

Структурные особенности нуклеопротеинов новых ортонайровирусов: Yezo virus, Songling virus и Beiji nairovirus

Гладышева А.*, Тюрин В., Цишевская А., Иматдинов И., Агафонов А.

Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора,
р.п. Кольцово, Россия

* gladysheva_av@vector.nsc.ru

Ключевые слова: РНК-содержащие вирусы; *Orthonairovirus*; вирусные белки; нуклеопротеин

Мотивация и цель: Вирусы являются причиной серьезных заболеваний во всем мире, особенно в развивающихся странах. В связи с ростом численности, плотности и мобильности населения увеличивается риск возникновения вспышек известных и неизвестных вирусных инфекций. В последние годы с использованием технологии NGS было выявлено множество новых патогенных для человека вирусов, переносимых клещами, в том числе ортонайровирусов: Songling virus (SGLV) [1], Yezo virus (YEZV) [2, 3] и Beiji nairovirus (BJNV) [4]. Имеющиеся ограниченные данные об этих вирусах указывают на то, что с высокой долей вероятности они могут быть возбудителями лихорадки у человека, что представляет новую угрозу для общественного здравоохранения. При этом информация о структурных особенностях вириона и белков новых ортонайровирусов, являющаяся необходимой для рационального создания эффективных вакцин и терапевтических противовирусных препаратов, на данный момент полностью отсутствует. Поэтому целью нашей работы стало изучение структурных особенностей нуклеопротеинов (NP) SGLV, YEZV и BJNV, которые играют важную роль в жизненном цикле ортонайровирусов и содержат иммунодоминантные регионы.

Методы и алгоритмы: Полноразмерные последовательности сегмента 1 SGLV, YEZV и BJNV были взяты из базы данных GenBank. Поиск близкородственных нуклеопротеинов, пространственные структуры которых уже были решены экспериментально, выполнен с помощью NCBI BLAST (National Center for Biotechnology Information Basic Local Alignment Search Tool) по базе данных PDB с использованием алгоритма blastp (protein-protein BLAST). Множественное выравнивание аминокислотных последовательностей выполнено в программе Unipro UGENE с использованием итеративного метода выравнивания MUSCLE. Функциональная аннотация вирусных белков произведена с помощью программы Protomenal и веб-сервиса HMMER. Поиск мотивов проводился по базам данных Pfam и NCBI CDD (Conserved Domains Database). Моделирование пространственных структур NP SGLV, NP YEZV и NP BJNV производили на сервере AlphaFold. Отбор пространственных моделей для дальнейшего теоретического анализа производился на основе коэффициента достоверности для каждой аминокислоты с учетом прогнозируемой локальной разности расстояний (pLDDT) AlphaFold 3. Визуализация пространственных моделей вирусных белков выполнена с помощью программы UCSF Chimera.

Синтетические кодон-оптимизированные ДНК копии генов, кодирующих NP SGLV, NP YEZV и NP BJNV, получали путем последовательного отжига частично

комплементарных олигонуклеотидов с последующей амплификацией методом ПЦР и клонированием в плазмиду pJET1.2/blunt. Для получения генетических конструкций, обеспечивающих синтез YEZV NP, SGLV NP и BJNV NP, производили амплификацию ДНК копий генов YEZV NP, SGLV NP и BJNV NP из pJET1.2/YEZV-NP, pJET1.2/SGLV-NP и pJET1.2/BJNV-NP с последующей обработкой ПЦР-продуктов эндонуклеазой рестрикции *Mall*I. Рекombинантные плазмиды, кодирующие YEZV NP, SGLV NP и BJNV NP, трансформировались в штамм KRX *Escherichia coli* и сверхэкспрессировались совместно со слитым белком убиквитин-подобного модификатора (SUMO) и сайтом протеолиза пикорнавируса A28 [5]. Очистку рекомбинантных белков осуществляли методом металл-хелатной аффинной хроматографии с использованием сорбента HisPur Ni-NTA Resin на хроматографической установке Bio-Lab 100. Анализ фракций рекомбинантных белков проводили методом SDS-PAGE. Концентрацию белка в препарате измеряли на флуориметре MAXLIFE Fluorimeter 2.0. Измерение динамического светорассеяния (DLS) от полученных биомолекул проводилось на анализаторе BeNano 180 Zeta Pro.

