

## Биоинформационный анализ структурно-функциональных свойств белков поверхностной зоны суставного хряща и белков сурфактанта

Герасимова Е.О.\*<sup>1</sup>, Третьякова А.В.<sup>1</sup>, Крылов П.А.<sup>1,2</sup>

<sup>1</sup>Волгоградский государственный университет

<sup>2</sup>ФНЦ Агроэкологии РАН

\*[lg6131602@gmail.com](mailto:lg6131602@gmail.com)

Остеоартроз – одно из распространенных болезней суставов, которое характеризуется болью в суставах и снижением подвижности. Лубрицин (PRG4) – это белок поверхностной зоны, основной функцией которого является способность к снижению уровня трения на соприкасающихся поверхностях суставного хряща. Функциональные способности у сурфактант-ассоциированных белков и лубрицина имеют схожие черты – сообщение с внешним пространством, наружное покрытие, образование защитных конструкций и смазка, а также противостояние воздействию давления. Исходя из вышесказанного, актуальность исследования заключается в сравнении структурно-функциональных свойств между лубрицином и белками сурфактанта с помощью биоинформационных методов, для выявления гомологии. Биоинформационный анализ аминокислотных последовательностей белка поверхностной зоны (лубрицин) и сурфактант-ассоциированных белков (гидрофильная фракция: SP-A, SP-D, гидрофобная фракция: SP-B, SP-C) был произведен с помощью ресурсов свободного доступа UniProt. В результате обнаружено, что некоторые функции исследуемых белков являются идентичными, так как большинство из них являются родственными. На основе полученных данных множественного выравнивания можно сделать вывод, что структурно аминокислотные последовательности довольно схожи между собой. Наиболее идентичны к белкам человека белки примата *Papio Anubis* (примерно 96 %), затем *Odobenus rosmarus divergens* (примерно 94 %) и *Rattus Norvegicus* (примерно 92 %).

*Ключевые слова:* биоинформационный анализ, множественное выравнивание, PRG4, SP-A, SP-B, SP-C, SP-D.

## Bioinformatic analysis of structural and functional properties of proteins of the surface zone and surfactant-associated proteins

Gerasimova E.O.<sup>1</sup>, Tretyakova A.V.<sup>1</sup>, Krylov P.A.<sup>1,2</sup>

<sup>1</sup>Volgograd State University

<sup>2</sup>Federal Research Center for Agroecology, Complex Amelioration, and Protective Afforestation, Russian Academy of Sciences

Osteoarthritis is one of the most common joint diseases, which is characterized by joint pain and decreased mobility. Lubricin (PRG4) is a surface zone protein whose main function is the ability to reduce the level of friction on the contact surfaces of the articular cartilage. The functional abilities of surfactant-associated proteins and lubricin have similar features - communication with the external space, outer coating, formation of protective structures and lubrication, as well as resistance to pressure. Based on the foregoing, the relevance of the study is to compare the structural and functional properties between lubricin and surfactant proteins using bioinformatic methods to identify homology. Bioinformatic analysis of the amino acid sequences of the surface zone protein (lubricin) and surfactant-associated proteins (hydrophilic fraction: SP-A, SP-D, hydrophobic fraction: SP-B, SP-C) was performed using UniProt's freely available resources. As a result, it was found that some functions of the studied proteins are identical, since most of them are related. Based on the obtained data of multiple alignment, it can be concluded that structurally the amino acid sequences are quite similar to each other. The primate *Papio Anubis* is the most identical to human proteins (about 96 %), followed by *Odobenus rosmarus divergens* (about 94%) and *Rattus Norvegicus* (about 92 %).

*Key words:* bioinformatic analysis, multiple alignment, PRG4, SP-A, SP-B, SP-C, SP-D.

## 1. Введение

Функциональные способности у сурфактант-ассоциированных белков и лубрицина с агреганом имеют схожие черты – сообщение с внешним пространством, наружное покрытие, образование защитных конструкций и смазка, а также противостояние воздействию давления.

