

Правительство Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(СПбГУ)

УДК 577.32

Рег. № НИОКТР АААА-А16-116102010033-6

УТВЕРЖДАЮ
Проректор по научной работе



С. В. Микушев

« 30 » / 10 / 2023 г.

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

ЛАБОРАТОРИЯ БИМОЛЕКУЛЯРНОГО ЯДЕРНОГО МАГНИТНОГО РЕЗОНАНСА
(промежуточный, этап 11)

Руководитель НИР:
руководитель лаборатории,
PhD

A handwritten signature in blue ink, consisting of a large, stylized initial 'S' followed by a horizontal line.

Н.Р. Скрынников

Санкт-Петербург 2023

СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель НИР, руководитель
лаборатории, PhD

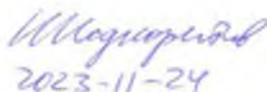


Н.Р. Скрынников
(весь отчёт)

24.11.2023

подпись, дата

Отв. исполнитель, зам. рук.
лаборатории

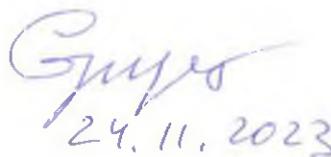

2023-11-24

И.С. Подкорытов
(разделы 4, 6, 7)

подпись, дата

Исполнители:

Ст. научн. сотр., канд. физ.-мат.
наук


24.11.2023

А.В. Струц
(введение, заключение)

подпись, дата

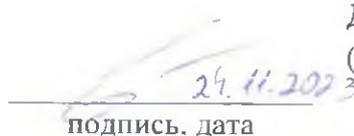
Ст. научн. сотр., канд. биол. наук


24.11.23

И.И. Тюрёва
(раздел 6)

подпись, дата

Научн. сотр.

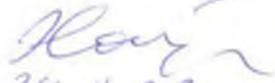

24.11.2023

Д.А. Лузик
(разделы 5, 6)

подпись, дата

Научн. сотр., канд. физ.-мат. наук,
Ph.D.

Б.Б. Харьков
(разделы 5, 6)


24.11.23

подпись, дата

Мл. научн. сотр.


24.11.23 С.А. Корбан
(раздел 5)

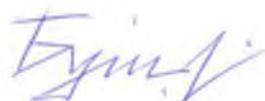
подпись, дата

Мл. научн. сотр.


24.11.2023 О.О. Лебеденко
(разделы 1, 2, 3, 4)

подпись, дата

Стажёр-исследователь.


24.11.2023 Е.И. Бушманова
(раздел 6)

подпись, дата

Стажёр-исследователь.


24.11.2023 П.С. Ваганова
(раздел 1)

подпись, дата

Стажёр-исследователь.


24.11.2023 А.Д. Лёвкина
(разделы 5, 6)

подпись, дата

Стажёр-исследователь, канд. биол.
наук.


24.11.23

А.Б. Матиев
(введение, раздел 6)

подпись, дата

Стажёр-исследователь, д-р. хим.
наук


24.11.23

С.И. Селиванов
(введение, заключение)

подпись, дата

Нормоконтроль


24.11.23

А.В. Вершинин

подпись, дата

РЕФЕРАТ

Отчет 40 с., 1 кн., 46 источн.

ЯМР СПЕКТРОСКОПИЯ, БЕЛКИ, СТРУКТУРА БЕЛКОВ, РАЗУПОРЯДОЧЕННЫЕ БЕЛКИ, МОЛЕКУЛЯРНАЯ ДИНАМИКА (МД), ДИФФУЗИЯ, ПАРАМАГНИТНАЯ РЕЛАКСАЦИЯ, РЕНТГЕНОСТРУКТУРНЫЙ АНАЛИЗ, НУКЛЕОСОМА, ДИНАМИКА ГИСТОНОВ, КОНСТАНТЫ СВЯЗЫВАНИЯ БЕЛКОВ, БЕЛКОВЫЕ ДИМЕРЫ, НЕЙРОСЕТИ

За отчётный период Лабораторией био-ЯМР СПбГУ были получены следующие основные результаты. (1) Была впервые продемонстрирована возможность предсказывать из первых принципов диффузионные параметры разупорядоченных белков (пептидов) на основе данных МД моделирования; было показано, что использование эмпирических инструментов (например, программы HYDROPRO) для подобных предсказаний может приводить к качественно неверным выводам. (2) Был разработан новый подход к интерпретации данных парамагнитной релаксации в (частично) разупорядоченных системах, основанный на использовании данных МД моделирования; в качестве приложения нового метода была исследована локализация N-концевого хвоста гистона H4 в нуклеосомной частице. (3) В рамках международного консорциума были разработаны стандартизованные протоколы для МД моделирования нуклеосомной частицы и анализа получаемых данных; в настоящее время ведётся запись траекторий общей длительностью 48 мкс. (4) Была депонирована в банк данных PDB полученная в нашей лаборатории структура малого белка LCB2; на примере этой структуры мы систематическим образом исследовали возможности для решения кристаллографических структур с использованием множественных компьютерных моделей белка в качестве моделей молекулярного замещения. (5) Проводится структурное исследование малого белка LCB3 T10Y с помощью методов гетероядерной ЯМР спектроскопии и рентгеновской кристаллографии; белок образует асимметричный димер с существенно видоизменённой структурой мономерных единиц. (6) Разработан новый нейросетевой алгоритм для предсказания констант связывания белков на основе структур соответствующих белок-белковых комплексов; в предварительных тестах новый алгоритм показывает более точные результаты чем программа PPI Affinity, считающаяся эталонной в данной области. (7) Был проделан анализ спектров ЯМР в новых металлоорганических люминесцентных соединениях; анализ потребовал внесения исправлений в стандартные программы для расчётов спиновой динамики.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	11
ОСНОВНЫЕ НАУЧНЫЕ РЕЗУЛЬТАТЫ, ПОЛУЧЕННЫЕ ЗА ОТЧЁТНЫЙ ПЕРИОД	17
1 Предсказание константы связывания для белок-белковых комплексов на основе структурных моделей	17
2 Создание консорциума для МД моделирования и экспериментального исследования методом ЯМР динамики гистоновых хвостов в составе нуклеосомной частицы.....	19
3 Пространственное распределение гистоновых хвостов в нуклеосоме по данным парамагнитной релаксации.....	22
4 Использование диффузионных данных ЯМР для валидации МД моделей нативно разупорядоченных белков	24
5 Использование множественных моделей молекулярного замещения при решении кристаллографической структуры белка.....	27
6 Структурное исследование предполагаемого асимметричного гомодимера белка LCB3 T10Y.....	29
7 Анализ и интерпретация ЯМР спектров металлоорганических люминофоров.....	30
ЗАКЛЮЧЕНИЕ.....	32
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	36

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

В настоящем отчете о НИР применяют следующие термины с соответствующими определениями:

- A_2 - спиновая система из двух ядер с одинаковой резонансной частотой
- $AA'X$ - спиновая система из трёх ядер, в которой ядра A и A' имеют равную резонансную частоту, а разница резонансных частот ядер типа A и типа X много больше любой из констант косвенного спин-спинового взаимодействия в системе
- $AA'XX'$ - спиновая система из четырёх ядер, в которой ядра A и A' а также X и X' попарно имеют равную резонансную частоту, а разница резонансных частот ядер типа A и типа X много больше любой из констант косвенного спин-спинового взаимодействия в системе
- AlphaFold - компьютерная программа для вычисления трёхмерной структуры белка по его аминокислотной последовательности,
<https://www.nature.com/articles/s41586-021-03819-2>
- AlphaFold2 - версия программы AlphaFold
- Amber - программный комплекс для моделирования движения биомолекул методом Молекулярной Динамики (МД), <https://ambermd.org/>
- bsc1 - один из типов силового поля для олигонуклеотидов в Молекулярной Динамике
- CCDC - Cambridge Crystallographic Data Centre, база данных рентгеновских структур, <https://www.ccdc.cam.ac.uk/>
- CHARMM - Chemistry at HARvard Macromolecular Mechanics, программный комплекс для молекулярного моделирования, <https://www.charmm.org>
- COVID-19 - COronaVIrus Disease 2019, коронавирусная инфекция 2019 г., вызываемая вирусом SARS-CoV-2
- DDfit - сервер для обработки данных диффузионных ЯМР экспериментов,
<https://ddfit.bio-nmr.spbu.ru>
- de novo* - снова (латынь), здесь используется в качестве определения к искусственно сконструированной биомолекуле (в противоположность природным соединениям)
- D_{rot} - коэффициент вращательной диффузии
- D_{tr} - коэффициент трансляционной диффузии
- ESM2 - компьютерная программа для вычисления трёхмерной структуры

белков,

https://www.science.org/doi/10.1126/science.ade2574?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed

ff14SB	- один из типов силового поля в программе Amber
ff19SB	- один из типов силового поля в программе Amber
ff99SB	- один из типов силового поля в программе Amber
ff99SB-disp	- один из типов силового поля в программе Amber
Github	- сайт разработчиков программного обеспечения, https://github.com
GROMACS	- программный комплекс для моделирования движения биомолекул методом Молекулярной Динамики (МД), https://www.gromacs.org/
H2A	- гистон H2A
H2B	- гистон H2B
H3	- гистон H3
H4	- гистон H4
HMR	- Hydrogen Mass Repartitioning, способ ускорения вычислений в программе Amber
HSQC	- Heteronuclear Single Quantum Coherence, вид двумерных гетероядерных ЯМР спектров
HYDROPRO	- программа для расчёта гидродинамических параметров макромолекул, http://leonardo.inf.um.es/macromol/
IDP	- Intrinsically Disordered Proteins, нативно разупорядоченные белки
I-Tasser (UM-TBM)	- компьютерная программа для вычисления трёхмерной структуры белка по его аминокислотной последовательности, https://zhanggroup.org/I-TASSER/
J	- константа косвенного спин-спинового взаимодействия в ЯМР
K_d	- константа диссоциации комплекса двух молекул
LCB2	- белок, хорошо связывается с шиповидным белком вируса SARS-CoV-2
LCB3	- белок, хорошо связывается с шиповидным белком вируса SARS-CoV-2
MolProbity	- сервис проверки достоверности структуры, http://molprobity.biochem.duke.edu/index.php?MolProbSID=riqsbp9n0jot03cuuulerik505&eventID=23
MR	- Molecular Replacement, молекулярное замещение, метод обработки данных в рентгеновской кристаллографии
N-H4	- N-терминальный фрагмент гистона H4

