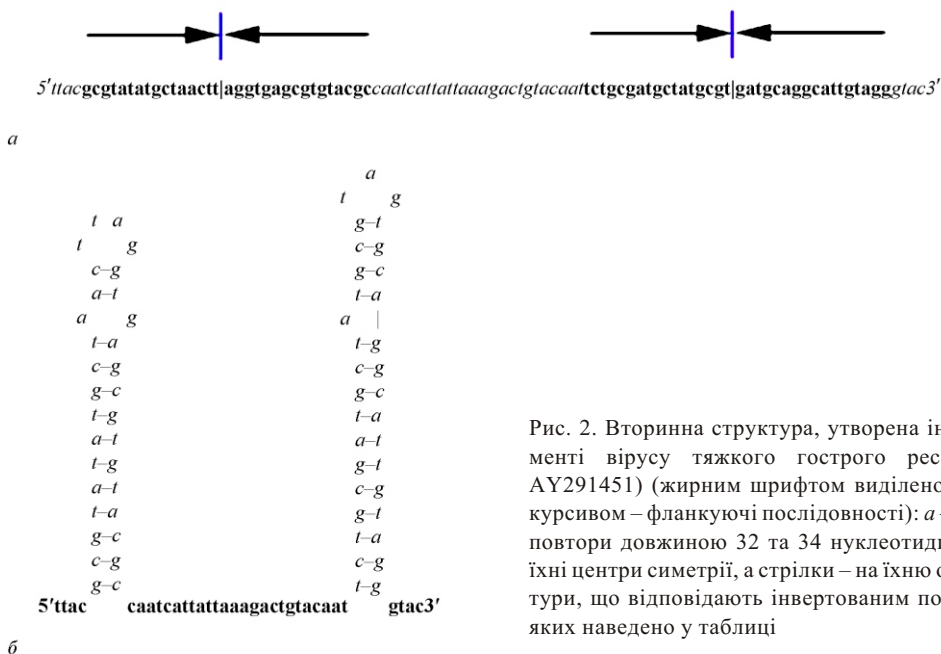


Рис. 2. Вторинна структура, утворена інвертованими повторами у фрагменті вірусу тяжкого гострого респіраторного синдрому (номер АУ291451) (жирним шрифтом виділено комплементарні послідовності; курсивом – фланкуючі послідовності): а – два недосконалих інвертованих повтори довжиною 32 та 34 нуклеотиди (вертикальні лінії вказують на їхні центри симетрії, а стрілки – на їхню орієнтацію); б – шпилькові структури, що відповідають інвертованим повторам N 16 та N 17, параметри яких наведено у таблиці

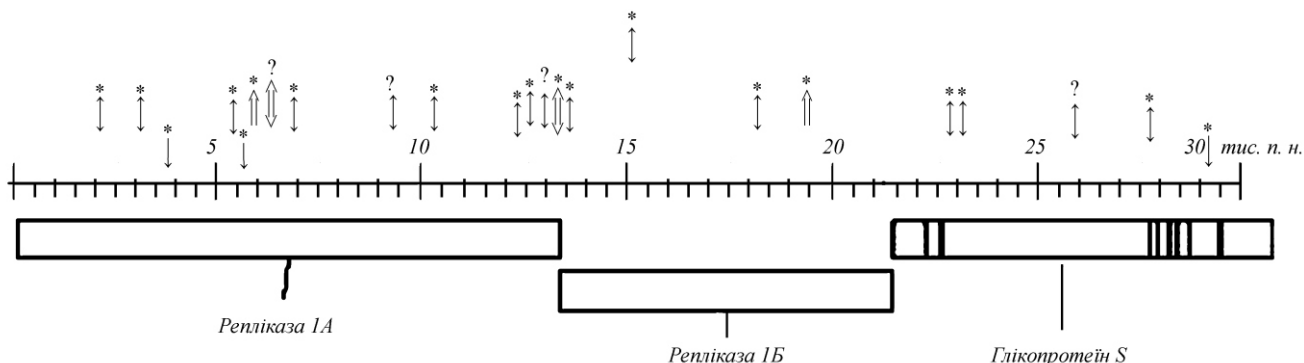


Рис. 3. Фізична карта бичачого коронавірусу (номер NC_003405) з наведеними позиціями відомих генів. Стрілки вказують на позиції шпилькових структур; зірочками позначено шпилькові структури, положення яких збігається з положенням останніх для іншого ізоляту бичачого коронавірусу (номер AF220295); знак запитання відповідає шпильковим структурам, положення яких не збігаються для двох зазначених ізолятів бичачого коронавірусу; \updownarrow , \updownarrow – недосконалі шпильки з енергією ($-G$) понад 10, 15 і 20 ккал/моль відповідно; \updownarrow – досконалі шпильки з енергією понад 10 ккал/моль

