

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра биохимии и биофизики

**АНАЛИЗ ГЕНОВ, ОТВЕТСТВЕННЫХ ЗА СИНТЕЗ О-АНТИГЕНОВ
БАКТЕРИЙ РОДА *AZOSPIRILLUM***

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 421 группы

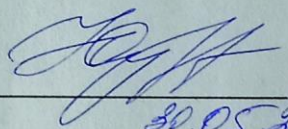
Направления подготовки бакалавриата 06.03.01

Биология Биологического факультета

Билкасумова Михаила Юрьевича

Научный руководитель:

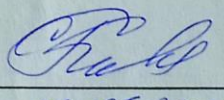
доцент, канд. биол. наук


30.05.2025

Ю.П. Федоненко

Заведующий кафедрой:

профессор, док. биол. наук


30.05.2025

С.А. Коннова

Саратов 2025

Введение: бактерии синтезируют широкий спектр полисахаридов, включая редкие сахара, обладающие разнообразной структурой и биологической функцией. Некоторые бактериальные гликополимеры, включая липополисахариды и тейхоевые кислоты, имеют промышленное и медицинское значение, а их структура используется в хемотаксономии [1–4]. Бактериальные полисахариды часто обладают антигенностью и применяются в вакцинологии [5]. Они состоят из повторяющихся олигосахаридных звеньев, обеспечивающих специфичность. Наряду с изучением их структуры, важной задачей остаётся анализ ферментов, участвующих в их биосинтезе, что позволяет не только предсказывать структуру полисахаридов, но и выявлять филогенетические связи между бактериями. Цель работы — провести биоинформатический анализ генов, вовлечённых в биосинтез О-антигенов у почвенных азотфиксирующих бактерий рода *Azospirillum*. Задачи исследования: Найти в базах данных гены, связанные с биосинтезом моносахаридов, гликозилтрансфераз и ферментов процессинга полисахаридов. Провести множественное выравнивание и филогенетический анализ нуклеотидных и аминокислотных последовательностей для выявления консервативных участков и оценки родства генов. Предсказать домены и возможные функции белков, кодируемых предполагаемыми гликозилтрансферазами.

Бакалаврская работа включает содержание, список сокращений, введение, 3 главы (обзор литературы, материалы и методы, результаты исследований и их обсуждение), заключение, выводы и список использованных источников, включающий 64 источника на русском и английском языках. Работа изложена на 58 страницах машинописного текста. Работа проиллюстрирована 17 рисунками, 5 листингами и 1 таблицей.

Основное содержание работы. Учитывая распространенность линейных олигосахаридных L-рамнанных мотивов в полисахаридах поверхности азоспирилл, а также лимитированность данных о ферментах,

задействованных в полимеризации полисахаридов и их процессинге, мы проанализировали представленные в доступных базах данных геномы бактерий рода *Azospirillum* для поиска предполагаемых L-рамнозилтрансфераз. Так как первичный анализ не дал достоверных результатов, мы сузили анализируемые данные, включив в анализ геномы *Azospirillum* spp. с установленными структурами ОПС и достаточно высоким качеством сборки геномов. Еще одним критерием поиска была кластеризация гена гликозилтрансферазы с генами биосинтеза L-рамнозы. Ген, удовлетворяющий заданным критериям поиска, был локализован в геноме типового штамма *A. agricola* CC-НІН038 (6327...7286;), структура основной цепи повторяющегося звена которого представлена линейным олигомером из трех 3-замещенных остатков L-рамнозы.

Биоинформатический анализ WP_209782629.1 не демонстрировал существенной гомологии с охарактеризованными генами рамнозилтрансфераз для бета- и гамма-протеобактерий (*Pseudomonas* spp., *Burkholderia* spp., *Neisseria* spp., *Bordetella* spp., *Acinetobacter baumannii*). Возможно, отсутствие выраженной гомологии является следствием вариабельности структур данного фермента у представителей филогенетически удаленных групп бактерий. Следует отметить, что один из используемых для классификации большого семейства гликозилтрансфераз критериев – гомология последовательностей – лежит в основе выделения 135 различных семейств этих ферментов (по состоянию на 15 апреля 2024 г.) в базе данных Carbohydrate Active enZymes [6]. Найденный нами ген предполагаемой гликозилтрансферазы относится по классификации CAZy к классу GT-2, куда отнесены рамнозилтрансферазы. Лимитированность данных о структурных и биохимических особенностях этих ферментов свидетельствует об актуальности их характеристики, в том числе с привлечением методов биоинформатики.

