

Original article

<https://doi.org/10.26565/2075-3810-2026-55-08>

УДК 619:616.98:578.831.31:636.4

### 3D МОДЕЛІ БІЛКІВ ТА КОНСЕРВАТИВНИХ G-КВАДРУПЛІ КСІВ В ГЕНОМІ ЦИРКОВІРУСУ СВИНІ

**О. Ю. Лиманська<sup>1,\*</sup>**, **О. В. Білойван<sup>1</sup>**, **О. К. Балак<sup>2</sup>**, **О. П. Лиманський<sup>3</sup>**<sup>1</sup>Національний науковий центр «Інститут експериментальної і клінічної ветеринарної медицини»,  
вул. Григорія Сковороди, 83, Харків, 61023, Україна;<sup>2</sup>Харківський національний медичний університет, пр. Науки, 4, Харків, 61022, Україна;<sup>3</sup>Інститут фізіологічно активних сполук, пр. Науки, 58, Харків, 61072, Україна\*e-mail: [olgaliman@ukr.net](mailto:olgaliman@ukr.net)

Надійшла до редакції 29 грудня 2025 р. Переглянута 06 травня 2026 р.

Прийнята до друку 07 травня 2026 р. Опублікована 25 червня 2026 р.

**Актуальність.** Неканонічні структури, що утворюються в молекулах нуклеїнових кислот, до яких відносять, зокрема, мультиспіральні з'єднання 3WJs (three-way junction), квадруплекси (G4s), є регуляторними елементами, які впливають на функціонування геному. Побудова 3D моделей цих структур і білків патогенів є першим етапом у розумінні їхніх функцій. У створеній компанією Google DeepMind разом з EMBL бази даних AlphaFold Protein Structure Database, яка містить понад 241 мільйон структур білків, 3D моделі білків цирковірусу свині типу 3 (ЦВС-3) є відсутніми.

**Мета роботи:** визначення та структурна характеристика потенційних G4s і 3WJ в геномі ЦВС-3, побудова 3D моделей білків ЦВС-3.

**Матеріали і методи.** 1138 ізолятів ЦВС-3 з повним геномом з GenBank використано для філогенетичного аналізу за допомогою пакету програм MEGA12. Для побудови графіків ентропії використано пакет програм BioEdit; пошуку мотивів G4s та визначення їх G-рахунків — програму QGRS Mapper; пошуку 100%-ої ідентичності нуклеотидних послідовностей — програму BLAST. 3D моделі білків та G4s побудовано за допомогою штучного інтелекту AlphaFold 3.

**Результати.** В геномі ЦВС-3 ідентифіковано 4 консервативні досконалі G4s, які утворено трьома тетрадами та підтверджено через побудову їхніх 3D моделей. Визначені послідовності G4s є консервативними мотивами, оскільки число ізолятів ЦВС-3 в GenBank з G4s у геномі становить понад 1000. В якості контролю коректності побудови цих моделей використовували 3D моделі G4s, утворення одного з яких експериментально визначено в геномі вірусу гепатиту Б. 3D моделі реплікази та білка капсиду побудовано для двох еволюційно найвіддаленіших ізолятів ЦВС-3, які визначено з філогенетичних дерев на основі генів *cap* та *rep*, встановлено їхні особливості. 3WJs в геномі ЦВС-3 не виявлено.

**Висновки.** Отримано 3D моделі реплікази, білка капсиду цирковірусу свині типу 3, а також чотирьох консервативних структурних мотивів G4s, які утворено трьома G-тетрадами. Збільшення довжини фрагментів, які фланкують G4s, не перешкоджає їхньому фолдингу, що свідчить про значущість цих G-багатих мотивів у життєвому циклі ЦВС-3. Існування G4s з трьох тетрад в геномі ЦВС-3 на відміну від двотетрадних G4s вірусу гепатиту Б у складі не тільки аптамерів, а й молекул збільшеної довжини підтверджено за допомогою AlphaFold 3.

**Як цитувати:** Лиманська ОЮ, Білойван ОВ, Балак ОК, Лиманський ОП. 3D моделі білків та консервативних G-квадруплексів в геномі цирковірусу свині. Біофізичний вісник. 2026;55:95–116. <https://doi.org/10.26565/2075-3810-2026-55-08>

**Citation:** Limanskaya OYu, Biloivan OV, Balak OK, Limanskii AP. 3D models of proteins and conservative G-quadruplexes in the genome of porcine circovirus. Biophysical Bulletin. 2026;55:95–116. <https://doi.org/10.26565/2075-3810-2026-55-08> (in Ukrainian)

**Open Access.** This article is licensed under a Creative Commons Attribution 4.0 <http://creativecommons.org/licenses/by/4.0/>

**КЛЮЧОВІ СЛОВА:** альтернативна структура; біоінформатика; AlphaFold 3; цирковірус свиней типу 3; G-квадруплекс (G4); структура білка; 3D модель.

G-квадруплекси (G4s) — це неканонічні вторинні структури, які утворюються у збагачених гуанідіном послідовностях молекул нуклеїнових кислот шляхом внутрішніх або міжмолекулярних взаємодій. G-квадруплекси складаються з двох або більше планарних G-тетрад, утворених чотирма гуанідінами через Хугстинівські водневі зв'язки. Стабільність G4s визначено стекінг-взаємодією G-тетрад, а також наявністю моно- або бівалентних катіонів за фізіологічних умов, які можуть бути розташованими як між, так і всередині G-тетрад [1]. G4s є регуляторними елементами та відіграють певну роль у таких важливих біологічних процесах, як трансляція, транскрипція, реплікація, мутагенез, рекомбінація та репарація ДНК, а також в організації геному [2, 3].

Утворення G4s спостерігають у випадку низки нейродегенеративних та онкологічних захворювань людини, що дозволяє розглядати G-квадруплекси не тільки як нові цільові мішені для лікарських засобів, а і як, наприклад, складові потенційних протипухлинних препаратів [4–6]. Наявність G-збагачених мотивів, що є схильними до утворення G4s, або існування G-квадруплексів продемонстровано для низки організмів, у тому числі, вірусів — збудників інфекційних захворювань сільськогосподарських тварин [7], зокрема, свиней. Так, показано, що негативний ланцюг геномної РНК вірусу репродуктивно-респіраторного синдрому свиней містить консервативний паралельний G4, який відіграє ключову роль у реплікації вірусу [8]. Дві послідовності з потенціалом утворення G4s визначено за результатами біоінформатичного аналізу у генах білка шипів S та неструктурного білка Nsp5, необхідного для процесу реплікації вірусу епідемічної діареї свиней. За допомогою спектроскопії кругового дихроїзму та флюоресценції підтверджено можливість фолдингу G-квадруплексів [9].

Консервативну G-збагачену послідовність, яка утворює G4, ідентифіковано у гені неструктурного білка E165R вірусу африканської чуми свиней (АЧС). Поряд з низкою інших неструктурних білків вірусу АЧС E165R відіграє певну роль як фермент метаболізму нуклеїнових кислот у процесі реплікації патогена [10]. Наявність G-квадруплексів у зазначених вище вірусів відкриває перспективи розвитку стратегії створення лікарських препаратів нового типу з урахуванням особливостей молекулярно-генетичної структури патогенів.

Цирковірус свиней типу 3 (ЦВС-3) — один із відомих чотирьох типів ЦВС, належить до роду *Circovirus* сімейства *Circoviridae*. Вперше ЦВС-3 було детектовано у 2016 р. у трьох свиней з патологією серця та мультисистемним запаленням [11]. ЦВС-3 поширений у популяціях свиней в усьому світі [12], в Україні вірус вперше виявлено у 2021 р. [13]. ЦВС-3 асоційований з гострим дерматитом свиней, синдромом нефропатії, репродуктивною недостатністю та, отже, є загрозою для розвитку та благополуччя свинарської галузі [14–17]. Основними резервуарами ЦВС-3 є свині та дикі кабани, проте можливою є міжвидова трансмісія вірусу [18].

Геном ЦВС-3 представлений одноланцюговою кільцевою молекулою ДНК довжиною 2000 нуклеотидів (н.) та містить три відкриті рамки зчитування ORFs. ORF1 та ORF2, які орієнтовано у протилежних напрямках, кодують репліказу та білок капсиду вірусу відповідно. Функція білка, що транслюється з ORF3, залишається нез'ясованою. Інформація щодо утворення неканонічних структур у геномі цирковірусів свиней є вельми обмеженою. Відомо, що на 5'-кінці міжгенної області локалізовано шпилькову структуру з довжиною петлі 9 н., яка є висококонсервативною для ЦВС незважаючи на виявлені незначні відмінності у послідовності петлі.

Вважають, що ця шпилька ініціює реплікацію ЦВС [19]. Інформація щодо можливості утворення G4s у геномі цирковірусів свиней, у тому числі ЦВС-3, у літературі на цей час є відсутньою.

Компанія Google DeepMind Technologies Limited (Лондон, Велика Британія) нещодавно випустила свою нову версію моделі штучного інтелекту (ШІ) AlphaFold 3 [20], яка передбачає 3D структуру білків з амінокислотної послідовності з порівняною для експериментальних методів точністю (0,1–2 нм для різних фрагментів білка).

Штучний інтелект AlphaFold 3 є єдиним серед численних біоінформатичних методів, за розробку якого через надзвичайно високу точність передбачення структури біополімерів його авторам Д. Хассабісу і Д. Джемперу (обидва — Google DeepMind Technologies Ltd) «за передбачення білкових структур», а також Д. Бейкеру (університет Вашингтону в Сіетлі) «за обчислювальний дизайн білків» присуджено Нобелівську премію з хімії 2024 року. Google DeepMind та Європейський Інститут біоінформатики разом з EMBL створили базу даних AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk>), випуск якої (станом на листопад 2025 року) у вільному доступі містить понад 241 мільйону структур білків, які передбачено на підставі амінокислотної послідовності, що веде до прискорення наукових досліджень. 3D моделі реплікази та білка капсиду ЦВС-3 у зазначеній базі даних на теперішній час є відсутніми.

Точність передбачення структури білка за допомогою першої версії — ШІ AlphaFold — становила 60%. При розробці другої версії — ШІ AlphaFold 2 — було змінено принцип її роботи. Згорткову нейронну мережу замінили на архітектуру трансформера. В якості вихідних даних користувач отримував безпосередньо координати атомів. Ці модифікації забезпечили точність передбачення структури білка 90%. Як результат, передбачення тривимірної структури білка за допомогою AlphaFold 2 стало можливим з точністю 0,1 нм, що є зіставленим з такою рентгеноструктурного аналізу.

Розвиток AlphaFold не зупинився на другій версії, і у травні 2024 року Google DeepMind представила AlphaFold 3 із значно розширеними можливостями. Крім передбачення тривимірної структури, AlphaFold 3 вміє оцінювати внесок посттрансляційних модифікацій в конформацію белкової молекули, а також передбачати взаємодію білків з іншими білками, нуклеїновими кислотами, іонами [21].

В даній роботі побудовано 3D моделі білків капсиду та реплікази еволюційно найвіддаленіших ізолятів цирковірусу свині типу 3, а також проведено ідентифікацію, характеризацію консервативних потенційних досконалих G-квадруплексів з трьох тетрад в геномі ЦВС-3.

## МАТЕРІАЛИ І МЕТОДИ

Нуклеотидні та амінокислотні послідовності 1138 ізолятів ЦВС-3 з повним геномом станом на листопад 2025 р. з різних географічних регіонів світу та різноманітного біологічного матеріалу від свиней та диких кабанів (легенів, ротової рідини, сироватки, фекалій, абортіваних плодів, лімфи) отримано у форматах \*.gb та \*.fasta через пошук за ключовим словом «porcine circovirus type 3 AND complete genome» (ідентифікатор таксону 1868221) в базі даних GenBank Національного центру біотехнологічної інформації (США). Пошук 100 %-ої ідентичності нуклеотидних послідовностей G-багатих мотивів ізолятів вірусу здійснювали за допомогою програми BLAST nucleotide (<https://blast.ncbi.nlm.nih.gov/>).