Остеоартроз – самое распространенное заболевание суставов невоспалительного дегенеративного характера, вызывает боль в суставах и снижение подвижности [1], которое затрагивает все компоненты сустава и связано с локальными, нарастающими со временем, потерями внеклеточного матрикса хряща, основными компонентами которого являются коллаген типа II и протеогликаны (ПГ) [2]. В России ОА страдают примерно 15 млн. человек, в США ОА болеет более 20 млн. За последние годы заболеваемость ОА возросла на 35% [14].

Избыточная нагрузка приводит к понижению трибологических характеристик синовиальной жидкости и суставных поверхностей. Вследствие этого в первую очередь начинается разрушение поверхностной зоны и гибель локализованных в ней хондроцитов (стирание, апоптоз), а далее уже промежуточной зоны суставного хряща со всеми вытекающими из этого последствиями (истончение, разволокнение, кластеризация и т. д.). Одним из ключевых компонентов хрящевого матрикса, синтезируемого хондроцитами как в промежуточной, так и поверхностной зонах являются протеогликаны: агреган и лубрицин соответственно.

Сурфактант-ассоциированные белки и белки поверхностной зоны сходны по функциональным свойствам и частично по структуре (имеются похожие участки). Они в основном секретируются в лёгких. Сурфактант, который был впервые идентифицирован в 1920-х годах, играет ключевую роль в снижении поверхностного натяжения в альвеолах легких, помогает снизить работу дыхания и предотвращает ателектазы. Белки поверхностно-активного вещества способствуют функционированию и стабильности пленки поверхностно-активного вещества [3].

Исходя из вышесказанного, целью данного исследования является определение структурно-функциональных свойств и нахождение связи между белком поверхностной зоны (лубрицин) и сурфактант-ассоциированными белками (биоинформационный анализ).

Задачи:

1. Применить методы биоинформатики для поиска структурно-функциональных свойств.
2. Сравнить последовательности у белка поверхностной зоны (лубрицин) и сурфактант-ассоциированных белков (гидрофильная фракция: SP-A, SP-D, гидрофобная фракция: SP-B, SP-C) человека и других млекопитающих.

## 2. Материалы и методы

На рисунке 1 представлен дизайн исследования.

Для выполнения исследования был составлен дизайн исследования с помощью программы Figma(онлайн-сервис для разработки интерфейсов, была

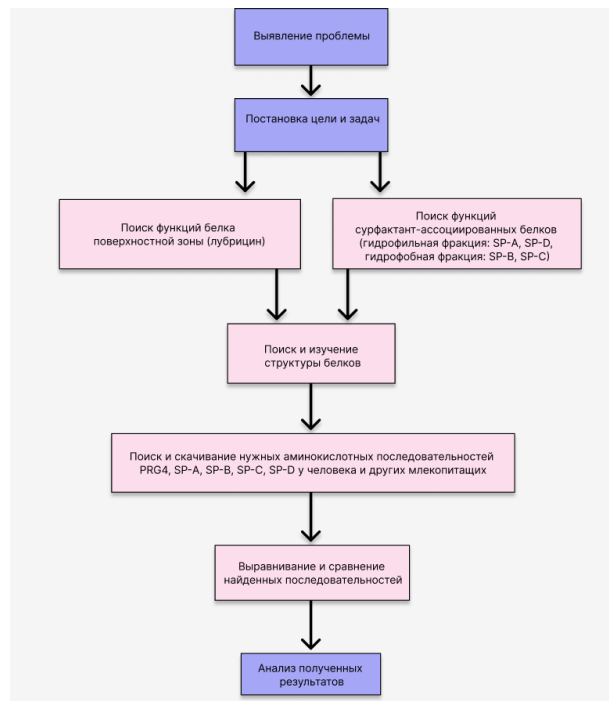


Рис. 1. Дизайн исследования.

создана в 2012 году [4]). Были определены следующие этапы выполнения: постановка цели и задач (указаны в введении), поиск функций с помощью PubMed (англоязычная текстовая база данных медицинских и биологических публикаций, созданная Национальным центром биотехнологической информации (NCBI) на основе раздела «биотехнология» Национальной медицинской библиотеки США (NLM) [5]) и NCBI (Национальный центр биотехнологической информации США [6]), поиск и изучение аминокислотных последовательностей. Для сравнения структур белков последовательности были взяты из базы данных UniProt (представляет собой набор последовательностей и аннотаций для более чем 120 миллионов белков во всех сферах жизни. Подробные аннотации, извлеченные из литературы опытными кураторами, были собраны для более чем полумиллиона этих белков [7]), а сравнивали в UGENE (многоплатформенное программное обеспечение с открытым исходным кодом, основной целью которого является оказание помощи в управлении, анализе и визуализации данных [8]).