NOE	- Nuclear Overhauser Effect, ядерный эффект Оверхаузера, перенос намагниченности ядерных спинов с одного ядра на другое в ЯМР
NVE	- ансамбль с фиксированным числом частиц N , фиксированным объёмом V и фиксированной энергией E
NVP	- ансамбль с фиксированным числом частиц N , фиксированным объёмом V и фиксированным давлением P
OPC	- одна из моделей молекулы воды, используемая в методе Молекулярной Динамики
PCA	- Principal Component Analysis, метод главных компонент, метод обработки экспериментальных данных
PDB	- Protein Data Bank, база данных структур биомолекул, https://www.rcsb.org/
PDBBind	- база данных об аффинности биомолекул, http://www.pdbbind.org.cn/
Phyre2	- компьютерная программа для вычисления трёхмерной структуры белка по его аминокислотной последовательности, http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index
PPI Affinity	- программа для предсказания констант диссоциации белок-белковых комплексов, https://protcal-suite.cbe.bci.tu-dortmund.de/PPIAffinity/
PRE	- Paramagnetic Relaxation Enhancement, увеличение скорости спиновой релаксации в ЯМР экспериментах за счёт парамагнитного эффекта
ProteinMPNN	- компьютерная программа для вычисления аминокислотной последовательности по заданной пространственной структуре белка, https://www.science.org/doi/10.1126/science.add2187
Quark	- компьютерная программа для вычисления трёхмерной структуры белка по его аминокислотной последовательности, https://zhanggroup.org/QUARK/
R_{free}	- критерий качества модели в рентгеновской кристаллографии
RMSD	- Root-Mean-Square Deviation, среднеквадратичное отклонение
Rosetta	- программный комплекс для моделирования биомолекул, https://new.rosettacommons.org/docs/latest/meta/Rosetta-Timeline
RoseTTAFold	- компьютерная программа для вычисления трёхмерной структуры белков и их комплексов, https://www.science.org/doi/10.1126/science.abj8754
R_{work}	- критерий качества модели в рентгеновской кристаллографии

SARS-CoV-2	- коронавирус человека, причина COVID-19
Spinach	- компьютерная программа для расчёта динамики взаимодействующих спинов, https://spindynamics.org/group/?page_id=12
Swiss-Model	- компьютерная программа для вычисления трёхмерной структуры белка по его аминокислотной последовательности, https://swissmodel.expasy.org/
TIP4P-D	- одна из моделей молекулы воды, используемая в методе Молекулярной Динамики
TIP4P-D-disp	- одна из моделей молекулы воды, используемая в методе Молекулярной Динамики
TIP4P-Ew	- одна из моделей молекулы воды, используемая в методе Молекулярной Динамики
trRosetta	- компьютерная программа для вычисления трёхмерной структуры белков, https://www.nature.com/articles/s41596-021-00628-9
Ub	- белок убиквитин
МД	- Молекулярная Динамика
ПТМ	- ПостТрансляционная Модификация
ЯМР	- Ядерный Магнитный Резонанс

ВВЕДЕНИЕ

В Лаборатории биомолекулярного ЯМР СПбГУ сложились и развиваются следующие основные направления исследований: (1) Применение спектроскопии ЯМР и других биофизических методов в сочетании с методами компьютерного моделирования для исследования фундаментальных вопросов науки о белках; (2) Разработка новых ЯМР экспериментов и методов компьютерного моделирования для исследования белков; (3) Клеточные эксперименты для исследования пептидов и белков с потенциальным применением в области онкотерапии, а также лечения вирусных заболеваний; (4) Разработка методов определения белковых структур с использованием данных рентгеновской кристаллографии и ЯМР спектроскопии.

В минувшем году мы продолжили наш цикл работ, посвящённых молекулярной диффузии. Ранее мы опубликовали работы посвящённые теории и практике диффузионных экспериментов ЯМР в образцах амилоидных фибрилл [Kharkov et al. *Angew. Chem. Int. Ed.* **60**, 15445-15451 (2021); Podkorytov & Skrynnikov *J. Magn. Reson.* **344**, 107303 (2022)]. В этом году в рамках **проекта #4** мы сосредоточили усилия на исследовании диффузии разупорядоченных белков, опубликовав по этому вопросу большую статью [Lebedenko et al. *Biophys. J.* in press (2023)]. В этой вычислительной и экспериментальной работе мы впервые продемонстрировали возможность успешного предсказания коэффициента трансляционной диффузии для разупорядоченного белка (пептида) при помощи метода Молекулярной Динамики (МД). При этом мы использовали ряд новых технических приёмов, позволяющих ускорить вычисления и выполнять их более корректным образом. Например, мы предложили схему, при которой траектория пептида в малой ячейке была использована в качестве источника начальных конформаций для моделирования пептида в виде набора коротких траекторий в ячейках среднего и крупного размера, что позволило в несколько раз ускорить расчёты. Нам удалось показать, что сопоставление расчётных и экспериментальных данных позволяет идентифицировать удачные и неудачные модели МД (например, известная модель воды TIP4P-Ew терпит неудачу при моделировании разупорядоченного пептида). Наконец, мы показали, что ряд эмпирических инструментов, широко применяемых для предсказания коэффициентов диффузии разупорядоченных белков (таких как, например, программа HYDROPRO) на самом деле не подходят для этой цели и могут приводить к качественно неверным заключениям. В будущем мы планируем продолжить эту работу и опубликовать статью, посвящённую новому методу для обработки диффузионных данных,

реализованному в виде веб-сервера DDfit, а также осветить некоторые другие вопросы, связанные с расчётами диффузионных параметров.

Основным объектом исследования в вышеописанном проекте был разупорядоченный пептид N-H4, аминокислотная последовательность которого соответствует последовательности N-концевого участка гистона H4. Так называемые "гистоновые хвосты" играют важнейшую роль в пространственной организации хроматина и регуляции транскрипции. В **проекте #3** мы поставили своей целью исследовать локализацию хвоста H4 в нуклеосомной частице. Для решения этой задачи нашими партнерами (лаборатория проф. Джароняка, Университет штата Огайо) были изготовлены образцы нуклеосомы с использованием рекомбинантного гистона H4, меченого изотопом ^{15}N , и гистона H3, экспрессированного в варианте с цистеиновой мутацией и конъюгированного с парамагнитной меткой. Полученные образцы были использованы для измерения эффекта парамагнитного усиления спиновой релаксации (PRE) для наблюдаемых спинов $^1\text{H}^{\text{N}}$ в подвижном хвосте H4. Для интерпретации полученных результатов нами были использованы данные МД моделирования нуклеосомной частицы, включая траектории, недавно записанные в нашей лаборатории [Rabdano et al. *Angew. Chem. Int. Ed.* **60**, 6480-6487 (2021)]. Расчёт скоростей PRE из сверхдлинных траекторий Молекулярной Динамики представляет собой новый подход, позволяющий более точно оценить структурную и динамическую информацию, содержащуюся в этих параметрах. Проведенный анализ позволил охарактеризовать пространственное распределение хвоста H4 и сделать вывод о том, что он взаимодействует с нуклеосомной ДНК по механизму "нечёткого взаимодействия" (fuzzy interaction), поддерживая контакт с ДНК, но сохраняя при этом латеральную подвижность. По результатам этой работы нами была опубликована статья [Sun et al. *J. Am. Chem. Soc.* DOI 10.1021/jacs.3c10340 (2023)].

В продолжение этой работы, в рамках **проекта #2** мы организовали международный консорциум для моделирования нуклеосомной частицы. В состав консорциума вошли исследовательские группы из США, Канады, России и Китая, включая ведущие лаборатории в области МД моделирования биомолекул, а также лаборатории, занимающиеся экспериментальными исследованиями в этой области. Путём коллективного обсуждения были отобраны комбинации силовых полей и моделей воды, сочтённые наиболее перспективными для моделирования нуклеосомы (которая представляет собой сложную систему, имеющую в своём составе как структурированные, так и разупорядоченные элементы). Силами сотрудников нашей лаборатории были адаптированы для совместного использования наборы параметров силового поля bsc1 для

олигонуклеотидов, ff99SB и ff99SB-disp для белков, и TIP4P-D и TIP4P-D-disp для воды. Помимо этого, параметры силового поля ff99-disp были преобразованы формат для применения в программе Amber. Нами также был разработан унифицированный протокол для моделирования нуклеосомы и подготовлены инструменты для последующего анализа данных МД (например, расчёта скоростей спиновой релаксации), предоставленные в коллективное пользование участников консорциума. В настоящее время силами пяти лабораторий записываются траектории нуклеосомной частицы планируемой общей продолжительностью 48 мкс. Предметом исследования служит влияние изменяющих заряд модификаций (ср. ацетилирование лизина) на взаимодействие гистоновых хвостов с нуклеосомной ДНК. Полученные результаты должны послужить лучшему пониманию функционала так называемого "гистонового кода".

Отдельное внимание в нашей работе уделяется структурным исследованиям. В 2023 г. мы депонировали в банк данных PDB решённую нами кристаллографическую структуру малого белка LCB2 (блокирующий лиганд шиповидного белка SARS-CoV-2). Данная структура, получившая идентификатор 8C3E (<https://doi.org/10.2210/pdb8C3E/pdb>), является первой и единственной структурой PDB, полученной в СПбГУ. На примере рентгеноструктурного анализа белка LCB2 нами было предпринято методологическое исследование, темой которого стало использование компьютерных моделей белковой структуры в качестве моделей молекулярного замещения (molecular replacement, MR). В рамках этого исследования, **проект #5**, мы независимо провели процедуру определения структуры LCB2, начиная с пяти различных стартовых моделей, полученных с помощью современных алгоритмов для предсказания структуры: AlphaFold2, Rosetta, MultiFold, RoseTTAFold, и trRosetta. Как было показано, все начальные модели достаточно значительно отклоняются от итоговой структуры и заметно разнятся между собой; в то же время все пять итоговых структур находятся в отличном согласии друг с другом. Таким образом, мы подтвердили сходимость процесса определения структуры независимо от отклонений, наблюдаемых в исходных компьютерных моделях. В свою очередь это позволило нам оценить точность полученной итоговой структуры, 0.30 Å, что соответствует ожидаемой для структуры с кристаллографическим разрешением 2.1 Å. Мы пришли к заключению, что использование множественных компьютерных MR моделей позволяет более успешно уточнять отдельные элементы структуры; например, в нашем случае использование MR моделей AlphaFold2 и RoseTTAFold позволяет реконструировать боковую цепь остатка R41, что оказывается невозможным при использовании остальных MR моделей.

Объектом структурного исследования в нашей лаборатории также стал малый белок LCB3, содержащий мутацию T10Y (**проект #6**). LCB3 относится к тому же классу искусственно полученных блокирующих лигандов шиповидного белка SARS-CoV-2, что и описанный выше мини-белок LCB2. Мутация T10Y была введена в состав белковой последовательности с тем, чтобы повысить аффинность белка к видоизменённым мишеням (например, S-белкам штаммов дельта+ и омикрон). При рассмотрении ЯМР спектра $^1\text{H}^{\text{N}}, ^{15}\text{N}$ -HSQC образца LCB3 T10Y было обнаружено, что данный спектр насчитывает вдвое больше пиков, чем можно ожидать для белка его размера. Чтобы объяснить этот феномен, мы предположили, что LCB3 T10Y образует в растворе асимметричный гомодимер. Это предположение было подтверждено данными релаксационных, диффузионных и обменных ЯМР экспериментов, а также данными изотермической титрационной калориметрии. В качестве следующего шага был записан набор 3D гетероядерных экспериментов и выполнено отнесение сигналов основной цепи для молекул А и В, образующих предполагаемый асимметричный димер. В то же время нам удалось успешно кристаллизовать LCB3 T10Y и получить рентгенодифракционные данные высокого разрешения; в настоящее время ведётся работа над получением кристаллографической структуры белка. На примере LCB3 T10Y мы планируем разработать новую малозатратную методику для получения структурной модели димера на основе ограниченных данных ЯМР (хим. сдвигов атомов основной цепи). Для этого мы намерены использовать методы белкового докинга и/или методы молекулярной динамики под контролем экспериментальных данных (хим. сдвигов). Для оценки получаемых моделей мы предполагаем использовать кристаллографическую структуру димера. По нашему замыслу, новая методика позволит построить достоверную модель димера, не прибегая к анализу данных NOE (nuclear Overhauser effect, ядерный эффект Оверхаузера), что является наиболее трудоёмким элементом стандартного структурного исследования ЯМР.