Показано, що в 3' UTR є три консервативних структурних мотиви – шпилькоподібні структури та єдина локальна вторинна шпилькоподібна структура, яка утворюється при згортанні 3'-кінцевих фрагментів РНК ТГРС-CoV.

Зазначимо, що структура шпильки, яку визначають за допомогою комп'ютерного аналізу, істотним чином залежить від алгоритму використаного пошуку, а також від параметрів шпильки. Тому на відміну від авторів роботи [13] ми зосередилися на пошуку термодинамічно стабільних досконалих і

недосконалих повторів, тобто тих, що реально спостерігали у попередніх експериментах. Ми не розглядали недосконалі повтори, у яких було більше двох пар неспарених нуклеотидів, а розмір петлі перевищував шість нуклеотидів. Прикладом такої термодинамічно стабільної шпильки є хрестоподібна структура, утворена двома шпильковими структурами у суперспіральній плазміді *pUC8* (рис. 5). Плазмід *pUC8* містить декілька інвертованих повторів, які можуть утворювати хрестоподібні структури. Вільна енергія G найстабільнішої з них

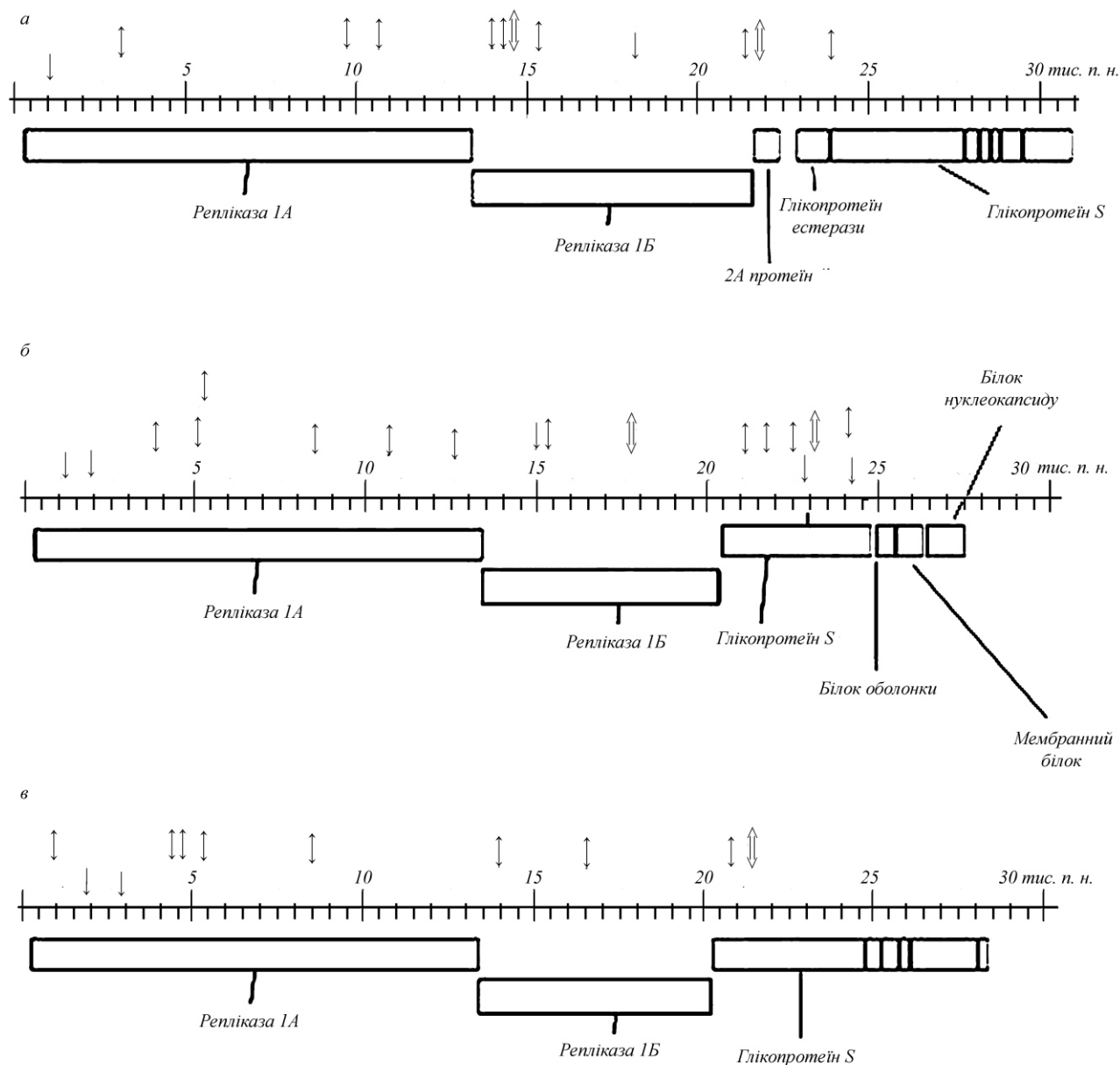


Рис. 4. Фізична карта вірусів гепатиту миші (NC_001846) (а), епідемічної діареї свині (номер NC_003436) (б) та трансмісивного гастроентериту (номер NC_002306) (в) з наведеними позиціями відомих генів. Стрілки вказують на позиції шпилькових структур; \updownarrow , \updownarrow – недосконалі шпильки з енергією понад 10 і 15 ккал/моль відповідно; \uparrow – досконалі шпильки з енергією понад 10 ккал/моль