Множинне вирівнювання, філогенетичний аналіз проведено за допомогою пакету програм Molecular Evolutionary Genetics Analysis MEGA12 (версія 12.0.9) [22]. Попередньо 1138 нуклеотидних послідовностей ЦВС-3 з повним геномом розділили на

6 груп: 5 груп по 200 ізолятів та групу з 138 ізолятів, для яких визначали матрицю попарних відстаней та проводили множинне вирівнювання. Потім з кожного з 6 побудованих філогенетичних дерев вибирали по 6 еволюційно найвіддаленіших ізолятів ЦВС-3, які сформували дерево з 36 ізолятів для визначення двох найвіддаленіших ізолятів в наборі з 1138 ізолятів ЦВС-3. Філогенетичні відношення визначено з використанням р-дистанцій, адаптивного бутстреп-аналізу (500 випадкових вибірок для оцінки достовірності дерева), моделі нуклеотидних замін Jukes – Cantor та методу максимальної правдоподібності або приєднання сусідів.

Для маніпулювання з нуклеотидними та амінокислотними послідовностями, мутаційного аналізу білків вірусів, побудови графіків ентропії Шеннона для визначення несинонімічних та синонімічних замін, побудови матриці відмінностей послідовностей для розрахунку швидкості мутацій використовували програму BioEdit (версія 7.2.5) [23].

Для визначення кількості стоп-кодонів застосовували програму HyPhy [24]. Визначення швидкості мутацій здійснювали згідно з [25].

Оцінку впливу природного відбору на гени шляхом визначення відношення частот несинонімічних (dN) до синонімічних (dS) замін на нуклеотидний сайт dN/dS проводили з використанням методу SLAC (single likelihood ancestor counting) на сервері Datamonkey ([www.datamonkey.org](http://www.datamonkey.org)).

Для пошуку мотивів внутрішньомолекулярних G-квадруплексів в ДНК ЦВС-3, вірусу гепатиту Б (ізолят V01460), а також для визначення G-рахунку (параметра, який характеризує стабільність G-квадруплексу у відносних одиницях) використано програму для аналізу квадруплексів QGRS Mapper на веб-сервері (<https://bioinformatics.ramapo.edu/QGRS/analyze.php>) [26, 27]. Пошук G4 проводили для ізоляту PP067099 за наступних параметрів: максимальна довжина G-квадруплексу — 45 н., мінімальний розмір G-тетради — 3, петля — від 1 до 20 н. Для пошуку потенційних досконалих G4s застосовано програму QGRS Mapper через найкращі показники порівняно з іншими програмами в частині помилково та правильно знайдених, помилково та правильно не знайдених G4s при тестуванні на одному й тому же G-мотиві [28].

3D моделі білків та G4s (з 10-20 іонами K<sup>+</sup>) побудовано за допомогою штучного інтелекту AlphaFold 3 [20]. Для оцінки надійності передбачення моделі в AlphaFold 3 використано параметр pLDDT (predicted local distance difference test), значення якого для високого рівня достовірності моделі має становити понад 90 за шкалою від 0 до 100. Значення pLDDT від 70 до 90 відповідає середньому рівню надійності моделі, а 70 > pLDDT > 50 — низькому рівню. Для аналізу вторинної структури білків, визначення кількості водневих зв'язків,  $\alpha$ -спіралей,  $\beta$ -листів,  $\beta$ -поворотів використовували програми PyMol 2.2 [29] та RasWin Molecular Graphics (версія 2.7.5.2; [www.openrasmol.org](http://www.openrasmol.org)). Визначення 3WJ-структур проводили за методом, що зазначено в роботі [30].

## РЕЗУЛЬТАТИ Й ОБГОВОРЕННЯ

Геном ЦВС-3 містить дві основні відкриті рамки зчитування ORF–ORF1, ORF2, — а також ORF3. ORF1 (положення 216-1106 на кільцевій ДНК ізоляту референтного штаму NC\_031753) кодує репліказу, ORF2, яка направлена протилежно по відношенню до ORF1 (положення 1336-1980), — білок капсиду, ORF3 — білок довжиною 231 амінокислотний залишок (а.з.), функція якого залишається невідомою.

### Консервативні G-квадруплекси у геномі ЦВС-3

На першому етапі за допомогою програми QGRS Mapper знайдено 4 потенційні G-квадруплекси, які утворено трьома G-тетрадами, в геномі випадково вибраного ізоляту ЦВС-3 PP067099 з повним геномом, G-рахунок яких становив від 58 до 72 (Табл. 1). Потім за допомогою програми BLAST nucleotide через пошук G-багатого мотиву в позиції 1164–1178 ізоляту PP067099 встановлено, що ця G-багата послідовність є висококонсервативним мотивом: число повногеномних ізолятів ЦВС-3 з цим мотивом становить 1138. Три інші G-багаті мотиви є також висококонсервативними для геному ЦВС-3, оскільки число повногеномних ізолятів вірусу з цими мотивами є більше 1000 для кожного мотиву.

Таблиця 1. Консервативні мотиви потенційних G-квадруплексів в геномі ізолятів ЦВС-3 з повним геномом (G-багаті фрагменти послідовностей підкреслено). Параметр pLDDT характеризує якість передбачення 3D моделі в AlphaFold 3 за шкалою від 0 до 100. Позиції наведено для ізоляту PP067099 ЦВС-3. N - число ізолятів ЦВС-3 в GenBank зі 100% рівнем ідентичності мотиву G4

Table 1. Conserved motifs of putative G-quadruplexes in the genome of PCV-3 isolates with complete genomes (G-rich sequence fragments are underlined). The pLDDT parameter characterizes the quality of the 3D model prediction in AlphaFold 3 on a scale from 0 to 100. The positions are shown for the PP067099 PCV-3 isolate. N is the number of PCV-3 isolates in GenBank with 100% identity to the G4 motif

Позиція, н.	Мотив G-квадруплексу	G-рахунок	pLDDT	N
645–680, ген <i>rep</i>	5'- <u>GGG</u> TTTGCCTGATTTTTGCG <u>GGG</u> TGAT <u>GGG</u> GTT <u>GGG</u> -3'	59	90 > pLDDT > 70	1172
802–833, ген <i>rep</i>	5'- <u>GGG</u> CCTTGGT <u>GGG</u> ATGGTTA TAAT <u>GGG</u> GA <u>GGG</u> -3'	63	70 > pLDDT > 50	1117
893–930, ген <i>rep</i>	5'- <u>GGG</u> ACAGGTACCCTCTGAG <u>GGT</u> TCCTGTAA <u>GGG</u> T <u>GGG</u> -3'	58	70 > pLDDT > 50	1011
1164–1178, промотор гена <i>cap</i>	5'- <u>GGG</u> T <u>GGG</u> T <u>GGG</u> T <u>GGG</u> -3'	72	pLDDT > 90	1138

На другому етапі для визначення еволюційно найвіддаленіших ізолятів серед 1138 ізолятів ЦВС-3 з повним геномом побудовано філогенетичні дерева на основі гена *cap*, що кодує білок капсиду (Рис. 1), гена *rep*, що кодує репліказу (Рис. 2), та на основі ізолятів ЦВС-3 з повним геномом (результат не показано).

Для проведення 3D моделювання структури білків та подальшого їх біоінформатичного аналізу вибрано ізоляти MT075517 та MG 372488, які належать до різних клад та є еволюційно найвіддаленішими за результатами філогенетичного аналізу на основі гена *cap*. З аналізу зазначених вище філогенетичних дерев випливає, що еволюційно найвіддаленішими ізолятами за геном *rep* є ізоляти MN491030 та MZ851986.

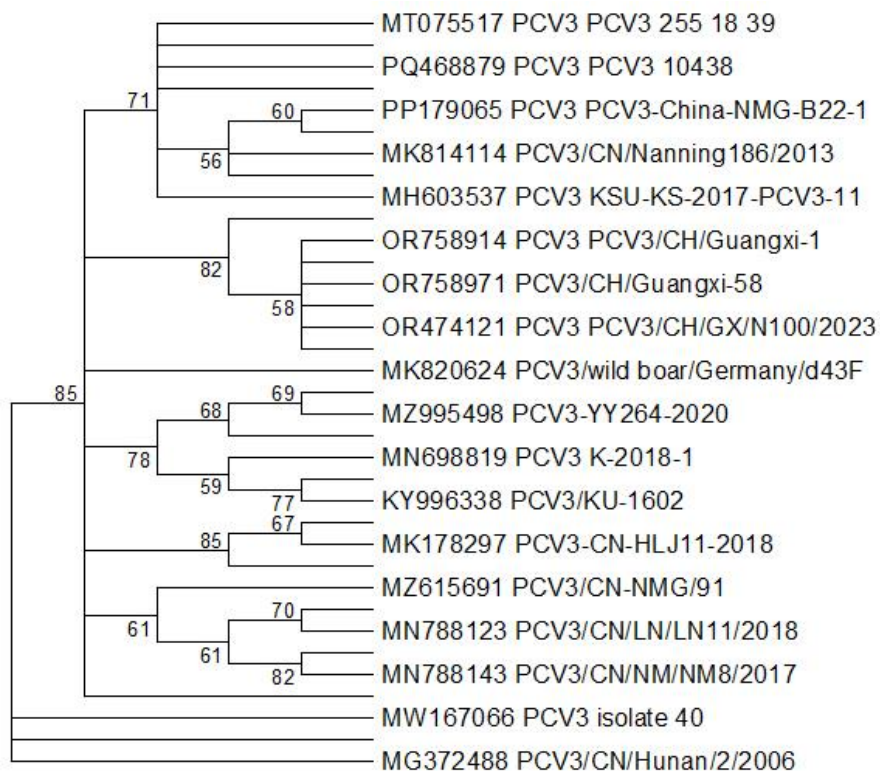


Рис. 1. Філогенетичні відношення між 36 еволюційно найвіддаленішими ізолятами (які визначено за результатами філогенетичного аналізу 1138 ізолятів ЦВС-3 з повним геномом; ізоляти, що перекриваються, не показано) на основі гена *cap*, що кодує білок капсиду. Наведено значення бутстреп-аналізу (500 повторів) для оцінки достовірності дерева, а також номери ізолятів в GenBank. Скорочене дерево побудовано з використанням р-дистанцій, методу приєднання сусідів шляхом об'єднання гілок з низькою достовірністю.

Fig. 1. Phylogenetic relationships between the 36 evolutionarily most distant isolates (determined from a phylogenetic analysis of 1138 PCV-3 isolates with complete genome; overlapping isolates not shown) based on the *cap* gene encoding the capsid protein. Bootstrap values (500 replicates) for assessing the reliability of the tree are shown, as well as the isolate numbers in GenBank. Condensed tree was constructed by p-distances, the neighbor joining method by merging low-confidence branches.

На третьому етапі використано AlphaFold 3 для побудови 3D моделей потенційних G4s в геномі ЦВС-3. В якості контролю коректності побудови цих моделей використовували 3D моделі G4s, утворення одного з яких експериментально визначено в прегеномній РНК (пгРНК) вірусу гепатиту Б (ВГБ) в роботі [30]. ПгРНК — це проміжна мРНК, що утворюється при реплікації ВГБ та служить матрицею для зворотної транскрипції і трансляції вірусних білків.

В геномі ВГБ авторами роботи [31] знайдено 8 потенційних G4s з петлею до 10 н., для яких значення G-рахунку варіювало від 15 до 20. У цьому дослідженні автори зосередилися на РНК-G-квадруплексі, розташованому в положенні 1884–1902 (позицію наведено для ізоляту V01460) пгРНК ВГБ, і експериментально підтвердили його роль у стимулюванні трансляції. За допомогою низки біофізичних (спектроскопії кругового дихроїзму, плавлення олігонуклеотидних дуплексів), імунологічних (ІФА) та молекулярно-генетичних (кількісна ПЛР) автори показали, що зазначений G4 відіграє життєво важливу роль у регуляції трансляції вірусної мРНК.