В связи с предшествующим проектом, укажем, что константа связывания для белкового димера (или, в более общем случае, белок-белкового комплекса) может определяться не только экспериментальным путём с помощью изотермической титрационной калориметрии или других методов, но и с помощью расчётов на основе структуры комплекса высокого разрешения. Традиционно для этой цели применялись т.н. физические модели; однако ввиду грубой параметризации точность предсказаний была и остаётся низкой. В рамках **проекта #1** мы создали новую программу для предсказания аффинности белок-белковых взаимодействий на основе нейросети. Алгоритм новой программы вкратце можно описать следующим образом. Для каждого из исследуемых

комплексов мы выделяем элементы структуры, образующие интерфейс связывания (определяется согласно отсечке в 5 Å, задающей максимальное расстояние между атомами взаимодействующих белков). Затем для всех остатков, образующих интерфейс, генерируется вектор признаков, кодирующий информацию о белковой последовательности; для этой цели мы воспользовались успешно зарекомендовавшей себя языковой моделью-автоэнкодером ESM2. Структуры комплексов были представлены в виде полносвязных графов, где каждой вершине соответствует C^α атом интерфейса с приписанным ему вектором признаков. Непосредственно для решения задачи о предсказании констант связывания была построена графовая нейросетевая модель с вниманием на вершинах и ребрах (EdgeConvNodeGATModel). Тренировка и тестирование нейросети проводились на основе базы данных PDBBind. Ввиду того, что эта база плохо аннотирована и не обновлялась в течение последних пяти лет, мы провели большую работу по её ревизии и обновлению. После того, как созданный нами алгоритм прошёл обучение на обновлённой и расширенной версии PDBBind, он показал более высокую точность в предсказании K_d , чем недавно разработанный алгоритм PPI Affinity, который можно считать эталоном в этой области. В ближайшем будущем мы планируем обнародовать полученные результаты и продолжить работу по совершенствованию базы данных PDBBind и разработанного нами алгоритма.

Разработка алгоритмов на основе нейросетей является для нашей лаборатории новой областью деятельности. В то же время наша группа сохраняет связь со своими истоками, которые лежат в сфере спектроскопии ЯМР. В рамках **проекта #7** мы продолжаем сотрудничество с лабораторией проф. С.П. Туника (СПбГУ), специализирующейся в области разработки металлоорганических люминофоров для применения в качестве функциональных зондов в биологии и экспериментальной медицине. В частности, в минувшем году мы приняли участие в исследованиях нового класса люминесцентных соединений, содержащих спиновую систему N₂-Pt-P-CH₂-P-Pt-N₂. Тщательный анализ спектров ³¹P полученных образцов выявил эффект дифференциального уширения спектральных линий, который и до этого наблюдался в такого рода соединениях, но не получил должного объяснения в литературе. Первоначально мы предположили, что наблюдаемый эффект является следствием хорошо известного явления кросс-коррелированной спиновой релаксации. Однако моделирование спектров с помощью программы спиновой динамики Spinach не подтвердило этого предположения. В то же время был идентифицирован новый сценарий, при котором эффект дифференциального уширения возникает благодаря динамической модуляции константы спинового спаривания между двумя спинами ¹⁹⁵Pt. Помимо этого нами была

обнаружена ошибка в программе Spinach в части касающейся моделирования т.н. скалярной релаксации второго рода, что не позволяет корректным образом учесть взаимодействие спинов ^{195}Pt и ^{14}N . В настоящее время мы работаем над созданием собственной программы для моделирования спектров, которая позволила бы максимально строгим и полным образом учесть все механизмы спектрального уширения в исследуемой системе.

ОСНОВНЫЕ НАУЧНЫЕ РЕЗУЛЬТАТЫ, ПОЛУЧЕННЫЕ ЗА ОТЧЁТНЫЙ ПЕРИОД

1 Предсказание константы связывания для белок-белковых комплексов на основе структурных моделей

Целью настоящего проекта является создание алгоритма для достоверного предсказания константы связывания двух белков на основе структуры соответствующего белок-белкового комплекса. Эта задача представляет собой не только фундаментальный, но и прикладной интерес. Мы вступаем в эпоху активного конструирования белков, включая белки – лиганды важных биомедицинских мишеней. Компьютерные алгоритмы, используемые для этой цели, генерируют реалистичные модели соответствующих белок-белковых комплексов [1]. Способность точно оценивать константу связывания на основе подобных структурных моделей позволяет отбирать наиболее перспективные белковые лиганды, повышая тем самым эффективность метода в целом.

Традиционно задача об определении константы диссоциации комплекса K_d или, что эквивалентно, свободной энергии связывания ΔG_{bind} , решалась на основе физических моделей. В такого рода моделях оценивается стабилизирующий вклад от различных взаимодействий: например, водородных связей или гидрофобной упаковки боковых цепей на интерфейсе связывания [2, 3]. Как показала практика, точность подобных предсказаний оказывается весьма невысокой. Фундаментальной причиной такого неуспеха является относительно грубая параметризация белок-белковых взаимодействий.

В качестве альтернативы мы исследуем возможность решения этой задачи на основе нейросетевых моделей, т.е. многомерных нелинейных функций, аппроксимирующих сложные зависимости в исследуемых данных. Преимущество нейросетевых подходов заключается в том, что они способны выявить и успешно смоделировать скрытые и трудноуловимые связи между координатами белкового комплекса и соответствующей константой диссоциации. В общем виде обучение (или, иначе говоря, настройка параметров) нейросетевой модели проводится на основе тренировочного набора данных, который включает в себя большое количество кристаллографических структур белок-белковых комплексов и соответствующие экспериментально измеренные значения K_d . С математической точки зрения обучение модели представляет собой минимизацию функции ошибок, определяющей меру расхождения экспериментальных данных и предсказаний, получаемых с помощью нейросетевой модели. Для проверки итоговой модели используется независимый тестовый набор данных, который не был задействован в процессе обучения.

В нашем случае для построения нейросетевой модели использовалась база данных PDBBind [4], содержащая 1975 структур белок-белковых комплексов с известным из эксперимента значением K_d . К сожалению, объём данных в этом наборе относительно невелик, а сама база данных плохо аннотирована (например, данные не классифицированы по типу экспериментальной методики, использовавшейся для измерения K_d). В этих условиях не представляется возможным использовать PDBBind для выработки вектора признаков, которые можно было бы привязать к отдельным элементам белковых структур (например, аминокислотным остаткам или отдельным атомам). Вместо этого мы воспользовались для этой цели успешно зарекомендовавшей себя языковой моделью-автоэнкодером ESM2 [5], которая вырабатывает содержательный вектор признаков на основе белковой последовательности. Иными словами, ESM2 позволяет закодировать последовательность белка в такой форме, которая несёт в себе информацию о скрытых в этой последовательности неявных закономерностях. В качестве альтернативы мы опробовали известный автоэнкодер ProteinMPNN [6], который также показал в наших тестах хорошие результаты.

Для обучения разрабатываемой нами нейросети из исходного набора данных PDBBind было отобрано 869 димерных комплексов. Затем получившаяся выборка была случайным образом разделена на три части – тренировочную, валидационную и тестовую – в соотношении 70%, 20% и 10%. Для каждого из исследуемых белок-белковых комплексов мы выделили элементы структуры, образующие интерфейс связывания (определялся согласно отсечке в 5 Å, задающей максимальное расстояние между атомами взаимодействующих белков). Для решения целевой задачи оценки K_d комплекса мы построили графовую нейросетевую модель с вниманием на вершинах и ребрах (EdgeConvNodeGATModel) [7]. Структуры комплексов были представлены в виде полносвязных графов, построенных на $C\alpha$ атомах, находящихся на интерфейсе взаимодействия. Каждая вершина графа характеризовалась вектором признаков, который был извлечен для соответствующей аминокислоты на подготовительном этапе с помощью языковой модели ESM2.

Построенная таким образом нейросетевая модель на тестовом наборе данных из PDBBind продемонстрировала уровень корреляции 0.56, сопоставимый с уровнем корреляции для лучших современных предикторов, включая недавно опубликованную модель PPI Affinity [8]. Следует отдельно отметить, что PPI Affinity также была построена с использованием набора данных PDBBind, однако авторы не указали, каким образом данные были поделены на тренировочную, валидационную и тестовую выборки. Для независимого сравнения нашей модели и PPI Affinity был составлен новый набор данных

из 30 структур белок-белковых комплексов с известным значением K_d , полученным с помощью наиболее точного экспериментального метода – изотермической титрационной калориметрии. На этом новом наборе данных PPI Affinity демонстрирует коэффициент корреляции 0.40, в то время как наша модель EdgeConvNodeGATModel – коэффициент корреляции 0.70. Дальнейшее улучшение нашей модели упирается в ограниченный размер набора данных PDBBind, который фактически не пополнялся после 2018 года. В ходе нашей работы мы планируем упорядочить и расширить эту базу данных. Поиск новых систем будет осуществляться на основе анализа структурной базы данных PDB, из которой будут извлечены структуры димерных белок-белковых комплексов. На новом улучшенном наборе данных мы планируем, в частности, протестировать другие современные языковые модели [9] помимо ESM2.

2 Создание консорциума для МД моделирования и экспериментального исследования методом ЯМР динамики гистоновых хвостов в составе нуклеосомной частицы

Нуклеосома представляет собой базовый элемент системы хранения, копирования и транскрипции генетической информации. Коровая нуклеосомная частица состоит из белковой сборки (по 2 копии гистонов H2A, H2B, H3 и H4) вокруг которой навивается примерно 1.7 оборота двухцепочечной ДНК. Гистоны содержат в своём составе разупорядоченные N- и C-терминальные фрагменты, известные как гистоновые хвосты. Эти хвосты играют роль своего рода "швартовых концов" к которым привязываются сотни различных хроматин-ассоциированных белков, ответственных за активацию/репрессию транскрипции, сборку и ремоделирование хроматина и т.д. Отдельную группу хроматин-ассоциированных белков представляют собой хроматин-модифицирующие ферменты, ответственные за посттрансляционные модификации (ПТМ) хвостов, такие как метилирование боковых цепей лизина и аргинина, ацетилирование лизина и т.д. Изменяемый паттерн ПТМ позволяет регулировать связывание хвостов с различными хроматин-ассоциированными белками.

