

Функциональная аннотация генома нового вируса Хасеки (Haseki tick virus)

Осинкина И.¹, Радченко Н.^{1, 2}, Гладышева А.^{1*}

¹ Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора, р.п. Кольцово, Россия

² Новосибирский государственный университет, Новосибирск, Россия

* gladysheva_av@vector.nsc.ru

Ключевые слова: РНК-содержащие вирусы; *Flaviviridae*; вирусные белки

Мотивация и цель: Согласно данным Ежегодного отчета ВОЗ за 2023 г., количество заболеваний, вызванных инфекционными агентами, выросло почти на 2 % по сравнению с 2022 г., даже несмотря на окончание пандемии COVID-19 [1]. Изучение новых вирусов – одна из важнейших задач противодействия вирусным угрозам и обеспечения национальной биологической безопасности [2]. Однако в решении данной задачи не всегда возможно применение классических вирусологических и молекулярно-биологических подходов, и на помощь приходят методы математического моделирования, синтетической и структурной биологии. Вирус Хасеки (Haseki tick virus, HSTV) – новый вирус, впервые выявленный методом NGS в сыворотках крови людей после укуса клеща в 2020 г. на территории РФ [3]. Инфекция, вызванная HSTV, у человека характеризуется 3–5-дневной лихорадкой, затрудненным дыханием и болью в горле. Геном HSTV представлен оцРНК(+) длиной ~16 тыс. нуклеотидов и содержит одну открытую рамку считывания, фланкированную нетранслируемыми регионами. Вирусные белки синтезируются как полипротеин, посттрансляционно расщепляемый протеазами. Однако информация о том, какие вирусные белки кодирует геном HSTV и какова их пространственная организация, полностью отсутствует, что делает фактически невозможным дальнейшее предсказание и изучение биологических свойств данного вируса. Поэтому целью нашей работы стало: 1) функциональная аннотация генома вируса Хасеки; 2) получение *de novo* предсказанных вирусных белков для последующего рентгеноструктурного анализа и изучения структуры.

Методы и алгоритмы: Белковые домены HSTV были предсказаны с использованием инструмента поиска консервативных доменов NCBI, программы MOTIF в базе данных PFAM, программы InterProScan и веб-сервиса HMMER. Программа TMHMM v2.0c и веб-сервер PredictProtein.org использовались для локализации трансмембранных доменов полипротеина HSTV. Предполагаемые сайты протеолиза определяли с использованием SignalP и PROSPER. Вторичную структуру 5'–3' UTR РНК HSTV прогнозировали с использованием серверов MFOLD 3.4 и Vienna RNA Package. Моделирование проводили при температуре 37 °С и в ионных условиях 1 М NaCl без двухвалентных ионов. Параметры «максимальное расстояние между парными основаниями» (MDBPB), «процент субоптимальности» (%S), «верхняя граница количества вычисленных складок» и «верхняя граница общего количества одноцепочечных оснований» были установлены по умолчанию, а «начальная свободная энергия ΔG » – на поиск

минимального значения. Моделирование пространственных структур белков HSTV производили с использованием AlphaFold 3 на сервере AlphaFold. Для анализа белковых структур использовали программу CrystalExplorer.

Сборку ДНК копий генов, кодирующих аминокислотные последовательности белков HSTV, осуществляли методом сборочной ПЦР с использованием высокоточной ДНК полимеразы Q5 и олигонуклеотидов длиной ~ 60 оснований и последующим клонированием продуктов амплификации в плазмидный вектор pJET. Для создания экспрессирующих генетических конструкций использовали плазмидные вектора pET200 и pEASY. Экспрессию осуществляли в штаммах BL-21(DE3) и KRX клеток *E. coli*. Растворимую фракцию рекомбинантных белков HSTV получали путем гомогенизирования клеточной биомассы ультразвуком. Клеточные лизаты анализировали путем электрофоретического разделения белков в 15 % полиакриламидном геле в денатурирующих условия по методу Лэммли.

