

# Молекулярні аспекти будови і експресії геному коронавірусу SARS

К. О. Одинець, О. І. Корнелюк

Інститут молекулярної біології і генетики НАН України  
Вул. Академіка Заболотного, 150, Київ, 03143, Україна  
E-mail: kornelyuk@imbg.org.ua

*Розглянуто молекулярні аспекти будови геному коронавірусу SARS-CoV — збудника атипової пневмонії, або тяжкого гострого респіраторного синдрому. Наведено характеристику коронавірусів і будову віріона. Аналізуються дані стосовно 36 повністю секвенованих геномів різних ізолятів SARS-CoV та результати молекулярного філогенетичного вивчення. Описано передбачені властивості восьми субгеномних мРНК та 14 відкритих рамок зчитування. Охарактеризовано синтез полібілків, їхній процесинг та зрілі білки SARS-CoV. Обговорюються властивості передбачених білків SARS-CoV та їхні функції. Поверхневий глікопротеїн S-білок є одним з основних антигенів SARS-CoV і відіграє важливу роль у взаємодії вірусу з клітинним рецептором. Описано також потенційні сайти зв'язування S-білка з можливим рецептором — амінопептидазою hAPN з використанням біоінформаційних та структурних підходів. Представлено будову основної M<sup>pro</sup> (3CL<sup>pro</sup>) протеїнази як потенційної мішені для антивірусної терапії, моделювання її просторової структури, будову активного центра, скринінг та дизайн потенційних інгібіторів SARS-CoV.*

Вступ. Перші випадки нового інфекційного захворювання людини SARS (severe acute respiratory syndrome — тяжкий гострий респіраторний синдром, або «атипова пневмонія»), були зареєстровані в середині листопада 2002 р. у провінції Гуандун (Китай) [1—6]. За короткий строк епідемія SARS охопила понад 32 країни світу, переважно Південної Азії. За переглянутими даними ВОЗ від 26 вересня 2003 р., за період з 1 листопада 2002 р. по 31 липня 2003 р. у світі зареєстровано 8098 ймовірних випадків SARS із 774 смертельними наслідками, що становить 9,6 % [7].

В Україні випадків атипової пневмонії не виявлено. У середині березня 2003 р. в декількох лабораторіях було виділено інфекційний агент, асоційований з SARS, і встановлено, що він є новим для людської популяції представником родини коронавірусів [1—6]. Цей вірус отримав назву SARS-асоційованого коронавірусу людини (SARS-CoV, інше скорочення SCoV). За короткий строк після ізоляції вірусу, 14 квітня 2003 р., було просеквено-

вано перші повні геноми SARS для ізолятів TOR2 (Canadian Michael Smith Genome Sciences Centre, Vancouver) [8] та Urbani (The Centers for Disease Control and Prevention, CDC) [9], і їхні послідовності депоновано в GenBank.

Детальний аналіз геномної послідовності SARS-CoV [10—12] показав, що, хоча в цілому організація геному дуже подібна до інших коронавірусів, SARS-CoV є раніше невідомим коронавірусом і суттєво відрізняється від усіх досі відомих представників цієї групи.

За даними секвенування, SARS-CoV людини може бути віднесений до ряду *Nidovirales* родини *Coronaviridae* роду *Coronavirus* як дистантний представник антигенної серогрупи II або стати першим представником нової четвертої групи [8—10]. Рід *Coronavirus* поділяють на три групи відповідно до їхньої антигенної крос-реактивності (серотипів) [13—15] і подібності геномних послідовностей [17—19], а також при аналізі найконсервативнішої ділянки гена реплікази [17—19]. Відомі коронавіруси груп I і II викликають захворювання у ссавців, а групи III — у птахів [13—16].

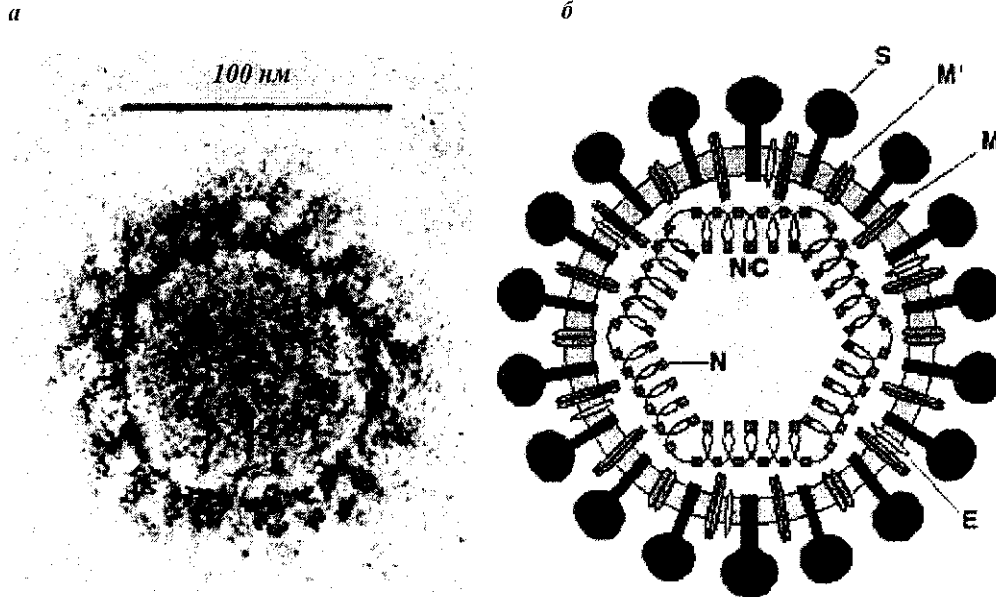


Рис. 1. Електронна мікрофотографія віріона коронавірусу SARS-CoV (а) і схема будови віріона (б). Показано розміщення чотирьох основних структурних білків S, E, M та N

Характеристика коронавірусів і будова віріона. Віріони коронавірусів і, зокрема SARS-CoV, мають сферичну або нерегулярну форму приблизно 60—220 нм у діаметрі, із зовнішньою мембранною оболонкою, на якій спостерігаються характерні вип'ячування — пепломери довжиною ~20 нм і шириною 10 нм на дистальному розширеному кінці (рис. 1, а). Така «короноподібна» структура оболонки дала назву цій родині вірусів.

Схематичне зображення віріона та локалізація основних структурних білків показані на рис. 1, б. Коронавіруси відносять до найбільших за розміром геному (27—31,5 тис. н.) та найскладніших за будовою несегментованих одноланцюгових (+)смыслових РНК-вмісних вірусів. Вірусна частинка містить одну молекулу геномної РНК та у віріоні відсутня РНК-полімераза. Ці віруси не мають ДНК-фази в циклі відтворення і реплікуються у цитоплазмі клітин хребетних (ссавців і птахів). Геномна РНК коронавірусів вміщує 5'-метильований кеп, 3'-poly(A)-последовність та може функціонувати безпосередньо як мРНК на відміну від (-)смыслових РНК-вмісних вірусів.

Секвенування та геномна організація SARS-CoV. За станом на 15 вересня 2003 р. в банку даних GenBank депоновано 36 геномів різних ізолятів SARS-CoV, секвенуваних повністю та вісім — частково (табл. 1). Довжина геномної послідовності SARS-CoV (за винятком ділянки 3'-poly(A)) варіює в межах 29705—29757 н. Під час накопичення SARS-CoV у культурі клітин знайдено делеційні мутанти, в яких були відсутні 45 н. (27670—27714 н.), що зменшило розмір геному до 29682 н. [11].

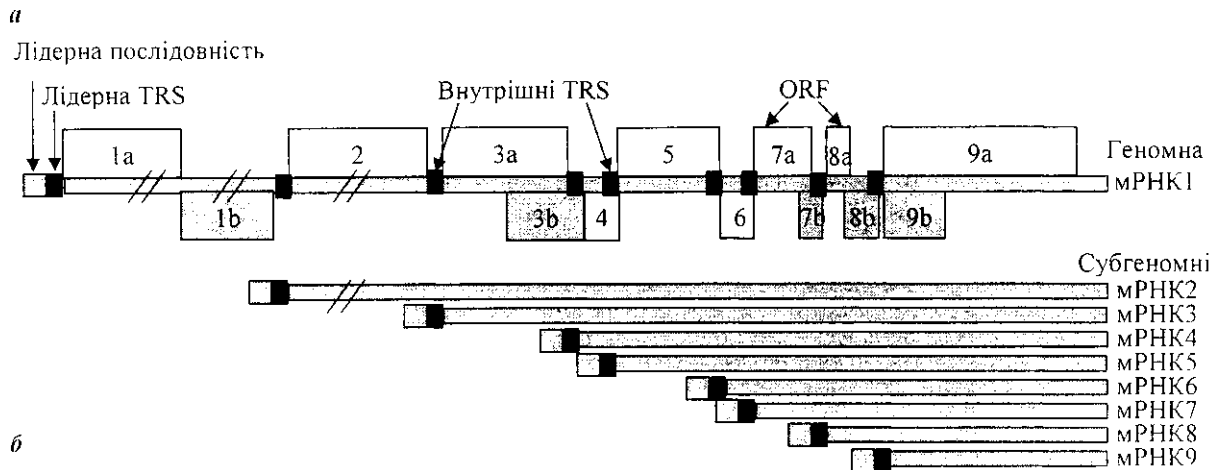
Ця делеція свідчить про те, що SARS-CoV може швидко змінюватися в культурі клітин [11].

Геном SARS-CoV має 41 %-й GC-склад, який варіює для геномів інших коронавірусів в межах 37—42 %. Виявлено 5'- і 3'-некодуючі ділянки (UTR) довжиною 265 і 342 н. відповідно. Геном включає 14 передбачених відкритих рамок зчитування (ORF) і має структурну організацію, подібну до інших коронавірусів (рис. 2, а). На 5'-кінці геному знаходиться ген реплікази *rep*, який містить ORF1a і ORF1b, що кодують великі за розміром полібілки *pp1a* і *pp1ab*. Останні надалі підлягають процесингу під дією специфічних вірусних протеїназ з формуванням 15 передбачених зрілих білків (рис. 2, б) [8—11]. Ближче до 3'-кінця знаходяться гени, які кодують структурні та додаткові білки, чотири з яких: S (шипа), E (оболонки), M (мембрани) і N (нуклеокапсиду) є характерними для всіх коронавірусів. Продукти гена *rep* синтезуються внаслідок прямої трансляції геномної РНК, тоді як для синтезу структурних вірусних білків необхідна попередня транскрипція із синтезом окремих субгеномних (sg) мРНК. Геном SARS-CoV не містить гена гемаглютининестерази (HE-білок), який присутній у коронавірусів груп II і III.