Ввиду разупорядоченного характера гистоновых хвостов, их исследование с помощью традиционных методов структурного анализа, таких как рентгеновская дифрактометрия, практически невозможно. В этой ситуации источником структурных моделей гистоновых хвостов в контексте хроматиновых волокон становится метод МД моделирования. Однако для применения этого метода необходимо использовать силовое поле и модель воды в равной мере подходящие для моделирования структурированных белков, разупорядоченных белков и ДНК (эта проблема не имеет в настоящее время

однозначного решения); для получения статистически достоверных результатов требуется также записывать длинные траектории (десятки микросекунд и более). Помимо этого, необходимым условием для использования метода МД моделирования является его валидация с помощью экспериментальных данных. На первый план здесь выступают данные спектроскопии ЯМР, такие как химические сдвиги и скорости спиновой релаксации (включая парамагнитную релаксацию).

Всё вышеописанное подтолкнуло нас в прошлом году к организации консорциума для МД моделирования нуклеосомной частицы. Помимо нашей группы, в состав консорциума вошли лаборатории проф. А. Онуфриева (Политехнический университет Виргинии, США), проф. А. Панченко (Университет Квинс, Канада), проф. Й Шу (Университет Циньхуа, Китай) и проф. А. Шайтана (МГУ). Экспериментальные данные будут получены в лаборатории проф. Джароньяка (Университет штата Огайо, США). Создание консорциума позволит объединить усилия и ресурсы участников для оптимизации протокола моделирования и получения массива из длинных траекторий нуклеосомной частицы. Консорциум ставит перед собой следующие задачи: (1) изучение динамики хвоста гистона H4 дикого типа и хвоста гистона H4 с точечной мутацией R3A в составе нуклеосомной частицы с помощью метода МД моделирования и по данным ЯМР спектроскопии (релаксация спинов ^{15}N); (2) совместная разработка оптимального протокола для МД моделирования нуклеосомной частицы; (3) публикация массива записанных траекторий в открытых публичных репозиториях для дальнейшего использования в других исследовательских проектах. Непосредственно нашей научной группой осуществляется координирование участников консорциума, а также подготовка данных и программных скриптов, необходимых для запуска и анализа траекторий МД. В 2023 г. силами нашей лаборатории был подготовлен набор скриптов для запуска и анализа траекторий МД, а также сопутствующая документация. Все подготовленные материалы распространены среди участников консорциума посредством репозитория

<https://github.com/bionmr-spbu-projects/nucleosome-md-simulation>

<https://github.com/OOLebedenko/nucleosome-trajectory-processing>.

В настоящее время ведётся запись траекторий МД силами исследовательских групп, участвующих в консорциуме, с использованием располагаемых вычислительных ресурсов.

Вкратце, исходная модель для МД моделирования построена на основе структуры 3LZ0. Отсутствующие в этой структуре гистоновые хвосты достраиваются в протяжённой конформации (полипролиновая спираль); вслед за этим двугранным углом отдельных

остатков в составе хвостов присваиваются значения, находимые в структуре 1AOI, что позволяет избежать стерических конфликтов с нуклеосомной ДНК.

Совместно с участниками консорциума нами были выбраны четыре комбинации силовое поле / модель воды: (1) ff99SB / OPC; (2) ff19SB / OPC; (3) ff14SB / TIP4P-D; (4) ff99SB-disp / TIP4P-D-disp. Сочетание параметров ff99SB / OPC рекомендовано разработчиками модели воды OPC, являющейся моделью общего назначения, подходящей для моделирования как свёрнутых, так и развёрнутых белков [10]. Сочетание ff19SB / OPC является новейшим стандартом, рекомендуемым к использованию в официальной документации пакета Amber [11]. Комбинация ff14SB / TIP4P-D включает в себя специализированную модель воды TIP4P-D, предназначенную для моделирования разупорядоченных белков; как было показано нами ранее, данная комбинация успешно предсказывает скорости спиновой релаксации в гистоновых хвостах [12]. И наконец, недавно предложенная комбинация ff99SB-disp / TIP4P-D-disp, наряду с ff99SB / OPC, претендует на универсализм при моделировании как свёрнутых, так и развёрнутых белков [13].

Во всех четырех комбинациях для моделирования ДНК был выбран наилучший на сегодняшний момент набор параметров bsc1 [14]. При этом силами сотрудников нашей лаборатории была проведена адаптация наборов параметров bsc1, ff99SB, ff99SB-disp, TIP4P-D и TIP4P-D-disp для устранения взаимных конфликтов в обозначениях типов атомов, что открывает возможность для их совместного использования. Также отметим, параметры силового поля ff99-disp были изначально опубликованы в формате, совместимом с дистрибутивом GROMACS. В нашей лаборатории была проведена работа по их преобразованию в формат, используемый дистрибутивом Amber. Результаты работы были апробированы разработчиками силового поля ff99SB-disp и включены в официальный репозиторий https://github.com/paulrobustelli/Force-Fields/tree/master/Amber_FF/a99SBdisp; также планируется, что параметры силового поля ff99SB-disp станут доступными для пользователей Amber, начиная со следующего официального релиза данного программного пакета.

Моделирование МД проводится в новых (по сравнению с нашими предыдущими работами) условиях, pH 7.2, 150 mM KCl, имитирующих физиологические условия и параметры образцов ЯМР. В качестве первого этапа записывается относительно короткая, 100 нс, траектория в ячейке большого размера (усечённый октаэдр с длиной ребра $L = 99$ Å). За это время гистоновые хвосты "конденсируются" на поверхности нуклеосомы. На следующем этапе система помещается в ячейку меньшего размера, после чего записывается в рабочем режиме длинная траектория. Размер уменьшенной ячейки ($L = 84$

А) был подобран нами на основе анализа данных МД моделирования, проводившихся в лабораториях наших партнёров [15].

Для распределения заданий среди лабораторий-участников был использован подписной лист. В настоящее время проводится запись траекторий общей длиной в 48 мкс (по 6 мкс на каждую комбинацию силовое поле / модель воды для гистона Н4 дикого типа и мутанта R3A). В лаборатории К. Джароньяка получены релаксационные данные для образцов, содержащих гистон Н4 дикого типа и мутант R3A.

3 Пространственное распределение гистоновых хвостов в нуклеосоме по данным парамагнитной релаксации

В предыдущем разделе мы охарактеризовали функциональную роль разупорядоченных гистоновых хвостов в хроматине. В данном контексте значительный интерес представляет вопрос о поведении гистоновых хвостов в составе нуклеосомной частицы. С одной стороны есть основания считать, что положительно заряженные хвосты взаимодействуют с отрицательно заряженной нуклеосомной ДНК и, таким образом, находятся в связанном состоянии на поверхности нуклеосомной частицы. С другой стороны, можно предположить, что хвосты полностью сольватированы и ведут себя подобно разупорядоченной полимерной цепочке в водном окружении. Если первый сценарий ограничивает возможности для взаимодействия гистоновых хвостов с хроматин-ассоциированными белками, то второй сценарий создает максимально благоприятные условия для таких взаимодействий.

Как уже было описано выше, большинство стандартных методов структурного анализа неприменимы для исследования гистоновых хвостов. В нашей предшествующей работе мы использовали данные спиновой релаксации ^{15}N чтобы получить представление о подвижности гистоновых хвостов в нуклеосомной частице [12]. Однако эти данные дают возможность лишь косвенно судить о локализации хвостов по отношению к "телу" нуклеосомной частицы. В прошлом году мы инициировали совместный проект с лабораторией К. Джароньяка (Университет штата Огайо), в котором мы используем эффект парамагнитного усиления спиновой релаксации (paramagnetic relaxation enhancement, PRE) для получения данных, непосредственно характеризующих пространственное распределение гистоновых хвостов.

В рамках этого проекта нашими партнёрами были изготовлены образцы нуклеосомных частиц, в которых гистоны Н4 были экспрессированы с изотопным обогащением спином ^{15}N , а гистоны Н3 были экспрессированы в варианте с цистеиновой мутацией K36C, L65C, K79C или Q125C и конъюгированы с парамагнитной меткой

MTSL. Используя эти образцы, сотрудники лаборатории Джароняка измерили скорость парамагнитной релаксации для спинов $^1\text{H}^{\text{N}}$ в N-терминальном хвосте гистона H4. Для интерпретации полученных данных мы использовали два набора МД траекторий, записанных с применением моделей воды TIP4P-D и OPC [12, 15]. Данные МД моделирования были использованы для предсказания скоростей парамагнитной релаксации. С этой целью применялся строгий количественный подход на основе теории Редфилда, где учитывалась не только вращательная динамика нуклеосомной частицы, но и трансляционное движение гистонового хвоста по отношению к телу нуклеосомы. Полученные результаты показали полуколичественное согласие с данными релаксационных измерений, тем самым подтвердив достоверный характер МД модели.

В качестве следующего шага мы использовали данные МД моделирования для того, чтобы реконструировать пространственное распределение хвоста H4 по отношению к телу нуклеосомной частицы. Полученные таким образом карты плотности подтвердили выдвинутое нами ранее предположение о том, что гистоновые хвосты взаимодействуют с нуклеосомной ДНК по механизму "нечёткого взаимодействия" (fuzzy interaction), т.е. формируют короткоживущие электростатические контакты, сохраняя при этом в значительной мере свою конформационную подвижность.

В 2023 году нами была написана и опубликована в *Journal of the American Chemical Society* статья с изложением полученных результатов [16]. В ходе этой работы нами был дополнительно исследован вопрос об альтернативных подходах к интерпретации данных PRE в рамках так называемого конформационного ансамбля. Мы показали, что такого рода статическая модель не содержит информации о характерных временах конформационной динамики, определяющих значения PRE, что в значительной мере увеличивает неопределённость анализа. Помимо этого, ярко выраженная зависимость PRE от расстояния между взаимодействующими спинами приводит к ситуации, когда один или несколько конформеров в составе конформационного ансамбля воспроизводят данные PRE, тогда как остальные конформеры отбираются фактически произвольным образом [17]. Мы проиллюстрировали такого рода эффект, сконструировав конформационный ансамбль, в составе которого два конформера со статистическим весом по 10% успешно воспроизводят данные PRE, тогда как остальные сорок конформеров лишь в минимальной степени отражают экспериментальные данные. Таким образом, следует отдать предпочтение разработанному нами новому подходу к интерпретации данных, основанному на прямом сравнении экспериментальных значений PRE и значений, рассчитанных из траекторий Молекулярной Динамики.