Результаты: Было предсказано, что геном HSTV содержит одну открытую рамку считывания (ORF), кодирующую полипротеин размером ~5100 аминокислот, который ко- и посттрансляционно расщепляется на структурные и неструктурные белки (рис. 1). ORF окружена короткой 5' UTR (~215 п.н.) и протяженной 3' UTR (~698 п.н.). Структура полипротеина HSTV схожа с полипротеином пестивирусов. Первым белком, кодируемым HSTV, является N-концевая протеаза (Npro) размером ~168 аминокислот. Npro предположительно является цистеиновой авто-протеазой, образующей N конец капсидного белка (C) HSTV. За белком C следуют два поверхностных белка E1 и E2, содержащие трансмембранные участки.

Было предсказано, что неструктурная часть генома HSTV содержит важные NS2-подобные (протеаза), NS3-подобные (протеаза и хеликаза) и RdRp-подобные (РНК-зависимая РНК-полимераза) домены. Размер NS2-подобной протеазы около 300 а.о. Это первый белок, закодированный в неструктурной части ORF и предположительно состоящий из двух структурных доменов. Предсказанные NS3-подобные и RdRp-подобные белки составляют примерно 570 и 486 а.о. в длину соответственно. Белок NS3 состоит из домена сериновой протеазы (~210 а.о.) и консервативного С-концевого домена геликазы (~260 а.о.). Множественные трансмембранные регионы расположены между структурными белками и в конце полипротеина. Предсказанная вторичная структура 5' UTR РНК HSTV состоит из большой стеблевой петли (SLA), которая включает боковую петлю, и Y-структуры, оканчивающейся инициатором трансляции AUG. Структура SLA высококонсервативна среди различных флавивирусов, и этой структуре предшествует 5'-m7GpppAmG кэп.

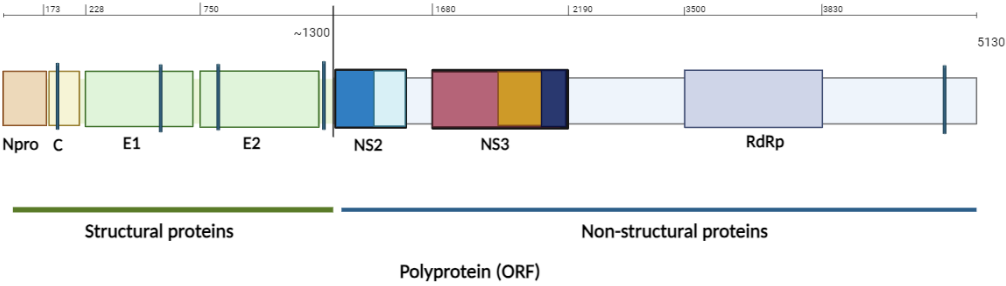


Рис. 1. Предполагаемая структура полипротеина вируса Хасеки

Модели пространственных структур белков HSTV в AlphaFold 3 имели низкий коэффициент достоверности (pLDDT Npro HSTV = 92; pLDDT SP2 HSTV = 47; pLDDT RdRp HSTV = 41) и, соответственно, не пригодны для дальнейшего анализа. Поэтому далее были получены синтетические ДНК копии генов, кодирующих предполагаемые белки Npro HSTV, надмембранную часть E2 (SP2) HSTV и RdRp HSTV, и экспрессирующие конструкции для наработки данных белков. Дизайн белков производили таким образом, чтобы получить максимальный выход белка в растворимой форме и увеличить вероятность их успешной кристаллизации. Для рекомбинантного белка Npro HSTV были созданы прокариотические штаммы-продуценты: BL-21/pET200/HSTV-Npro с оптимальными условиями наработки при культивировании в течение 16–20 часов, 20 °C и 180 об/мин с добавлением 1 mM ИПТГ и KRX/pET200/HSTV-SP1 – культивирование 20 часов при 16 °C и 210 об/мин с индукцией 1 mM ИПТГ и 0.1 % рапмузы.

Выводы: В данной работе мы впервые описали структуру генома нового вируса Хасеки, которая свидетельствует об эволюционном сходстве вируса Хасеки с вирусами рода *Pestiviruses*. Далее планируется кристаллизация и рентгеноструктурный анализ собранных *de novo* рекомбинантных белков Npro, SP2 и RdRp вируса Хасеки. Это позволит экспериментально подтвердить структуру белков Npro, SP2 и RdRp, установить их роль в реализации генетической информации HSTV и поможет разработать специальные тесты для обнаружения, профилактики и контроля предполагаемых инфекций HSTV у клещей и людей.