Молекулярний філогенетичний аналіз ізолятів SARS-CoV. Проведене вирівнювання повних геномів 30 ізолятів SARS-CoV засвідчило їхню високу схожість (рис. 3, а). Ідентифіковано 137 варіацій в геномах різних ізолятів SARS-CoV, з яких тільки 18 зустрічаються двічі або більше разів [22—24]. Геном SARS-CoV має типову для коронавірусів організацію, проте з певними унікаль-

Таблиця 1  
Секвеновані геноми різних ізолятів коронавірусу SARS-CoV станом на 21 вересня 2003 р. (наведено 36 геномів, які секвеновано повністю, та ще вісім — частково)

Ізолят	Код геномної послідовності	GI-код	Розмір, нуклеотиди	Дата	Країна
Референтна геномна послідовність (TOR2)	NC_004718.3	gi:30271926	29751	11.09.03	Канада
<i>Повністю секвеновані геноми</i>					
GZ50	AY304495.1	gi:34482146	29720	05.09.03	Гонконг
SZ16	AY304488.1	gi:34482139	29731	05.09.03	Гонконг
SZ3	AY304486.1	gi:34482137	29741	05.09.03	Гонконг
HKU-39849	AY278491.2	gi:30023963	29742	29.08.03	Гонконг
FRA	AY310120.1	gi:33578015	29740	18.08.03	Німеччина
GD01	AY278489.2	gi:31416290	29757	18.08.03	Китай
TWC3	AY362699.1	gi:33518725	29727	13.08.03	Тайвань
TWC2	AY362698.1	gi:33518724	29727	13.08.03	Тайвань
ZMY1	AY351680.1	gi:33304219	29749	03.08.03	Китай
TWY	AP006561.1	gi:33411459	29727	02.08.03	Тайвань
TWS	AP006560.1	gi:33411444	29727	02.08.03	Тайвань
TWK	AP006559.1	gi:33411429	29727	02.08.03	Тайвань
TWJ	AP006558.1	gi:33411414	29725	02.08.03	Тайвань
TWH	AP006557.1	gi:33411399	29727	02.08.03	Тайвань
Taiwan TC3	AY348314.1	gi:33188324	29573	29.07.03	Тайвань
Taiwan TC2	AY338175.1	gi:32493130	29573	28.07.03	Тайвань
Taiwan TC1	AY338174.1	gi:32493129	29573	28.07.03	Тайвань
HSR1	AY323977.2	gi:33115118	29751	22.07.03	Італія
TWC	AY321118.1	gi:31873092	29725	26.06.03	Тайвань
Frankfurt 1	AY291315.1	gi:31581502	29727	11.06.03	Німеччина
BJ04	AY279354.2	gi:31416306	29732	05.06.03	Китай
BJ03	AY278490.3	gi:31416305	29740	05.06.03	Китай
BJ02	AY278487.3	gi:31416292	29745	05.06.03	Китай
ZJ01	AY297028.1	gi:30910859	29715	19.05.03	Китай
TW1	AY291451.1	gi:30698326	29729	14.05.03	Тайвань
SIN2774	AY283798.1	gi:30468046	29711	09.05.03	Сингапур
SIN2748	AY283797.1	gi:30468045	29706	09.05.03	Сингапур
SIN2679	AY283796.1	gi:30468044	29711	09.05.03	Сингапур
SIN2677	AY283795.1	gi:30468043	29705	09.05.03	Сингапур
SIN2500	AY283794.1	gi:30468042	29711	09.05.03	Сингапур
CUHK-Su10	AY282752.1	gi:30421451	29736	07.05.03	Гонконг
BJ01	AY278488.2	gi:30275666	29725	01.05.03	Китай
TOR2	AY274119.3	gi:30248028	29751	30.04.03	Канада
CUHK-W1	AY278554.2	gi:30027610	29736	28.04.03	Китай
HKU-39849	AY278491.2	gi:30023963	29742	18.04.03	Гонконг
Urbani	AY278741.1	gi:30027617	29727	21.04.03	США



**б**

мРНК1	62	UCUAAACGAACUUUAA	77
мРНК2	21481	ACUAAACGAACAUGUU	21496
мРНК3	25201	CUUAAACGAACUUUUG	25270
мРНК4	26101	UGAGUACGAACUUUUG	26119
мРНК5	26341	UCUAAACGAACUUAACU	26358
мРНК6	26908	ACAUCACGAACGCUUU	26923
мРНК7	27262	AUAAAACGAACAUGAA	27277
мРНК8	27768	UCUAAACGAACAUGAA	27783
мРНК9	28101	AAUAAAACGAACAAAUU	28116

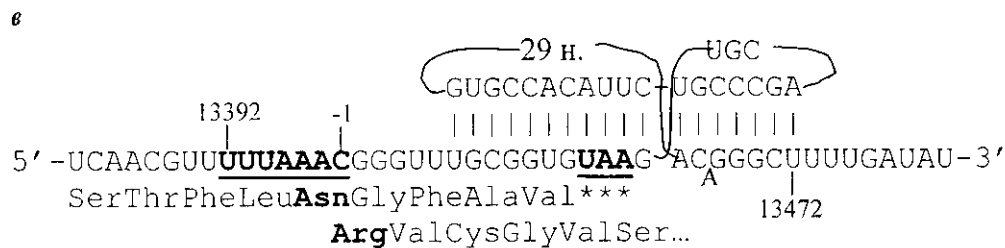


Рис. 2. Генетична карта (схема локалізації) відкритих рамок зчитування (ORF) та транскрипційних регуляторних послідовностей (TRS) відносно геномної послідовності коронавірусу SARS-CoV (показано схему утворення восьми субгеномних мРНК; потенційні додаткові ORF (1b, 3b, 7b, 8b та 9b) помічено сірим кольором (а) та вирівнювання послідовностей лідерної та восьми внутрішніх TRS (консенсусну послідовність (ACGAAC) виділено сірим кольором) (б), а також передбачена структура ділянки 1 зсуву рибосоми при трансляції полібілка рr1ab (виділено послідовність 13392-UUUAAAC, на якій відбувається переключення рамки трансляції, та структура псевдовузла, розташованого за нею; наведено фрагменти відповідних амінокислотних послідовностей у двох рамках зчитування) (в)

ними відмінностями. Аналіз філогенії геномних послідовностей SARS-CoV з 12 [25], 14 [22], 17 [24] та 20 [26] різних ізолятів показав дуже високу ідентичність усіх секвенованих геномів SARS-CoV, які утворюють єдину монофілетичну групу, що потребує використання нових підходів для дослідження їхньої подібності, таких як ультраметричні дистанційні матриці [24] та ін. Найб-

лижчою відомою дистантною групою є торовіруси (інший рід родини *Nidovirales*) [16, 27–30], порівняння з якими допомагає здійснювати філогенетичний аналіз.

Топологія філогенетичного дерева свідчить про те, що лінія SARS-CoV досить рано відокремилася від гілки групи II, проте пізніше, ніж розділилися три основні групи (рис. 3, б, в). Тому запропонува-

**Референтна послідовність**

1.	TWC2	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
2.	TWH	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
3.	TWC3	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
4.	TWK	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
5.	TWJ	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
6.	TWY	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
7.	TWS	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
8.	BJ02	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
9.	TOR2	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
10.	HSR	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
11.	HKU-39849	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
12.	FRA	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
13.	Frankfurt	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
14.	Urban1	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
15.	TW1	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
16.	TWC	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
17.	ZMY	TTATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	60
18.	BJ03	----TAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	56
19.	ZJ01	-----CCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	46
20.	CUHK-W1	-----CTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	45
21.	CUHK-Sul0	-----CTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	45
22.	GD01	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	44
23.	BJ04	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	44
24.	Sin2677	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	44
25.	Sin2500	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	44
26.	Sin2679	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	44
27.	Sin2774	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	44
28.	Sin2748	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	44
29.	BJ01	-----CCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	41

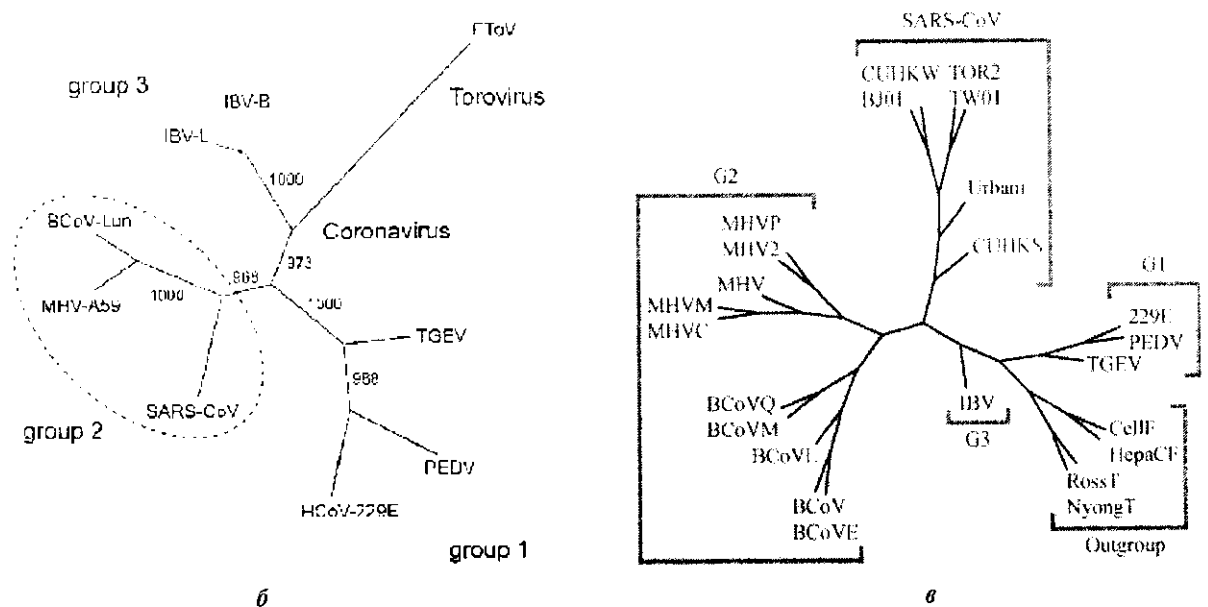


Рис. 3. Множинне вирівнювання 30 повних геномних послідовностей різних ізолятів SARS-CoV, проведене за допомогою сервера ClustalW 1.82 (показано лише початок вирівнювання); послідовності є майже ідентичними (а) та різні версії філогенетичного дерева коронавірусів, побудовані за порівнянням їхніх секвенованих геномів (показано три групи коронавірусів та гілка, яка відповідає SARS-CoV; як дистантну групу (outgroup) взято геноми тога- та флавівірусів (б, в))

но віднести SARS-CoV до підгрупи IIb в межах групи II [10]. Дуже ймовірно, що цей вірус нещодавно перетнув видовий бар'єр тварина—людина. Іншим подібним прикладом є пара коронавірусів HCoV-OC43 (людина) і BCoV (велика рогата худоба) [31, 32].

Аналіз мутацій вказує на те, що ізоляти SARS-CoV можна розділити на дві групи (генотипи), інфікування якими збігається з відомими клініч-

ними ланцюгами інфікування (табл. 2). Перша група, яка формує окремий кластер, асоційована з пацієнтами, інфікованими в «готелі М» у Гонконзі. Саме з ним пов'язано поширення епідемії SARS. Для чотирьох локусів, які містять п'ять чи більше варіантів С/Т- (положення 9404, 22222 та 27827) і G/Т-поліморфізму (положення 17564), виявлено, що в пов'язаних з «готелем М» ізолятах зустрічаються відповідно нуклеотиди Т:Т:Т:Т.

Таблиця 2  
Мутації в геномах різних ізолятів SARS-CoV

Позиція в геномі	Генотип T.T.T.T За 'язок з «готелем М»								Генотип C.G.C.C Немає зв'язку з «готелем М»					
	SIN2500	SIN2748	SIN2774	SIN2677	SIN2679	TOR2	URBANI	HKU39849	CUHKW1	GZ01	BJ01	BJ02	BJ03	BJ04
2601	T	T	T	T	T	C	T	C	T	T	T	T	T	-
7919	C	C	C	C	C	C	T	C	C	C	C	C	T	C
8559	T	T	T	T	T	T	T	T	T	C	T	T	T	A
8572	G	G	G	G	G	G	G	G	G	T	G	T	G	G
<b>9404</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>		<b>C</b>	<b>C</b>	<b>C</b>	<b>C</b>	-
9479	T	T	T	T	T	T	T	T	C	C	T	T	T	-
9854	C	C	C	C	C	C	C	C	C	C	T	T	T	-
<b>17564</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>		<b>G</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>G</b>
19064	A	A	A	A	A	A	G	A	G	A	A	A	A	A
19084	T	T	T	T	C	C	C	C	C	C	C	C	C	C
19838	A	A	A	A	A	A	A	A	A	A	G	G	G	G
21721	G	G	G	G	G	G	G	G	A	-	-	A	-	-
<b>22222</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>		<b>C</b>	<b>C</b>	<b>C</b>	<b>C</b>	<b>N</b>
27243	C	C	C	C	C	C	C	C	C	C	T	T	N	T
<b>27827</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>	<b>T</b>		<b>C</b>	<b>C</b>	<b>C</b>	<b>C</b>	<b>C</b>
29279	A	A	A	A	A	A	A	A	A	C	A	C	A	A

В ізолятах, які не мають подібної асоціації, знайдено відповідно C:G:C:C [22]. Цікаво, що мутація в положенні 22222 призводить до амінокислотної заміни Val/Ile у важливій антигенній ділянці S-білка. Це може свідчити про досить високий рівень мутабельності геному SARS-CoV [22]. Мутації в позиціях 9404 та 17564 також спричиняють амінокислотні заміни (Val/Ala та Asp/Glu відповідно).

