

Идентификация РНК-связывающего участка белка с доменом холодового шока CspA из *Mycobacterium tuberculosis* методами молекулярного докинга и молекулярной динамики

Панкратова П.*, Леконцева Н., Смольянова Н., Никулин А.

Институт белка РАН, Пушкино, Россия

* pankratova.p.y@gmail.com

Ключевые слова: белки холодового шока; CspA; малые регуляторные РНК; РНК-белковые комплексы; *Mycobacterium tuberculosis*

Мотивация и цель: Транс-кодируемые малые регуляторные РНК (мрРНК) являются посттранскрипционными регуляторами трансляции, что позволяет бактериями быстро адаптироваться к изменениям окружающей среды и воздействиям различных стрессов [1, 2].

Бактерия *Mycobacterium tuberculosis* является возбудителем туберкулеза человека и входит в десятку наиболее важных причин смертности во всем мире. С помощью биоинформатического анализа в геноме *M. tuberculosis* выявлено около 2000 участков ДНК, которые могут кодировать малые нетранслируемые РНК [3]. Однако возможное участие нетранслируемых РНК в регуляции экспрессии генов и развитии стадий инфекции *M. tuberculosis* исследованы слабо, и функционально охарактеризованы только 8 мрРНК.

Взаимодействию между мрРНК и их мишенями часто способствуют белковые РНК-шапероны. В *M. tuberculosis* не обнаружены классические бактериальные РНК-шапероны Hfq и ProQ. Претендентом на роль РНК-шаперонов в *M. tuberculosis* могут служить белки с доменом холодового шока Csp (Cold shock protein), которые выполняют широкий спектр биологических функций в клетках различных организмов [4–6].

В нашей работе была поставлена цель создать модель взаимодействия белка CspA из *M. tuberculosis* (далее MtbCspA) с мрРНК методами молекулярного докинга и молекулярной динамики.

Методы и алгоритмы: Для предсказания структуры белка MtbCspA использовали аминокислотную последовательность из базы UNIPROT. Поиск гомологов проводился при помощи метода BLAST. Пространственная структура белка MtbCspA промоделирована с помощью программы AlphaFold2. Полученная модель MtbCspA подвергалась стабилизации с помощью молекулярной динамики в GROMACS с силовым полем CHARMM в сочетании с явной моделью воды TIP3P в 0.1 М растворе NaCl при температуре 300 К. Длина траектории составила 300 нс.

Для формирования моделей комплексов MtbCspA с РНК были взяты как короткие последовательности одноцепочечной РНК, так и несколько шпилечных структур мрРНК. Вторичная структура исследуемых мрРНК была рассчитана программами RNAfold и MXfold2. Пространственная структура шпилечных РНК была

промоделирована в программе VfoldLA. Структуры РНК-белковых комплексов были получены макромолекулярным докинггом в программе HDOCK.

Результаты: Проанализирован ряд комплексов между *MtbCspA* и фрагментами РНК длиной от 12 до 20 рибонуклеотидов. Нами идентифицированы две области связывания исследуемых РНК. Первый сайт представляет собой область аминокислотных остатков Trp8, Phe28, Arg57-Gly58-Pro59 и соответствует общему мотиву для семейства белков холодового шока, ранее описанному для белков группы CSP из *Bacillus subtilis* и *Escherichia coli*. Второй участок, в области петель белка $\beta 3$ - $\beta 4$ и $\beta 4$ - $\beta 5$, ранее не был описан и идентифицирован нами впервые. Одноцепочечные РНК связывались с поверхностью белка *MtbCspA* как с первым участком, так и со вторым, причем были стабильны на протяжении траектории MD 300 нс. Шпильчатые РНК имеют преимущественное сродство ко второй области, однако, такие комплексы менее стабильны по сравнению с комплексами оцРНК-*MtbCspA*.

Выводы: На поверхности белка CspA из *M. tuberculosis* идентифицированы две предполагаемые области связывания одноцепочечных РНК и шпильчатых РНК, причем вторая область взаимодействия нами была идентифицирована впервые.

Финансирование: Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 24-24-00071.

Identification of the RNA-binding region of the cold shock protein CspA from *Mycobacterium tuberculosis* using molecular docking and molecular dynamics methods

Pankratova P.*, Lekontseva N., Smolyanova N., Nikulin A.