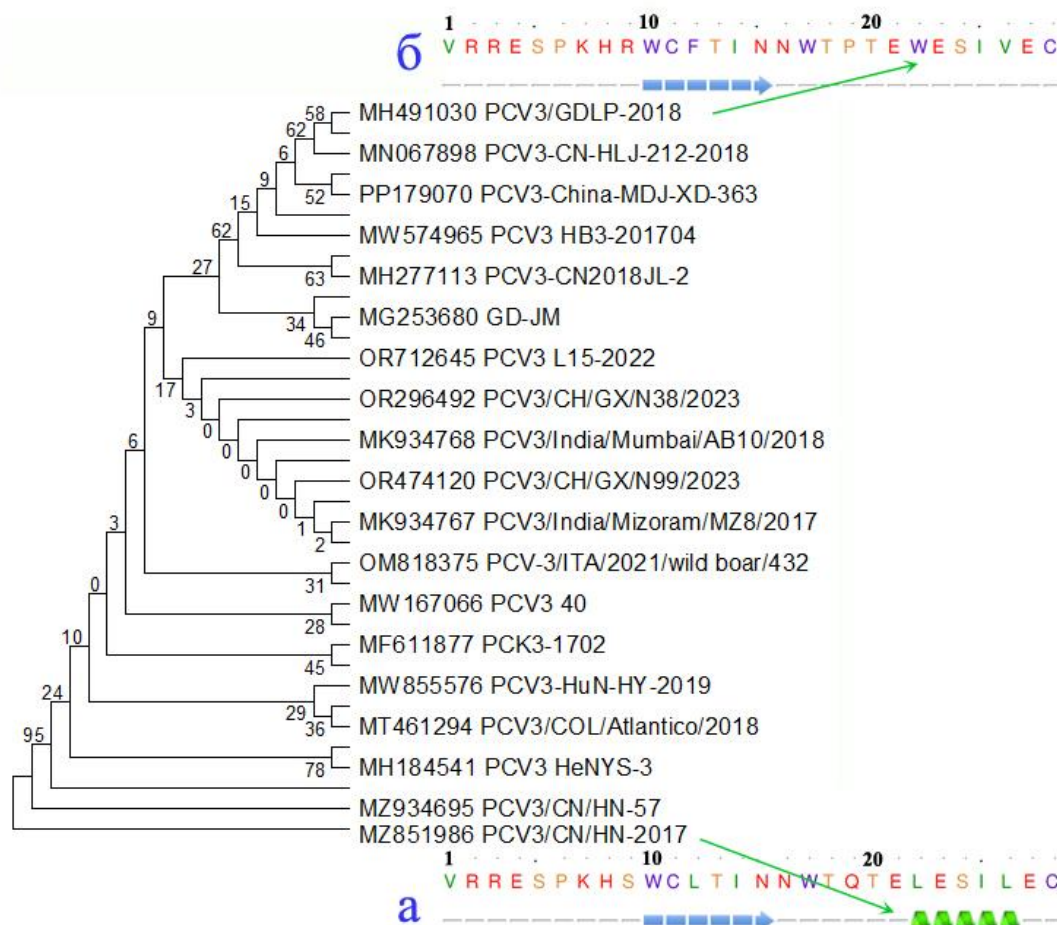


Рис. 2. Філогенетичні відношення між 36 еволюційно найвіддаленішими ізолятами ЦВС-3 (які визначено за результатами філогенетичного аналізу 1138 ізолятів ЦВС-3 з повним геномом; ізоляти, що перекриваються, не показано) на основі гена *rep*, що кодує репліказу. Наведено значення бутстреп-аналізу (500 повторів) для оцінки достовірності дерева, а також номери ізолятів в GenBank. Скорочене дерево побудовано з використанням р-дистанцій, методу приєднання сусідів шляхом об'єднання гілок з низькою достовірністю. Стрілки вказують на утворення  $\alpha$ -спіралі в пентапептиді (позиції 22–26 амінокислотних залишків) реплікази ізоляту MZ851986 (a) порівняно з її відсутністю в репліказі ізоляту MH491030 (b).

Fig. 2. Phylogenetic relationships between 36 most distant isolates of PCV-3 (determined from a phylogenetic analysis of 1138 PCV-3 isolates with complete genome; overlapping isolates not shown) based on the *rep* gene encoding the replicase. Bootstrap values (500 replicates) for assessing the reliability of the tree are indicated, as well as the isolate numbers in GenBank. Condensed tree was constructed by p-distances, the neighbor joining method by merging low-confidence branches. Arrows indicate the formation of  $\alpha$ -helix in the pentapeptide (22–26 positions of amino acid residues) in the replicase of MZ851986 isolate (a) in comparison with its absence in the replicase of MH491030 isolate (b).

З 8-ми знайдених в роботі [31] G4s, що утворено двома G-тетрадами, 3D моделі G4s нами побудовано тільки для трьох, послідовності яких наведено в Табл. 2 (моделі для двох з них наведено на Рис. 3а – Рис. 3г). Для інших мотивів, в тому числі, й для мотиву в позиції 2987–3009, побудовано 3D моделі дуплексів або аптамерів (однониткових олігонуклеотидів), а не G-квадруплексів (результати не показано). Відсутність третьої тетради в G4s ВГБ веде до менш стабільних структур порівняно з тритетрадними G4s, але утворення G4s в низці випадків є ще можливим.

Таблиця 2. Консервативні мотиви потенційних G-квадруплексів в РНК вірусу гепатиту Б (G-багаті фрагменти послідовностей підкреслено). Послідовності наведено для ізоляту V01460 ВГБ. G4 в РНК ВГБ в позиції 1884-1902 експериментально перевірено в роботі [31], G4 в ДНК ВГБ\* в позиції 1202–1220 — в роботі [32]

Table 2. Conserved motifs of putative G-quadruplexes in hepatitis B virus RNA (G-rich sequence fragments are underlined). Sequences are given for an V01460 HBV isolate. RNA G4 at position 1884-1902 and DNA G4 at position 1202–1220 were experimentally tested in [31] and [32] respectively.

Позиція, н.	Мотив G-квадруплексу	G-раху- нок	pLDDT
1202-1220 ген Р	5'- <u>GGCUGGG</u> GCUUG <u>GUCAUGG</u> -3'	20	pLDDT > 90
	5'- <u>GGCTGGGG</u> GCTT <u>GGT</u> TCAT <u>GG</u> -3' (*)	20	90 > pLDDT > 70
1774-1795 ген С	5'- <u>GGAGG</u> CUGUA <u>GG</u> CAUAAAU <u>UGG</u> -3'	14	90 > pLDDT > 70
1884-1902 ген С	5'- <u>GGGU</u> <u>GG</u> CUUU <u>GGGG</u> CAU <u>GG</u> -3'	18	90 > pLDDT > 70
2987-3009 ген S	5'- <u>GGAG</u> CU <u>GGAG</u> CAUUC <u>GGGCUGGG</u> -3'	18	-

З наведених на Рис. 3 3D моделей G4s тільки G4 в положенні 1202-1220 РНК ВГБ, який має найвищий параметр pLDDT (pLDDT > 90, Рис. 3а та Рис. 3б), утворено двома досконалими тетрадами (G13-G7-G2-G19 та G15-G6-G1-G18). G4 в цій же позиції в ДНК ВГБ (існування якого експериментально підтверджено в роботі [31]) є недосконалим, оскільки утворено однією досконалою тетрадою (G1-G7-G13-G19) та однією недосконалою структурою (G2-G6-G18), характеризується порівняно з РНК G4 меншим параметром pLDDT (90 > pLDDT > 70). З порівняння 3D моделей G4s в РНК (Рис. 3а та Рис. 3б) та ДНК (Рис. 3г та Рис. 3д) ВГБ в позиції 1202–1220 впливає, що вони також відрізняються сіткою водневих зв'язків. Зокрема, для G18 в РНК визначено три водневих зв'язки (G1 N<sub>7</sub> ... HN<sub>2</sub> G18 (на який вказує стрілка на Рис. 3б), G1 O<sub>6</sub> ... HN<sub>2</sub> G18, G18 N<sub>1</sub>H... O<sub>6</sub> G1) та координаційний зв'язок з іоном K<sup>+</sup> G18 O<sub>6</sub> ... K<sup>+</sup>, в той час як для G18 в РНК — тільки координаційний зв'язок з іоном K<sup>+</sup> (dG18 O<sub>6</sub> ... K<sup>+</sup>, на який вказує стрілка на Рис. 3д).

З теоретично визначених на першому етапі чотирьох консервативних G-мотивів в геномній ДНК ЦВС-3 для всіх побудовано 3D моделі G4s, які містять три G-тетради (Рис. 4, Рис. 5).

Додатково в якості контролю коректності 3D моделей G4s ЦВС-3 нами побудовано 3D модель G-багатого мотиву в промоторі гена hTERT (human telomerase reverse transcriptase) (результат не показано). Цей фрагмент довжиною 68 н. містить три G4s з трьох тетрад, утворення яких експериментально підтверджено за допомогою методів кругового дихроїзму, термічної денатурації, ЯМР, розсіювання рентгенівського випромінювання [33].

Останні досягнення в розробці експериментальних методів (кріоелектронної мікроскопії, ЯМР, рентгенструктурного аналізу тощо) та обробці даних дозволили охарактеризувати структури в нуклеїнових кислотах з високою роздільною здатністю (0,3–0,8 нм). Так, за допомогою FRET-аналізу визначено кристалічну структуру 3WJ (PDB: 4KZ2) з роздільною здатністю 0,3 нм, яку знайдено в РНК пакувального мотору

бактеріофага phi29, що зчеплений з пакуванням ДНК [34]. В роботах [35] та [36] за допомогою модифікації ЯМР визначено існування структури 3WJ в геномі ВІЛ (PDB: 7lva).

Раніше авторами в роботі [30] проведено оцінку якості передбачення 3D моделі біополімерів за допомогою AlphaFold 3 через порівняння з цими експериментально визначеними структурами з атомарною роздільною здатністю. Показано, що 3D моделі вторинних 3WJ-структур в геномі ВІЛ та РНК бактеріофага phi29, які передбачено за допомогою AlphaFold 3, мають приблизно таку ж точність, як і 3WJs, структуру яких визначено за допомогою експериментальних методів та координати занесено в базу даних PDB.

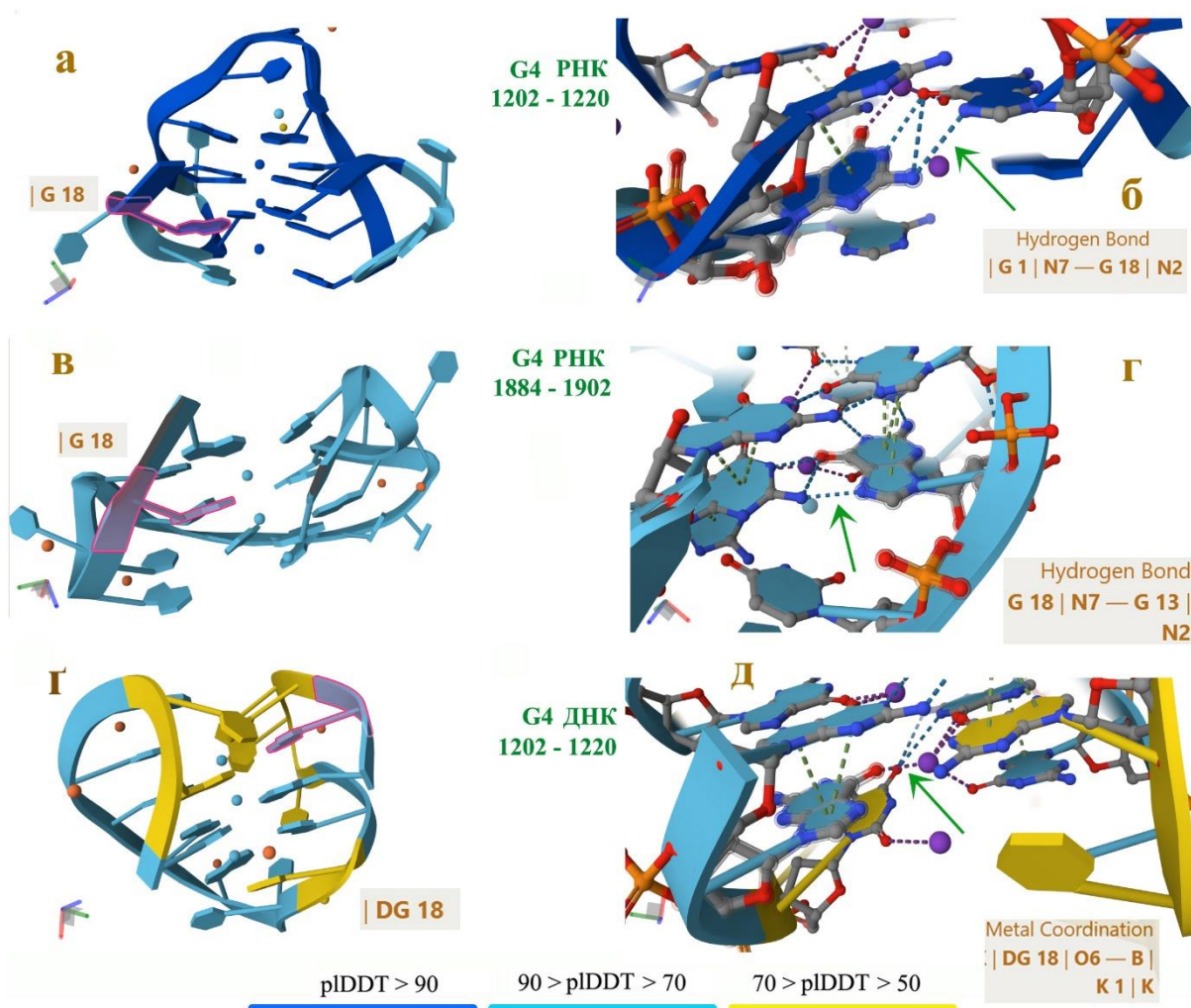


Рис. 3. 3D моделі G-багатих фрагментів в РНК (а-г) та ДНК (е-г) вірусу гепатиту Б з потенціалом утворення G-квадруплексів двома G-тетрадами. Позицію наведено для ізоляту V01460 ВГБ. а, б — pLDDT > 90, позиція 1202–1220; в, г — 90 > pLDDT > 70; позиція 1884–1902; е, г — положення 1202–1220. Стрілками показано водневі зв'язки (б — G1 N7 ... HN<sub>2</sub> G18, г — G18 N7 ... HN<sub>2</sub> G13) та координаційний зв'язок з іоном K<sup>+</sup> (д — dG18 O<sub>6</sub> ... K<sup>+</sup>).