**Проблема походження SARS-CoV.** Питання про походження SARS-CoV залишається відкритим, хоча вже виявлено потенційний попередник цього вірусу серед коронавірусів тварин. Коронавіруси з 99 %-ю подібністю послідовності до поверхневого S-білка ізолятів SARS-CoV виділено в провінції Гуангдонг (Китай) з кішкоподібних ссавців *Paguma larvata* [23]. В результаті мутації або рекомбінації вірус набув здатності інфікувати, викликати захворювання та розповсюджуватися серед людей.

**Транскрипція сгмРНК.** Експресія геному коронавірусів починається з трансляції великих полібілків з ORF1a і ORF1ab безпосередньо з геномної РНК. Крім того, коронавіруси (разом з *Arteriviruses*) використовують унікальну експресійну стратегію синтезу низки сгмРНК з однаковими 5'- та 3'-кінцевими послідовностями, ідентичними кінцям

геномної послідовності (рис. 2, а) [33]. Присутність цих сгмРНК підтверджено експериментально та визначено їхню приблизну довжину (включаючи 5'-лідерну та 3'-poly(A)-послідовності) — 8,4; 4,6; 3,8; 3,5; 3,0; 2,6; 2,1 та 1,8 тис. н. відповідно [10]. Кожна з цих мРНК містить коротку (приблизно 72 н.) 5'-кінцеву «лідерну» послідовність [9], яка походить з 5'-кінця геному [13, 14, 34, 35]. Перед 5'-кінцем більшості ORF структурних білків знаходяться специфічні міжгенні послідовності, які регулюють транскрипцію (TRS). Поєднання цих послідовностей з лідерною РНК відбувається, як вважають, під час синтезу мінус-ланцюга [13, 14, 35—41].

Крім TRS на 3'-кінці лідерної послідовності, інші TRS також знаходяться перед окремими ORF у 3'-кінцевій частині геному («body» TRS) [13, 14, 40]. Для SARS-CoV підтверджено використання такої стратегії «розірваної» транскрипції виявленням відповідних сгмРНК методом Нозерн-блотингу та секвенуванням послідовностей мРНК у ділянках, в яких спільна 5'-лідерна TRS з'єднана з різними внутрішніми TRS. За допомогою Нозерн-блот-аналізу poly(A)-вмісних РНК, виділених із SARS-CoV-інфікованих клітин, та зонду, специфічного для 3'-кінцевих 794 н., виявлено синтез не менше, ніж дев'яти мРНК, де мРНК1 відповідає

повному вірусному геному довжиною 29,7 тис. н. [10]. Розміри цих сгмРНК визначено з використанням раніше охарактеризованих мРНК коронавірусу HCoV-229E як маркерів [10]. На 5'-кінці геномної РНК знаходиться консервативна для коронавірусів лідерна TRS-послідовність 5'-ACGAAC-3' [8, 9, 11]. У SARS-CoV — це 5'-CUAAACGAACUUU-3'. Ці сигнали регулюють «розірвану» транскрипцію восьми сгмРНК. TSR-послідовності коронавірусів розташовані між ORF і взаємодіють з транскриптазою і клітинними факторами так, що відбувається злиття лідерної послідовності з початком основних рамок зчитування.

Таким чином, сгмРНК продукуються не шляхом сплайсингу, а власне самою РНК-полімеразою під час транскрипції. Кількість ідентичних нуклеотидів у лідерній TRS та внутрішніх TRS варіює в межах 6—11 [8, 9] (рис. 2, б), але немає чіткої кореляції між ступенем їхньої комплементарності та рівнем експресії даної сгмРНК, що вказує на вплив додаткових факторів, які регулюють синтез цих мРНК.

**Особливості некодуючих ділянок геному SARS-CoV.** 3'-Нетрансльована ділянка (3'-UTR) містить сегмент довжиною 32 н., який відповідає консервативному структурному мотиву s2m (stem-loop II-like motif) [8, 9]. Мотив s2m є універсальною рисою астровірусів, він ідентифікований також у коронавірусів групи III (вірусу пташиного інфекційного бронхіту, IBV) та у риновірусу ERV-2. В той же час мотив s2m відсутній в РНК ряду коронавірусів груп I і II. Високий рівень консервативності між мотивами s2m цих різних вірусів та їхня еволюційна відстань свідчать про те, що IBV та ERV-2 могли набути мотив s2m завдяки процесам рекомбінації між різними вірусами [8, 9]. Крім мотиву s2m, немає інших свідочств обміну генетичним матеріалом між SARS-CoV та вірусами за межами родини *Coronaviridae*.

Детально досліджено вторинну і третинну структури консервативних мотивів у першій половині 3'-UTR геномної РНК SARS-CoV методами комп'ютерного моделювання [42]. Описано структуру псевдовузла, який є найподібнішим до псевдовузлів коронавірусів групи I, однак значно відрізняється за розмірами петель та стебел у коронавірусу групи III (IBV). Псевдовузли в РНК різних коронавірусів є подібними за формою, однак відрізняються за первинною послідовністю. Таким чином, з одного боку SARS-CoV містить мотив s2m, схожий з коронавірусами групи III, а з іншого боку — псевдовузол, подібний до коронавірусів групи I, що вказує на його унікальні властивості [42, 43].

**Трансляція полібілків та зсув рамки зчитування.** Як показано для інших коронавірусів, експресія ORF1b з геномної РНК включає стадію 1 рибосомного зсуву рамки зчитування (frameshifting) [44, 45], при цьому синтезується полібілок pp1ab. Цей процес залежить від наявності двох елементів, відомих як «slippery»-послідовність, 13392-UUU-AAAC-13398, та складної структури РНК псевдовузла [11] (рис. 2, в), що було підтверджено в експериментах із сайт-спрямованого мутагенезу [11]. Точці зсуву відповідає пептидний зв'язок Asn4378-Arg4379 [10].

**Полібілки, їхній процесинг та зрілі білки SARS-CoV.** Подібно до інших коронавірусів, ген реплікази/полімерази SARS-CoV кодує два великих неструктурних полібілки (pp1a та pp1ab), для посттрансляційного процесингу яких потрібні дві специфічні вірусні протеїнази — PL2<sup>pro</sup> (PHP2) і 3CL<sup>pro</sup> [33] (табл. 3). ORF1a (4382 амінокислотних залишки, а. з., 265—13398 н.) і ORF1ab (7073 а. з., 265—21485 н.) гена реплікази/транскриптази займають 21,2 тис. н. (приблизно 71 %) геному SARS-CoV.

Ці ORF є консервативними за довжиною і амінокислотними послідовностями відносно репліказних білків інших коронавірусів. При розщепленні полібілків SARS-CoV вивільняються приблизно 15 зрілих неструктурних білків — NSP1—NSP16, які продукуються при їхньому протеолітичному розщепленні [11, 46—48] (табл. 3 і 4).

Всі 15 сайтів розщеплення присутні і в коронавірусів MHV та BCoV, які демонструють схожу схему протеолітичного процесингу та доменної організації полібілків. Для багатьох з цих білків виявлено дистантні гомологи серед різних вірусних та клітинних білків, зокрема, РНК-процесуючих ферментів.

Взаємне розташування передбачених двох полібілків і 23 зрілих білків, які кодуються геномом SARS-CoV, показано на рис. 4, а і б. Виявлено приблизно дев'ять дистантних гомологів ферментів, які розташовані в полібілках у такому порядку: аденозиндифосфо-рибозо-1'-фосфатаза (ADRP) [49], папаїн-подібна протеїназа 2 (PL2<sup>pro</sup> або PHP2) [56—64], основна протеїназа M<sup>pro</sup> (3CL<sup>pro</sup>) [48, 65—71], РНК-залежна РНК-полімераза (RdRp) [10, 50], АТФ-хеліказа (HEL1) [51—55], poly(U)-специфічна ендорибонуклеаза (XendoU) [10, 72], 3'-5'-екзонуклеаза (ExoN) [10, 73], S-аденозилметіонін-залежна рибозо-2'-O-метилтрансфераза (2'-O-MT) [10, 74—76] та циклічна фосфодіестераза (CPD) [49, 77, 78]. Функції інших зрілих білків ще не встановлено. Ідентифікація

Таблиця 3

Положення транскрипційних регуляторних послідовностей (TRS), передбачених відкритих рамок зчитування (ORF), довших за 40 амінокислотних залишків (а. з.), та зрілих неструктурних білків (NSP), які є продуктами розщеплення полібілків, відносно геномної послідовності SARS-CoV

ORF і білки	Положення ORF	Розмір, нуклеотиди	Розмір, а. з.
<i>Полібілки</i>			
gp1, ORF1a, полібілок 1a	265—13398	13149	4382
gp1, ORF1b, полібілок 1ab	265—13398, 13398—21485	21222	7073
<i>NSP, вивільняються при розщепленні полібілків</i>			
NSP1, лідерний білок	265—804	537	180
NSP2, білковий партнер M1V p65	805—2718	1917	638
NSP3, передбачена фосфостераза, папаїн-подібна PL2 <sup>pro</sup> протеїназа, аденозиндифосфатрибозо-1'-фосфатаза (ADPR)	2719—8484	5766	1922
NSP4	8485—9984	1500	500
NSP5, M <sup>pro</sup> (3C1 <sup>pro</sup> ) протеїназа	9985—10902	918	306
NSP6	10903—11772	870	290
NSP7	11773—12021	249	83
NSP8	12022—12615	594	198
NSP9	12616—12954	339	113
NSP10, білок, подібний до фактора росту	12955—13371	417	139
NSP11, короткий С-кінцевий пептид	13372—13410	39	13
NSP12, РНК-залежна РНК-полімераза (Pol/Rep)	13372—13398, 13398—16166	2796	932
NSP13, хеліказа (Hel) NTPase	16167—17969	1803	601
NSP14, гомолог екзонуклеази EchoN	17970—19550	1581	527
NSP15, гомолог ендо-РНКазы XendoU	19551—20588	1038	346
NSP16, можлива рибозо-2'-О-метилтрансфераза	20589—21482	894	298
<i>Структурні і передбачені допоміжні білки</i>			
gp2, ORF2, S (Spike)-білок	21492—25259	3768	1255
gp3, ORF3a	25268—26092	825	274
gp4, ORF3b	25689—26153	465	154
gp5, ORF4, E (Envelope)-білок	26117—26347	231	76
gp6, ORF5, M (Membrane)-білок	26398—27063	666	221
gp7, ORF6	27074—27265	192	63
gp8, ORF7a	27273—27641	369	122
gp9, ORF7b	27638—27772	135	44
gp10, ORF8a	27779—27898	120	39
gp11, ORF8b	27864—28118	255	84
gp12, ORF9a, N (Nucleocapside)-білок	28120—293788	1269	422
gp13, ORF9b	28130—28426	297	98



Закінчення табл. 3

ORF і білки	Положення ORF	Розмір, нуклеотиди	Розмір, а. з.
<i>Нетрансльовані та регуляторні послідовності</i>			
5'-UTR	1—294	264	—
3'-UTR	29389—29751	363	—
TRS лідерна	67—72	—	—
TRS		—	—
МРНК2	21486—21491	—	—
МРНК3	25260—25265	—	—
МРНК4	26109—26114	—	—
МРНК5	26348—26353	—	—
МРНК6	26913—26918	—	—
МРНК7	27267—27272	—	—
МРНК8	27773—27778	—	—
МРНК9	28106—28111	—	—

П р и м і т к а. Нумерацію всіх нуклеотидних ділянок наведено за референтною послідовністю NC\_004718.3, яка створена по геномній послідовності ізоляту TOR2 (GenBank, код AY274119.3); ORF та білки представлено в порядку, в якому вони кодуються геномом SARS-CoV.