4 Использование диффузионных данных ЯМР для валидации МД моделей нативно разупорядоченных белков

Нативно разупорядоченные белки (IDP) представляют собой обширный класс белков, многие из которых тесным образом связаны с различными патологическими процессами в человеческом организме. Широко известна ключевая роль IDP в нейродегенеративных заболеваниях, таких как болезнь Альцгеймера; целый ряд разупорядоченных белков играет ключевые роли в онкогенезе. В этом контексте приобретает особую актуальность структурно-динамическая характеристика IDP. Если в случае свернутых белков подобный анализ опирается прежде всего на структуры, полученные методом рентгеновской дифрактометрии, то в случае разупорядоченных белков аналогичную роль могут выполнять траектории МД, которые способны во всей полноте отразить конформационное многообразие IDP, включая динамические параметры (т.е. скорости переходов между различными конформационными состояниями). Однако необходимым условием для применения моделей МД является их тщательная валидация путём сопоставления с экспериментальными данными. При этом круг доступных экспериментальных данных достаточно ограничен и сводится в основном к различным параметрам, измеряемым с помощью жидкостной спектроскопии ЯМР. Среди них одним из наиболее значимых является коэффициент трансляционной диффузии D_{tr} , измеряемый с помощью ЯМР экспериментов стимулированного эха с применением импульсных градиентов магнитного поля. Тем не менее до настоящего времени в литературе отсутствуют примеры полноценного расчёта D_{tr} на основе данных МД моделирования IDP и последующего сравнения расчётных значений с экспериментальными. Это и стало целью описываемой работы.

В качестве основного исследуемого объекта был избран пептид N-H4 (25 аминокислотных остатков) с последовательностью, воспроизводящей последовательность N-концевого фрагмента гистона H4. Дополнительно в качестве модельной системы был исследован глобулярный белок убиквитин (Ub). Данные по трансляционной диффузии N-H4 и Ub были получены в рамках данной работы сотрудниками нашей лаборатории; в дополнение к этому данные по вращательной диффузии убиквитина были позаимствованы из литературы. Обработка диффузионных данных была проведена с помощью сервера DDfit, разработанного и поддерживаемого в нашей лаборатории (<https://ddfit.bio-nmr.spbu.ru/>).

В ходе исследования был разработан ряд технических приёмов для моделирования диффузии и расчёта значений D_{tr} и D_{rot} из данных молекулярной динамики. Во-первых, следует отметить, что для определения диффузионных параметров необходимо записать

несколько траекторий в ячейках разного размера [18]. Нами был разработан протокол, в котором моделирование пептида в ячейке малого размера осуществлялось стандартным способом, а затем набор конформаций, полученных в ходе данного моделирования, использовался в качестве начальных условий для записи множества коротких траекторий в ячейках среднего и большого размера. Такой подход помогает в несколько раз сократить время вычислений за счёт параллельного использования ресурсов (карт с графическими процессорами). Во-вторых, повсеместно используемый термостат Ланжевена не позволяет с количественной точностью воспроизводить динамические параметры системы, включая диффузионные параметры. В качестве альтернативы нами был запрограммирован для платформы Amber стохастический термостат Бусси-Паринелло [19] (опция *ntt* = 11 в текущем дистрибутиве Amber). В третьих, в настоящий момент не имеет однозначного проблема подбора модели воды для моделирования разупорядоченных белков. В нашей работе была задействована классическая модель TIP4P-Ew, а также наиболее перспективные модели TIP4P-D и OPC [20-22].

Общая длительность траекторий N-H4, записанных в рамках данной работы, составила 75 мкс. При этом часть траекторий была записана с применением ансамбля NPT, а часть – ансамбля NVE, что позволило оценить применимость термостатирующего алгоритма Бусси-Паринелло. Совокупная длина траекторий убиквитина составила 30 мкс. В дополнение к этому был также записан массив траекторий для определения диффузии в воде, представленной различными моделями (перечислены выше).

Анализ полученных данных показал, что в случае убиквитина успех предсказаний D_{tr} и D_{rot} на основе данных МД моделирования целиком определяется достаточно тривиальным эффектом – а именно, вязкостью воды (заметно завышена для модели TIP4P-Ew, несколько занижена для модели TIP4P-D и почти точно воспроизводится в модели OPC). С другой стороны, для разупорядоченного пептида N-H4 картина оказалась более сложной. В этом случае на результаты влияет не только вязкость воды, но и характеристики конформационного ансамбля, получаемого в ходе моделирования. Как выяснилось, использование классической модели TIP4P-Ew ведёт к тому, что пептид принимает излишне компактные конформации, что приводит к переоценке коэффициента диффузии. С другой стороны, молекулярная динамика с использованием воды TIP4P-D и OPC успешно воспроизвела экспериментальные значения D_{tr} (с поправкой на тривиальный эффект вязкости). Дальнейшая попытка определить какая из этих двух моделей воды является предпочтительной на основе релаксационных данных ^{15}N также не дала результата – обе модели с практически одинаковой степенью точности воспроизвели экспериментальные результаты.

Если в нашей работе диффузионный коэффициент D_{tr} был рассчитан по данным МД моделирования "из первых принципов", то в ряде недавно опубликованных статей для этой цели применяется упрощённый инструментарий, а именно программа для расчёта гидродинамических параметров HYDROPRO [23]. Прделанные нами вычисления с использованием HYDROPRO показали, что такой подход может привести к качественно неверным заключениям. Например, в нашем случае данная стратегия приводит к выводу о том, что модель TIP4P-Ew является лучшей для моделирования разупорядоченного пептида, тогда как фактически она является худшей из всех рассматриваемых в настоящей работе. Причиной неверных предсказаний является тот факт, что HYDROPRO предназначен для работы со свёрнутыми белками, т.е. белками, обладающими достаточно жёсткой структурой. Применение данной программы к полипептидным цепочкам с высокой конформационной подвижностью приводит к значительным ошибкам.

Для того, чтобы удостовериться в вышеописанной интерпретации, мы записали отдельную серию траекторий N-H4 с наложенными ограничениями. В этих траекториях пептид диффундировал без изменения конформационного состояния, как жёсткая молекула. Полученные таким образом результаты подтвердили нашу гипотезу – пренебрежение конформационной динамикой пептида ведёт к существенной систематической недооценке D_{tr} . Аналогичным образом мы продемонстрировали, что ряд других эмпирических подходов для расчёта D_{tr} по данным МД моделирования также ведёт к неудовлетворительным результатам [24-26].

Наконец, отметим, что моделирование диффузии IDP методом молекулярной динамики требует использования ячеек очень большого размера и в силу этого является достаточно времязатратным. Для того, чтобы ускорить вычисления мы опробовали протокол с увеличенным шагом интегрирования, 4 фс, в сочетании со схемой перераспределения массы атомов водорода (hydrogen mass repartitioning, HMR). При этом выяснилось, что если мы следовали предписаниям руководства для пользователей программы Amber [11] и прикладывали схему HMR исключительно к белковой молекуле, то результаты моделирования оказывались неудовлетворительными. Однако в том случае, когда схема HMR прикладывалась не только к молекулам белка, но и к молекулам воды, коэффициенты диффузии удалось предсказать с хорошей точностью. Хотя быстрое движение атомов водорода в молекулах воды частично ограничено алгоритмом SETTLE, более общая схема HMR необходима для того, чтобы избежать ошибок интегрирования [27].

Все изложенные выше результаты были описаны в нашей статье, которая в настоящее время находится в печати в журнале *Biophysical Journal* [28].

5 Использование множественных моделей молекулярного замещения при решении кристаллографической структуры белка

В 2023 г. мы депонировали в банк данных PDB решённую нами кристаллографическую структуру малого белка LCB2. Данная структура, получившая идентификатор 8C3E (<https://doi.org/10.2210/pdb8C3E/pdb>), характеризуется следующими параметрами: кристаллографическое разрешение 2.1 Å, $R_{work} = 0.211$, $R_{free} = 0.246$, 100-й перцентиль MolProbity. Следует отметить, что эта структура является первой и единственной структурой PDB, полученной в СПбГУ.

На примере рентгеноструктурного анализа белка LCB2 нами было проведено методологическое исследование, темой которого стало использование компьютерных моделей белковой структуры в качестве моделей молекулярного замещения (molecular replacement, MR). Метод молекулярного замещения широко применяется при решении рентгеновских структур; к настоящему времени более 80% всех биомолекулярных структур в банке данных PDB получены этим методом. В качестве моделей MR традиционно используются структуры гомологичных (или в тех случаях, когда это возможно, идентичных) белков. Исторически компьютерные модели белков, полученные с помощью алгоритмов для предсказания белковых структур, не применялись для этой цели ввиду их низкого качества. Однако в последние годы появилось несколько принципиально новых алгоритмов (в частности, с использованием нейросетей), способных предсказывать структуры с высокой точностью, достаточной для их использования в качестве MR моделей.

Использование MR моделей, сгенерированных компьютерными программами, имеет ряд преимуществ. Во-первых, подобные модели могут быть получены в том числе для тех белков, для которых отсутствуют гомологичные структуры (это, например, актуально для белков, полученных методом белковой инженерии). Во-вторых, в отличие от гомологов, компьютерные модели точно соответствуют аминокислотной последовательности исследуемого белка и, следовательно, не нуждаются в "переделке" для преобразования в MR модель. В третьих, существует несколько современных высокоэффективных алгоритмов для предсказания белковых структур, что позволяет использовать при решении кристаллографической структуры множественные MR модели.

В этой связи открываются новые возможности для оценки качества различных компьютерных моделей, для изучения влияния исходных моделей на получаемую структуру, а также для оценки точности получаемой структуры. В рамках данного исследования нами были исследованы компьютерные модели полученные с помощью алгоритмов AlphaFold2 [29], Rosetta [30], MultiFold [31], RoseTTAFold [32], trRosetta [33],

Phyre2 [34], Swiss-Model [35], I-Tasser (UM-TBM) [36] и Quark [37]. Первые пять моделей оказались успешными, позволив в ходе многостадийной итеративной процедуры определить структуру LCB2. Напротив, модели с шестой по десятую, сконструированные с помощью устаревших алгоритмов, потерпели неудачу.

В процессе решения структуры методом MR следует иметь в виду так называемый эффект фазовой ошибки (phase bias). Вкратце, использование неточной MR модели может приводить к погрешностям в получаемом решении. Как правило, такие погрешности считаются малыми. Однако они могут заметным образом проявляться при недостатке опыта или невнимательности кристаллографа, излишне полагающегося при решении структуры на автоматизированные методы. В данном случае нам представляется возможность оценить эффект phase bias путём сравнения структур LCB2, полученных из различных начальных моделей. При этом потенциальные проявления phase bias дополнительно усугубляются эффектом двойникования кристалла (как показал наш анализ, двойникование кристалла ведет к небольшому, но статистически значимому снижению качества кристаллографических структур, что способствует проявлению phase bias).

С этой целью мы предприняли систематический сравнительный анализ десяти структурных моделей: пяти стартовых моделей и пяти полученных из них итоговых структур. Анализ проводился на основе метрик RMSD (среднеквадратичное отклонение координат), GDT-HA (высокоточный глобальный тест на расстояния) и LDDT (локальный тест на разницу в расстояниях) [38]. Также нами был использован сравнительный анализ с помощью метода главных компонент (principal component analysis, PCA). Как было показано, все начальные модели значительно отклоняются от итоговой структуры (до 1.13 Å) и заметно разнятся между собой (до 0.93 Å). В тоже время все пять итоговых структур находятся в отличном согласии друг с другом (от 0.13 до 0.30 Å).