### 3. Результаты исследования

#### 3.1. SP-A

Этот ген кодирует белок поверхностно-активного вещества легких, который является членом подсемейства лектинов С-типа, называемых коллектинами. Кодируемый белок связывает специфические углеводные фрагменты, обнаруженные на липидах и на поверхности микроорганизмов. Этот белок играет важную роль в гомеостазе сурфактанта и в защите от респираторных патогенов. Мутации в этом гене связаны с идиопатическим легочным фиброзом. Экспрессия SP-A происходит в основном в лёгких, так же немного в двенадцатиперстной кишке, желчном пузыре, печени, коже, яичниках и слюнной железе. Белок имеет следующие функции: обеспечивает связывание углеводов, белков, участвует в газообмене дыхательной системы. Имеет два домена: коллагеноподобный и лектин типа С.

#### 3.2. SP-B

Этот ген кодирует связанный с легкими сурфактантный белок В (SPB), амфипатический сурфактантный белок, необходимый для функции легких и гомеостаза после рождения. Сурфактант секретируется альвеолярными клетками легких и поддерживает стабильность легочной ткани за счет снижения поверхностного натяжения жидкостей, покрывающих легкие. SPB повышает скорость растекания и повышает стабильность монослоев поверхностно-активного вещества *in vitro*. Идентифицированы множественные мутации в этом гене, которые вызывают дисфункцию метаболизма сурфактанта в легких 1 типа, также называемую легочным альвеолярным протеинозом из-за дефицита протеина Белок экспрессируется так же в основном в лёгких и также в двенадцатиперстной кишке, пищеводе, печени, поджелудочной железе и мочевом пузыре. Имеет следующие функции: обеспечивает связывание белков, участвует в морфогенезе органом и газообмене дыхательной системы. У SP-B присутствуют четыре домена: сапозин А типа и сапозин В типа 1, 2 и 3. Сапозины представляют собой небольшие лизосомальные белки, которые служат активаторами различных лизосомальных ферментов, расщепляющих липиды [9].

#### 3.3. SP-C

Этот ген кодирует связанный с легкими сурфактантный белок С (SPC), чрезвычайно гидрофобный сурфактантный белок, необходимый для функции легких и гомеостаза после рождения. Были идентифицированы множественные мутации в этом гене, которые вызывают дисфункцию метаболизма сурфактанта в легких 2 типа, также называемую легочным

альвеолярным протеинозом из-за дефицита протеина С сурфактанта, и связаны с интерстициальным заболеванием легких у младенцев старшего возраста, детей и взрослых. Экспрессия SP-C так же происходит в основном в легких, но также и в эндометрии, печени, яичниках, поджелудочной железе и слюнной железе. Основные функции: обеспечивает связывание белков и участвует в газообмене дыхательной системы. Присутствует один домен (BRICHOS). белок сурфактанта С (SP-C), связан с респираторным дистресс-синдромом (RDS) [10].

#### 3.4. SP-D

Белок, кодируемый этим геном, является частью врожденного иммунного ответа, защищающего легкие от вдыхаемых микроорганизмов и химических веществ [11]. Кодируемый белок также может участвовать в метаболизме сурфактанта. SP-D так же больше всего экспрессируется в лёгких, но и присутствует в костном мозге, коже и щитовидной железе. Также секретируется эпителиальными клетками почечных канальцев [12]. Имеет следующие функции: обеспечение связывания белков и углеводов, вовлечен во врожденный иммунитет, участвует в защитной реакции на бактерии. SP-D также обладает способностью модулировать функцию лейкоцитов, а в некоторых случаях усиливать их уничтожение микроорганизмами [15]. У SP-D присутствуют два домена, такие же как и у SP-A, но у него коллагеноподобный домен находится на отрезке от 46 до 222 нуклеотида, лектин типа С – от 260 до 375 домена.