Таблиця 4

Сайти специфічного протеолітичного розщеплення неструктурних полібілків pp1a (4382 а. з.) та pp1ab (7073 а. з.) коронавірусу SARS-CoV, їхнє вирівнювання та розміри зрілих білків

Границі білка	Розмір, а. з.	Сайт розщеплення	Фланкуючі послідовності
M1—G180	180	—	54321^1'
A181—G818	638	G180^A	174-TREIN <sup>■</sup> ^AVTRYVD
A819—G2740	1922	G818^A	812-VFR <sup>■</sup> K <sup>■</sup> ^APIKGVV
K2741—Q3240	500	G2740^K	2734-KIS <sup>■</sup> K <sup>■</sup> ^KIVSTCF
S3241—Q3546	306	Q3240^S	3234-ITSAV <sup>■</sup> LQ^SGFRKMA
G3547—Q3836	290	Q3546^G	3540-CSGVTFQ^GKFKKIV
S3837—Q3919	83	Q3836^S	3830-IKVATVQ^SKMSDVK
A3920—Q4117	198	Q3919^A	3913-DNRATLQ^AIASEFS
N4118—Q4230	113	Q4117^N	4111-NSAVK <sup>■</sup> LQ^NNELSPV
A4231—Q4369	139	Q4230^A	4224-AATVRLQ^AGNATEV
S4370—Q5301	932	Q4369^S	4353-LREPLM <sup>■</sup> Q^SADASTF
A5302—Q5902	601	Q5301^A	5295-TPHTV <sup>■</sup> LQ^AVGACVL
A5903—Q6429	527	Q5902^A	5896-RNVAT <sup>■</sup> LQ^AENVTGL
S6430—Q6775	346	Q6429^S	6423-NTFT <sup>■</sup> RLQ^SLENVAY
A6776—N7073	298	Q6775^A	6769-TFYPK <sup>■</sup> LQ^ASQAWQP
	Всього:		
	7073		

П р и м і т к а. Символом «^» позначено пептидний зв'язок, який розщеплюється. Розщеплення за першими трьома сайтами відбувається під дією пепсин-подібної протеїнази PL2<sup>pro</sup>. Наступні 11 сайтів (4—14) розщеплюються під дією основної M<sup>pro</sup> протеїнази. Консервативні амінокислотні залишки глутаміну в положенні P1, лейцину в положенні P2, а також однакові залишки в положеннях P4 і P1' затемнено.

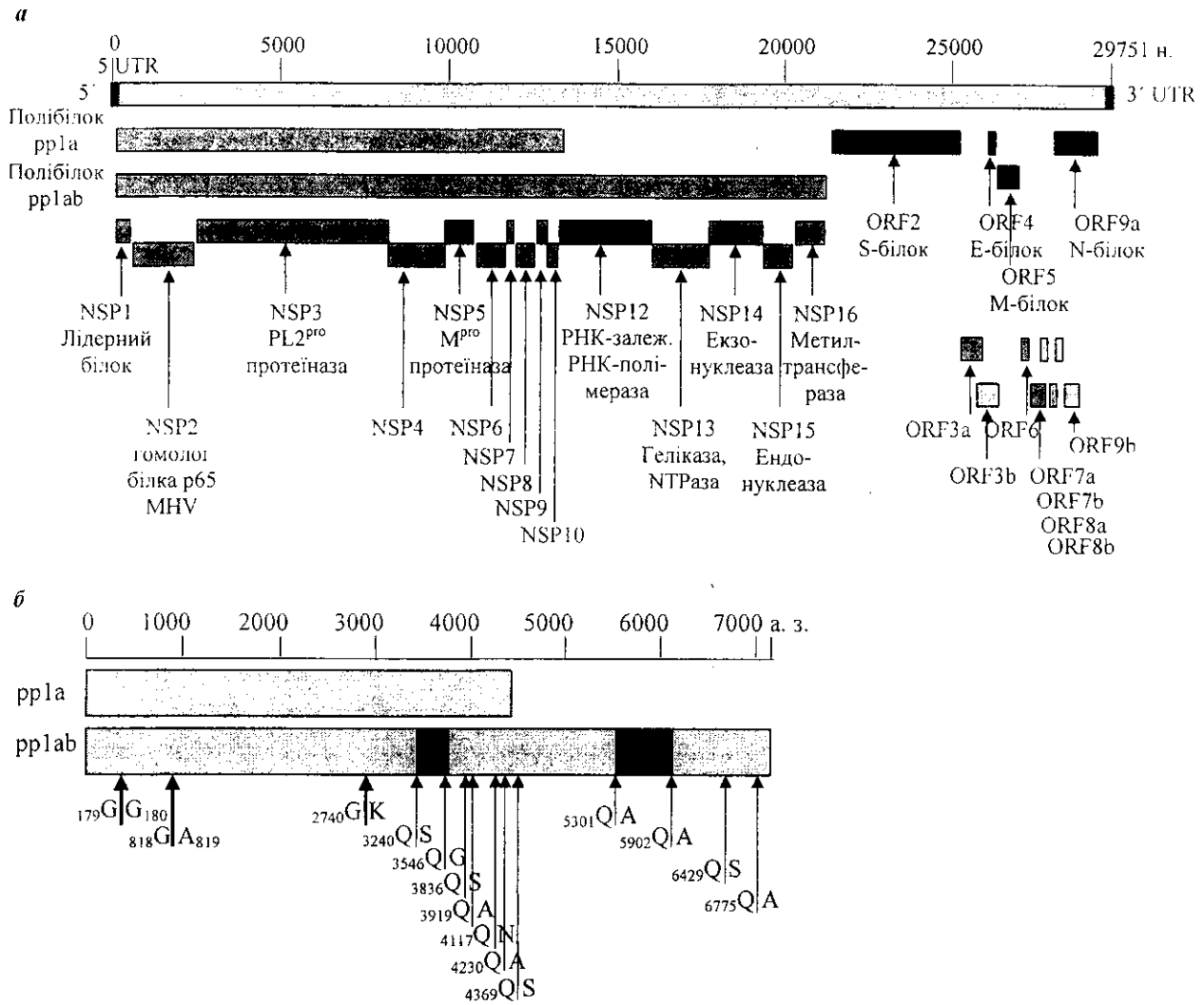


Рис. 4. Загальна схема положення відкритих рамок зчитування та приблизно 28 передбачених білків SARS-CoV (а) та схема посттрансляційного протеолітичного процесингу полібілків ppla і pplab двома вірусними протеїназами PL2<sup>pro</sup> та M<sup>pro</sup> (3CL<sup>pro</sup>) (протеїназа PL2<sup>pro</sup> розщеплює по трьох сайтах в N-кінцевій половині полібілків (товсті стрілки), а протеїназа M<sup>pro</sup> — по 11 сайтах у С-кінцевій половині (тонкі стрілки); при цьому вивільняється 15 зрілих білків та один короткий пептид) (б)

вірусних ферментів дозволяє в'яснити механізми синтезу РНК у коронавірусу SARS-CoV, ці білки є потенційними мішенями для дизайну антивірусних препаратів.

Трансляція полібілка ppla SARS-CoV та його подовженої С-кінцевої версії, білка pplab, поєднана з їхнім інтенсивним протеолітичним процесингом вірусними цистеїновими протеїназами — основною M<sup>pro</sup> (3CL<sup>pro</sup>) протеїназою та пепсинподібною протеїназою PL2<sup>pro</sup> [79]. За винятком вірусів IBV, всі раніше охарактеризовані коронавіруси кодують дві паралогічні папаїн-подібні цис-

теїнові протеїнази (PL1<sup>pro</sup> та PL2<sup>pro</sup>), які розщеплюють N-кінцеві ділянки полібілків у трьох сайтах [10, 56, 62]. Основна цистеїнова протеїназа (M<sup>pro</sup> або 3CL<sup>pro</sup>) розщеплює центральні і С-кінцеві ділянки в 11 консервативних сайтах [11, 46—48]. Консервативність як положення, так і послідовностей коронавірусних сайтів розщеплення полібілків ppla/pplab дозволяє передбачити ці сайти і в SARS-CoV. Білок NSP3 SARS-CoV є найбільшим серед зрілих білків (1922 а. з.) і складається з окремих доменів (Ac, X, SUD, PL2<sup>pro</sup>, Y). Доменна організація цього білка відрізняється від такої го-

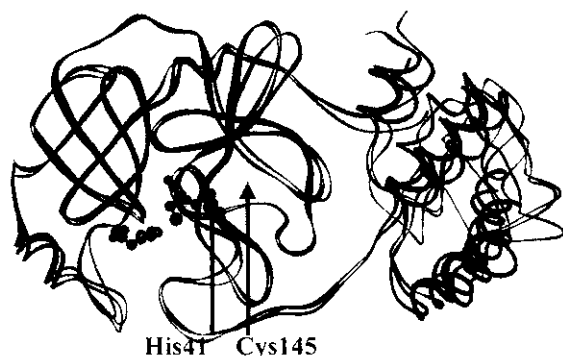


Рис. 5. Суперпозиція моделі просторової структури  $M^{pro}$  протеїнази SARS-CoV та гомологічних протеїназ коронавірусів групи 1 — трансмісивного вірусу гастроентериту свиней (TGEV) та коронавірусу людини HCoV-229E. Показано амінокислотні залишки каталітичної діади His41 та Cys145 [88]

мологічних білків інших коронавірусів. Так, NSP3 SARS-CoV не має домену, який відповідає протеїназі PL1<sup>pro</sup> [10, 57, 59], проте має унікальну вставку SUD (SARS-CoV unique domain) довжиною 375 а. з., яка знаходиться між доменами X та PL2<sup>pro</sup> [10]. Передбачається значна мультифункціональність білка NSP3, який може бути залучений до контролю синтезу субгеномних мРНК [10, 59, 60].

Основна  $M^{pro}$  (3CL<sup>pro</sup>, тип гідролази, EC 3.4.24) протеїназа та моделювання її просторової структури. Основна протеїназа SARS-CoV —  $M^{pro}$  (3CL<sup>pro</sup>) — є важливим ферментом циклу відтворення SARS-CoV, який контролює інші активності реплікаційного комплексу [80—84]. Завдяки цьому  $M^{pro}$  є перспективною мішенню для розробки специфічних інгібіторів, які можуть стати новими лікарськими препаратами від SARS та інших коронавірусних інфекцій.  $M^{pro}$  належить до цистеїнових протеїназ, є гомодимером і каталізує специфічне розщеплення неструктурних полібілків pp1a і pp1ab за 11 відповідними сайтами (табл. 4).  $M^{pro}$  протеїназа SARS-CoV складається з 306 а. з., які відповідають ділянці Ser3241-Gln3546 полібілка pp1ab. У геномній послідовності ізоляту TOR2 (код GenBank NC\_004718.3) зріла протеїназа кодується ділянкою 9985—10902 н. і анотована (код GenPept NP\_828863.2) як коронавірусний білок NSP5.