Результаты: Исследуемые белки NP SGLV, NP YEZV и NP BJNV не имеют известных аналогов с решенной пространственной структурой в PDB. Уровень идентичности нуклеопротеинов этих вирусов с белками с известными структурами составил около 30 %. С помощью AlphaFold 3 были получены модели структур NP SGLV, NP YEZV и NP BJNV с коэффициентами достоверности pLDDT NP SGLV = 86.45, pLDDT NP YEZV = 86.85, pLDDT NP BJNV = 81.30. Все модели являются высокоструктурированными и имеют два пространственно разделенных домена, в основном представленных альфа-спиралями и соединенных между собой свободными петлями, что указывает на их пригодность для кристаллизации и PCA. Модели белков NP SGLV, NP YEZV и NP BJNV демонстрируют высокий уровень топологического сходства между собой и ортонайровирусом ККГЛ, несмотря на отсутствие гомологов первичной структуры. Пространственная организация белков напоминает ракету, состоящую из домена плеча «стебля» (домен 1) и глобулярного домена «головы» (домен 2) (рис. 1). Распределение электростатического потенциала на поверхности нуклеопротеинов демонстрирует наличие РНК-связывающего сайта в районе центральной щели и в домене 1.

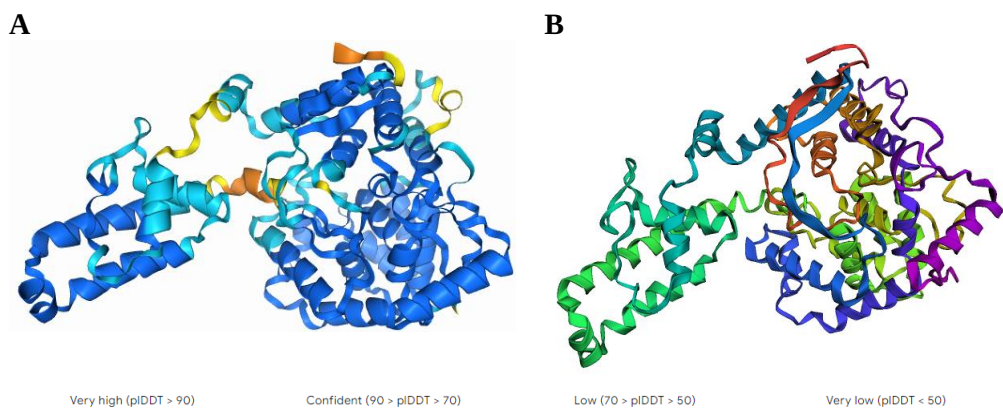


Рис. 1. Модели пространственной структуры мономера нуклеопротеина YEZV.

А – структура изображена в цветовом кодировании коэффициента pLDDT; В – структура в ленточном представлении

На втором этапе работы были получены *de novo* ДНК копии генов, кодирующих NP SGLV (1458 нуклеотидов), NP YEZV (1509 нуклеотидов) и NP BJNV (1659 нуклеотидов). Дизайн генетических конструкций производили таким образом, чтобы получить полноразмерные нуклеопротеины и отдельно их домены в растворимой форме. Из нуклеопротеинов были исключены неструктурированные со стороны С и N концов области. Далее после наработки биомассы и двух последовательных хроматографических очисток были получены рекомбинантные NP YEZV молекулярной массой $56.2 \cdot 10^3$ Да, домен 1 NP YEZV – $14.2 \cdot 10^3$ Да, NP BJNV – $61 \cdot 10^3$ Да и NP SGLV – $54.2 \cdot 10^3$ Да без аффинных меток и белков слияния с чистотой более 95 %, пригодные для дальнейших структурных исследований экспериментальными методами, в частности PCA. Методом DLS установлено, что размер биомолекулы рекомбинантного домена 1 NP YEZV варьирует от 2.26 до 3.84 нм, что соответствует ожидаемому. В настоящее время ведутся работы по подбору начальных условий кристаллизации NP SGLV, NP YEZV, домена 1 NP YEZV и NP BJNV.

Выводы: В ходе данного исследования на основе вычислительных моделей были выявлены структурные особенности нуклеопротеинов трех новых ортонайровирусов, а также разработаны подходы для экспериментального исследования структуры рекомбинантных белков малоизученных потенциально патогенных для человека вирусов. Дальнейшее исследование структуры белков новых вирусов даст представление о фундаментальных механизмах инфекций, вызванных данными вирусами, и позволит таргетно подходить к разработке эффективных вакцин и противовирусных препаратов для обеспечения национальной биологической безопасности.

Финансирование: Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации (соглашение от 12.10.2021 № 075-15-2021-1355) в рамках реализации отдельных мероприятий Федеральной научно-технической программы развития синхротронных и нейтронных исследований и исследовательской инфраструктуры на 2019–2027 гг.

Structural features of nucleoproteins of the novel orthonairoviruses: Yezo virus, Songling virus and Beiji nairovirus

Gladysheva A.*, Tuyrin V., Tsishevskaya A., Imatdinov I., Agafonov A.