#### 3.5. PRG4

Белок, кодируемый этим геном, представляет собой крупный протеогликан, который синтезируется хондроцитами, расположенными на поверхности суставного хряща, и некоторыми клетками синовиальной оболочки. Этот белок содержит как хондроитинсульфат, так и кератансульфатгликозаминогликаны. Он действует как граничная смазка на поверхности хряща и способствует эластичному поглощению и диссипации энергии синовиальной жидкости. Наибольшая экспрессия протеогликана происходит в основном в печени, жире и легких. Имеет следующие функции: обеспечивает устойчивость к сжатию, активирует макрофаги, участвует в иммунном ответе. У протеогликана 4 присутствует два домена: SMB 1 и 2.

#### 3.6. Сравнение структур белков

Для сравнения были взяты четыре организма: *Homo Sapiens* (референсная последовательность), *Papio Anubis*, *Rattus Norvegicus* и *Odobenus*

*rosmarus divergens* (Морж). Найденные последовательности добавляем в UGENE, затем выравниваем с помощью ClustalW (классический метод множественного выравнивания [13])

(матрица BLOSUM62). Результаты сравнения представлены в Таблице 1.

**Таблица 1.** Результаты выравнивания аминокислотных последовательностей PRG4 и сурфактант-ассоциированных белков

	PRG4	SP-A	SP-B	SP-C	SP-D	SP-C/SP-B	SP-B/PRG4	SP-C\PRG4	SP-A\PRG4	SP-D\PRG4	SP-A\SP-D
<i>Papio anubis</i>	88 %	98 %	98 %	99 %	99 %	13 %	13 %	17 %	10 %	9 %	38 %
<i>Rattus Norvegicus</i>	76 %	97 %	95 %	98 %	96 %	14 %	14 %	18 %	10 %	7 %	34 %
<i>Odobenus rosmarus divergens</i>	82 %	97 %	96 %	98 %	97 %	13 %	13 %	18 %	9 %	9 %	36 %
<i>Homo sapiens</i>	–	–	–	–	–	13 %	14 %	18 %	10 %	7%	38 %

Примечание\* Последовательность *Homo Sapiens* является референсной и в сравнении белков между организмами не указан, только в сравнении между белками.

По результатам можно сделать вывод, что сурфактант-ассоциированные белки более схожи у млекопитающих, чем белки поверхностной зоны. При сравнении всех белков с человеком наиболее приближенным к человеку является примат *Papio Anubis*, затем *Odobenus rosmarus divergens* и *Rattus Norvegicus*.

При сравнении белков между собой большая разница в схожести структуры белков обусловлена в различной длине аминокислотной последовательности данных белков. У лубрицина она составляет 1404 аминокислотных остатка, у сурфактант-ассоциированного белка SP-A 248 аминокислотный остаток, у сурфактант-ассоциированного белка SP-D 375, у сурфактант-ассоциированного белка SP-B 381 аминокислотный остаток, у сурфактант-ассоциированного белка SP-C 197 аминокислотных остатков, из-за чего присутствует большая погрешность. SP-A и SP-D более схожи, чем SP-C и SP-B, ведь у них одинаковые домены и функции схожи.

#### 4. Заключение

С помощью упомянутых выше баз данных были изучены структурные и функциональные свойства белков поверхностной зоны (лубрицин) и сурфактант-ассоциированных белков (гидрофильная фракция: SP-A, SP-D, гидрофобная фракция: SP-B, SP-C). Некоторые функции исследуемых белков являются идентичными. Можно сделать вывод, на основе полученных результатов множественного выравнивания, что структурно последовательности довольно схожи между собой. Возможно, применение сурфактант-ассоциированных белков в качестве добавки с существующими хондропротекторами смягчит воздействие негативных факторов на

поверхностную зону хрящей за счет улучшения смазочных функций как матрикса, так и синовиальной жидкости, повлияет на синтетическую деятельность клеток поверхностной зоны, что, возможно, увеличит экспрессию SZP (или же подавит - заместит). Это может привести к восстановлению структуры хрящевого матрикса и обеспечить нормальное функционирование суставного хряща.