Методом рентгеноструктурного аналізу визначено просторові (3D) структури  $M^{pro}$  протеїназ двох інших коронавірусів. Структуру протеїнази трансмісивного вірусу гастроентериту свиней (TGEV) встановлено як для вільного білка (Protein Data Bank ID код 1LVO, три гомодимери в кристалографічній комірі, ланцюги A—F) [65], так і для його комплексу з субстрат-подібним інгібітором

Cbz-ValAsnSerThrLeuGln-CMK (PDB код 1P9U), де мономери В і F знаходяться в комплексі з інгібітором [48]. Структуру димеру  $M^{pro}$  протеїнази коронавірусу людини штаму 229E (HCoV-229E) визначено для вільного ферменту (код PDB 1P9S) [48]. На середину серпня 2003 р. різними науковими групами побудовано за гомологією ряд моделей  $M^{pro}$  протеїнази SARS-CoV, виходячи з експериментально визначених структур протеїназ коронавірусів TGEV та HCoV-229E. Ці моделі депоновано в розділі PDB для теоретичних моделей: PDB-коди 1P76, 1P9T [48], 1PA5, 1PUK і 1Q1X. Моделі 1P76 та 1Q1X побудовано для комплексів протеїнази з субстрат-подібними пептидами, AlaThrValArgLeuGln<sup>Ala</sup> та ValAsnSerThrLeuGln<sup>Ala</sup> відповідно, а модель 1PUK містить чотири додаткових С-кінцевих залишки з сайту розщеплення, 307-<sup>GlyLysPheLys</sup>. Нами побудовано модель просторової (3D) структури  $M^{pro}$  протеїнази SARS-CoV методом моделювання за гомологією (рис. 5) [88] виходячи із встановленої структури гомологічної протеїнази TGEV (PDB коди 1LVO та 1P9U [48, 65]). Моделювання здійснено з використанням веб-серверу Swiss-Model 3.5 з урахуванням різних варіантів положення трьох коротких вставок. Досліджено будову активного центра та динамічні властивості мономеру  $M^{pro}$  протеїнази методом симуляції молекулярної динаміки у водному середовищі в часовому інтервалі 1 нс [88].

Аналіз будови активного центра моделі SARS- $M^{pro}$  показав, що більшість амінокислотних залишків, які формують сайти зв'язування S5-S1' для залишків білкових субстратів у положеннях P5-P1', є консервативними для  $M^{pro}$  всіх коронавірусів. Ці залишки можна за положенням у поліпептидному ланцюгу віднести до п'яти сегментів: 25—27, 41—52, 140—145, 163—166, 187—189. Крім каталітичної діади, у формуванні активного центра беруть участь такі амінокислотні залишки, більшість з яких (16 з 24) відносяться до домену II: Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, His164, Met165, Glu166, Leu167, His172, Asp187, Arg188, Gln189, Gln192, а також залишки домену I: Thr25, Thr26, Leu27, His41, Cys44, Thr45, Pro52, Tyr54. Показано, що  $\beta$ -барелі доменів I і II є відносно жорсткими структурами і практично не змінюють своєї конформації в процесі динаміки [88]. Жорсткість  $\beta$ -барелів визначає в цілому низьку рухливість каталітичного модуля  $M^{pro}$  протеїнази. С-Кінцевий домен III демонструє значно більшу конформаційну рухливість. Дані щодо будови і динаміки активного центра SARS- $M^{pro}$  можуть бути використані для дизайну специфічних інгібіторів  $M^{pro}$  протеїнази та отримання в перспек-

тиві лікарських препаратів від коронавірусних інфекцій.

У кінці липня в PDB депоновано першу кристалографічну структуру димеру SARS-M<sup>pro</sup> у вільному стані з розділенням 1,86 Å (PDB-код 1Q2W). В обох субодиницях відсутні координати 11 а. з., для ланцюга «А» — це 1—2, 45—48 та 302—306, а для ланцюга «В» — 46—51 та 302—306. Чотири інші експериментальні структури, отримані при різних значеннях рН (PDB-коди 1UJ1, 1UK2, 1UK3), та комплекс протеїнази з інгібітором (код 1UK4) вже депоновано, але ще не відкрито.

**Структурні білки.** Структура сгмРНК SARS-CoV вказує на те, що п'ять з дев'яти цих мРНК можуть бути функціонально біцистронними (1a—1b, 3a—3b, 7a—7b, 8a—8b, 9a—9b відповідно) [10] і використовувати ще не встановлений механізм для експресії своїх внутрішніх додаткових ORF [89, 90]. Кількість 3'-проксимальних ORF значно варіює у різних коронавірусів [13, 14, 91]. Відкриті рамки зчитування основних структурних білків були ідентифіковані за гомологією до відомих білків інших коронавірусів [13—16, 34]. Три структурних білки, які містить вірусна оболонка — це білок шипів, Spike (S)-білок, глікопротеїд, який зв'язується з клітинним рецептором, що призводить до злиття віріона з клітиною, і є головним вірусним антигеном. Малий мембранний, Envelope (E)-білок — це білок, асоційований з мембранною оболонкою. Мембранний, Membrane (M)-білок — трансмембранний білок, зв'язаний з формуванням оболонки. Основний нуклеокапсидний білок, Nucleocapsid (N)-білок — це позитивно заряджений фосфопроїд, з яким зв'язана геномна РНК. Такий самий порядок розташування ORF для чотирьох структурних білків (S-E-M-N) виявлено і в коронавірусу BCoV групи II. У коронавірусів деяких видів також існує п'ятий структурний білок — HE (Haemagglutinin-esterase), однак у SARS-CoV ген цього білка відсутній. Основні дані по всіх передбачених білках SARS-CoV наведено в табл. 2 і 3 та на рис. 2. Здійснений для цих білків біоінформаційний пошук включав визначення гомологічних послідовностей (сервери BLAST і FASTA), передбачення сигнальної послідовності і мембранного якоря (сервер SignalP), трансмембранних спіралей (сервер TM<sup>pred</sup>) і топології орієнтації білка відносно мембрани (сервер TMHMM).

S-Білок (ORF2) кодує попередник основного поверхневого глікопротеїну з передбаченою довжиною 1255 а. з. Встановлено, що три молекули S-білка формують характерні пепломери, або коронаподібні структури навколо віріона. Відомо, що мутації в цьому гені корелюють з різним рівнем

патогенності та вірулентності у інших коронавірусів. У SARS-CoV зрілий S-білок інтегрований з вірусною оболонкою так, що більша частина білка є експонованою на поверхні вірусних частинок [92—94]. Аналіз S-білка виявляє високу (0,99) імовірність існування сигнального пептиду із точкою розщеплення між залишками 13 і 14. Показано наявність сильного трансмембранного домену біля С-кінця білка. Разом ці дані свідчать про те, що він є мембранним білком типу I з N-кінцем і більшою частиною (залишки 14—1195), розташованою ззовні вірусної частинки, що збігається з даними про S-білки інших коронавірусів. Нещодавно для віріонів HCoV-229E показано, що залишки 417—546 потрібні для зв'язування з клітинним рецептором амінопептидази N.

Відомо, що різні коронавіруси використовують різні рецептори, отже, ймовірно, що при цьому задіяні різні сайти зв'язування рецепторів. В S-білках багатьох коронавірусів знайдено сайт розщеплення на субодиниці S1 і S2 (RRFR, RRSRR, RRSRR, RSRR, RARS, RARR в IBV, BCoV, HCoV-OC43, PHEV, MHV та коронавірусі шурів відповідно), однак у S-білка SARS-CoV гомологічного сайту не виявлено.

S-Білок є одним з основних антигенів SARS-CoV і відіграє важливу роль у взаємодії вірусу з клітинним рецептором, проникненні вірусу до клітини шляхом злиття вірусної і клітинної мембран та в тканинному тропізмі вірусу [94]. Передбачено потенційні сайти зв'язування S-білка з можливим рецептором — білком CD13 (амінопептидаза людини, hAPN) з використанням цілої низки біоінформаційних та структурних підходів [94]. Здійснено моделювання білково-білкових взаємодій мотиву Asp757-Arg761 S-білка і домену Pro585-Ala653 CD13. Передбачено, що цей мотив S-білка може комплементарно зв'язуватися із заглибленням на поверхні домену CD13, утворюючи ряд специфічних контактів.

**E-Білок.** Малий білок оболонки E (ORF4) відповідає передбаченому білку з 76 а. з. і має значну подібність до малих мембранних білків інших коронавірусів. E-білок є компонентом вірусної оболонки, його консервативні послідовності також знайдено в інших вірусах, у тому числі вірусі гастроентериту свиней і вірусі гепатиту мишей.

Передбачено наявність трансмембранного якоря (імовірність 0,94) та трансмембранного домену в положенні 17—34. Це мембранний білок типу II з гідрофільним доменом (46 а. з.) та С-кінцем, локалізованим на поверхні вірусної частинки. У деяких коронавірусів (таких як TGEV) E-білок є важливим для реплікації вірусу [10, 11, 95, 96].

**М-Білок.** Мембранний глікопротеїд М (ORF5), містить 221 а. з. і дуже подібний до інших коронавірусних М-білків. Асоціація S-білка з М-білком є важливим етапом у формуванні вірусної оболонки і в накопиченні обох білків у сайтах збирання віріонів [97]. Передбачено сигнальну послідовність (імовірність 0,93), яка, вірогідно, не відщеплюється. Передбачено також присутність трьох трансмембранних спіралей, локалізованих приблизно по залишках 15—37, 50—72 та 77—99, зі 121 а. з. гідрофільним доменом всередині вірусної частинки, де він взаємодіє з нуклеокапсидом. Аналіз гомології показав подібність до домену PF01635 у базі даних PFAM та до 85 інших послідовностей в цій базі даних, які містять цей домен.

**N-Білок** довжиною 422 а. з. кодується ORF9a і добре вирівнюється з нуклеокапсидними білками інших коронавірусів, хоча його коротка, багата на лізін ділянка (KTFPPTPKKDKKKKTDEAQ) [10, 11, 98] є унікальною для SARS-CoV. Ця ділянка є ймовірним біпартитним сигналом ядерної локалізації (InterProDomain IPR001472). Очевидно, N-білок SARS-CoV має додаткову ядерну функцію, яка може відігравати певну роль у патогенезі цього вірусу. Крім того, основна природа згаданої ділянки свідчить про те, що вона може брати участь у зв'язуванні РНК.

Інші передбачені кодуєчі білок ORF. Крім чотирьох основних структурних білків, у складі восьми субгеномних мРНК, частина з яких, скоріш за все, є біцистронними, передбачено вісім відкритих рамок читування, які можуть кодувати певні допоміжні білки. Проаналізовано всі ORF, якщо вони кодують більше, ніж 40 а. з., незважаючи на можливу відсутність консенсусної послідовності TRS перед потенційним ініціюючим кодоном. Більшість з передбачених білків не має істотної подібності до інших відомих білків. ORF3a кодує передбачений білок з 274 а. з. Аналіз перших 70 N-кінцевих амінокислотних залишків дав слабе передбачення (імовірність 0,54) наявності сигнального пептиду та сайту його відщеплення. Передбачено існування трьох трансмембранних ділянок, які відповідають залишкам 34—56, 77—99 та 103—125. Найімовірніша модель, що з цього випливає, полягає в тому, що С-кінець і великий (149 а. з.) N-кінцевий домен локалізовані всередині вірусної частинки. С-кінцева внутрішня ділянка цього білка подібна до домену з АТФ-зв'язуючими властивостями (PD037277). ORF3b кодує передбачений білок з 154 а. з. і цілком перекривається з ORF3a та ORF4 для Е-білка. Аналіз не виявив потенційної TRS послідовності на 5'-кінці ORF4. Однак, ймовірно, що цей білок експресується з ORF3b мРНК,

використовуючи внутрішній сайт зв'язування рибосоми. У складі цього білка передбачено одну трансмембранну спіраль. ORF6 кодує передбачений білок з 63 а. з. Аналіз передбачає можливу трансмембранну спіраль між залишками 3—22 та N-кінець, локалізований ззовні вірусної частинки. ORF7a кодує передбачений білок з 122 а. з., який також не має істотних збігів з відомими білками. Аналіз показав наявність сигналу розщеплення (імовірність 0,99) з точкою розщеплення між залишками 15 і 16. Передбачається трансмембранна спіраль, локалізована приблизно в ділянці 99—117 а. з. Разом ці дані вказують на те, що ORF7a, ймовірно, є мембранним білком типу I з основним гідрофільним доменом (залишки 16—98) та N-кінцем, орієнтованим всередину ламел ендоплазматичного ретикулу, або на поверхні вірусної частинки в залежності від мембранної локалізації цього білка.