Fig. 3. 3D models of G-rich fragments in hepatitis B virus RNA (a–d) and DNA (e–f) with the potential for G-quadruplex formation by two G-tetrads. Positions are indicated for V01460 HBV isolate. a, b — pLDDT > 90, 1202–1220; c, d — 90 > pLDDT > 70; b — 1884–1902 position; e, f — 1202–1220 position. The arrows indicate hydrogen bonds (b — G1 N7 ... HN<sub>2</sub> G18, d — G18 N7 ... HN<sub>2</sub> G13) and the coordination bond with K<sup>+</sup> ion (f — dG18 O<sub>6</sub> ... K<sup>+</sup>).

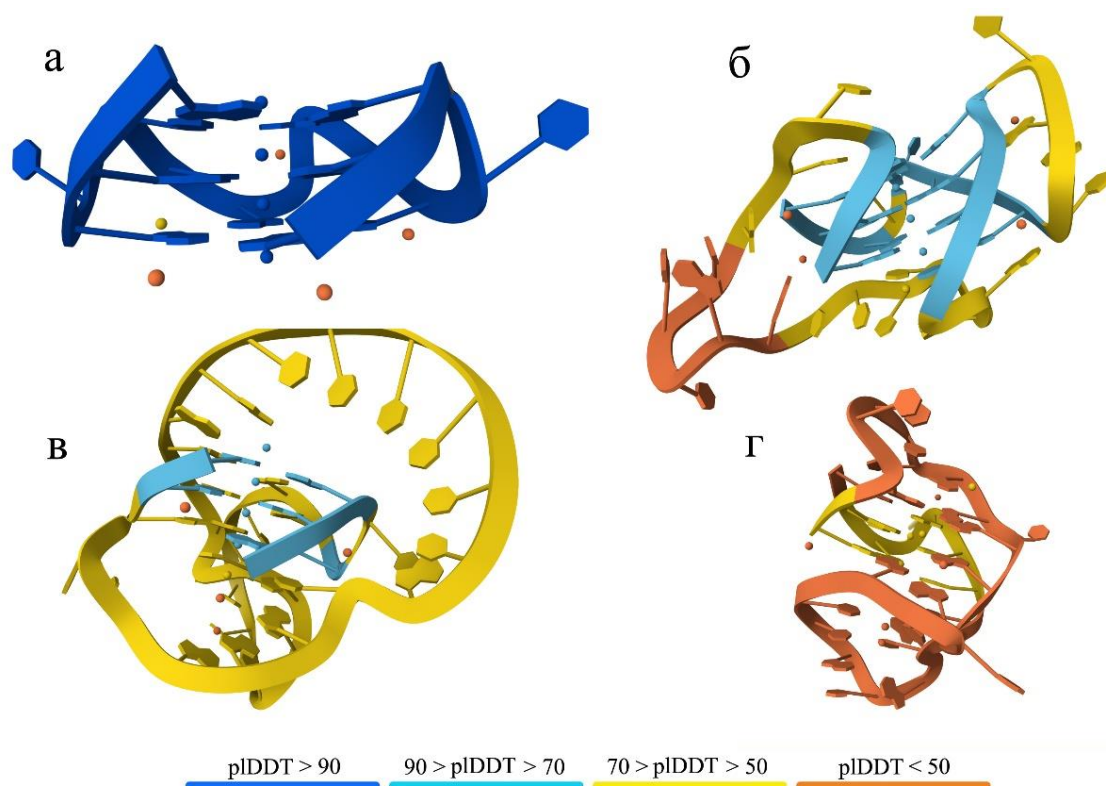


Рис. 4. 3D моделі консервативних G-квадруплексів, які утворено трьома G-тетрадами, в одноланцюговій ДНК ізолятів ЦВС-3 для фрагмента довжиною 15 н. в положенні 1164–1178 промотора гена *cap* (а), довжиною 36 н. в положенні 645–680 (б), довжиною 32 н. в положенні 802–833 (в), довжиною 38 н. в положенні 893–930 (г) гена, що кодує репліказу, з 10 іонами  $K^+$ . (а) —  $pLDDT > 90$ , (б, в)  $90 > pLDDT > 70$ , (г) —  $70 > pLDDT > 50$ .

Fig. 4. 3D models of conservative G-quadruplexes, which are formed by three G-tetrads, in single-stranded DNA of PCV-3 isolates for a fragment with length of 15 nt at 1164–1179 position of gene *cap* promoter (a), fragment with length of 36 nt of the gene encoding the replicase at 645–680 position (b), fragment with length of 32 nt at 802–833 position (c), fragment with length of 38 nt at 893–930 position (d) with 10  $K^+$  ions (a) —  $pLDDT > 90$ , (b, c)  $90 > pLDDT > 70$ , (d) —  $70 > pLDDT > 50$

В роботі [37] промоторну область визначили як область, яка локалізована на 500 н. вище сайту старту транскрипції. Всі промотори безпосередньо ранніх генів вірусу простого герпеса першого типу містять консервативні послідовності, які можуть утворювати G4s [38], що підкреслює можливу роль G4s в транскрипційному та трансляційному регулюванні вірусних генів, а також в пошкодженні та відновленні провірусної ДНК [2].

З чотирьох визначених потенційних G4s в геномі ЦВС-3 G-квадруплекс в положенні 1164–1178 знаходиться в області промотору гена *cap*, а інші три G4s — всередині гена *rep*. Для порівняння, з трьох потенційних G4s в геномі ВГБ G-квадруплекс в положенні 1776–1797 є локалізованим в області промотору гена X. Інші два G4s знаходяться всередині генів P та C.

Для підтвердження структури G4s, в тому числі, з високою роздільною здатністю за допомогою ЯМР, криоелектронної мікроскопії [39], дослідники, як правило, використовують аптамери довжиною 12–46 н. [30, 33, 40, 41]. Проте в геномі патогенів

G4s *in vivo* та *in vitro*, якщо вони є термодинамічно вигідними, мають існувати в складі макромолекул, а не тільки аптамерів. З метою перевірки цієї тези нами побудовано 3D моделі послідовностей аптамерів, які містять G-мотиви збільшеної довжини на 10–20 н. (по 5–10 н. з 5'- та 3'-кінця) як для G-багатих фрагментів в геномі ВГБ (Табл. 2), так і для таких ЦВС-3 (Табл. 1).

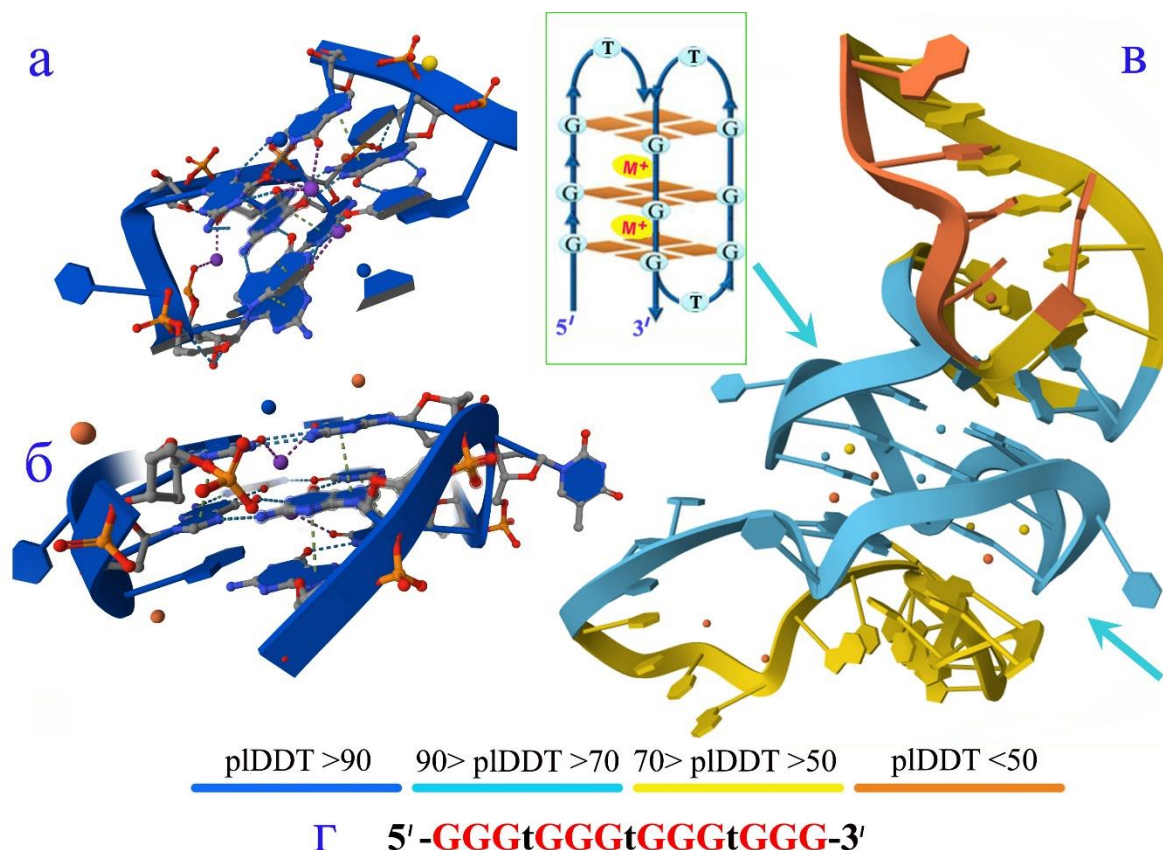


Рис. 5. 3D моделі консервативних G-квадруплексів, які утворено трьома G-тетрадами, в одноланцюговій ДНК ізолятів ЦВС-3 для фрагмента довжиною 15 н. в положенні 1164–1178 з 8 Na<sup>+</sup> (а, б) та довжиною 57 н. в положенні 1144–1200 з 20 іонами Na<sup>+</sup> (в; стрілки вказують на G4) в промоторі гена *cap*. (а, б) — pLDDT > 90, (в) — 90 > pLDDT > 70. (г) — мотив G4. Пунктиром показано водневі зв'язки між комплементарними нуклеотидами в тетраді та стекінг взаємодію між нуклеотидами паралельних тетрад в різних площинах (а, б). На вставці — схема консервативного внутрішньомолекулярного антипаралельного G-квадруплексу, що утворено стекінг взаємодією трьох G-тетрад, довжиною 15 н. у позиції 1164–1178 ДНК ЦВС-3 та стабілізовано центральним сольватованим моновалентним катіоном (K<sup>+</sup> > Na<sup>+</sup> > Li<sup>+</sup>). G-квартет, який містить чотири копланарних гуанідини, стабілізовано водневими зв'язками.