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.

Functional annotation of the novel Haseki tick virus genome

Osinkina I.¹, Radchenko N.^{1, 2}, Gladysheva A.^{1*}

¹ State Research Center of Virology and Biotechnology “Vector” Rospotrebnadzor, Koltsovo, Russia

² Novosibirsk State University, Novosibirsk, Russia

* gladysheva_av@vector.nsc.ru

Key words: RNA viruses; *Flaviviridae*; viral proteins

Motivation and Aim: According to the WHO Annual Report for 2023, the number of diseases caused by infectious agents increased by almost 2 % compared to 2022, even despite the end of the COVID-19 pandemic [1]. The study of new viruses is one of the most important tasks in countering viral threats and ensuring national biological safety [2]. However, in solving this problem it is not always possible to use classical virological and molecular biological approaches, and methods of mathematical modeling, synthetic and structural biology come to the rescue. Haseki tick virus (HSTV) is a new virus, first identified by NGS in human blood sera after a tick bite in 2020 in the Russian Federation [3]. HSTV infection in humans is characterized by 3–5 days of fever, difficulty breathing and sore throat. The HSTV genome is represented by ssRNA(+) with a length of

~16 kilobase and contains one open reading frame flanked by untranslated regions. Viral proteins are synthesized as a polyprotein that is posttranslationally cleaved by proteases. However, information about which viral proteins the HSTV genome encodes and what their spatial organization is completely missing. It is making further prediction and study of the HSTV biological properties is impossible. Therefore, the purpose of this work was: (1) functional annotation of the Haseki virus genome and (2) obtaining *de novo* predicted viral proteins for subsequent X-ray diffraction analysis and structure studies.

Methods and Algorithms: The protein domains of HSTV were predicted using the NCBI conserved domain search tool, MOTIF Search in the PFAM database, InterProScan tool and HMMER web service. TMHMM v2.0c tool and PredictProtein.org web server were used to localize transmembrane domains for polyproteins. Putative proteolysis sites were determined using SignalP and PROSPER. The secondary structure of the 5' UTR RNA HSTV was predicted using the MFOLD 3.4 server and Vienna RNA Package. Simulation was performed at a folding temperature of 37°C and under ionic conditions of 1 M NaCl without divalent ions. The parameters of maximal distance between paired bases (MDBPB), percent suboptimality (%S), the upper bound on the number of computed folds, and the upper bound on the total number of single-stranded bases that are allowed in a bulge or interior loop were set as default, and the initial free energy ΔG was set as the minimum value. The structures of HSTV proteins were modeled using AlphaFold 3 on the AlphaFold server. CrystalExplorer program was used to analyze protein structures.

Assembly of DNA copies of genes encoding HSTV proteins was carried out by overlap-PCR using high-precision Q5 DNA polymerase and oligonucleotides ~60 bases long, and subsequent cloning into the pJET plasmid. pET200 and pEASY vectors were used to create expression genetic constructions. Expression was carried out in strains BL-21(DE3) and KRX of *E. coli* cells. The soluble fraction of recombinant HSTV proteins was obtained by homogenizing cell biomass with ultrasound. Cell lysates were analyzed by electrophoretic separation of proteins in a 15 % polyacrylamide gel under denaturing conditions using the Laemmli method.