Відкрита рамка ORF7b кодує передбачений білок з 44 а. з., для якого передбачено існування однієї трансмембранної спіралі з невеликою різницею для альтернативних моделей, у яких N-кінець локалізований всередині чи ззовні вірусної частинки. Відкрита рамка ORF8a кодує передбачений білок з 39 а. з., для якого передбачено трансмембранну спіраль з N-кінцем, локалізованим всередині вірусної частинки. Ділянка геному перед ORF8a дуже подібна до TRS-консенсусу. Останнє свідчить про те, що транскрипт починається з цього сайту.

ORF8b кодує передбачений білок з 84 а. з., який має лише короткі збіги (9—10 а. з.) з ділянкою попередника глікопротеїду Е2 коронавірусу людини (починаючи з залишка 801). Аналіз, здійснений за допомогою SignalP та TMHMM, передбачає, що цей білок є розчинним. Так само, як і у випадку ORF8a, знайдено подібність до TRS консенсусної послідовності (рис. 2, б). Відкрита рамка ORF9b кодує передбачений білок з 98 а. з., для якого не виявлено подібних послідовностей та не передбачено трансмембранних спіралей.

Скринінг та дизайн потенційних інгібіторів. Доступність геномної послідовності забезпечує сприятливі умови для швидкого розвитку нових і чутливих діагностичних тестів, антивірусних агентів та вакцин. Інформація про послідовність геному дозволить прискорити дослідження патогенезу цього нового коронавірусу.

На сьогодні проводиться інтенсивний пошук інгібіторів різних білків SARS-CoV, у першу чергу M<sup>pro</sup> протеїнази. Ці інгібітори мають бути стабільними і не пригнічувати клітинні протеїнази. Опубліковано дані про експериментальний та вірту-

альний (симуляція докінгу малого ліганда *in silico*) скринінг потенційних інгібіторів  $M^{pro}$  протеїнази. На основі віртуального докінгу з 16 відомих інгібіторів протеїназ відібрано чотири з найбільшою спорідненістю до активного центра ферменту: каланолід А, невірапін, віразол та гліковір [87]. Експериментальним методом доведено пригнічуючу активність гліциризину [100].

На основі значної консервативності сайтів зв'язування субстратів передбачається анти-SARS- $M^{pro}$ -активність відомого інгібітора риновірусної 3C<sup>pro</sup> протеїнази [48]. Інгібітор KZ7088, який є похідним від AG7088, досліджено в роботі [85]. Зараз здійснюється експериментальна перевірка цих препаратів у клінічній практиці.

**Висновки.** За короткий строк проведено інтенсивне і різномічне дослідження нового коронавірусу SARS-CoV з використанням як експериментальних методів (секвенування геному, клонування ряду ферментів, ідентифікація ряду сайтів розщеплення полібілків, рентгеноструктурний аналіз  $M^{pro}$  протеїнази), так і численних біоінформаційних підходів. Серед останніх слід зазначити ідентифікацію за гомологією цілого ряду білків SARS-CoV, передбачення субгеномних мРНК і відкритих рамок зчитування, моделювання структури S-білка і активного центру  $M^{pro}$  протеїнази, віртуальний скринінг їхніх потенційних інгібіторів і лігандів. Таким чином, постгеномна стратегія дослідження природи нових вірусів з використанням сучасних методів біоінформатики дає ефективний поштовх для розробки нових антивірусних препаратів і боротьби з інфекційними захворюваннями.

*K. A. Odynets, A. I. Kornelyuk*

Molecular aspects of organization and expression of SARS-CoV coronavirus genome

#### Summary

*The molecular aspects of SARS-CoV coronavirus genome structure—the ethiological agent of atypical pneumonia or Severe Acute Respiratory Syndrome are discussed. The general characterization of coronaviruses and virion structure are considered in the review. The data about 36 completely sequenced genomes of different SARS-CoV isolates and their phylogenetic analysis are discussed. The predicted properties of eight subgenomic mRNAs and 14 open reading frames are described, as well as a synthesis of polyproteins, their processing and mature SARS-CoV proteins. The properties of the viral proteins and their functions are analyzed. The surface S-protein is one of the most important antigens of SARS-CoV and it plays an important role in the virus interaction with cell receptor. The potential sites of S-protein binding with a putative receptor-aminopeptidase hAPN are predicted using both bioinformatics and structural methods. The main  $M^{pro}$  (3CL<sup>pro</sup>) proteinase is a potential protein target for antiviral therapy. The modeling of 3D structure of  $M^{pro}$  (3CL<sup>pro</sup>) proteinase and its active site organization is considered in terms of design of potential SARS-CoV inhibitors.*

*K. A. Odynets, A. I. Kornelyuk*

Molecular aspects of structure and expression of the genome of the SARS coronavirus

#### Резюме

*Рассмотрены молекулярные аспекты строения генома коронавируса SARS-CoV — возбудителя атипичной пневмонии, или тяжелого острого респираторного синдрома. Приведена характеристика коронавирусов и строения вириона. Анализируются данные по 36 полностью секвенированным геномам разных изолятов SARS-CoV и данные молекулярного филогенетического исследования. Описаны предсказанные свойства восьми субгеномных мРНК и 14 открытых рамок считывания. Охарактеризованы синтез полибелков, их процессинг и зрелые белки SARS-CoV. Обсуждаются свойства предсказанных белков SARS-CoV и их функции. Поверхностный S-белок является одним из основных антигенов SARS-CoV и играет важную роль во взаимодействии вируса с клеточным рецептором. Предсказаны потенциальные сайты связывания S-белка с вероятным рецептором — аминопептидазой hAPN — с использованием биоинформационных и структурных подходов. Представлены строение основной  $M^{pro}$  (3CL<sup>pro</sup>) протеиназы как потенциальной мишени для антивирусной терапии, моделирование ее пространственной структуры, строение активного центра, скрининг и дизайн потенциальных ингибиторов SARS-CoV.*

#### PERELIK LITERATURY

1. Peiris J. S., Lai S. T., Poon L. L., Guan Y., Yam L. Y., Lim W., Nicholls J., Yee W. K., Yan W. W., Cheung M. T., Cheng V. C., Chan K. H., Tsang D. N., Yung R. W., Ng T. K., Yuen K. Y. Coronavirus as a possible cause of severe acute respiratory syndrome // *Lancet*.—2003.—361, N 9366.—P. 1319—1325.
2. Peiris J. S., Chu C. M., Cheng V. C., Chan K. S., Hung I. F., Poon L. L., Law K. I., Tang B. S., Hon T. Y., Chan C. S., Chan K. H., Ng J. S., Zheng B. J., Ng W. L., Lai R. W., Guan Y., Yuen K. Y. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study // *Lancet*.—2003.—361, N 9371.—P. 1767—1772.
3. Drosten C., Gunther S., Preiser W., van der Werf S., Brodt H. R., Becker S., Rabenau H., Panning M., Kolesnikova L., Fouchier R. A., Berger A., Burguiere A. M., Cinatl J., Eickmann M., Escriou N., Grywna K., Kramme S., Manuguerra J. C., Muller S., Rickerts V., Sturmer M., Vieth S., Klenk H. D., Osterhaus A. D., Schmitz H., Doerr H. W. Identification of a novel coronavirus in patients with severe acute respiratory syndrome // *N. Engl. J. Med.*—2003.—348, N 20.—P. 1967—1976.
4. Drosten C., Preiser W., Gunther S., Schmitz H., Doerr H. W. Severe acute respiratory syndrome: identification of the etiological agent // *Trends Mol. Med.*—2003.—9, N 8.—P. 325—327.
5. Ksiazek T. G., Erdman D., Goldsmith C. S., Zaki S. R., Peret T., Emery S., Tong S., Urbani C., Comer J. A., Lim W., Rollin P. E., Dowell S. F., Ling A. E., Humphrey C. D., Shieh W. J., Guarner J., Paddock C. D., Rota P., Fields B., DeRisi J., Yang J. Y., Cox N., Hughes J. M., LeDuc J. W., Bellini W. J., Anderson L. J. A novel coronavirus associated with severe acute respiratory syndrome // *N. Engl. J. Med.*—2003.—348, N 20.—P. 1953—1966.
6. Fouchier R. A., Kuiken T., Schutten M., van Amerongen G., van Doornum G. J., van den Hoogen B. G., Peiris M., Lim