(показаної стрілками на рис. 5) становить $-17,8$ ккал/моль, стебло шпильки утворюють 11 п. н., а петля має довжину 4 нуклеотиди. Проте в програмі RNA2 пакета GeneVee, яку ми використали для передбачення вторинної структури РНК коронавірусів, пара G–T не вважається некомплементарною. Дійсно, утворення неутсон-криківської пари G–T стає можливим за рахунок формування рідкісних таутомерних енольних та іміноформ нук-

леотидів [14]. Наведене АСМ-зображення хрестоподібної структури ДНК *pUC8* показує, що серед декількох паліндромів, які містить ДНК *pUC8*, за умов *in vitro* реалізується лише одна хрестоподібна структура, термодинамічно найстабільніша.

Для геномної РНК бичачого коронавірусу показано можливість утворення 23 таких термодинамічно стабільних консервативних структурних мотиви (рис. 3). Положення переважної більшості з

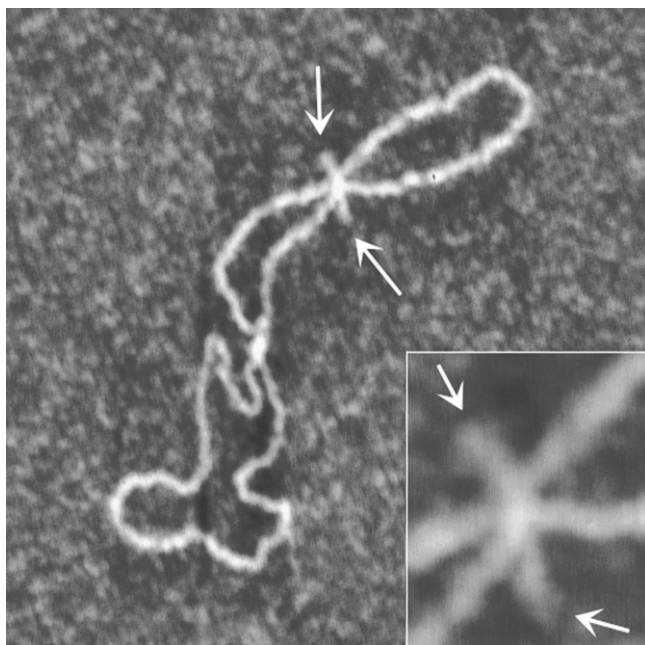


Рис. 5. АСМ-зображення суперспіральної плазмиди *pUC8* у повітрі. Розмір кадра – 372 нм × 372 нм. Стрілками показано дві шпильки, які утворили хрестоподібну структуру. Розмір вставки із збільшеним зображенням шпильок 55 нм × 62 нм

них збігається з положенням шпильок для іншого ізоляту бичачого коронавірусу (позначено на рис. 3), як це встановлено і для ізолятів вірусу ТГРС.