Fig. 5. 3D models of conservative G-quadruplexes formed by three G-tetrads in single-stranded DNA of PCV-3 isolates for a fragment with length of 15 nt at 1164–1178 position with 8 Na<sup>+</sup> (a, b) and fragment with length of 57 nt at 1144–1200 position with 20 Na<sup>+</sup> ions (c; arrows point to G4) in the gene *cap* promoter. (a, b) — pLDDT > 90, (c) — 90 > pLDDT > 70. (d) — G4 motif. The dotted line shows hydrogen bonds between complementary nucleotides in the tetrad and stacking interaction between nucleotides of parallel tetrads in the different planes (a, b). Inset — a scheme of a conservative intramolecular antiparallel G-quadruplex formed by stacking interaction of three G-tetrads with length of 15 nt at 1164–1178 position of the PCV-3 DNA and stabilized by a central solvated monovalent cation (K<sup>+</sup> > Na<sup>+</sup> > Li<sup>+</sup>). The G-quartet that contains four coplanar guanidines is stabilized by hydrogen bonds.

3D моделювання трьох аптамерів збільшеної довжини на 10–20 н., які містять потенційні G4s в РНК ВГБ, показало відсутність G4s у теоретично визначених G-багатих фрагментах (послідовності яких наведено в Табл. 2).

Для G4s в геномній ДНК ЦВС-3 ситуація є іншою. Незважаючи на збільшення довжини аптамерів структура трьох визначених G4s (послідовності яких наведено в Табл. 1) зберігається. На Рис. 5в наведено 3D модель для фрагмента довжиною 57 н. (в положенні 1144–1200 н.), який містить G-багатий мотив довжиною 15 н. (в положенні 1164–1178). 3D моделі двох інших визначених потенційних G4s в геномі ЦВС-3 збільшеної довжини на 20 н. (по 10 н. з 5'- та 3'-кінця) не наведено.

Запропоноване та апробоване нами дослідження не коротких G-багатих мотивів, а аптамерів збільшеної довжини узгоджується з наведеними нижче висновками в роботі [2].

Для коротких нуклеотидних послідовностей з потенціалом утворення G4s отримано структури з високою роздільною здатністю за допомогою дифракції рентгенівських променів [42], ЯМР-спектроскопії [43], криоелектронної мікроскопії [38]. Проте структура таких аптамерів часто відрізняється від їхньої нативної конформації через низку чинників, в тому числі, через вплив оточуючих нуклеотидів. Важливою перевагою AlphaFold 3 над зазначеними методами є можливість отримувати тривимірні моделі для нуклеїнових кислот довжиною до 5 т.н. з роздільною здатністю, яка є порівняною з такою експериментальних методів.

Беручи до уваги наведене вище, нами отримано 3D модель фрагмента довжиною 129 н., який містить два G4s, які знаходяться на відстані понад 100 н. (Табл. 1, в положеннях 802–833 та 893–930). Проте надійність 3D моделі для цих двох G4s в цьому випадку є нижчою порівняно з мономерними G4s:  $70 > \text{pLDDT} > 50$  та  $50 > \text{pLDDT}$  для кожного з двох G4s відповідно (3D модель не наведено).

Крім того, проведено пошук внутрішньомолекулярних 3WJ (three-way junctions) в мРНК ізоляту MT075518 ЦВС-3 через побудову 3D моделей за допомогою AlphaFold 3.

Структура 3WJ є мультиспіральним з'єднанням, яке утворено трьома з'єднаними у точці зв'язування дуплексами. Попередньо послідовність мРНК було поділено на фрагменти, що не перекриваються, довжиною 300 н., зі зсувом на 150 н. На відміну від спумавірусу великої рогаатої худоби (СВ ВРХ), для якого авторами знайдено для набору з 37 ізолятів СВ ВРХ консервативний мотив 3WJ довжиною 73 н. зі 100 %-им рівнем подібності [36], в мРНК ЦВС-3 структуру 3WJ не знайдено.

### 3D моделі білків ЦВС-3

Тривимірна структура надає важливу інформацію для розуміння біологічних функцій білків. Google DeepMind надає безкоштовний доступ до AlphaFold 3 для некомерційних досліджень через платформу AlphaFold Server, що дозволило отримати наведені в роботі результати.

Для 3D моделювання структури білків вибрано пари ізолятів MH491030 і MZ851986 та MT075517 і MG372488 ЦВС-3, які є еволюційно найвіддаленішими за результатами філогенетичного аналізу на основі гена *cap*, що кодує білок капсиду (Рис. 1), та на основі гена *rep*, що кодує репліказу (Рис. 2).

Фермент реплікази Rep містить 296 амінокислотних залишків (а.з.), а білок капсиду Cap — 214 а.з. З побудованих 3D моделей білків еволюційно найвіддаленіших ізолятів ЦВС-3 за *rep* та *cap* генами (Рис. 6) встановлено їхні особливості, а саме кількість  $\alpha$ -спіралей,  $\beta$ -листів,  $\beta$ -поворотів та водневих зв'язків (Табл. 3). Всі передбачені структури характеризуються вбудованими в AlphaFold 3 показниками для оцінки достовірності 3D моделі — pLDDT, pTM та ipTM.

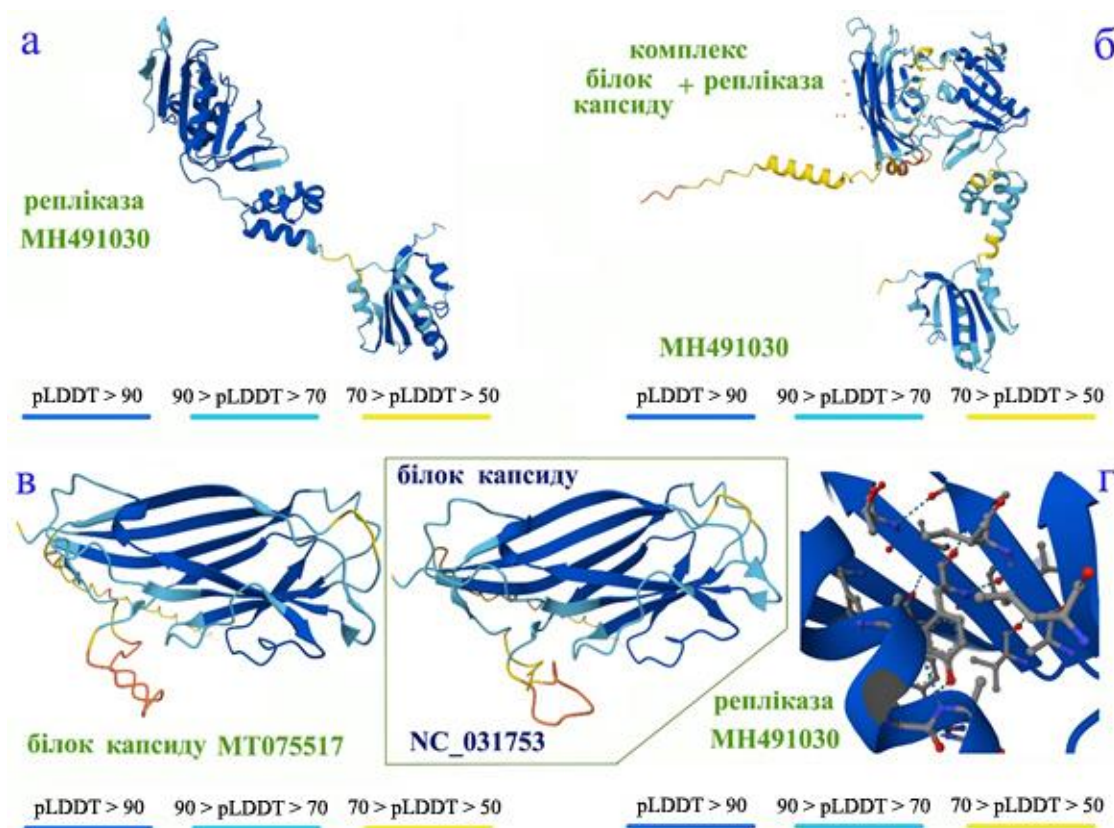


Рис. 6. 3D модель реплікази довжиною 296 а.з. для ізоляту MN491030 (а), білка капсиду довжиною 214 а.з. ізоляту MT075517 (в), комплексу білок капсиду — репліказа з 25 іонами  $\text{Na}^+$  для ізоляту MN491030 (б),  $\alpha$ -спіралі та  $\beta$ -листів реплікази ізоляту MN491030 (г). На вставці — 3D модель білка капсиду референтного штаму NC\_031753.

Fig. 6. 3D model replicase with length of 296 aa for MN491030 isolate (a), capsid protein with length of 214 aa of MT075517 isolate (c), capsid protein — replicase complex with 25  $\text{Na}^+$  ions for MN491030 isolate (b),  $\alpha$ -helix and  $\beta$ -sheets of the replicase of MN491030 isolate (d). Inset — 3D model capsid protein of NC\_031753 reference strain.

Параметр pLDDT визначає достовірність локальної структури, оцінюючи, наскільки добре 3D модель узгоджуватиметься з експериментальною структурою. Для білків цей параметр є аналогічним параметру IDDT- $\text{Ca}$ , але має більшу дискретність, оскільки може змінюватися для кожного атома, а не для кожного залишку. Передбачувана структура порівнюється з справжньою структурою з бази даних PDB за допомогою тесту IDDT, який визначає точність домену без необхідності сегментації ланцюгових структур. Відстані обчислюються або між усіма важкими атомами (IDDT), або лише між атомами  $\text{Ca}$  для вимірювання точності остова (IDDT- $\text{Ca}$ ).

Для переважної частини кожного з трьох доменів реплікази ізолятів ЦВС-3, які містять  $\beta$ -листи та  $\alpha$ -спіралі (Рис. 6а, Табл. 3), значення pLDDT перевищує 90, а для незначної частини доменів та неупорядкованої структури реплікази – в інтервалі від 70 до 90. Отримані значення pLDDT вказують на високу достовірність і точне передбачення 3D моделі реплікази ЦВС-3. NC\_031753 – референтний штам [44].

Білок капсиду еволюційно найвіддаленіших ізолятів ЦВС-3 містить  $\beta$ -листи, значення pLDDT для переважного числа яких перевищує 90, а неупорядкована структура характеризується значеннями pLDDT від рахунку менше 50 до не більше 70 (Рис. 6в).

Таблиця 3. Параметри білків Rep та Cap еволюційно найвіддаленіших ізолятів МН491030 і МZ851986 та МТ075517 і МG372488 з 1138 ізолятів ЦВС-3 з повним геномом згідно з філогенетичними відношеннями для повновимірних генів *rep*, який кодує репліказу, та *cap*, який кодує білок капсиду. NC\_031753 — референтний штаб [44]

Table 3. Parameters of Rep and Cap proteins of the evolutionarily most distant МН491030 and МZ851986, МТ075517 and МG372488 isolates from 1138 of PCV-3 isolates with a complete genome according to phylogenetic relationships for the full-length *rep* gene which encodes the replicase and *cap* gene which encodes the capsid protein. NC\_031753 — reference strain [44]

Білок	Ізолят	Число α-спіралей, %	Число β-листів, %	Число β-поворотів, %	Число водневих зв'язків
Репліказ	МН491030	21	26	32	188
	МZ851986	23	27	31	189
	NC_031753	12	16	32	190
Капсидний	МТ075517	0	43	22	129
	МG372488	15	35	19	124
	NC_031753	1	16	22	126

Параметри рТМ та ірТМ показують оцінку передбаченої матриці моделювання (рТМ; predicted template modeling) та оцінку передбаченого інтерфейсу матриці моделювання (ірТМ; interface predicted template modeling). Обидві ці оцінки виводяться з показника матриці моделювання (ТМ), яка визначає точність всієї структури [45, 46]. Рахунок рТМ вище 0,5 означає, що загальний передбачений фолдинг має бути подібним до справжньої структури. Параметр ірТМ вимірює точність передбаченого відносного положення субодиниць всередині комплексу.

Рахунок рТМ для 3D моделей реплікази та білка капсиду ЦВС-3 (Рис. 6) становить 0,7, що задовольняє умові достовірності даних моделей. ірТМ показує точність передбаченого відносного положення всередині комплексу, і тому цей показник для зазначених вище моделей не визначено.