Экспериментальное подтверждение данных о наличии и последовательности гена GT (WP_209782629.1) в типовом штамме бактерий *A. agricola* CC-HIN038 было осуществлено методом ПЦР с последующим разделением продуктов амплификации в агарозном геле, экстракцией соответствующего продукта и секвенированием его последовательности.

На следующем этапе исследования был проведен комплексный поиск в геномах бактерий рода *Azospirillum* с целью выявления гомологичных WP_209782629.1 последовательностей и их филогенетического родства. Для реализации этих задач использовались инструменты биоинформатики, включая библиотеку Biopython и методы BLAST-поиска. Импорт необходимых библиотек, включая Biopython (модули Bio.Blast, Bio.Seq, Bio.Align и Bio.Phylo), а также библиотеки для визуализации данных (Matplotlib, Seaborn и NumPy), позволил автоматизировать процесс анализа последовательностей и построения графиков, упрощая дальнейшую обработку данных.

Осуществлённый поиск гомологичных последовательностей с использованием программы blastp с запросом в формате XML позволил в дальнейшем обработать результаты с помощью Biopython (листинг 3.1). Полученные результаты BLAST-сопоставления были сохранены в файл для дальнейшего анализа.

Следующим этапом стало извлечение и обработка полученных BLAST-результатов. Использовался модуль NCBIXML для парсинга XML-файла и извлечения релевантных данных, включая идентификаторы гомологичных последовательностей, уровень идентичности и степень покрытия выравнивания. Эти данные позволили определить степень родства между исследуемыми последовательностями и выявить наиболее схожие белки у других бактерий.

Листинг 3.1 – Код для выполнения BLAST-запроса (Python)

```
# ID белка
```

```

protein_id = "WP_209782629.1"
# Выполняем BLAST-запрос на сервер NCBI
print("Запуск BLAST-запроса...")
result_handle = NCBIWWW.qblast(
    program="blastp",
    database="nr",
    sequence=protein_id,
    format_type="XML"
)
# Сохраняем результат
with open("blast_results.xml", "w") as out_handle:
    out_handle.write(result_handle.read())
print("Результаты BLAST сохранены в 'blast_results.xml'.")

```

После сбора данных было выполнено сохранение найденных последовательностей в формате FASTA, что упростило их дальнейшее использование при анализе. FASTA-файл содержал выровненные последовательности, которые затем были обработаны инструментами множественного выравнивания. Для выравнивания последовательностей применялся модуль Align из Biopython, который позволил выполнить множественное выравнивание найденных гомологов (листинги 3.3).

Листинг 3.3 – Код для выравнивания в MUSCLE (Python)

```

muscle = r"muscle.exe" # Путь к MUSCLE
cmd = f"muscle -align {input_file} -output {output_file}"
results = subprocess.run(cmd, shell=True,
stdout=subprocess.PIPE, text=True)

```

Это дало возможность оценить консервативность последовательностей и определить ключевые различия между ними. Для проверки качества выравнивания использовались методы визуализации, в частности, тепловые карты, построенные с использованием библиотеки Seaborn .

Использование визуализации помогло оценить степень различий между последовательностями и выделить наиболее консервативные регионы.

Следующим важным этапом исследования стало построение филогенетического дерева.

Для этого использовались методы расчёта эволюционных расстояний (DistanceCalculator) и алгоритмы конструирования деревьев (DistanceTreeConstructor) из Biopython (листинг 3.4). Построенное филогенетическое дерево позволяло проследить эволюционные связи между исследуемыми последовательностями и определить их относительное положение среди других бактерий (рисунок 5). Визуализация дерева осуществлялась с помощью модуля Phylo, что позволило представить результаты в удобочитаемой форме.

Для получения пространственной структуры гликозилтрансферазы штамма *A. agricola* CC-HIN038 (белок WP_209782629.1) была построена 3D-модель с помощью алгоритма AlphaFold3. Этот метод использует нейросетевые предсказания, основанные на экспериментальных структурах, и позволяет моделировать белки с высокой точностью без необходимости лабораторного анализа. Полученная модель продемонстрировала высокое качество: средний показатель pLDDT составил более 88, что свидетельствует о надёжности предсказания, особенно в области каталитического домена. Также был рассчитан индекс RAe, подтвердивший низкую ошибку выравнивания ключевых участков.