За допомогою аналізу геномної РНК вірусу гепатиту миші (рис. 4, а) у даній послідовності знайдено 12 шпилькоподібних структур. У результаті вивчення геномної РНК вірусу епідемічної діареї свині визначено можливість утворення 18 шпильок (рис. 4, б), причому 10 із них локалізовані в області гена, який кодує репліказу.

Дослідження повної послідовності геномної РНК іншого коронавірусу свині – вірусу трансмісивного гастроентериту – свідчить про можливість існування 11 шпилькоподібних структур, дев'ять з яких також розташовані у гені, що кодує репліказу (рис. 4, в).

Складення карт розподілу інвертованих повторів підіймає декілька запитань. По-перше, для ініціації та термінації транскрипції достатньо двох шпильок на 5'- та 3'-кінцях матричного ланцюга ДНК. У той же час, зокрема, послідовність гена, що кодує репліказу А бичачого коронавірусу, містить 14 шпильок. Отже, біологічна функція переважної більшості знайдених шпильок є невизначеною.

По-друге, відсутність шпильок, наприклад, на 5'-кінці гена реплікази А вірусу ТГРС (рис. 1), бичачого коронавірусу (рис. 3) свідчить про те, що поряд із шпильками інші неканонічні структури ДНК (зокрема триплекси) можуть слугувати сигналами для зв'язування ферментів. По-третє, вірус ТГРС відрізняється від інших коронавірусів не тільки за якісним характером розподілу шпильок, а й за їхніми кількісними параметрами. Наприклад, кількість високостабільних недосконалих інвертованих повторів з енергією – G понад 15 ккал/моль для вірусу ТГРС становить сім, у той час як для бичачого коронавірусу знайдено лише два повтори з вільною енергією – G понад 20 ккал/моль та один повтор з енергією понад 15 ккал/моль. Вищевикладене свідчить про можливість використання розподілу термодинамічно стабільних інвертованих повторів для структурної диференціації вірусів.

Таким чином, проведенням комп'ютерним аналізом ізолятів різних коронавірусів тварин та ТГРС визначено, що основними сайтами локалізації потенційно можливих консервативних структурних мотивів є гени реплікази та глікопротеїнів шипів коронавірусів. Ми вважаємо, що порівняння розподілу шпилькових структур може слугувати ще одним інструментом (поряд з філогенетичним аналізом) дослідження еволюційних взаємовідносин та геномної організації не тільки коронавірусів, а й представників інших видів. Крім того, отримані карти розподілу шпильок можуть стати важливими для подальшого вивчення молекулярних механізмів і регуляторної ролі таких альтернативних структур, як шпилькові та хрестоподібні структури, які інтенсивно обговорюють в літературі [15, 16].

Роботу частково підтримано грантами АМН 47/2002 та АМН 72/2007 від Академії медичних наук України.

O. Yu. Limanskaya

Bioinformatic analysis for inverted repeats of coronavirus genome

Summary

Thermodynamically stable inverted repeats forming hairpin-loop structures that can arise in the genomic RNA of the human and animal coronaviruses are determined. The diagrams of hairpin localization (that are one of the chains of genome signaling) on the

genome of coronaviruses are obtained. The hairpins are shown to be conservative structural elements inside the set of coronavirus isolates for one species.

Keywords: severe acute respiratory syndrome, coronavirus, hairpin-loop structure, inverted repeat.