State Research Center of Virology and Biotechnology “Vector” Rospotrebnadzor, Koltsovo, Russia

* gladysheva_av@vector.nsc.ru

Key words: RNA viruses; *Orthonairovirus*; viral proteins; nucleoprotein

Motivation and Aim: Viruses cause serious illness throughout the world, especially in developing countries. As populations grow in size, density, and mobility, the risk of outbreaks of known and unknown viral infections increases. In recent years, using NGS technology, many new tick-borne viruses pathogenic to humans have been identified, including orthonairoviruses: Songling virus (SGLV) [1], Yezo virus (YEZV) [2, 3] and Beiji nairovirus (BJNV) [4]. The limited data available on these viruses indicate that they are highly likely to cause fever in humans, representing an emerging public health

threat. At the same time, information about the structural features of the virion and proteins of new orthonairoviruses, which is necessary for the rational creation of effective vaccines and therapeutic antiviral drugs, is currently completely absent. Therefore, the goal of this work was to study the structural features of the nucleoproteins of SGLV, YEZV and BJNV, which play an important role in the life cycle of orthonairoviruses and contain immunodominant regions.

Methods and Algorithms: Full-length segment 1 sequences of SGLV, YEZV, and BJNV were obtained from the GenBank database. A search for closely related nucleoproteins whose spatial structures have already been solved experimentally was performed using NCBI BLAST (National Center for Biotechnology Information Basic Local Alignment Search Tool) against the PDB database using the blastp (protein–protein BLAST) algorithm. Multiple alignment of amino acid sequences was performed in the Unipro UGENE program using the MUSCLE iterative alignment method. This program was also used to search and map amino acid substitutions and analyze their location. Functional annotation of viral proteins was performed using Protomenal programs and HMMER web-service. The search for motifs was carried out using the Pfam and NCBI CDD (Conserved Domains Database) databases. Modeling of the NP SGLV, NP YEZV, and NP BJNV structures was carried out on the AlphaFold server. The selection of spatial models for further theoretical analysis was based on the confidence coefficient for each amino acid considering the predicted per-residue model confidence score (pLDDT) AlphaFold 3. Visualization of spatial structures of viral proteins was performed using the UCSF Chimera program.

Synthetic codon-optimized DNA copies of the genes encoding SGLV NP, YEZV NP, and BJNV NP were obtained by sequential annealing of partially complementary oligonucleotides, followed by PCR and cloning into the pJET1.2/blunt plasmid. DNA copies of the YEZV NP, SGLV NP and BJNV NP genes were amplified from pJET1.2/YEZV-NP, pJET1.2/SGLV-NP, and pJET1.2/ BJNV-NP, followed by treatment of PCR products with restriction endonuclease *Mall* in order to obtain genetic constructs providing the synthesis of YEZV NP, SGLV NP, and BJNV NP. Recombinant plasmids encoding YEZV NP, SGLV NP, and BJNV NP were transformed into *Escherichia coli* strain KRX and overexpressed together with a small ubiquitin-like modifier (SUMO) fusion protein and the picornaine 3C proteolysis site of rhinovirus A28 [5]. Purification of recombinant proteins was carried out by metal chelate affinity chromatography using HisPur Ni-NTA Resin sorbent on a Bio-Lab 100 chromatography unit. Analysis of recombinant protein fractions was carried out by SDS-PAGE. The protein concentration in the preparation was measured using a MAXLIFE Fluorimeter 2.0. Dynamic light scattering (DLS) measurements from the obtained biomolecules were carried out on a BeNano 180 Zeta Pro analyzer.

Results: The SGLV NP, YEZV NP, and BJNV NP proteins have no known analogues with a solved spatial structure in the PDB. The identity level of the nucleoproteins of these viruses with proteins with known structures in the PDB was about 30 %. Models of the SGLV NP, YEZV NP, and BJNV NP structures were obtained with confidence coefficients: pLDDT NP SGLV = 86.45, pLDDT NP YEZV = 86.85, pLDDT NP BJNV = 81.30 using AlphaFold 3. All models are highly structured and have two spatially separated domains, mainly represented by α -helices and interconnected by free loops, which indicates their suitability for crystallization and X-ray diffraction analysis. The SGLV NP, YEZV NP, and BJNV NP protein models demonstrate a high level of topological similarity between themselves and Crimean–Congo hemorrhagic fever virus,

despite poor primary sequence similarity. The spatial organization of proteins resembles a racket-shaped overall structure, consisting of a “stalk” arm domain (domain 1) and a globular “head” domain (domain 2) (Fig. 1). The distribution of electrostatic potential on the surface of SGLV NP, YEZV NP, and BJNV NP demonstrates the presence of an RNA-binding site in the central cleft region and in the domain 1.