Полученная модель была сопоставлена с экспериментальными структурами гомологичных белков из базы PDB, что подтвердило корректность укладки и наличие консервативных структурных элементов, характерных для гликозилтрансфераз. Высокое качество модели также позволило оценить потенциальное влияние аминокислотных замен на стабильность и активность фермента. Таким образом, использование AlphaFold3 дало надёжную 3D-модель рамнозилтрансферазы *A. agricola*, что в сочетании с результатами BLAST и аннотациями UniProt обеспечивает основу для дальнейшего функционального и филогенетического анализа.

Для уточнения функции предсказанной гликозилтрансферазы из *A. agricola* CC-HIH038T проведён BLASTp-анализ аминокислотной последовательности WP_209782629.1 (319 аминокислот) против базы UniProtKB. Обнаружен широкий спектр гомологов, преимущественно из семейства Acetobacteraceae, а также из родственных и отдалённых таксонов. Идентифицировано более 150 гомологов, большинство аннотированы как гликозилтрансферазы семейства 2 или специфические рамнозилтрансферазы, переносящие рамнозу на различные акцепторы.

Анализ E-value показал высокое сходство исследуемой гликозилтрансферазы с белками рода *Azospirillum* (например, *A. brasilense*, *A. lipoferum*) и близкородственным *Roseicella aerolata* — значения варьировали от $6,1 \times 10^{-78}$ до $5,2 \times 10^{-41}$, что подтверждает функциональную и эволюционную консервативность фермента в семействе Azospirillaceae. Белки из семейства Rhodospirillaceae (*Rhodospirillum rubrum*, *Magnetospirillum gryphiswaldense*) имели менее тесное родство (E-value 10^{-10} – 10^{-6}), но сохранили ключевые каталитические мотивы DxD и HxH, важные для связывания и координации молекул в активном центре. Большинство белков аннотированы как гликозилтрансферазы GT2, участвующие в биосинтезе полисахаридов клеточной стенки, липополисахаридов, гликозилировании белков и липидов, а также синтезе экзополисахаридов, важных для колонизации растений и формирования биоплёнок. Около 75 % содержат дополнительные С-концевые домены, которые, вероятно, регулируют специфичность субстрата, взаимодействия с белками или мембранную локализацию, что указывает на участие ферментов в мультибелковых комплексах экстраклеточных метаболитов. В нижней части списка BLAST выявлены более удалённые гомологи (E-value 10^{-3} – 10^{-1}) из филогенетически отдалённых таксонов, включая грамположительные бактерии и дрожжи. Несмотря на низкое сходство, они сохраняют отдельные структурные элементы GT, что может отражать

конвергентную эволюцию каталитических механизмов, но не влияет существенно на функциональную интерпретацию изучаемого фермента. Таким образом, BLAST-анализ подтверждает высокую эволюционную консервативность рамнозилтрансферазы *A. agricola* и надёжность модели, построенной с помощью AlphaFold3. Наличие большого числа близких гомологов с низкими E-value и ключевых каталитических мотивов GT2 указывает на сохранённую каталитическую функцию белка.

Для идентификации структурных гомологов предсказанной модели гликозилтрансферазы *Azospirillum agricola* проведена аннотация с помощью Dali Server [7], основанная на пространственном расположении α -углеродов в структуре, полученной AlphaFold3. Выявлены белки с высоким пространственным сходством (Z-скор >16), в основном гликозилтрансферазы и ферменты биосинтеза полисахаридов, что подтверждает функцию исследуемого белка как рамнозилтрансферазы. Несовпадение между гомологами, найденными по последовательности и по структуре, объясняется рядом факторов.