Results: The HSTV genome contains one open reading frame (ORF) encoding a polyprotein of ~5100 amino acids that is co- and post-translationally cleaved into structural and nonstructural proteins (Fig. 1). The ORF is flanked by a short 5' UTR (~215 bp) and a long 3' UTR (~698 bp). The HSTV polyprotein structure is similar to pestiviruses. The first protein encoded by HSTV is an N-terminal protease (Npro) of ~168 amino acids in length. HSTV Npro is a cysteine autoprotease that forms the N terminus of the HSTV capsid protein (C). Protein C is followed by two surface proteins E1 and E2 containing transmembrane regions. It was predicted that the HSTV genome contains important pestivirus NS2-like (protease), NS3-like (protease and helicase), and RdRp-like (RNA-dependent RNA polymerase) domains. The size of NS2-like protease is about 300 aa. It is the first nonstructural protein encoded in the ORF. NS2-like protease consists of two structural domains. The predicted NS3-like and RdRp-like proteins are approximately 570 aa and 486 aa in length respectively. The NS3-like protein derived into the serine protease domain (~210 aa) and the helicase conserved C-terminal domain (~260 aa). Multiple transmembrane domains located at the end of the polyprotein. The predicted secondary structure of 5' UTR RNA HTV consists of a large stem-loop (SLA) that includes a side loop, and Y-structure terminating in the translation initiator AUG. The SLA structure is highly conserved among different flaviviruses, and this structure is preceded by a 5'-m7GpppAmG cap.

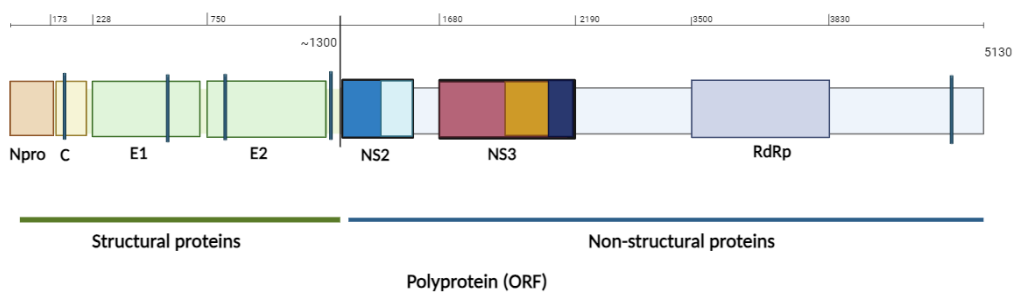


Fig. 1. Structure of the Haseki tick virus polyprotein

Models of HSTV proteins spatial structures in AlphaFold 3 had a low per-residue model confidence score (pLDDT Npro HSTV = 92; pLDDT SP2 HSTV = 47; pLDDT RdRp HSTV = 41). These protein models are not suitable for further analysis. Therefore, we obtained synthetic DNA copies of the genes encoding the putative Npro HSTV, extracellular part of E2 (SP2) HSTV and RdRp HSTV proteins, and expression constructs for the production of these proteins. Proteins were designed to maximize protein yield in soluble form and increase the likelihood of successful crystallization. Prokaryotic producer strains of recombinant Npro HSTV were created. Optimal production conditions for BL-21/pET200/HSTV-Npro are cultivation for 16-20 hours, 20 °C and 180 rpm with the addition of 1 mM IPTG and for KRX/pET200 /HSTV-SP1 are cultivation for 20 hours at 16 °C and 210 rpm with induction of 1 mM IPTG and 0.1 % rhamnose.

Conclusion: In this work, we described for the first time the genome structure of the new Haseki tick virus. This result indicates the evolutionary similarity of the Haseki tick virus with viruses of the *Pestiviruses* genus. Next, crystallization and X-ray diffraction analysis of the *de novo* assembled recombinant Npro, SP2, and RdRp Haseki tick virus proteins are planned. This will allow us to experimentally confirm the structure of the Npro, SP2 and RdRp proteins, establish their role in the implementation of HSTV genetic information and help to develop specific tests for the detection, prevention and control of putative HSTV infections in ticks and humans.

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. <https://www.who.int/data/gho/publications/world-health-statistics>
2. Ergunay K. et al. The expanding range of emerging tick-borne viruses in Eastern Europe and the Black Sea Region. *Sci Rep.* 2023;13(1):19824. doi 10.1038/s41598-023-46879-2
3. Kartashov M.Y., Gladysheva A.V., Shvalov A.N., Tupota N.L., Chernikova A.A., Ternovoi V.A., Loktev V.B. Novel Flavi-like virus in ixodid ticks and patients in Russia. *Ticks Tick Borne Dis.* 2023;14(2):102101. doi 10.1016/j.ttbdis.2022.102101