- W., Stohr K., Osterhaus A. D. Aetiology: Koch's postulates fulfilled for SARS virus // *Nature*.—2003.—423, N 6937.—P. 240.
7. Fung W. K., Yu P. L. SARS case-fatality rate // *Can. Med. Assoc. J.*—2003.—169, N 4.—P. 277—278.
  8. Marra M. A., Jones S. J., Astell C. R., Holt R. A., Brooks-Wilson A., Butterfield Y. S., Khattra J., Asano J. K., Barber S. A., Chan S. Y., Cloutier A., Coughlin S. M., Freeman D., Girn N., Griffith O. L., Leach S. R., Mayo M., McDonald H., Montgomery S. B., Pandoh P. K., Petrescu A. S., Robertson A. G., Schein J. E., Siddiqui A., Smailus D. E., Stott J. M., Yang G. S., Plummer F., Andonov A., Artsob H., Bastien N., Bernard K., Booth T. F., Bowness D., Czub M., Drebot M., Fernando L., Flick R., Garbutt M., Gray M., Grolla A., Jones S., Feldmann H., Meyers A., Kabani A., Li Y., Normand S., Stroher U., Tipples G. A., Tyler S., Vogrig R., Ward D., Watson B., Brunham R. C., Kraiden M., Petric M., Skowronski D. M., Upton C., Roper R. I. The Genome sequence of the SARS-associated coronavirus // *Science*.—2003.—300, N 5624.—P. 1399—1404.
  9. Rota P. A., Oberste M. S., Monroe S. S., Nix W. A., Campagnoli R., Icenogle J. P., Penaranda S., Bankamp B., Maher K., Chen M. H., Tong S., Tamin A., Lowe L., Frace M., DeRisi J. L., Chen Q., Wang D., Erdman D. D., Peret T. C., Burns C., Ksiazek T. G., Rollin P. E., Sanchez A., Liffick S., Holloway B., Limor J., McCaustland K., Olsen-Rasmussen M., Fouchier R., Gunther S., Osterhaus A. D., Drosten C., Pallansch M. A., Anderson L. J., Bellini W. J. Characterization of a novel coronavirus associated with severe acute respiratory syndrome // *Science*.—2003.—300, N 5624.—P. 1394—1399.
  10. Snijder E. J., Bredenbeek P. J., Dobbe J. C., Thiel V., Ziebuhr J., Poon L. L., Guan Y., Rozanov M., Spaan W. J., Gorbalenya A. E. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage // *J. Mol. Biol.*—2003.—331, N 5.—P. 991—1004.
  11. Thiel V., Ivanov K. A., Putics A., Hertzog T., Schelle B., Bayer S., Weissbrich B., Snijder E. J., Rabenau H., Doerr H. W., Gorbalenya A. E., Ziebuhr J. Mechanisms and enzymes involved in SARS coronavirus genome expression // *J. Gen. Virol.*—2003.—84, N 9.—P. 2305—2315.
  12. Zeng F. Y., Chan C. W., Chan M. N., Chen J. D., Chow K. Y., Hon C. C., Hui K. H., Li J., Li V. Y., Wang C. Y., Wang P. Y., Guan Y., Zheng B., Poon L. L., Chan K. H., Yuen K. Y., Peiris J. S., Leung F. C. The complete genome sequence of severe acute respiratory syndrome coronavirus strain HKU-39849 (HK-39) // *Exp. Biol. Med.*—2003.—228, N 7.—P. 866—873.
  13. Lai M. M. C., Holmes K. V. Coronaviruses // *Fields Virology* / Eds D. M. Knipe, P. M. Howley.—Lippincott, 2001.—P. 1163—1185.
  14. Siddell S. G. The Coronaviridae // *The Viruses* / Eds H. Fraenkel-Conrat, R. R. Wagner.—New York: Plenum press, 1995.—418 p.
  15. Sanchez C. M., Jimenez G., Laviada M. D., Correa I., Sune C., Bullido M. Antigenic homology among coronaviruses related to transmissible gastroenteritis virus // *Virology*.—1990.—174, N 2.—P. 410—417.
  16. Enjuanes L., Spaan W. J. M., Snijder E. J., Cavanagh, D. Order nidovirales // *Virus Taxonomy* / Eds M. H. V. Regenmortel, C. M. Fauquet, D. H. L. Bishop.—New York: Acad. press, 2000.—P. 827—834.
  17. Chouljenko V. N., Lin X. Q., Storz J., Kousoulas K. G., Gorbalenya A. E. Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia // *J. Gen. Virol.*—2001.—82, N 12.—P. 2927—2933.
  18. Herold J., Raabe T., Schelle-Prinz B., Siddell S. G. Nucleotide sequence of the human coronavirus 229E RNA polymerase locus // *Virology*.—1993.—195, N 2.—P. 680—691.
  19. Stephensen C. B., Casebolt D. B., Gangopadhyay N. N. Phylogenetic analysis of a highly conserved region of the polymerase gene from 11 coronaviruses and development of a consensus polymerase chain reaction assay // *Virus Res.*—1999.—60, N 2.—P. 181—189.
  20. Nicholls J. M., Poon L. L., Lee K. C., Ng W. F., Lai S. T., Leung C. Y., Chu C. M., Hui P. K., Mak K. L., Lim W., Yan K. W., Chan K. H., Tsang N. C., Guan Y., Yuen K. Y., Peiris J. S. Lung pathology of fatal severe acute respiratory syndrome // *Lancet*.—2003.—361, N 9371.—P. 1773—1778.
  21. Kuiken T., Fouchier R. A., Schutten M., Rimmelzwaan G. F., van Amerongen G., van Riel D., Laman J. D., de Jong T., van Doornum G., Lim W., Ling A. E., Chan P. K., Tam J. S., Zambon M. C., Gopal R., Drosten C., van der Werf S., Escriviou N., Manuguerra J. C., Stohr K., Peiris J. S., Osterhaus A. D. Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome // *Lancet*.—2003.—362, N 9380.—P. 263—270.
  22. Ruan Y. J., Wei C. L., Ee A. L., Vega V. B., Thoreau H., Su S. T., Chia J. M., Ng P., Chiu K. P., Lim L., Zhang T., Peng C. K., Lin E. O., Lee N. M., Yee S. L., Ng L. F., Chee R. E., Stanton L. W., Long P. M., Liu E. T. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection // *Lancet*.—2003.—361, N 9371.—P. 1779—1785.
  23. Guan Y., Zheng B. J., He Y. Q., Liu X. L., Zhuang Z. X., Cheung C. L., Luo S. W., Li P. H., Zhang L. J., Guan Y. J., Butt K. M., Wong K. L., Chan K. W., Lim W., Shortridge K. F., Yuen K. Y., Peiris J. S., Poon L. L. Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China // *Science*.—2003.—302, N 5643.—P. 276—278.
  24. Gao L., Qi J., Wei H., Sun Y., Hao B. Molecular phylogeny of coronaviruses including human SARS-CoV // *Chinese Sci. Bull.*—2003.—48, N 12.—P. 1170—1174.
  25. Qi Z., Hu Y., Li W., Chen Y., Zhang Z., Sun S., Lu H., Zhang J., Bu D., Ling L., Chen R. Phylogeny of SARS-CoV as inferred from complete genome comparison // *Chin. Sci. Bull.*—2003.—48, N 12.—P. 1175—1178.
  26. Hu L. D., Zheng G. Y., Jiang H. S., Xia Y., Zhang Y., Kong X. Y. Mutation analysis of 20 SARS virus genome sequences: evidence for negative selection in replicase ORF1b and spike gene // *Acta pharmacol. Sin.*—2003.—24, N 8.—P. 741—745.
  27. Snijder E. J., Horzinek, M. C. Toroviruses: replication, evolution and comparison with other members of the coronavirus-like superfamily // *J. Gen. Virol.*—1993.—74, N 11.—P. 2305—2316.
  28. Snijder E. J., den Boon J. A., Bredenbeek P. J., Horzinek M. C., Rijnbrand R., Spaan W. J. M. The carboxyl-terminal part of the putative Berne virus polymerase is expressed by ribosomal frameshifting and contains sequence motifs which indicate that toro- and coronaviruses are evolutionarily related // *Nucl. Acids Res.*—1990.—18, N 15.—P. 4535—4542.
  29. Snijder E. J., den Boon J. A., Spaan W. J. M., Weiss M., Horzinek M. C. Primary structure and post-translational processing of the Berne virus peplomer protein // *Virology*.—1990.—178, N 2.—P. 355—363.
  30. den Boon J. A., Snijder E. J., Krijns Locker J., Horzinek M. C., Rottier P. J. M. Another triple-spanning envelope protein

- among intracellularly budding RNA viruses: the torovirus E protein // *Virology*.—1991.—182, N 2.—P. 655—663.
31. Zhang X. M., Herbst W., Kousoulas K. G., Storz J. Biological and genetic characterization of a hemagglutinating coronavirus isolated from a diarrhoeic child // *J. Med. Virol.*—1994.—44, N 2.—P. 152—161.
  32. Sasseville A. M., Boutin M., Gelinas A. M., Dea S. Sequence of the 30-terminal end (8.1 kb) of the genome of porcine haemagglutinating encephalomyelitis virus: comparison with other haemagglutinating coronaviruses // *J. Gen. Virol.*—2002.—83, N 10.—P. 2411—2416.
  33. Hofmann M. A., Chang R. Y., Ku S., Brian D. A. Leader-mRNA junction sequences are unique for each subgenomic mRNA species in the bovine coronavirus and remain so throughout persistent infection // *Virology*.—1993.—196, N 1.—P. 163—171.
  34. Snijder E. J., Meulenbergh J. J. M. Arteriviruses // *Fields Virology* / Eds D. M. Knipe, P. M. Howley.—Lippincott, 2001.—P. 1205—1220.
  35. Pasternak A. O., van den Born E., Spaan W. J. M., Snijder E. J. Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis // *EMBO J.*—2001.—20, N 24.—P. 7220—7228.
  36. van Vliet A. L., Smits S. L., Rottier P. J., de Groot R. J. Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus // *EMBO J.*—2002.—21, N 23.—P. 6571—6580.
  37. Thiel V., Siddell S. G. Internal ribosome entry in the coding region of murine hepatitis virus mRNA 5 // *J. Gen. Virol.*—1994.—75, N 11.—P. 3041—3046.
  38. Nagy P. D., Simon A. E. New insights into the mechanisms of RNA recombination // *Virology*.—1997.—235, N 1.—P. 1—9.
  39. Sawicki S. G., Sawicki D. L. A new model for coronavirus transcription // *Adv. Exp. Med. and Biol.*—1998.—440.—P. 215—219.
  40. Sawicki S. G., Sawicki D. L. Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands // *Adv. Exp. Med. and Biol.*—1995.—380.—P. 499—506.
  41. Sethna P. B., Hung S. L., Brian D. A. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons // *Proc. Nat. Acad. Sci. USA.*—1989.—86, N 14.—P. 5626—5630.
  42. Зарудная М. И., Потягайло А. Л., Говорун Д. Н. Консервативные структурные мотивы в 3' нетранслируемой области геномной РНК вируса SARS-CoV // *Биополимери і клітина.*—2003.—19, № 3.—С. 298—303.
  43. Qin L., Xiong B., Luo C., Guo Z. M., Hao P., Su J., Nan P., Feng Y., Shi Y. X., Yu X. J., Luo X. M., Chen K. X., Shen X., Shen J. H., Zou J. P., Zhao G. P., Shi T. L., He W. Z., Zhong Y., Jiang H. L., Li Y. X. Identification of probable genomic packaging signal sequence from SARS-CoV genome by bioinformatics analysis // *Acta pharmacol. Sin.*—2003.—24, N 6.—P. 489—496.
  44. Brierley I., Digard P., Inglis S. C. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot // *Cell.*—1989.—57, N 4.—P. 537—547.
  45. Brierley I. Ribosomal frameshifting viral RNAs // *J. Gen. Virol.*—1995.—76, N 8.—P. 1885—1892.
  46. Ziebuhr J., Snijder E. J., Gorbalenya A. E. Virus-encoded proteinases and proteolytic processing in the Nidovirales // *J. Gen. Virol.*—2000.—81, N 4.—P. 853—879.
  47. Gorbalenya A. E., Koonin E. V., Donchenko A. P., Blinov V. M. Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis // *Nucl. Acids Res.*—1989.—17, N 12.—P. 4847—4861.
  48. Anand K., Ziebuhr J., Wadhvani P., Mesters J. R., Hilgenfeld R. Coronavirus main proteinase (3CL<sup>pro</sup>) structure: Basis for design of anti-SARS drugs // *Science.*—2003.—300, N 5626.—P. 1763—1767.
  49. Martzen M. R., McCraith S. M., Spinelli S. L., Torres F. M., Fields S., Grayhack E. J., Phizicky E. M. A biochemical genomics approach for identifying genes by the activity of their products // *Science.*—1999.—286, N 5442.—P. 1153—1155.
  50. Campanacci V., Egloff M. P., Longhi S., Ferron F., Rancurel C., Salomoni A., Durousseau C., Tocque F., Bremond N., Dobbe J. C., Snijder E. J., Canard B., Cambillau C. Structural genomics of the SARS coronavirus: cloning, expression, crystallization and preliminary crystallographic study of the Nsp9 protein // *Acta crystallogr. D. Biol. crystallogr.*—2003.—59, N 9.—P. 1628—1631.
  51. Tanner J. A., Watt R. M., Chai Y. B., Lu L. Y., Lin M. C., Malik Peiris J. S., Poon L., Kung H. F., Huang J. D. The SARS coronavirus NTPase/helicase belongs to a distinct class of 5' to 3' viral helicases // *J. Biol. Chem.*—2003.—278, N 41.—P. 39578—39582.
  52. Seybert A., Hegyi A., Siddell S. G., Ziebuhr J. The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5'-to-3' polarity // *RNA.*—2000.—6, N 7.—P. 1056—1068.
  53. van Dinten L. C., van Tol H., Gorbalenya A. E., Snijder E. J. The predicted metal-binding region of the arterivirus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion biogenesis // *J. Virol.*—2000.—74, N 11.—P. 5213—5223.
  54. Gorbalenya A. E., Koonin E. V., Donchenko A. P., Blinov V. M. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes // *Nucl. Acids Res.*—1989.—17, N 12.—P. 4713—4730.
  55. Seybert A., van Dinten L. C., Snijder E. J., Ziebuhr J. Biochemical characterization of the equine arteritis virus helicase suggests a close functional relationship between arterivirus and coronavirus helicases // *J. Virol.*—2000.—74, N 20.—P. 9586—9593.
  56. Kanjanahaluethai A., Baker S. C. Identification of mouse hepatitis virus papain-like proteinase 2 activity // *J. Virol.*—2000.—74, N 17.—P. 7911—7921.
  57. Baker S. C., Yokomori K., Dong S., Carlisle R., Gorbalenya A. E., Koonin E. V., Lai M. M. Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus // *J. Virol.*—1993.—67, N 10.—P. 6056—6063.
  58. Lim K. P., Ng L. F., Liu D. X. Identification of a novel cleavage activity of the first papain-like proteinase domain encoded by open reading frame 1a of the coronavirus Avian infectious bronchitis virus and characterization of the cleavage products // *J. Virol.*—2000.—74, N 4.—P. 1674—1685.
  59. Herold J., Siddell S. G., Gorbalenya A. E. A human RNA viral cysteine proteinase that depends upon a unique Zn<sup>2+</sup>-binding finger connecting the two domains of a papain-like fold // *J. Biol. Chem.*—1999.—274, N 21.—P. 14918—14925.
  60. Tijms M. A., van Dinten L. C., Gorbalenya A. E., Snijder E. J. A zinc finger-containing papain-like protease couples subgenomic mRNA synthesis to genome translation in a positive-stranded RNA virus // *Proc. Nat. Acad. Sci. USA.*—2001.—98, N 4.—P. 1889—1894.
  61. Ziebuhr J., Thiel V., Gorbalenya A. E. The autocatalytic release of a putative RNA virus transcription factor from its polyprotein precursor involves two paralogous papain-like pro-