Таким образом можно заключить, что компьютерные модели последнего поколения отвечают требованиям для использования в качестве MR моделей. Связанный с их использованием эффект фазовой ошибки (phase bias) носит минимальный характер. Точность получаемых структур можно оценить как 0.30 Å, что соответствует ожидаемой для структуры с кристаллографическим разрешением 2.1 Å. Отметим, что использование нескольких компьютерных MR моделей при решении кристаллографических структур позволяет более успешно уточнять отдельные элементы структуры (как, например, в нашем случае боковую цепь остатка R41).

6 Структурное исследование предполагаемого асимметричного гомодимера белка LCB3 T10Y

LCB3 представляет собой сконструированный *de novo* небольшой трёхспиральный белок, обладающий ингибирующей способностью по отношению к S-белку коронавируса SARS-CoV2 [39]. В попытке улучшить взаимодействие с мутантными вариантами белка-мишени нами был исследован вариант LCB3 с заменой треонина 10 на тирозин, LCB3 T10Y. При характеристике методом ЯМР спектроскопии мы обнаружили, что ^1H - ^{15}N HSQC спектр данного белка содержит набор пиков практически вдвое больший, чем можно ожидать из аминокислотной последовательности (106 спектральных пиков вместо ожидаемых 64). Данное наблюдение привело нас к предположению о том, что LCB3 T10Y принимает в растворе форму асимметричного гомодимера.

Для проверки этой гипотезы нами был поставлен ряд экспериментов. Во-первых, обменный (zz-exchange) ЯМР эксперимент показал, что молекула А и молекула В в составе димера находятся в состоянии обмена с константой времени ок. 1 сек. Во-вторых, измерения методом изотермической титрационной калориметрии показали, что константа диссоциации димера составляет ок. 50 мкМ. В-третьих, измерения скоростей спиновой релаксации ^{15}N подтвердили, что сигналы как от молекулы А, так и от молекулы В можно классифицировать как сигналы димера. Этот результат также был подтверждён данными диффузионного эксперимента с использованием импульсных градиентов магнитного поля.

В качестве дальнейшего шага нами были выполнены гетероядерные 3D эксперименты для спектрального отнесения данного образца и получено частичное отнесение сигналов основной цепи для молекул А и В. Ввиду значительного перекрытия спектральных сигналов получение полного отнесения на основе данных, записанных на спектрометре с рабочей частотой 500 МГц, не представлялось возможным. Это побудило нас обратиться с предложением о сотрудничестве к проф. Льюису Кэю (Университет Торонто, Канада). В его лаборатории были записаны данные на спектрометре с рабочей частотой 1 ГГц, что позволило успешно завершить отнесение спектральных сигналов.

Помимо этого, силами сотрудников нашей лаборатории белок LCB3 T10Y был успешно кристаллизован. Для одного из кристаллов были получены рентгенодифракционные данные высокого качества. В настоящее время проводится обработка этих данных с целью дальнейшего определения структуры.

Собранные к настоящему времени данные рисуют картину асимметричного гомодимера, в котором наблюдаются существенные различия между молекулами А и В.

Об этом свидетельствуют, например, заметные различия в профилях релаксации R_2 ^{15}N . Косвенно, об этом также говорят обменные параметры системы – судя по всему, скорость формирования димера k_{on} составляет ок. $10^4 \text{ с}^{-1}\text{M}^{-1}$, что заставляет предположить определённую перестройку молекул при связывании. В этой связи можно вспомнить такой класс систем, как замещённые димеры (swapped dimers) [40].

Мы планируем использовать пример LCB3 T10Y для решения методологической задачи о получении структурной модели гомодимера на основе ограниченных данных ЯМР (хим. сдвигов атомов основной цепи). С этой целью мы намерены использовать методы белкового докинга и/или методы молекулярной динамики (под управлением экспериментальных данных хим. сдвигов) для построения модели димера. Для оценки получаемых моделей мы рассчитываем использовать кристаллографическую структуру димера. Таким образом, мы задаёмся целью разработать "дешёвую" методику для определения структуры димера без спектрального отнесения боковых цепей и анализа данных NOE (nuclear Overhauser effect, ядерный эффект Оверхаузера), представляющих собой наиболее трудозатратные элементы стандартного структурного исследования ЯМР.

7 Анализ и интерпретация ЯМР спектров металлоорганических люминофоров

Наша лаборатория в течение многих лет сотрудничает с группой проф. С.П. Туника (хим. факультет СПбГУ), работающей в области химии и фотофизики люминесцентных соединений, предназначенных для применения в качестве функциональных зондов (сенсоры, метки для люминесцентной микроскопии) в биологии и экспериментальной медицине [41]. В рамках этого сотрудничества, мы приняли участие в исследованиях нового класса люминесцентных соединений, содержащих спиновую систему $\text{N}_2\text{-Pt-P-CH}_2\text{-P-Pt-N}_2$. При анализе ^{31}P ЯМР спектров этих молекул мы обратили внимание на интересную особенность этих спектров, которая будет описана ниже.

В первом приближении данные спектры представляют собой суперпозицию трёх симметричных субспектров, соответствующих трём изотопомерам по отношению к ^{195}Pt (естественное содержание 34%):

- (1) Спектр типа A_2 ($A = ^{31}\text{P}$), соответствующий изотопомерам, в которых отсутствует ^{195}Pt . Синглет, расположенный в центре спектра.
- (2) Спектр $AA'X$ ($A = ^{31}\text{P}$, $X = ^{195}\text{Pt}$) от изотопомеров с единственным атомом ^{195}Pt . Включает сателлиты синглета A_2 в центре спектра, а также сателлиты по краям спектра. Положение линий у центра спектра диктуется константами косвенного спин-спинового взаимодействия $^3J(^{31}\text{P}\text{-}^{195}\text{Pt}) \sim 100 \text{ Гц}$ и $^2J(^{31}\text{P}\text{-}^{31}\text{P}) \sim 30 \text{ Гц}$, а

положение внешних сателлитов дополнительно определяется константой $^1J(^{31}\text{P}-^{195}\text{Pt}) \sim 4000$ Гц, за счёт которой эти пики и выносятся на края спектра.

- (3) Спектр типа AA'XX' от изотопомеров с двумя атомами ^{195}Pt . Состоит из сателлитов по краям спектра (по три линии на каждом крае). Положение пиков определяется всеми тремя вышеперечисленными константами. Отметим, что частоты всех пиков в экспериментальных спектрах хорошо описываются в предположении, что $J(^{195}\text{Pt}-^{195}\text{Pt}) \sim 0$.

При анализе экспериментальных данных выяснилось, что все шесть пиков ^{31}P в субспектре AA'XX' заметно уширены по сравнению с пиками субспектров A_2 и AA'X. При этом и сами пики AA'XX' имеют разную ширину: четыре из них (по два на каждом крае спектра) гораздо шире двух остальных. Предпринятый нами поиск по базе данных CCDC обнаружил тридцать рентгеновских структур молекул, содержащих одну цепочку атомов Pt-P-CH₂-P-Pt. В трёх статьях, описывающих данные структуры, приведены ^{31}P ЯМР спектры, в которых наблюдаются схожие эффекты дифференциального уширения [42-44]. Однако природа этого явления авторами либо не обсуждалась, либо приписывалась "эффектам второго порядка", что, на наш взгляд, не проясняет природу наблюдаемого явления.

Первоначально мы предположили, что дифференциальное уширение сигнала связано с хорошо известным эффектом кросс-корреляций между диполь-дипольными взаимодействиями и/или взаимодействиями с анизотропией хим. сдвига [45]. Однако исчерпывающе полное моделирование спектра с помощью пакета для расчётов спиновой динамики Spinach [46] не позволило выявить подобного механизма. Наиболее удачными оказались попытки введения обменного процесса, который модулирует во времени величину скалярного спин-спинового взаимодействия между ядрами платины, $J(^{195}\text{Pt}-^{195}\text{Pt})$. Но и в этом случае различия между моделью и экспериментом явно превышает уровень шума. В ходе исследований выяснилось, что программа Spinach содержит ошибку в части кода, описывающего так называемую скалярную релаксацию второго рода, что не позволяет корректным образом учесть взаимодействие платины-195 с азотами-14. В настоящее время мы работаем над собственной программой для моделирования спектров системы N₂-Pt-P-CH₂-P-Pt-N₂ с учётом всех возможных механизмов спиновой релаксации.

ЗАКЛЮЧЕНИЕ

В этом заключительном разделе мы постараемся осветить значение проводимых нами исследований в более широком контексте развития соответствующих областей биофизики, структурной биологии, компьютерного моделирования и биоинформатики.

Опубликованная нами работа по молекулярной диффузии стала третьей нашей статьей по этой теме за последние три года; в следующем году мы планируем опубликовать ещё одну-две работы, посвящённых этому вопросу. Мы надеемся, что публикация серии из нескольких статей будет иметь синергический эффект, сделав их более заметными для сообщества учёных, работающих в данной области. В последней работе нам удалось показать, что метод МД моделирования может быть с успехом использован для предсказания коэффициента трансляционной диффузии разупорядоченного белка (пептида). Следует отметить, что коэффициент трансляционной диффузии не слишком чувствителен к деталям конформационного равновесия в разупорядоченных системах. Тем не менее, он является одним из немногих информативных параметров, доступных для IDP (наряду с хим. сдвигами и скоростями релаксации). В нашей работе мы показали, что сопоставление данных МД моделирования с экспериментальными результатами позволяет прийти к некоторым содержательным выводам - например, мы показали (подтвердили), что классическая модель воды TIP4P-Ew не пригодна для моделирования разупорядоченных белков. Отдельно отметим, что в литературе опубликовано не менее десятка статей, где для расчёта D_{tr} разупорядоченных белков используется программа HYDROPRO. Нам удалось показать, что этот подход порочен по своей сути, поскольку HYDROPRO предназначена для жестких молекул и не учитывает внутреннюю подвижность, свойственную IDP. Мы продемонстрировали, что применение HYDROPRO в контексте разупорядоченных белков ведёт к ошибочным выводам относительно моделей IDP. То же самое можно сказать о ряде эмпирических подходов, разработанных в лаборатории К. Линдорф-Ларсена (Университет Копенгагена, Дания).