О. Ю. Лиманская

Биоинформатический анализ инвертированных повторов генома коронавируса

Резюме

Цель. Создание карт локализации совершенных и несовершенных потенциальных шпилечных структур в геноме коронавируса человека и животных. **Методы.** Биоинформатический анализ нуклеотидных последовательностей коронавируса, атомно-силовая микроскопия. **Результаты.** Определены термодинамически стабильные совершенные и несовершенные инвертированные повторы, образующие шпилечные структуры, которые могут возникать в геномной РНК коронавируса человека и животных – вирусов тяжелого острого респираторного синдрома, гепатита мыши, эпидемической диареи свиньи, трансмиссивного гастроэнтерита и бычьего коронавируса. Созданы карты локализации шпилек (являющихся одной из цепей сигнальных механизмов функционирования генома) на геноме коронавируса. **Выводы.** Основными сайтами локализации потенциально возможных консервативных структурных мотивов служат гены репликазы и гликопротеинового шипа коронавируса. Шпилечные структуры являются консервативными элементами внутри набора изолятов одного вида коронавируса.

Ключевые слова: вирус тяжелого острого респираторного синдрома, коронавирус, шпилечная структура, инвертированный повтор.

ПЕРЕЛІК ЛІТЕРАТУРИ

1. McClellan J., Boublikova P., Palecek E., Lilley D. Superhelical torsion in cellular DNA response directly to environmental and genetic factors // Proc. Nat. Acad. Sci. USA.–1990.–**87**, N 21.–P. 8373–8377.
2. Bagga R., Ramesh N., Brahmachari S. Supercoil-induced unusual DNA structures as transcriptional block // Nucl. Acids Res.–1990.–**18**, N 11.–P. 3363–3369.
3. Ward G., McKenzie R., Zannis-Hadjopoulos M., Price G. The dynamic distribution and quantification of DNA cruciforms

in eukaryotic nuclei // Exp. Cell Res.–1990.–**188**, N 2.–P. 235–246.

4. Bessler J. DNA inverted repeats and human disease // Frontiers in Biosci.–1998.–N 3.–P. d408–d418.
5. Poon L., Guan Y., Nicholls J., Yuen K., Peiris J. The aetiology, origins, and diagnosis of severe acute respiratory syndrome // Lancet Infect. Dis.–2004.–**4**, N 11.–P. 663–671.
6. Rychlik W., Spencer W., Rhoads R. Optimization of the annealing temperature for DNA amplification *in vitro* // Nucl. Acids Res.–1990.–**18**, N 21.–P. 6409–6412.
7. Brodsky L., Drachev A., Tatuzov R., Chumakov K. The package of programs for biopolymer sequence analysis: GeneBee // Biopolymers and Cell.–1991.–**7**, N 1.–P. 10–14.
8. Limanskii A. P. Study of cruciform structure in supercoiled pUC8 plasmid DNA by atomic force microscopy and computer modelling // Biopolymers and Cell.–2002.–**18**, N 5.–P. 401–405.
9. pGEMEX-1 and pGEMEX-2 vectors // Techn. Bull., Promega.–2000.–N 253.–P. 1–13.
10. Limanskii A. P. Visualization of DNA–T7 RNA polymerase complex by atomic force microscopy // Biopolymers and Cell.–2007.–**23**, N 1.–P. 3–13.
11. Panayotatos N., Fontaine A. A native cruciform DNA structure probed in bacteria by recombinant T7 endonuclease // J. Biol. Chem.–1987.–**262**, N 23.–P. 11364–11368.
12. Panyutin I., Klishko V., Lyamichev V. Kinetics of cruciform formation and stability of cruciform structure in superhelical DNA // J. Biomol. Struct. and Dyn.–1984.–**1**, N 4.–P. 1311–1324.
13. Zarudnaya M. I., Potyahaylo A. L., Hovorun D. M. Conservative structural motifs in the 3' untranslated region of SARS coronavirus // Biopolymers and cell.–2003.–**19**, N 3.–P. 298–303.
14. Saenger W. Principles of nucleic acids structure.–New York etc.: Springer, 1984.–556 p.
15. Odynets K. A., Kornelyuk A. I. Molecular aspects of organization and expression of SARS-CoV coronavirus genome // Biopolymers and Cell.–2003.–**19**, N 5.–P. 414–431.
16. Voineagu I., Narayanan V., Lobachev K., Mirkin S. Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins // Proc. Nat. Acad. Sci. USA.–2008.–**105**, N 29.–P. 9936–9941.

УДК 577.2:577.32
Надійшла до редакції 15.10.08