№	PDB	Цепь	Z-скор	RM SD	Длина выравнивания (lali)	Размер белка (nres)	% идентичности	Описание
1	<i>Streptomyces</i> sp. AVP053U2	A	21.3	2.9	229	306	19%	GALNAC(5)-DINACBAC-PP-UNDECAPRENOL BETA-1,3-GLUCOSYLTRANSFERASE

2	<i>Streptococcus parasanguinis</i> FW213	A	20.8	2.7	215	276	16%	PUTATIVE GLYCOSYLTRANSFERASE (GALT1)
3	<i>Listeria monocytogenes</i>	C	19.7	3.8	242	608	14%	GLYCOSYLTRANSFERASE
4	<i>Bacillus subtilis</i>	A	19.6	3.1	218	240	11%	SPORE COAT POLYSACCHARIDE BIOSYNTHESIS PROTEIN

Таблица 1 — Гомологи предполагаемой гликозилтрансферазы бактерии *A. agricola* CC-HIH038 по данным структурного анализа в Dali Server [7]

BLAST и аналогичные алгоритмы выявляют гомологию по сходству первичной аминокислотной последовательности, эффективно при идентичности >25–30%, но теряют чувствительность при удалённом родстве. В отличие от них, DALI и другие структурные методы анализируют пространственную конфигурацию белка, что позволяет обнаруживать структурные аналоги даже при низкой последовательной идентичности. Пространственная структура белка эволюционно более консервативна, чем последовательность. Белки с разным последовательным сходством могут сохранять общий фолдинг, обеспечивающий схожие функции. Особенно это важно для ферментов, таких как рамнозилтрансферазы, где каталитическое ядро остаётся неизменным при изменениях окружающих участков. Некоторые белки могут независимо приобретать схожую структуру и

функцию — конвергентная эволюция. В таких случаях DALI выявляет структурное сходство, хотя прямого происхождения может не быть. Модели, полученные с помощью AlphaFold3, точнее отражают консервативные регионы, тогда как BLAST анализирует только последовательности из баз данных. Кроме того, базы PDB и UniProt содержат разные наборы белков, что влияет на результаты. Иллюстрация этих выводов показана на наложении 3D-структур гликозилтрансферазы *A. agricola* и гомологов.

На следующем этапе проведена идентификация предполагаемой рамнозилтрансферазы на основе аминокислотной последовательности *Azospirillum brasilense* (OPH21365.1). Этот выбор обусловлен результатами BLAST-анализа UniProt, показавшими наивысший уровень гомологии по E-value и проценту идентичности среди всех исследованных белков.

В ходе BLASTp-запроса среди гомологов рамнозилтрансферазы *A. brasilense* выявлен белок ferritin-like domain-containing protein штамма *Azospirillum agricola* (WP_209777021.1) с E-value $2e-153$ и идентичностью 80,61 %. Это позволяет рассматривать данный белок как предполагаемую рамнозилтрансферазу исследуемого штамма. Далее проведён анализ консервативных доменов с использованием баз InterPro, CAZy и NCBI Conserved Domains.

Во всех базах данных был найден консервативный домен с идентификатором DUF455 (protein of unknown function) и с неизвестными функциями, однако последующий сравнительный анализ показал, что данный домен часто встречается в составе известных рамнозилтрансфераз у других видов бактерий, включая представителей рода *Azospirillum*. Это свидетельствует о возможной роли DUF455 в обеспечении рамнозилтрансферазной активности.

Для подтверждения принадлежности белка WP_209777021.1 к рамнозилтрансферазам смоделирована его 3D-структура с помощью AlphaFold3. Сравнение моделей рамнозилтрансферазы и

гликозилтрансферазы *Azospirillum agricola* показало сходную укладку и расположение аминокислот. Это морфологическое соответствие позволяет заключить, что белок WP_209777021.1, аннотированный в NCBI как ferritin-like domain-containing protein, на самом деле является рамнозилтрансферазой *Azospirillum agricola*, замаскированной неспецифической аннотацией.

3.6 Верификация амплификации целевого гена методом ПЦР

Для подтверждения амплификации целевого гена проведена ПЦР с использованием специфичных праймеров. ПЦР-продукты проанализированы агарозным гель-электрофорезом. В лунке М — молекулярный маркер с фрагментами 3000, 2000, 1500 и 1200 п.н. В лунке 1 — продукт ПЦР гена гликозилтрансферазы. На геле наблюдается чёткая полоса между 1200 и 1000 п.н., соответствующая ожидаемой длине. Это подтверждает успешную амплификацию гена рамнозилтрансферазы *Azospirillum agricola*.