Institute of Protein Research, RAS, Pushchino, Russia

* pankratova.p.y@gmail.com

Key words: cold shock proteins; CspA, sRNA; RNA-protein complexes; *Mycobacterium tuberculosis*

Motivation and Aim: Transcribed small regulatory RNAs (sRNAs) are post-transcriptional translational regulators that allow bacteria to adapt rapidly to environmental changes and various stresses [1, 2]. *Mycobacterium tuberculosis* is the causative agent of human tuberculosis and ranks among the top ten leading causes of mortality worldwide. Through bioinformatics analysis, approximately 2000 DNA loci have been identified in the *M. tuberculosis* genome that have the potential to encode small non-coding RNAs [3]. However, the potential involvement of non-coding RNAs in regulating gene expression and the developmental *M. tuberculosis* infection has been poorly investigated, and only 8 sRNAs have been functionally characterized.

The interaction between sRNAs and their targets is often facilitated by proteins called RNA chaperones. Notably, the classical RNA chaperones Hfq and ProQ have not been identified in *M. tuberculosis*. Potential candidates for RNA chaperone in *M. tuberculosis* could be proteins containing the Cold Shock Protein (Csp) domain, which perform a wide range of biological functions in cells of various organisms [4–6]. Our work aimed to create a model of the interaction of the protein CspA from *M. tuberculosis* (hereinafter *MtbCspA*) with sRNA using molecular docking and molecular dynamics methods.

Methods and Algorithms: To predict the structure of the *MtbCspA* protein, the amino acid sequence from the UNIPROT database was used. The search for homologs was carried out using the BLAST method. The 3D structure of the *MtbCspA* protein was modeled using the AlphaFold2 program. The resulting *MtbCspA* model was subjected to molecular dynamics (MD) stabilization in GROMACS with the CHARMM force field coupled with an explicit TIP3P water model in 0.1 M NaCl solution at 300 K. The trajectory length was 300 ns.

To model the *MtbCspA* complexes with RNA, both short sequences of single-stranded RNA and several hairpin structures of sRNA were taken. The secondary structure of the studied sRNAs was calculated using the RNAfold and MXfold2 programs. The spatial structure of hairpin RNAs was modeled using the VfoldLA program. The structures of RNA-protein complexes were obtained by macromolecular docking using the HDock program.

Results: A number of complexes between *MtbCspA* and RNA fragments from 12 to 20 ribonucleotides in length were analyzed. We have identified two RNA-binding sites on the protein. The first site is a region of amino acid residues Trp8, Phe28, Arg57-Gly58-Pro59 and corresponds to a common motif for the cold shock protein family previously described for CSP proteins from *Bacillus subtilis* and *Escherichia coli*. The second site, including protein loops β 3- β 4 and β 4- β 5, has not been previously described and was identified by us for the first time. Single-stranded RNAs bound to the surface of the *MtbCspA* protein both in the first and the second RNA-binding sites, and were stable throughout the MD trajectory of 300 ns. Hairpin RNAs have a preferential affinity for the second RNA-binding site; however, such complexes are less stable compared to ssRNA-*MtbCspA* complexes.

Conclusions: We identified two putative RNA-binding sites on the surface of the CspA protein from *M. tuberculosis*, with the second RNA binding site being described for the first time.

Both *MtbCspA* RNA-binding sites had an affinity for both single-stranded RNA and hairpin RNA.

Funding: The study is supported by the RSF (24-24-00071).

Список литературы/References

1. Gong C.C., Klumpp S. Modeling sRNA-Regulated Plasmid Maintenance. *PLoS One*. 2017;12(1):e0169703. doi 10.1371/journal.pone.0169703
2. Jørgensen M.G., Pettersen J.S., Kallipolitis B.H. sRNA-mediated control in bacteria: An increasing diversity of regulatory mechanisms. *Biochim Biophys Acta Gene Regul Mech*. 2020;1863(5):194504. doi 10.1016/j.bbagr.2020.194504
3. Острик А.А., Ажикина Т.Л., Салина Е.Г. Малые некодирующие РНК и их роль в патогенезе *Mycobacterium tuberculosis*. *Успехи биологической химии*. 2021;61:229-252 [Ostrik A.A., Azhikina T.L., Salina E.G. Small non-coding RNAs and their role in the pathogenesis of *Mycobacterium tuberculosis*. *Advances in Biological Chemistry*. 2021;61:229-252 (in Russian)]
4. Corley M., Burns M.C., Yeo G.W. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Mol Cell*. 2020;78(1):9-29. doi 10.1016/j.molcel.2020.03.011
5. Roy A., Ray S. An in-silico study to understand the effect of lineage diversity on cold shock response: unveiling protein-RNA interactions among paralogous CSPs of *E. coli*. *3 Biotech*. 2023;13(7):236. doi 10.1007/s13205-023-03656-2
6. Zhang P., Wu W., Chen Q., Chen M. Non-Coding RNAs and their Integrated Networks. *J Integr Bioinform*. 2019;16(3):20190027. doi 10.1515/jib-2019-0027