- teases that cleave the same peptide bond // *J. Biol. Chem.*—2001.—276, N 35.—P. 33220—33232.
62. Gorbalenya A. E., Koonin E. V., Lai M. M. Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses // *FEBS Lett.*—1991.—288, N 1—2.—P. 201—205.
  63. Bonilla P. J., Hughes S. A., Weiss S. R. Characterization of a second cleavage site and demonstration of activity in trans by the papain-like proteinase of the murine coronavirus mouse hepatitis virus strain A59 // *J. Virol.*—1997.—71, N 2.—P. 900—909.
  64. Herold J., Gorbalenya A. E., Thiel V., Schelle B., Siddell S. G. Proteolytic processing at the amino terminus of human coronavirus 229E gene 1-encoded polyproteins: identification of a papain-like proteinase and its substrate // *J. Virol.*—1998.—72, N 2.—P. 910—918.
  65. Anand K., Palm G. J., Mesters J. R., Siddell S. G., Ziebuhr J., Hilgenfeld R. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain // *EMBO J.*—2002.—21, N 13.—P. 3213—3224.
  66. Hegyi A., Friebe A., Gorbalenya A. E., Ziebuhr J. Mutational analysis of the active centre of coronavirus 3C-like proteases // *J. Gen. Virol.*—2002.—83, N 3.—P. 581—593.
  67. Ziebuhr J., Heusipp G., Siddell S. G. Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase // *J. Virol.*—1997.—71, N 5.—P. 3992—3997.
  68. Ziebuhr J., Herold J., Siddell S. G. Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity // *J. Virol.*—1995.—69, N 7.—P. 4331—4338.
  69. Hegyi A., Ziebuhr J. Conservation of substrate specificities among coronavirus main proteases // *J. Gen. Virol.*—2002.—83, N 3.—P. 595—599.
  70. Seybert A., Ziebuhr J., Siddell S. G. Expression and characterization of a recombinant murine coronavirus 3C-like proteinase // *J. Gen. Virol.*—1997.—78, N 1.—P. 71—75.
  71. Ziebuhr J., Siddell S. G. Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to pp1a and pp1ab // *J. Virol.*—1999.—73, N 1.—P. 177—185.
  72. Laneve P., Altieri F., Fiori M. E., Scaloni A., Bozzoni J., Caffarelli E. Purification, cloning, and characterization of XendoU, a novel endoribonuclease involved in processing of intron-encoded small nucleolar RNAs in *Xenopus laevis* // *J. Biol. Chem.*—2003.—278, N 15.—P. 13026—13032.
  73. Zuo Y., Deutscher M. P. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution // *Nucl. Acids Res.*—2001.—29, N 5.—P. 1017—1026.
  74. Bugl H., Fauman E. B., Staker B. L., Zheng F., Kushner S. R., Saper M. A. RNA methylation under heat shock control // *Mol. Cell.*—2000.—6, N 2.—P. 349—360.
  75. Eglhoff M. P., Benarroch D., Selisko B., Romette J. L., Canard B. An RNA cap (nucleoside-2'-O)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization // *EMBO J.*—2002.—21, N 11.—P. 2757—2768.
  76. Feder M., Pas J., Wyrwicz L. S., Bujnicki J. M. Molecular phylogenetics of the RrmJ/fibrillarin superfamily of ribose 2'-O-methyltransferases // *Gene.*—2003.—302, N 1—2.—P. 129—138.
  77. Nasr F., Filipowicz W. Characterization of the *Saccharomyces cerevisiae* cyclic nucleotide phosphodiesterase involved in the metabolism of ADP-ribose 100, 200-cyclic phosphate // *Nucl. Acids Res.*—2000.—28, N 8.—P. 1676—1683.
  78. Culver G. M., Consaul S. A., Tycowski K. T., Filipowicz W., Phizicky E. M. TRNA splicing in yeast and wheat germ. A cyclic phosphodiesterase implicated in the metabolism of ADP-ribose 100, 200-cyclic phosphate // *J. Biol. Chem.*—1994.—269, N 40.—P. 24928—24934.
  79. Liu D. X., Tibbles K. W., Cavanagh D., Brown T. D., Brierley J. Involvement of viral and cellular factors in processing of polyprotein encoded by ORF1a of the coronavirus IBV // *Adv. Exp. Med. and Biol.*—1995.—380.—P. 413—421.
  80. van der Meer Y., Snijder E. J., Dobbe J. C., Schleich S., Denison M. R., Spaan W. J. M., Krijnse Locker J. Localization of mouse hepatitis virus nonstructural proteins and RNA synthesis indicates a role for late endosomes in viral replication // *J. Virol.*—1999.—73, N 9.—P. 7641—7657.
  81. Gosert R., Kanjanahaluethai A., Egger D., Bienz K., Baker S. C. RNA replication of Mouse Hepatitis Virus takes place at double-membrane vesicles // *J. Virol.*—2002.—76, N 8.—P. 3697—3708.
  82. Allmang C., Kufel J., Chanfreau G., Mitchell P., Petfalski E., Tollervey D. Functions of the exosome in rRNA, snoRNA and snRNA synthesis // *EMBO J.*—1999.—18, N 19.—P. 5399—5410.
  83. Bost A. G., Carnahan R. H., Lu X. T., Denison M. R. Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly // *J. Virol.*—2000.—74, N 7.—P. 3379—3387.
  84. Ng M. L., Tan S. H., See E. E., Ooi E. E., Ling A. E. Early events of SARS coronavirus infection in *vero* cells // *J. Med. Virol.*—2003.—71, N 3.—P. 323—331.
  85. Chou K. C., Wei D. Q., Zhong W. Z. Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS // *Biochem. and Biophys. Res. Commun.*—2003.—308, N 1.—P. 148—151.
  86. Xiong B., Gui C. S., Xu X. Y., Luo C., Chen J., Luo H. B., Chen L. L., Li G. W., Sun T., Yu C. Y., Yue L. D., Duan W. H., Shen J. K., Qin L., Shi T. L., Li Y. X., Chen K. X., Luo X. M., Shen X., Shen J. H., Jiang H. L. A 3D model of SARS-CoV 3CL proteinase and its inhibitors design by virtual screening // *Acta pharmacol. Sin.*—2003.—24, N 6.—P. 497—504.
  87. Lee V. S., Wittayanarakul K., Remsungnen T., Parasuk V., Sompornpisut P., Chantratita W., Sangma C., Vannarat S., Srichaikul P., Hannongbua S., Saparpakorn P., Treesuwan W., Aruksakulwong O., Pasomsab E., Promsri S., Chuakheaw D., Hannongbua S. Structure and dynamics of SARS coronavirus proteinase: The primary key to the designing and screening for anti-SARS drugs // *ScienceAsia.*—2003.—29.—P. 181—188.
  88. Одынець К. А., Каниболоцкий Д. С., Корнелюк А. И. Модель пространственной структуры и строение активного центра M<sup>pro</sup> протеиназы коронавируса SARS // *Анали Мечниківського інституту.*—2003.—№ 4—5.—С. 123.
  89. Senanayake S. D., Brian D. A. Bovine coronavirus I protein synthesis follows ribosomal scanning on the bicistronic N mRNA // *Virus Res.*—1997.—48, N 1.—P. 101—105.
  90. Liu D. X., Inglis S. C. Internal entry of ribosomes on a tricistronic mRNA encoded by infectious bronchitis virus // *J. Virol.*—1992.—66, N 10.—P. 6143—6154.
  91. de Haan C. A., Masters P. S., Shen X., Weiss S., Rottier P. J. The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host // *Virology.*—2002.—296, N 1.—P. 177—189.

92. Bosch B. J., van der Zee R., de Haan C. A., Rottier P. J. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex // *J. Virol.*—2003.—77, N 16.—P. 8801—8811.
93. Spiga O., Bernini A., Ciutti A., Chiellini S., Menciaci N., Finetti F., Causarone V., Anselmi F., Prischi F., Niccolai N. Molecular modelling of S1 and S2 subunits of SARS coronavirus spike glycoprotein // *Biochem. and Biophys. Res. Commun.*—2003.—310, N 1.—P. 78—83.
94. Yu X. J., Luo C., Lin J. C., Hao P., He Y. Y., Guo Z. M., Qin L., Su J., Liu B. S., Huang Y., Nan P., Li C. S., Xiong B., Luo X. M., Zhao G. P., Pei G., Chen K. X., Shen X., Shen J. H., Zou J. P., He W. Z., Shi T. L., Zhong Y., Jiang H. L., Li Y. X. Putative hAPN receptor binding sites in SARS-CoV spike protein // *Acta pharmacol. Sin.*—2003.—24, N 6.—P. 481—488.
95. Filipowicz W., Pogacic V. Biogenesis of small nucleolar ribonucleoproteins // *Curr. Opin. Cell Biol.*—2002.—14, N 3.—P. 319—327.
96. Shen X., Xue J. H., Yu C. Y., Luo H. B., Qin L., Yu X. J., Chen J., Chen L. L., Xiong B., Yue L. D., Cai J. H., Shen J. H., Luo X. M., Chen K. X., Shi T. L., Li Y. X., Hu G. X., Jiang H. L. Small envelope protein E of SARS: cloning, expression, purification, CD determination, and bioinformatics analysis // *Acta pharmacol. Sin.*—2003.—24, N 6.—P. 505—511.
97. de Haan C. A., de Wit M., Kuo L., Montalto-Morrison C., Haagmans B. L., Weiss S. R., Masters P. S., Rottier P. J. The glycosylation status of the murine hepatitis coronavirus M protein affects the interferogenic capacity of the virus *in vitro* and its ability to replicate in the liver but not the brain // *Virology.*—2003.—312, N 2.—P. 395—406.
98. Lin Y., Shen X., Yang R. F., Li Y. X., Ji Y. Y., He Y. Y., Shi M. D., Lu W., Shi T. L., Wang J., Wang H. X., Jiang H. L., Shen J. H., Xie Y. H., Wang Y., Pei G., Shen B. F., Wu J. R., Sun B. Identification of an epitope of SARS-coronavirus nucleocapsid protein // *Cell Res.*—2003.—13, N 3.—P. 141—145.
99. Holmes K. V., Enjuanes L. *Virology*. The SARS coronavirus: a postgenomic era // *Science.*—2003.—300, N 5624.—P. 1377—1378.
100. Cinatl J., Morgenstern B., Bauer G., Chandra P., Rabenau H., Doerr H. W. Glycyrrhizin, an active component of liquorice roots, and replication of SARS-associated coronavirus // *Lancet.*—2003.—361, N 9374.—P. 2045—2046.

УДК 577.322

Надійшла до редакції 06.07.03