Выводы: в геноме *Azospirillum agricola* был выявлен предполагаемый белок рамнозилтрансферазы WP_209782629.1. Обнаружение гомологов этого белка у других штаммов рода *Azospirillum* подтверждает его сохранённость в пределах рода. В последовательности белка выявлены консервативные аминокислотные участки, характерные для гликозилтрансфераз. Анализ филогенетического положения показал, что белок формирует кластер с другими рамнозилтрансферазами *Azospirillum*, что указывает на их общее происхождение. С помощью AlphaFold3 была получена надёжная трёхмерная модель белка с высоким уровнем уверенности предсказания (pLDDT > 90). Пространственный анализ показал сходство с известными рамнозилтрансферазами, что подтверждает структурную близость и поддерживает предполагаемую функцию. Были определены характерные домены гликозилтрансферазного типа и структурные особенности, указывающие на функциональную активность фермента. Экспериментально проведённая амплификация целевого гена с помощью ПЦР и последующий гель-электрофорез подтвердили наличие продукта длиной около 1100 пар

оснований, соответствующего ожидаемому размеру гена рамнозилтрансферазы.

Заключение: гены, кодирующие гликозилтрансферазы, остаются наименее изученными среди генов, вовлечённых в биосинтез поверхностных полисахаридов у бактерий рода *Azospirillum*. Эти ферменты играют ключевую роль в синтезе экзополисахаридов, липополисахаридов и других компонентов клеточной поверхности, однако для многих из них отсутствуют функциональная аннотация и экспериментальные данные. Для идентификации и филогенетического анализа гомологов рамнозилтрансфераз в геномах различных видов *Azospirillum* был разработан специализированный скрипт на Python с использованием библиотеки Biopython, который автоматизировал поиск, фильтрацию и классификацию последовательностей. С помощью этого инструмента и биоинформационных ресурсов BLAST, PDB, UniProt, InterPro, CAZy, DALI и AlphaFold3 проведена функциональная аннотация предполагаемой рамнозилтрансферазы WP_209777021.1 и гликозилтрансферазы WP_209782629.1. Анализ последовательности, доменной структуры и сравнение с другими ферментами выявили консервативный домен, характерный для рамнозилтрансфераз. Построенная с помощью AlphaFold3 трёхмерная модель позволила оценить укладку и каталитические центры белка, подтверждая его принадлежность к данному семейству. Эти результаты создают основу для дальнейших исследований функциональной роли фермента в биосинтезе полисахаридов у *Azospirillum*.

Список использованных источников

- 1 Dong J. Bacterial exopolysaccharides: Characteristics and antioxidant mechanism / J. Dong [et al.] // *International Journal of Biological Macromolecules*. — 2024. — Art. 138849.
- 2 Kumari J. Microbial exopolysaccharides: Classification, biosynthetic pathway, industrial extraction and commercial production to unveil its bioprospection: A comprehensive review / J. Kumari [et al.] // *International Journal of Biological Macromolecules*. — 2025. — Art. 139917.
- 3 Yao X. A Review on the Extraction, Structural Characterization, Function, and Applications of Peptidoglycan / X. Yao, Z. Yi, M. Xu, Y. Han // *Macromolecular Rapid Communications*. — 2025. — Art. 2400654.
- 4 Gorman, A. Lipopolysaccharide structure and the phenomenon of low endotoxin recovery / A. Gorman, A. P. Golovanov // *European Journal of Pharmaceutics and Biopharmaceutics*. — 2022. — V. 180. — P. 289-307.
- 5 Sheikhi, A. Analysis of immunogenicity and purification methods in conjugated polysaccharide vaccines: a new approach in fighting pathogenic bacteria / A. Sheikhi, M. Shirmohammadpour, N. Mahdei Nasirmahalleh, B. Mirzaei // *Frontiers in Immunology*. — 2024. — V. 15. — Art. 1483740.
- 6 Cantarel B. L. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics / B. L. Cantarel [et al.] // *Nucleic Acids Research*. 2009. Vol. 37, Database issue. P. D233–D238. URL: <https://www.cazy.org/> (дата обращения: 03.06.2025)
- 7 Holm, L. (2020). DALI and the persistence of protein shape. *Protein Science*, 29(1), 128-140. Dali, URL: <http://ekhidna2.biocenter.helsinki.fi/dali/> (дата обращения: 25.05.2025) Яз. Англ.