Опубликованная нами в этом году работа по характеристике разупорядоченных "гистоновых хвостов" в составе нуклеосомной частицы на основе данных парамагнитной релаксации (PRE) также является продолжением нашей предыдущей работы в этой области. Эксперимент PRE традиционно рассматривается как эффективный инструмент структурного анализа для сложных в изучении биомолекулярных систем. Однако, как мы показываем в нашей работе, для полностью или частично разупорядоченных систем интерпретация данных PRE сильно осложняется тем обстоятельством, что структурные и

динамические параметры, определяющие скорость PRE, плохо поддаются разделению. В этом случае интерпретация может основываться на расчёте PRE из данных МД моделирования. При этом следует признать, что такого рода расчёты предъявляют крайне высокие требования к сходимости данных моделирования. Тем не менее можно предвидеть, что по мере дальнейшего совершенствования вычислительной техники и алгоритмов МД такой тип исследования будет всё более востребованным. В нашей работе удалось показать, что экспериментальные данные PRE совместимы с моделью МД, которая предсказывает, что гистоновые хвосты "конденсируются" на поверхности нуклеосомы, не утрачивая при этом свой подвижности.

В условиях, когда сходимость данных МД моделирования имеет критическое значение, разумным решением представляется объединение вычислительных ресурсов нескольких исследовательских групп с тем, чтобы увеличить совокупную длительность траекторий. Наша лаборатория взяла на себя роль организатора такого неформального "консорциума". Таким образом нам удалось привлечь к работе ряд крупнейших специалистов в области методологии МД моделирования и, конкретно, в области моделирования хроматина (проф. Онуфриев, США; проф. Панченко, Канада). В данный момент идёт запись массива траекторий нуклеосомной частицы совокупной длиной 48 мкс; по завершении первого этапа мы планируем отобрать наиболее успешные / перспективные сочетания силового поля и модели воды и использовать их для записи второго пакета траекторий аналогичной длительности. Целью данного исследования является изучение эффекта посттрансляционных модификаций на динамику гистонового хвоста и его взаимодействие с нуклеосомной ДНК. В качестве простейшей модели ПТМ используется мутация R3A на N-концевом участке гистона H4. Наши предшествующие данные позволяют предположить, что эта мутация должна существенным образом ослабить взаимодействие гистонового хвоста с ДНК и тем самым повысить конформационную подвижность хвоста. Для проверки этой гипотезы будут использованы данные МД моделирования, проводимого в рамках консорциума, а также экспериментальные данные спиновой релаксации ^{15}N , получаемые в лаборатории проф. Джароньяка (США).

В 2023 г. нами была депонирована в банк данных PDB кристаллографическая структура малого белка LCB2, которой присвоен идентификатор 8c3e. Это первая и единственная белковая структура PDB, полученная в СПбГУ (если быть точным, то в PDB можно найти ещё одну структуру с аффилиацией СПбГУ, 7q3v, которая также была депонирована нашей лабораторией, однако представляет собой результат уточнения ранее полученного набора координат). Отметим, что в базе данных также содержится две

структуры с аффилиацией МГУ, однако руководитель этой работы базируется в США. С учётом того, что банк данных PDB в настоящее время содержит 212 тыс. структур и представляет собой фундамент современной структурной биологии, это даёт представление о длине того пути, который нам предстоит пройти на этом направлении научных исследований. При решении структуры LCB2 мы использовали в качестве моделей молекулярного замещения различные компьютерные модели, полученные с помощью программ для предсказания белковых структур. Само по себе использование компьютерных моделей в качестве моделей MR не является чем-то новым. Однако мы впервые предприняли систематическое исследование с привлечением различных моделей. При этом выяснилось, что успешные модели могут быть получены не только с помощью новейших программ AlphaFold2 и RoseTTAFold, но и других программ. Мы показали, что использование множественных моделей MR помогает уточнить некоторые плохо разрешённые элементы структуры, а также оценить точность полученных координат.

С точки зрения методологии структурных исследований интерес представляет также наше исследование малого белка LCB3, несущего мутацию T10Y. Согласно полученным данным, этот белок образует асимметричный гомодимер. В настоящее время в нашем распоряжении имеется высококачественный набор данных рентгеновской дифракции для кристалла LCB3 T10Y, а также полное отнесение спектральных сигналов основной цепи для молекул А и В, образующих димер. Из имеющихся данных можно предположить, что формирование димера приводит к существенной структурной перестройке в сравнении с мономерной формой LCB3. Возможно, речь идёт о варианте, напоминающем т.н. замещённый димер (swapped dimer). В этой связи мы задаёмся вопросом: возможно ли построить модель димера, основываясь на известной структуре мономера и используя лишь ограниченный набор данных ЯМР (а именно, хим. сдвиги основной цепи)? Если нам удастся реализовать эту повестку, то это будет означать, что архитектуру димера можно воссоздать с помощью относительно "дешёвых" ЯМР экспериментов, не прибегая к трудоёмким экспериментам по отнесению сигналов боковых цепей и анализу данных NOE.

Отдельно остановимся на наших исследованиях в области предсказания константы связывания белков на основе структурных данных высокого разрешения. На первый взгляд может показаться, что эта задача представляет чисто академический интерес. Действительно, в тех случаях, когда структура комплекса известна и доступна через банк данных PDB, как правило известна также и константа связывания белков, экспериментально измеренная тем или иным способом. Однако такое заключение будет чересчур поспешным. Во-первых, легко представить себе ситуацию, когда известна

структура комплекса и константа связывания для белка дикого типа, и по этим данным требуется предсказать константу связывания для мутанта (подобная задача актуальна, например, для постоянно мутирующих вирусных белков). Во-вторых, речь может идти о компьютерной модели белок-белкового комплекса, которые за последние годы достигли высокой степени точности. В этом случае экспериментальные данные относительно константы связывания, как правило, отсутствуют. В такой ситуации возможность получить достаточно точную оценку K_d на основе структурной модели представляется достаточно важной, позволяя, например, оценить потенциальную ценность лиганда, сконструированного методами белковой инженерии. В нашей работе мы воспользовались самыми современными техническими решениями в данной области, создав алгоритм для предсказания K_d при помощи нейросети. По предварительным оценкам, реализованный нами алгоритм обладает более высокой точностью по сравнению с программой PPI Affinity, которую можно считать эталонной в данной области. Помимо этого мы также существенно упорядочили, обновили и расширили базу данных PDBBind, которая служит основой для обучения и тестирования подобных алгоритмов. В ближайшие полгода мы приложим максимальные усилия для того, чтобы опубликовать результаты данной работы.

Наконец, кратко коснёмся нашей работы по интерпретации спектров ЯМР металлоорганических люминофоров. В исследуемых спектрах обнаруживаются эффекты, которые не поддаются интерпретации с применением программы Spinach, которая считается наиболее полной и строгой программой для моделирования ЯМР спектров. В ходе работы над этой задачей мы выявили ошибку в программе Spinach и в настоящее время работаем над её устранением (в том числе посредством создания собственного кода для корректного моделирования спектров). Устранение ошибки в программе, которая считается эталонной для такого рода исследований, и выяснение природы наблюдаемых спектральных эффектов станет вкладом в инструментарий классической ЯМР спектроскопии.

В заключение отметим, что наша лаборатория опирается на достаточно обширную сеть коллабораций как в России, так и за рубежом. В материалах отчёта упоминаются совместные инициативные исследования с научными группами проф. К. Джароняка и А. Онуфриева (США), А. Панченко и Л. Кея (Канада), Й Шу (Китай), А. Шайтана и С. Туника (Россия). Руководитель проекта в этом году представил три стендовых доклада на ежегодном симпозиуме Protein Society (Бостон, США), а также выступил с приглашённым докладом на конференции Международного общества ЯМР спектроскопистов ISMAR (Брисбен, Австралия).

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Cao L., Coventry B., Goreshnik I., Huang B., Sheffler W., Park J. S., Jude K. M., Marković I., Kadam R. U., Verschueren K. H. G., Verstraete K., Walsh S. T. R., Bennett N., Phal A., Yang A., Kozodoy L., DeWitt M., Picton L., Miller L., Strauch E.-M., DeBouver N. D., Pires A., Bera A. K., Halabiya S., Hammerson B., Yang W., Bernard S., Stewart L., Wilson I. A., Ruohola-Baker H., Schlessinger J., Lee S., Savvides S. N., Garcia K. C., Baker D. Design of protein-binding proteins from the target structure alone // *Nature*. – 2022. – Т. 605, № 7910. – С. 551-560.
2. Horton N., Lewis M. Calculation of the free energy of association for protein complexes // *Protein Sci*. – 1992. – Т. 1, № 1. – С. 169-81.
3. Miller B. R., McGee T. D., Swails J. M., Homeyer N., Gohlke H., Roitberg A. E. MMPBSA.py: an efficient program for end-state free energy calculations // *Journal of Chemical Theory and Computation*. – 2012. – Т. 8, № 9. – С. 3314-3321.
4. Su M., Yang Q., Du Y., Feng G., Liu Z., Li Y., Wang R. Comparative Assessment of Scoring Functions: The CASF-2016 Update // *Journal of Chemical Information and Modeling*. – 2019. – Т. 59, № 2. – С. 895-913.
5. Lin Z., Akin H., Rao R., Hie B., Zhu Z., Lu W., Smetanin N., Verkuil R., Kabeli O., Shmueli Y., dos Santos Costa A., Fazel-Zarandi M., Sercu T., Candido S., Rives A. Evolutionary-scale prediction of atomic-level protein structure with a language model // *Science*. – 2023. – Т. 379, № 6637. – С. 1123-1130.
6. Dauparas J., Anishchenko I., Bennett N., Bai H., Ragotte R. J., Milles L. F., Wicky B. I. M., Courbet A., de Haas R. J., Bethel N., Leung P. J. Y., Huddy T. F., Pellock S., Tischer D., Chan F., Koepnick B., Nguyen H., Kang A., Sankaran B., Bera A. K., King N. P., Baker D. Robust deep learning-based protein sequence design using ProteinMPNN // *Science*. – 2022. – Т. 378, № 6615. – С. 49-56.
7. Veličković P., Cucurull G., Casanova A., Romero A., Liò P., Bengio Y. Graph Attention Networks // *arXiv [stat.ML]*. – 2018.
8. Romero-Molina S., Ruiz-Blanco Y. B., Mieres-Perez J., Harms M., Münch J., Ehrmann M., Sanchez-Garcia E. PPI-Affinity: A Web Tool for the Prediction and Optimization of Protein–Peptide and Protein–Protein Binding Affinity // *Journal of Proteome Research*. – 2022. – Т. 21, № 8. – С. 1829-1841.
9. Romero-Romero S., Lindner S., Ferruz N. Exploring the Protein Sequence Space with Global Generative Models // *arXiv [q-bio.BM]*. – 2023.