Репліказа ЦВС — це вірусний білок, необхідний для реплікації одноланцюгової кільцевої ДНК вірусу, що працює за механізмом кільця, що котиться, і взаємодіє з іншими вірусними білками, такими як капсидний білок Cap [47]. Репліказа (Rep) та білок капсиду (Cap) взаємодіють, причому Rep безпосередньо зв'язується з Cap, утворюючи комплекси, які впливають на реплікацію, складання та патогенез вірусу, часто за участю факторів хазяїна. ірТМ комплексу білок капсиду — репліказа за 25 іонів Na<sup>+</sup> для ізоляту МН491030 (Рис. 6) становить 0,11, а рТМ — 0,39.

В роботі [48] розглянуто точність передбачення тривимірних структур за допомогою AlphaFold 3 через порівняння прогнозів 3D моделей з експериментально визначеними структурами аптамерів у базі даних білків PDB, а також зі структурами аптамерів, які не включено до PDB та охарактеризовано іншими методами. Показано, що AlphaFold 3 ефективно моделював низку аптамерів, для яких є структура в базі даних PDB, включаючи G4s. Для аптамерів, щодо яких відсутні відомості в PDB, прогнози AlphaFold 3 характеризувалися меншою достовірністю, проте демонстрували достатнє збігання з експериментальними даними, точно прогнозуючи конформацію G4s.

Структурне порівняння 3D моделей білків двох еволюційно віддалених ізолятів ЦВС-3 показало, що їхня просторова організація відрізняється кількістю водневих

зв'язків,  $\alpha$ -спіралей,  $\beta$ -листів,  $\beta$ -поворотів (структур, що представляють собою повернену на  $180^\circ$  петлю) (Табл. 3). Визначена різниця може бути пов'язаною з замінами амінокислотних залишків різних класів.

### Мутаційний та еволюційний аналіз білків та генів ЦВС-3

За допомогою аналізу графіків ентропії Шеннона — залежності міри мінливості амінокислотних залишків в заданому положенні від їхньої позиції — виявлено поліморфні сайти для білка капсиду та реплікази еволюційно віддалених ізолятів ЦВС-3. Радикальні та консервативні заміни виявлено для обох білків, які транскрибуються з відкритих рамок зчитування ЦВС-3.

Нами ідентифіковано 20 та 21 амінокислотний варіабельний сайт для еволюційно віддалених ізолятів ЦВС-3 (Рис. 7). Відносна кількість варіабельних сайтів для реплікази та білка капсиду, що транскрибуються з відкритих рамок зчитування ORF1 та ORF2, становила: *Rep* — 6,7 % (20/296), *Cap* — 9,8 % (21/214).

Згідно з класичною шкалою гідрофобності Кайта-Дуліттла [49], яку широко застосовують для аналізу структури білків, до гідрофобних амінокислот належать ізолейцин, валін, лейцин, фенілаланін, цистеїн, метіонін, аланін (по порядку зменшення гідрофобності). Решта амінокислот належать до числа гідрофільних.

При порівнянні амінокислотних послідовностей реплікази цих двох ізолятів ЦВС-3 (Табл. 4) виявлено 19 різних замін, з яких 2 консервативні заміни E→D у позиціях 77 та 260. Найбільшу кількість замін (5) встановлено за участю гідрофільного серину (S), дві із яких є замінами на гідрофобні залишки аланіну (A) та ізолейцину (I). Чотири заміни виявлено для гідрофільного аргініну (R), одна із яких є заміною на гідрофобний залишок валіну (V).

Зазначені вище амінокислотні заміни ведуть до зростання гідрофобності і зміни поверхневих властивостей білків. Наприклад, заміна амінокислотного залишку Q→P (в положенні 125 для ізолятів MG372488 і MT075517, Табл. 5) веде до зниження полярності білка.

Оскільки найчастішою молекулярно-генетичною подією в ході молекулярної еволюції патогенів є мутації, нами проведено мутаційний аналіз генів реплікази *rep* та білка капсиду *cap*, 5'-нетрансльованої області (НТО, позиції 1–222), міжгенної області (позиції 1114–1342) повногеномних ізолятів ЦВС-3, а також геномної ДНК вірусу, розраховано швидкість мутацій (Табл. 6).

Для білка капсиду ізолятів ЦВС-3 виявлено 20 замін, із яких 9 є неконсервативними. Три однакові консервативні заміни (R→K) визначено у позиціях 27, 82, 148. Найбільшу кількість замін (6) виявлено за участю гідрофільних залишків лізину (K) та аргініну (R) (Табл. 5).

Встановлено, що найвищу швидкість мутацій має 5'-НТО кільцевої ДНК ЦВС-3, а найменшу — ген *rep*, що дозволяє розглядати його як потенційний маркер для детекції ЦВС-3 шляхом ПЛР.

Для встановлення напряму природного відбору, що діє на гени та фрагменти геному ЦВС-3, визначено відношення частот несинонімічних замін (dN) до синонімічних (dS) на нуклеотидний сайт dN/dS для генів *rep* та *cap*, 5'-НТО, міжгенної області повногеномних ізолятів ЦВС-3. За цим параметром встановлено, що тільки послідовність 5'-НТО перебуває під тиском позитивного відбору (Табл. 6), що узгоджується з встановленою для цієї області найвищою швидкістю мутацій.

З аналізу заряджених амінокислотних залишків в парі ізолятів MH491030 → MZ851986 випливає, що в репліказі ізоляту MH491030 є 5 позитивно заряджених а.з., 2 негативно заряджених а.з. Репліказа ізоляту MZ851986 містить 3 позитивно заряджених а.з., 2 негативно заряджених а.з., а також 1 полярний а.з., який позитивно заряджений.

Аналіз замін амінокислот, виходячи з їхньої полярності, показує, що репліказа ізоляту МН491030 характеризується збільшеною полярністю (4 а.з.) порівняно з такою ізоляту МZ851986 (3 а.з., один з яких, а саме, гістидин, несе позитивний заряд).

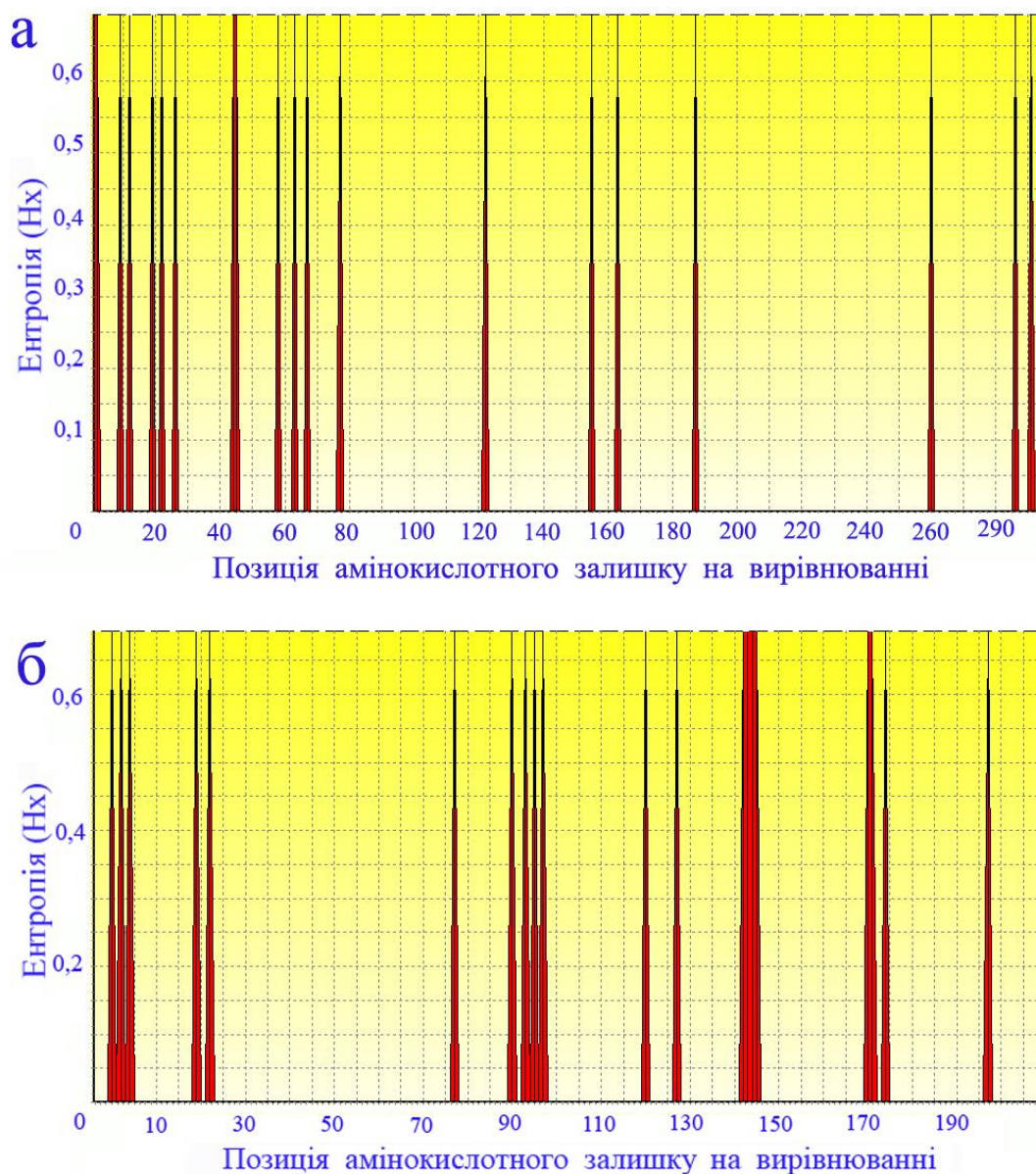


Рис. 7. Графік ентропії Шеннона для реплікази довжиною 296 а.з. еволюційно найвіддаленіших ізолятів МН491030 та МZ851986 (а) та для білка капсиду довжиною 214 а.з. еволюційно найвіддаленіших ізолятів МТ075517 та МG372488 (б). Міру мінливості кожного сайту отримано на підставі множинного вирівнювання амінокислотних послідовностей білків наведених вище ізолятів ЦВС-3.

Fig. 7. Shannon entropy plot for the replicase with a length of 296 aa of the evolutionarily most distant isolates МН491030 and МZ851986 (a) and the capsid protein with a length of 214 aa of the evolutionarily most distant isolates МТ075517 and МG372488 (b). The measure of variability of every site was obtained based on multiple alignment of the amino acid sequences of the proteins of the above mentioned PCV-3 isolates.

Таблиця 4. Амінокислотні заміни для реплікази еволюційно найвіддаленіших ізолятів МН491030→MZ851986 ЦВС-3 з локалізацією у  $\alpha$ -спіралях ( $\alpha$ ),  $\beta$ -листах ( $\beta$ ), неупорядкованих ділянках (-). Позицію наведено відповідно до графіка ентропії Шеннона на Рис. 7 а. Підкреслено заряджені, курсивом виділено полярні амінокислоти. Виділено позиції, заміна залишків в яких веде до утворення  $\alpha$ -спіралі в пентапептиді (позиції з 22 по 26 а.з.) реплікази ізоляту MZ851986

Table 4. Amino acid substitutions for the replicase of the evolutionarily most distant isolates МН491030→MZ851986 of PCV-3 with localization in  $\alpha$ -helices,  $\beta$ -sheets, disordered fragments. The position is indicated according to Shannon entropy plot in Fig. 7 a. Charged amino acids are underlined, polar amino acids are shown by italic. Positions are highlighted, the substitution of residues in which leads to the formation of  $\alpha$ -helix in the pentapeptide (22–26 aa position) in the replicase of MZ851986 isolate

Позиція	1	2	9	12	19	22	26
Заміна амінокислоти	<u>V</u> → <u>R</u> - → -	<u>R</u> →де- леція	<u>R</u> →S - → -	F→L $\beta$ → $\beta$	P→ <u>Q</u> - → -	W→L - → $\alpha$	V→L - → $\alpha$
Позиція	44	45	58	63	67	77	122
Заміна амінокислоти	<u>K</u> → <u>R</u> - → -	S→G - → -	N→S - → -	S→I - → -	<u>R</u> →P - → -	<u>E</u> → <u>D</u> $\beta$ → $\beta$	A→S - → -
Позиція	155	163	187	260	286	291	-
Заміна амінокислоти	P→A - → -	<u>Y</u> → <u>H</u> $\beta$ → $\beta$	<u>Q</u> → <u>K</u> $\beta$ → $\beta$	<u>E</u> → <u>D</u> - → -	<u>K</u> → <u>D</u> - → -	V→M - → -	-

Заміна тріптофану (W), ароматичної гетероциклічної амінокислоти, та валіну (V), аліфатичної амінокислоти, на лейцин (L), аліфатичну амінокислоту, в пентапептиді (позиції 22 а.з. та 26 а.з.; Табл. 4) веде до утворення  $\alpha$ -спіралі в пентапептиді реплікази ізоляту MZ851986 (позиції з 22 по 26 а.з.; Рис. 2а) на відміну від ізоляту МН491030 (Рис. 2б). Через наявність у вторинній структурі  $\alpha$ -спіралей та  $\beta$ -листів репліказа ізоляту MZ851986, як, доречі, й капсидні білки, може відноситися до білків « $\alpha+\beta$ » класу. Репліказу МН491030 через наявність у вторинній структурі тільки  $\beta$ -листів (Табл. 4) можна віднести до білків « $\beta$ »-класу (Табл. 4).