#### 5. Благодарность

Работа была выполнена при поддержке гранта Президента РФ для молодых ученых-кандидатов наук МК-199.2022.1.4.

#### 6. Список литературы

1. Fan X., Wu X., Crawford R., Xiao Y., Prasadam I. Macro, Micro, and Molecular. Changes of the Osteochondral Interface in Osteoarthritis Development. *Front. Cell Dev. Biol.* 2021. V. 9. P. 659654.
2. Wu Y., Wu T., Xu B., Xu X., Chen H., Li X. Oxytocin prevents cartilage matrix destruction via regulating matrix metalloproteinases. *Biochem. Biophys. Res. Commun.* 2017. V. 486. No. 3. P. 601–606.
3. Singh J., Jaffe A., Schultz A., Selvadurai H. Surfactant protein disorders in childhood interstitial lung disease. *Eur. J. Pediatr.* 2021. V. 180. No. 9. P. 2711–2721. doi: [10.1007/s00431-021-04066-3](https://doi.org/10.1007/s00431-021-04066-3)
4. Хвостенко Т.М., Велисар Д.С. Фигма – перспективный инструмент современного веб-дизайнера. *Вестник образовательного консорциума Среднерусский университет. Информационные технологии.* 2019. Т. 2. № 14. С. 7–10.

5. National Center for Biotechnology Information. URL: [PubMed Help. PMID \[PMID\]](#) (accessed 21.09.2022).
6. Global Research Identifier Database URL: <https://www.grid.ac/institutes/grid.419234.9> (accessed 21.09.2022).
7. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019. V. 47. No. D1. P. D506–D515. doi: [10.1093/nar/gky1049](https://doi.org/10.1093/nar/gky1049)
8. Okonechnikov K., Golosova O., Fursov M. The UGENE team, Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* 2012. V. 28. No. 8. P. 1166–1167. doi: [10.1093/bioinformatics/bts091](https://doi.org/10.1093/bioinformatics/bts091)
9. Sun Y., Grabowski G.A. Altered autophagy in the mice with a deficiency of saposin A and saposin B. *Autophagy.* 2013. V. 9. No. 7. P. 1115. doi: [10.4161/auto.24919](https://doi.org/10.4161/auto.24919)
10. Willander H., Hermansson E., Johansson J., Presto J. BRICHOS domain associated with lung fibrosis, dementia and cancer--a chaperone that prevents amyloid fibril formation? *FEBS J.* 2011. V. 278. No. 20. P. 3893–3904. doi: [10.1111/j.1742-4658.2011.08209.x](https://doi.org/10.1111/j.1742-4658.2011.08209.x)
11. Ortega F.J., Pueyo N., Moreno-Navarrete J.M., Sabater M., Rodriguez-Hermosa J.I., Ricart W., Tinahones F.J., Fernández-Real J.M. The lung innate immune gene surfactant protein-D is expressed in adipose tissue and linked to obesity status. *Int. J. Obes. (Lond).* 2013. V. 37. No. 12. P. 1532. doi: [10.1038/ijo.2013.23](https://doi.org/10.1038/ijo.2013.23)
12. Liu J., Li G., Li L., Liu Z., Zhou Q., Wang G., Chen D. Surfactant protein-D (SP-D) gene polymorphisms and serum level as predictors of susceptibility and prognosis of acute kidney injury in the Chinese population. *BMC Nephrol.* 2017. V. 18. P. 67. doi: [10.1186/s12882-017-0485-x](https://doi.org/10.1186/s12882-017-0485-x)
13. Мавропуло-Столяренко Г., Тулуб А., Стефанов В. *Биоинформатика. Учебник для академического бакалавриата*. Litres, 2021. 253 с.
14. Алексеева Л.И. Препараты замедленного действия в лечении остеоартроза. *Русский Медицинский Журнал.* 2012. Т. 16. № 7. С. 389–391.
15. Crouch E.C. Structure, biologic properties, and expression of surfactant protein D (SP-D). *Biochim. Biophys. Acta.* 1998. V. 1408. No. 2–3. P. 278–289. doi: [10.1016/s0925-4439\(98\)00073-8](https://doi.org/10.1016/s0925-4439(98)00073-8)