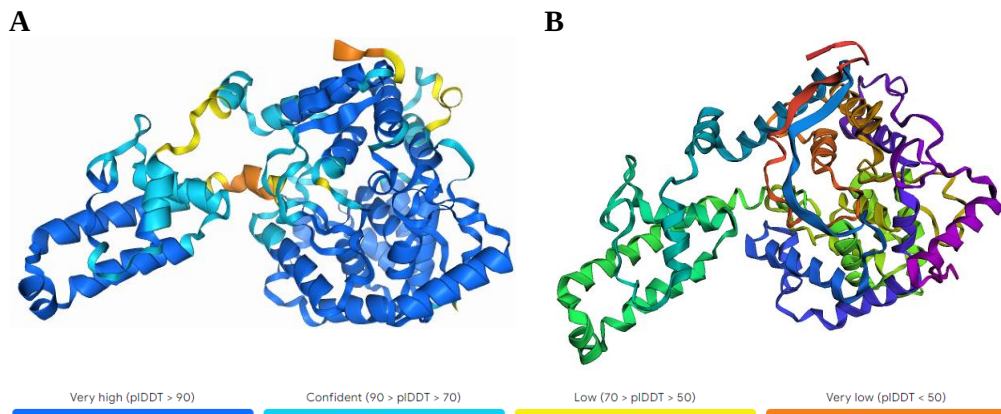


Fig. 1. Structure of monomer YEZV NP.

A – overall structure in pLDDT color representation; B – overall structure in cartoon representation

At the second stage of the work, *de novo* DNA copies of the genes encoding SGLV NP (1458 nucleotides), YEZV NP (1509 nucleotides), and BJNV NP (1659 nucleotides) were obtained. The design of genetic constructions was carried out in such a way as to obtain full-length nucleoproteins and separately their domains in soluble form. Unstructured regions at the C and N ends of SGLV NP, YEZV NP, and BJNV NP were excluded from the recombinant nucleoproteins. Then, recombinant YEZV NP with a molecular weight of $56.2 \cdot 10^3$ Da, domain 1 YEZV NP – $14.2 \cdot 10^3$ Da, BJNV NP – $61 \cdot 10^3$ Da and SGLV NP – $54.2 \cdot 10^3$ Da were obtained after production of biomass and two sequential chromatographic purifications. Recombinant SGLV NP, YEZV NP, and BJNV NP had not contain affinity tags and fusion proteins, had a purity of more than 95 %, and were suitable for further structural studies by experimental methods, in particular X-ray diffraction. Using the DLS method, it was found that the size of the biomolecule of the recombinant domain 1 YEZV NP varies from 2.26 to 3.84 nm, which corresponds to the expected value. Currently, research is underway to select the initial conditions of SGLV NP, YEZV NP, domain 1 YEZV NP, and BJNV NP crystallization. *Conclusion:* In the course of this study, based on computational models, the structural features of the nucleoproteins of three new orthonairoviruses were identified, and approaches were developed for the experimental study of the structure of recombinant proteins of poorly studied viruses that are potentially pathogenic for humans. Further study of the protein structure of new viruses will provide insight into the fundamental mechanisms of infections caused by these viruses and will allow for a targeted development of effective vaccines and antiviral drugs to ensure national biological safety.

Funding: The study was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-15-2021-1355 dated October 12, 2021) as a part of the implementation of certain activities of the Federal Scientific and Technical Program for the Development of Synchrotron and Neutron Research and Research Infrastructure for 2019–2027.

Список литературы/References

1. Ma J., Lv X.L. et al. Identification of a new orthonairovirus associated with human febrile illness in China. *Nat Med.* 2021;27(3):434-439. doi 10.1038/s41591-020-01228-y
2. Kodama F., Yamaguchi H. et al. A novel nairovirus associated with acute febrile illness in Hokkaido, Japan. *Nat Commun.* 2021;12(1):5539. doi 10.1038/s41467-021-25857-0
3. Lv X., Liu Z. et al. Yezo Virus Infection in Tick-Bitten Patient and Ticks, Northeastern China. *Emerg Infect Dis.* 2023;29(4):797-800. doi 10.3201/eid2904.220885
4. Wang Y.C., Wei Z. et al. A new nairo-like virus associated with human febrile illness in China. *Emerg Microbes Infect.* 2021;10(1):1200-1208. doi 10.1080/22221751.2021.1936197
5. Tishin A.E., Gladysheva A.V., Pyatavina L.A. et al. Preparation and Crystallization of Picornain 3C of Rhinovirus A28. *Crystallogr Rep.* 2023;68:924-930. doi 10.1134/S1063774523601119