За будь-яких структурних переходів відбувається зміна загальної енергії системи, на яку, як відомо, впливають ентальпійний та ентропійний фактори, внесок яких є різним при формуванні тих чи інших елементів вторинної структури білка. В альфа-спіралях амінокислотні залишки утворюють більше внутрішньомолекулярних зв'язків, ніж у бета-листах [50]. Ентропія при утворенні альфа-спіралей знижується більшою мірою, ніж при утворенні бета-структури, що має сприяти переходу з альфа-спіралі до бета-листа. Через більшу кількість зв'язків між залишками ентальпія при утворенні альфа-спіралей знижується більшою мірою, ніж при формуванні бета-листів, що має сприяти переходу з бета-листа до альфа-спіралі.

Варто зазначити, що положення еволюційно найвіддаленіших ізолятів ЦВС-3, у тому числі, ізолятів MZ851986 та MG372488, на філогенетичному дереві визначено нами коректно за допомогою методу приєднання сусідів, що підтверджено положенням ізоляту MG372488 на дереві, яке визначено для повногеномних ізолятів ЦВС-3 за допомогою методу максимальної правдоподібності [51].

Капсидний білок ізоляту MG372488 має 7 позитивно заряджених амінокислотних залишків, 1 негативно заряджений та 6 полярних незаряджених залишків, в той час як білок капсиду ізоляту МТ075517 має 4 позитивно заряджених і 4 полярних незаряджених залишків.

Таблиця 5. Амінокислотні заміни для білка капсиду еволюційно найвіддаленіших ізолятів MG372488→MT075517 ЦВС-3 з локалізацією у  $\alpha$ -спіралях ( $\alpha$ ),  $\beta$ -листах ( $\beta$ ), неупорядкованих ділянках (-). Положення наведено відповідно до графіка ентропії Шеннона на Рис. 7 б. Підкреслено заряджені, курсивом виділено полярні амінокислоти

Table 5. Amino acid substitutions for the capsid protein of the evolutionarily most distant isolates MG372488→MT075517 of PCV-3 with localization in  $\alpha$ -helices,  $\beta$ -sheets, disordered fragments. The position is shown according to Shannon entropy plot in Fig. 7 b. Charged amino acids are underlined, polar amino acids are shown by italic

Позиція	5	7	9	24	27	82	95
Заміна амінокислоти	<u>P</u> →A - → -	<u>Y</u> →F - → -	<u>K</u> → <u>R</u> - → -	A→V $\alpha$ → $\beta$	<u>R</u> → <u>K</u> - → -	<u>R</u> → <u>K</u> $\beta$ → $\beta$	L→S $\beta$ → -
Позиція	98	100	102	125	132	147	148
Заміна амінокислоти	<u>K</u> → <u>Q</u> - → $\beta$	<u>K</u> →T - → $\beta$	L→T - → $\beta$	<u>Q</u> →P - → -	<u>G</u> → <u>R</u> - → -	L→P $\alpha$ → -	<u>R</u> → <u>K</u> $\alpha$ → -
Позиція	149	150	175	176	179	202	-
Заміна амінокислоти	<u>S</u> →P $\alpha$ → -	L→I $\alpha$ → $\beta$	<u>R</u> →T - → -	F→Y - → -	V→T $\alpha$ → -	<u>D</u> →G - → -	-

Таблиця 6. Швидкість мутацій та відношення частот несинонімічних замінь до синонімічних на нуклеотидний сайт dN/dS для генів і фрагментів геному ЦВС-3

Table 6. Mutation rate and frequencies ratio of nonsynonymous substitutions to synonymous ones per nucleotide site dN/dS for PCV-3 genes and genome fragments

Ген/фрагмент геному	<i>rep</i>	<i>cap</i>	5'-НТО	Міжгенна область	Геномна ДНК
Швидкість мутацій, %	0,003	0,009	0,032	0,025	0,021
dN/dS	0,309	0,628	3,230	0,927	-

Примітка. dN/dS >1 вказує на позитивний відбір, dN/dS <1 — очищуючий відбір.

Note. dN/dS >1 indicates positive selection, dN/dS <1 indicates purifying selection.

З наведеного вище випливає, що капсидний білок еволюційно віддалених ізолятів ЦВС-3 відрізняється розподілом заряду, полярністю залишків, що веде до різної здатності утворювати водневі зв'язки та, як наслідок, до різного числа  $\alpha$ -спіралей та  $\beta$ -листів. Репліказа цих ізолятів ЦВС-3 є значно консервативнішою структурою порівняно з капсидним білком через майже ідентичні значення наведених вище параметрів.

## ВИСНОВКИ

За допомогою біоінформатичних методів ідентифіковано потенційні G-квадруплекси в геномі цирковірусу свині третього типу, а також визначено особливості структурної організації білків, що транскрибуються з двох відкритих рамок зчитування, — реплікази та капсидного білка.

Проведено порівняння двох еволюційно найвіддаленіших ізолятів ЦВС-3 на основі побудованих за допомогою AlphaFold 3 3D моделей та мутаційного аналізу білків, еволюційного аналізу генів. 3D моделі реплікази, білка капсиду цирковірусу свині типу

3, а також чотирьох консервативних структурних мотивів — G4s, — які утворено трьома G-тетрадами, мають, як випливає з показників AlphaFold 3 [20], порівняну до експериментальних методів точність (0,1–2 нм для різних фрагментів білка, зокрема).

Отримані результати показують, що 3D моделі реплікази двох ізолятів ЦВС-3, які перебувають на максимальній еволюційній дистанції, незначно відрізняються числом  $\alpha$ -спіралей,  $\beta$ -листів,  $\beta$ -поворотів та водневих зв'язків на відміну від 3D моделей капсидного білка таких ізолятів, для яких виявлено значно більшу різницю між зазначеними вище параметрами. Мутаційний та еволюційний аналізи реплікази і капсидного білка та генів, що їх кодують, віддалених ізолятів ЦВС-3 вказують на значно більшу варіабельність гена *cap* та більше число варіабельних сайтів для капсидного білка порівняно з геном *rep* та репліказою відповідно. Зазначені молекулярно-генетичні відмінності генів та білків двох ізолятів ЦВС-3 ведуть до відмінностей тривимірних структур їхніх білків.

Найстабільніший G4 з чотирьох визначених в геномі ЦВС-3, який утворено трьома G-тетрадами, локалізовано у промоторі гена *cap*. Збільшення довжини фрагментів, які фланкують цей G4, не перешкоджає його фолдингу, що свідчить про його значущість у життєвому циклі ЦВС-3. Існування G4s з трьох тетрад в геномі цирковірусу на відміну від двотетрадних G4s вірусу гепатиту Б у складі не тільки аптамерів, а й молекул збільшеної довжини підтверджено за допомогою AlphaFold 3.

### КОНФЛІКТ ІНТЕРЕСІВ

Автори повідомляють про відсутність конфлікту інтересів.

### ПОДЯКА


Роботу частково підтримано грантом 34.02.01.01 Ф Національної академії аграрних наук України.

Автори висловлюють подяку проф. Yuchen Xia (State Key Laboratory of Virology and Hubei Province Key Laboratory of Allergy and Immunology, Institute of Medical Virology, TaiKang Center for Life and Medical Sciences, TaiKang Medical School, Wuhan University, Wuhan, China) за інформаційну підтримку (додаткову інформацію щодо ізоляту ВГБ) при підготовці статті.


Окремо автори щиро дякують невідомому рецензенту за надзвичайно ретельний, критичний аналіз статті та низку корисних зауважень, рекомендацій, порад, направлених на покращення статті.

### Authors' ORCID ID

О. Ю. Лиманська  <https://orcid.org/0000-0002-6022-0342>

О. В. Білойван  <https://orcid.org/0000-0002-9973-4551>

О. К. Балак  <https://orcid.org/0009-0000-8745-8089>

О. П. Лиманський  <https://orcid.org/0009-0007-9922-4619>

### REFERENCES

- Zareie AR, Dabral P, Verma SC. G-quadruplexes in the regulation of viral gene expressions and their impacts on controlling infection. *Pathogens*. 2024;13(1):60. <http://doi.org/10.3390/pathogens13010060>
- Monsen RC, Trent JO, Chaires JB. G-quadruplex DNA: a longer story. *Acc Chem Res*. 2022;55(22):3242–52. <http://doi.org/10.1021/acs.accounts.2c00519>
- Wulfridge P, Sarma K. Intertwining roles of R-loops and G-quadruplexes in DNA repair, transcription and genome organization. *Nat Cell Biol*. 2024;26(7):1025–36. <http://doi.org/10.1038/s41556-024-01437-4>
- Teng F-Y, Jiang Z-Z, Guo M, Tan X-Z, Chen F, Xi X-G, et al. G-quadruplex DNA: a novel target for drug design. *Cell Mol Life Sci*. 2021;78(19-20):6557–83. <http://doi.org/10.1007/s00018-021-03921-8>
- Balasubramanian S, Neidle S. G-quadruplex nucleic acids as therapeutic targets. *Curr Opin Chem Biol*. 2009;13(3):345–53. <http://doi.org/10.1016/j.cbpa.2009.04.637>

6. Teng Y, Girvan AC, Casson LK, Pierce WM, Qian M, Thomas SD, et al. AS1411 alters the localization of a complex containing protein arginine methyltransferase 5 and nucleolin. *Cancer Res.* 2007;67(21):10491–500. <http://doi.org/10.1158/0008-5472.CAN-06-4206>
7. Zhang X, Xu H, Sun R, Xiong G, Shi X. An insight into G-quadruplexes: Identification and potential therapeutic targets in livestock viruses. *Eur J Med Chem.* 2024;279:116848. <http://doi.org/10.1016/j.ejmech.2024.116848>
8. Fang P, Xie C, Pan T, Cheng T, Chen W, Xia S, et al. Unfolding of an RNA G-quadruplex motif in the negative strand genome of porcine reproductive and respiratory syndrome virus by host and viral helicases to promote viral replication. *Nucleic Acids Res.* 2023;51(19):10752–67. <https://doi.org/10.1093/nar/gkad759>
9. Li Y, Zhu Y, Wang Y, Feng Y, Li D, Li S, et al. Characterization of RNA G-quadruplexes in porcine epidemic diarrhea virus genome and the antiviral activity of G-quadruplex ligands. *Int J Biol Macromol.* 2023;231:123282. <http://doi.org/10.1016/j.ijbiomac.2023.123282>
10. Liu W, He X, Zhu Y, Li Y, Wang Z, Li P, et al. Identification of a conserved G-quadruplex within the E165R of African swine fever virus (ASFV) as a potential antiviral target. *J Biol Chem.* 2024;300(7):107453. <http://doi.org/10.1016/j.jbc.2024.107453>
11. Phan TG, Giannitti F, Rossow S, Marthaler D, Knutson T, Li L, et al. Detection of a novel circovirus PCV3 in pigs with cardiac and multi-systemic inflammation. *Virol J.* 2016;13:184. <https://doi.org/10.1186/s12985-016-0642-z>
12. Da Silva RR, da Silva DF, da Silva VH, de Castro AM. Porcine circovirus 3: a new challenge to explore. *Front Vet Sci.* 2024;10:1266499. <https://doi.org/10.3389/fvets.2023.1266499>
13. Rudova N, Lymanska O, Stegnyy B, Bolotin V, Solodiankin O, Gerilovych A. First detection of porcine circovirus type 3 in Ukraine. *Agricult Sci Pract.* 2021;8(2):16–23. <https://doi.org/10.15407/agrisp8.02.016>
14. Ha Z, Li J, Xie C, Yu C, Hao P, Zhang Y, et al. Prevalence, pathogenesis, and evolution of porcine circovirus type 3 in China from 2016 to 2019. *Vet Microbiol.* 2020;247:108756. <http://doi.org/10.1016/j.vetmic.2020.108756>
15. Chang C-C, Wu C-W, Chang Y-C, Wu C-Y, Chien M-S, Huang C. Detection and phylogenetic analysis of porcine circovirus type 3 in Taiwan. *Arch Virol.* 2021;166(1):259–63. <http://doi.org/10.1007/s00705-020-04870-6>
16. Turlewicz-Podbielska H, Augustyniak A, Pomorska-Mól M. Novel porcine circoviruses in view of lessons learned from porcine circovirus type 2-epidemiology and threat to pigs and other species. *Viruses.* 2022;14(2):261. <http://doi.org/10.3390/v14020261>
17. Assao VS, Santos MR, Pereira CER, Vannucci F, Silva-Junior A. Porcine circovirus 3 in North and South America: Epidemiology and genetic diversity. *Transbound Emerg Dis.* 2021;68(6):2949–56. <http://doi.org/10.1111/tbed.14238>
18. Zhai S-L, Lu S-S, Wei W-K, Lv D-H, Wen X-H, Zhai Q, et al. Reservoirs of porcine circoviruses: a mini review. *Front Vet Sci.* 2019;6:319. <https://doi.org/10.3389/fvets.2019.00319>
19. Gao Y-Y, Wang Q, Li H-W, Zhang S, Zhao J, Bao J, et al. Genomic composition and pathomechanisms of porcine circoviruses: A review. *Virulence.* 2024;15(1):2439524. <http://doi.org/10.1080/21505594.2024.2439524>
20. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature.* 2024;630(8016):493-500. <https://doi.org/10.1038/s41586-024-07487-w>
21. Nobel Prize in Chemistry. Press release 2024 [homepage on the Internet]. [9 October 2024]. Available from <https://www.nobelprize.org/prizes/chemistry/2024/press-release/>
22. Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S, Tamura K. MEGA12: molecular evolutionary genetics analysis version 12 for adaptive and green computing. *Mol Biol Evol.* 2024;41:1–9. <https://doi.org/10.1093/molbev/msae263>
23. Hall TA. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 1999;41:95–8.
24. Kosakovsky Pond SL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 2005;22(5):1208–22. <http://doi.org/10.1093/molbev/msi105>
25. Gupta S, Gupta D, Bhatnagar S. Analysis of SARS-CoV-2 genome evolutionary patterns. *Microbiol Spectr.* 2024;12(2):e0265423. <http://doi.org/10.1128/spectrum.02654-23>
26. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 2006;34(2):W676–W82. <http://doi.org/10.1093/nar/gkl253>
27. Menendez C, Frees S, Bagga PS. QGRS-H Predictor: a web server for predicting homologous quadruplex forming G-rich sequence motifs in nucleotide sequences. *Nucleic Acids Res.* 2012;40(W1):W96–W103. <http://doi.org/10.1093/nar/gks422>
28. Lombardi EP, Londono-Vallejo A. A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res.* 2020;48(1):1–15. <http://doi.org/10.1093/nar/gkz1097>

29. Powell HR, Islam SA, David A, Sternberg MJE. Phyre2.2: A community resource for template-based protein structure prediction. *J Mol Biol.* 2025;437(15):168960. <https://doi.org/10.1016/j.jmb.2025.168960>
30. Limanskaya OYu, Balak OK, Limanskii AP. Noncanonical structures in the genome of bovine foamy virus. *Cytology and Genetics.* 2025;59(5):503–15. <http://doi.org/10.3103/S009545272505007X>
31. Wang J, Huang H, Zhao K, Teng Y, Zhao L, Xu Z, et al. G-quadruplex in hepatitis B virus pregenomic RNA promotes its translation. *J Biol Chem.* 2023;299(9):105151. <http://doi.org/10.1016/j.jbc.2023.105151>
32. Molnár OR, Végh A, Somkuti J, Smeller L. Characterization of a G-quadruplex from hepatitis B virus and its stabilization by binding TMPyP4, BRACO19 and PhenDC3. *Sci Rep.* 2021;11:23243. <https://doi.org/10.1038/s41598-021-02689-y>
33. Monsen RC, DeLeeuw L, Dean WL, Gray RD, Sabo TM, Chakravarthy S, et al. The hTERT core promoter forms three parallel G-quadruplexes. *Nucleic Acids Res.* 2020;48(10):5720–34. <http://doi.org/10.1093/nar/gkaa107>
34. Zhang H, Endrizzi JA, Shu Y, Haque F, Sauter C, Shlyakhtenko LS, et al. Crystal structure of 3WJ core revealing divalent ion-promoted thermostability and assembly of the Phi29 hexameric motor pRNA. *RNA.* 2013;19(9):1226–37. <http://doi.org/10.1261/rna.037077.112>
35. Chu C-C, Plangger R, Kreutz C, Al-Hashimi HM. Dynamic ensemble of HIV-1 RRE stem IIB reveals non-native conformations that disrupt the Rev-binding site. *Nucleic Acids Res.* 2019;47(13):7105–17. <http://doi.org/10.1093/nar/gkz498>
36. Song Z, Gremminger T, Singh G, Cheng Y, Li J, Qiu L, et al. The three-way junction structure of the HIV-1 PBS-segment binds host enzyme important for viral infectivity. *Nucleic Acids Res.* 2021;49(10):5925–42. <https://doi.org/10.1093/nar/gkab342>
37. Stefos GC, Theodorou G, Politis I. Genomic landscape, polymorphism and possible LINE-associated delivery of G-quadruplex motifs in the bovine genes. *Genomics.* 2022;114(2):110272. <http://doi.org/10.1016/j.ygeno.2022.110272>
38. Frasson I, Nadai M, Richter SN. Conserved G-quadruplexes regulate the immediate early promoters of human Alphaherpesviruses. *Molecules.* 2019;24(13):2375. <http://doi.org/10.3390/molecules24132375>
39. Monsen RC, Chua EYD, Hopkins JB, Chaires JB, Trent JO. Structure of a 28.5 kDa duplex-embedded G-quadruplex system resolved to 7.4 Å resolution with cryo-EM. *Nucleic Acids Res.* 2023;51(4):1943–59. <https://doi.org/10.1093/nar/gkad014>
40. Wolfe AL, Singh K, Zhong Y, Drewe P, Rajasekhar VK, Sanghvi VR, et al. RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature.* 2014;513(7516): 65–70. <http://doi.org/10.1038/nature13485>
41. Wang SR, Zhang QY, Wang JQ, Ge X-Y, Song Y-Y, Wang Y-F, et al. Chemical targeting of a G-quadruplex RNA in the Ebola virus L gene. *Cell Chem Biol.* 2016;23(9):1113–22. <http://doi.org/10.1016/j.chembiol.2016.07.019>
42. Krauss IR, Spiridonova V, Pical A, Napolitano V, Sica F. Different duplex/quadruplex junctions determine the properties of anti-thrombin aptamers with mixed folding. *Nucleic Acids Res.* 2016;44(2):983–91. <http://doi.org/10.1093/nar/gkv1384>
43. Tsukakoshi K, Yamagishi Y, Kanazashi M, Nakama K, Oshikawa D, Savory N, et al. G-quadruplex-forming aptamer enhances the peroxidase activity of myoglobin against luminol. *Nucleic Acids Res.* 2021;49(11):6069–81. <https://doi.org/10.1093/nar/gkab388>
44. Palinski R, Piceyro P, Shang P, Yuan F, Guo R, Fang Y, et al. A novel porcine circovirus distantly related to known circoviruses is associated with porcine dermatitis and nephropathy syndrome and reproductive failure. *J Virol.* 2017;19(10): e01879-16. <https://doi.org/10.1128/jvi.01879-16>
45. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins.* 2004;57(4):702–10. <http://doi.org/10.1002/prot.20264>
46. Xu J, Zhang Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics.* 2010;26(7):889–95. <http://doi.org/10.1093/bioinformatics/btq066>
47. Cheung AK. Rolling-circle replication of an animal circovirus genome in a theta-replicating bacterial plasmid in *Escherichia coli*. *J Virol.* 2006;80(17):8686–94. <http://doi.org/10.1128/JVI.00655-06>
48. Ochoa S, Milam VT. Direct modeling of DNA and RNA aptamers with AlphaFold 3: A promising tool for predicting aptamer structures and aptamer–target interactions. *ACS Synth Biol.* 2025;14(8):3049–64. <https://doi.org/10.1021/acssynbio.5c00196>
49. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982;157(1):105–32. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
50. Abrusán G, Marsh JA. Alpha helices are more robust to mutations than beta strands. *PLoS Comput Biol.* 2016;12(12):e1005242. <https://doi.org/10.1371/journal.pcbi.1005242>
51. Zhang M, Liu CC, Huang Y, Hill JE, Araya MB, Ojkic D, et al. Phylogenetic analysis of porcine circovirus 3 circulating in Canadian pigs. *Vet Med Sci.* 2022 Sep;8(5):1969–74. <http://doi.org/10.1002/vms3.851>

### 3D MODELS OF PROTEINS AND CONSERVATIVE G-QUADRUPLEXES IN THE GENOME OF PORCINE CIRCOVIRUS

O. Yu. Limanskaya<sup>1,\*</sup>, O. V. Biloivan<sup>1</sup>, O. K. Balak<sup>2</sup>, A. P. Limanskii<sup>3</sup>

<sup>1</sup>National Scientific Center "Institute of Experimental and Clinical Veterinary Medicine",  
83 Hryhoriia Skovorody St., Kharkiv, 61023, Ukraine;

<sup>2</sup>Kharkiv National Medical University, 4 Nauka Ave., Kharkiv, 61022, Ukraine;

<sup>3</sup>Institute of Physiological Active Compounds, 58 Nauka Ave., Kharkiv, 61072, Ukraine

\*e-mail: [olgaliman@ukr.net](mailto:olgaliman@ukr.net)

Submitted December 29, 2025; Revised May 06, 2026;

Accepted May 7, 2026; Published June 25, 2026

**Background:** Non-canonical structures formed in nucleic acid molecules, which include, in particular, multihelical connections 3WJs (three-way junctions), quadruplexes (G4s), are regulatory elements that affect the functioning of the genome. Building 3D models of these structures and pathogen proteins is the first step in understanding their functions. AlphaFold Protein Structure Database was created by Google DeepMind together with EMBL and it contains more than 241 million protein structures. 3D models of porcine circovirus type 3 (PCV-3) proteins have not been identified in this database.

**The objective** was to identify and characterize potential G4s and 3WJs in the PCV-3 genome, to build 3D models of PCV-3 proteins.

**Materials and Methods:** 1138 PCV-3 isolates with complete genomes from GenBank were applied for phylogenetic analysis by MEGA12 software. The BioEdit software was used to construct entropy plots; the QGRS Mapper program was used to search for G4s motifs and determination of their G-scores; the BLAST software was used to search for 100% nucleotide sequence identity. 3D models of proteins and G4s were designed by AlphaFold 3 artificial intelligence.

**Results:** 4 conserved perfect G4s were identified in the PCV-3 genome, which are formed by three tetrads and confirmed by designing their 3D models. The determined G4s sequences are conservative structural motifs, since the number of PCV-3 isolates in GenBank with G4s in the genome is over 1000. 3D models of G4s the formation of one of which was experimentally determined in the genome of the hepatitis B virus were used as a control of the correctness of these models design. 3D models of the replicase and capsid protein were designed for the two evolutionarily most distant PCV-3 isolates, which were determined from phylogenetic trees based on the *cap* and *rep* genes, and the features of their structures were established. 3WJs were not identified in the PCV-3 genome.

**Conclusions:** 3D models of the replicase, capsid protein of porcine circovirus type 3 as well as four conservative G4s, which are formed by three G-tetrads, were obtained. Increasing the length of the fragments flanking G4s does not prevent their folding, which indicates the importance of these G-rich motifs in the PCV-3 life cycle. The existence of G4s with three tetrads in PCV-3 genome, in contrast to two-tetrad G4s in HBV, in the composition of not only aptamers, but also in molecules of increased length was confirmed by AlphaFold 3.

**KEY WORDS:** alternative structure; bioinformatics; AlphaFold 3; porcine circovirus type 3; G-quadruplex (G4); protein structure; 3D model.