10. Shabane P. S., Izadi S., Onufriev A. V. General purpose water model can improve atomistic simulations of intrinsically disordered proteins // *Journal of Chemical Theory and Computation*. – 2019. – T. 15, № 4. – C. 2620-2634.
11. Amber 2023. / Case D. A., Aktulga H. M., Belfon K., Ben-Shalom I. Y., Berryman J. T., Brozell S. R., Cerutti D. S., Cheatham T. E., Cisneros G. A., Cruzeiro V. W. D., Darden T. A., Forouzesh N., Giambasu G., Giese T., Gilson M. K., Gohlke H., Goetz A. W., Harris J., Izadi S., Izmailov S. A., Kasavajhala K., Kaymak M. C., King E., Kovalenko A., Kurtzman T., Lee T. S., Li P., Lin C., Liu J., Luchko T., Luo R., Machado M., Man V., Manathunga M., Merz K. M., Miao Y., Mikhailovskii O., Monard G., Nguyen H., O'Hearn K. A., Onufriev A., Pan F., Pantano S., Qi R., Rahnamoun A., Roe D. R., Roitberg A., Sagui C., Schott-Verdugo S., Shajan A., Shen J., Simmerling C. L., Skrynnikov N. R., Smith J., Swails J., Walker R. C., Wang J., Wang J., Wei H., Wu X., Wu Y., Xiong Y., Xue Y., York D. M., Zhao S., Zhu Q., Kollman P. A. – San Francisco: University of California, 2023.
12. Rabdano S. O., Shannon M. D., Izmailov S. A., Gonzalez Salguero N., Zandian M., Purusottam R. N., Poirier M. G., Skrynnikov N. R., Jaroniec C. P. Histone H4 Tails in Nucleosomes: a Fuzzy Interaction with DNA // *Angewandte Chemie International Edition*. – 2021. – T. 60, № 12. – C. 6480-6487.
13. Robustelli P., Piana S., Shaw D. E. Developing a molecular dynamics force field for both folded and disordered protein states // *Proceedings of the National Academy of Sciences of the United States of America*. – 2018. – T. 115, № 21. – C. E4758-E4766.
14. Ivani I., Dans P. D., Noy A., Pérez A., Faustino I., Hospital A., Walther J., Andrio P., Goñi R., Balaceanu A., Portella G., Battistini F., Gelpí J. L., González C., Vendruscolo M., Laughton C. A., Harris S. A., Case D. A., Orozco M. Parmbsc1: a refined force field for DNA simulations // *Nature Methods*. – 2016. – T. 13, № 1. – C. 55-58.
15. Peng Y., Li S., Onufriev A., Landsman D., Panchenko A. R. Binding of regulatory proteins to nucleosomes is modulated by dynamic histone tails // *Nature Communications*. – 2021. – T. 12, № 1.
16. Sun W., Lebedenko O. O., Salguero N. G., Shannon M. D., Zandian M., Poirier M. G., Skrynnikov N. R., Jaroniec C. P. Conformational and Interaction Landscape of Histone H4 Tails in Nucleosomes Probed by Paramagnetic NMR Spectroscopy // *Journal of the American Chemical Society*. – 2023.10.1021/jacs.3c10340.
17. Ganguly D., Chen J. H. Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states // *Journal of Molecular Biology*. – 2009. – T. 390, № 3. – C. 467-477.

18. Yeh I.-C., Hummer G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions // *The Journal of Physical Chemistry B*. – 2004. – T. 108, № 40. – C. 15873-15879.
19. Bussi G., Donadio D., Parrinello M. Canonical sampling through velocity rescaling // *Journal of Chemical Physics*. – 2007. – T. 126, № 1. – C. 014101.
20. Horn H. W., Swope W. C., Pitera J. W., Madura J. D., Dick T. J., Hura G. L., Head-Gordon T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew // *Journal of Chemical Physics*. – 2004. – T. 120, № 20. – C. 9665-9678.
21. Piana S., Donchev A. G., Robustelli P., Shaw D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states // *Journal of Physical Chemistry B*. – 2015. – T. 119, № 16. – C. 5113-5123.
22. Izadi S., Anandakrishnan R., Onufriev A. V. Building water models: a different approach // *Journal of Physical Chemistry Letters*. – 2014. – T. 5, № 21. – C. 3863-3871.
23. de la Torre J. G., Huertas M. L., Carrasco B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure // *Biophysical Journal*. – 2000. – T. 78, № 2. – C. 719-730.
24. Nygaard M., Kragelund B. B., Papaleo E., Lindorff-Larsen K. An Efficient Method for Estimating the Hydrodynamic Radius of Disordered Protein Conformations // *Biophysical Journal*. – 2017. – T. 113, № 3. – C. 550-557.
25. Pesce F., Newcombe E. A., Seiffert P., Tranchant E. E., Olsen J. G., Grace C. R., Kragelund B. B., Lindorff-Larsen K. Assessment of models for calculating the hydrodynamic radius of intrinsically disordered proteins // *Biophysical Journal*. – 2023. – T. 122, № 2. – C. 310-321.
26. Tranchant E. E., Pesce F., Jacobsen N. L., Fernandes C. B., Kragelund B. B., Lindorff-Larsen K. Revisiting the use of dioxane as a reference compound for determination of the hydrodynamic radius of proteins by pulsed field gradient NMR spectroscopy // *bioRxiv* 2023.06.02.543514. – 2023.10.1101/2023.06.02.543514.
27. Jung J., Kobayashi C., Sugita Y. Optimal Temperature Evaluation in Molecular Dynamics Simulations with a Large Time Step // *Journal of Chemical Theory and Computation*. – 2019. – T. 15, № 1. – C. 84-94.
28. Lebedenko O. O., Salikov V. A., Izmailov S. A., Podkorytov I. S., Skrynnikov N. R. Using NMR diffusion data to validate MD models of disordered proteins: test case of N-terminal tail of histone H4 // *Biophysical Journal*. – 2023. – T. in press.
29. Jumper J., Evans R., Pritzel A., Green T., Figurnov M., Ronneberger O., Tunyasuvunakool K., Bates R., Žídek A., Potapenko A., Bridgland A., Meyer C., Kohl S. A. A., Ballard A. J., Cowie A., Romera-Paredes B., Nikolov S., Jain R., Adler J., Back T., Petersen S., Reiman D.,

- Clancy E., Zielinski M., Steinegger M., Pacholska M., Berghammer T., Bodenstein S., Silver D., Vinyals O., Senior A. W., Kavukcuoglu K., Kohli P., Hassabis D. Highly accurate protein structure prediction with AlphaFold // *Nature*. – 2021. – T. 596, № 7873. – C. 583-589.
30. Ovchinnikov S., Park H., Kim D. E., DiMaio F., Baker D. Protein structure prediction using Rosetta in CASP12 // *Proteins*. – 2018. – T. 86 Suppl 1, № Suppl 1. – C. 113-121.
31. McGuffin L. J., Edmunds N. S., Genc A. G., Alharbi S. M. A., Salehe Bajuna R., Adiyaman R. Prediction of protein structures, functions and interactions using the IntFOLD7, MultiFOLD and ModFOLDdock servers // *Nucleic Acids Research*. – 2023. – T. 51, № W1. – C. W274-W280.
32. Baek M., DiMaio F., Anishchenko I., Dauparas J., Ovchinnikov S., Lee G. R., Wang J., Cong Q., Kinch L. N., Schaeffer R. D., Millán C., Park H., Adams C., Glassman C. R., DeGiovanni A., Pereira J. H., Rodrigues A. V., van Dijk A. A., Ebrecht A. C., Opperman D. J., Sagmeister T., Buhlheller C., Pavkov-Keller T., Rathinaswamy M. K., Dalwadi U., Yip C. K., Burke J. E., Garcia K. C., Grishin N. V., Adams P. D., Read R. J., Baker D. Accurate prediction of protein structures and interactions using a three-track neural network // *Science*. – 2021. – T. 373, № 6557. – C. 871-876.
33. Du Z., Su H., Wang W., Ye L., Wei H., Peng Z., Anishchenko I., Baker D., Yang J. The trRosetta server for fast and accurate protein structure prediction // *Nature Protocols*. – 2021. – T. 16, № 12. – C. 5634-5651.
34. Kelley L. A., Mezulis S., Yates C. M., Wass M. N., Sternberg M. J. The Phyre2 web portal for protein modeling, prediction and analysis // *Nat Protoc*. – 2015. – T. 10, № 6. – C. 845-58.
35. Schwede T., Kopp J., Guex N., Peitsch M. C. SWISS-MODEL: An automated protein homology-modeling server // *Nucleic Acids Res*. – 2003. – T. 31, № 13. – C. 3381-5.
36. Yang J., Zhang Y. Protein Structure and Function Prediction Using I-TASSER // *Curr Protoc Bioinformatics*. – 2015. – T. 52. – C. 5.8.1-5.8.15.
37. Xu D., Zhang Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field // *Proteins*. – 2012. – T. 80, № 7. – C. 1715-35.
38. Olechnovič K., Monastyrskyy B., Kryshtafovych A., Venclovas Č. Comparative analysis of methods for evaluation of protein models against native structures // *Bioinformatics*. – 2019. – T. 35, № 6. – C. 937-944.
39. Cao L., Goreshnik I., Coventry B., Case J. B., Miller L., Kozodoy L., Chen R. E., Carter L., Walls A. C., Park Y.-J., Strauch E.-M., Stewart L., Diamond M. S., Veelsler D., Baker D. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors // *Science*. – 2020. – T. 370, № 6515. – C. 426-431.

40. Ahuja U., Rozhkova A., Glockshuber R., Thöny-Meyer L., Einsle O. Helix swapping leads to dimerization of the N-terminal domain of the c-type cytochrome maturation protein CcmH from *Escherichia coli* // *FEBS Letters*. – 2008. – T. 582, № 18. – C. 2779-2786.
41. Titov A. A., Filippov O. A., Smol'yakov A. F., Godovikov I. A., Shakirova J. R., Tunik S. P., Podkorytov I. S., Shubina E. S. Luminescent Complexes of the Trinuclear Silver(I) and Copper(I) Pyrazolates Supported with Bis(diphenylphosphino)methane // *Inorganic Chemistry*. – 2019. – T. 58, № 13. – C. 8645-8656.
42. Kui S. C. F., Chui S. S.-Y., Che C.-M., Zhu N. Structures, Photoluminescence, and Reversible Vapoluminescence Properties of Neutral Platinum(II) Complexes Containing Extended π -Conjugated Cyclometalated Ligands // *Journal of the American Chemical Society*. – 2006. – T. 128, № 25. – C. 8297-8309.
43. Gerisch M., Kirschbaum K., Bruhn C., Schmidt H., Davies J. A., Steinborn D. Synthesis and characterization of novel dinuclear and cationic cyclobutadiene platinum complexes // *Journal of Organometallic Chemistry*. – 1998. – T. 570, № 1. – C. 129-139.
44. Berti B., Cesari C., Conte F., Ciabatti I., Femoni C., Iapalucci M. C., Vacca F., Zacchini S. Synthesis of $[\text{Pt}_{12}(\text{CO})_{20}(\text{dppm})_2]^{2-}$ and $[\text{Pt}_{18}(\text{CO})_{30}(\text{dppm})_3]^{2-}$ Heteroleptic Chini-type Platinum Clusters by the Oxidative Oligomerization of $[\text{Pt}_6(\text{CO})_{12}(\text{dppm})]^{2-}$ // *Inorganic Chemistry*. – 2018. – T. 57, № 13. – C. 7578-7590.
45. Kumar A., Christy Rani Grace R., Madhu P. K. Cross-correlations in NMR // *Progress in Nuclear Magnetic Resonance Spectroscopy*. – 2000. – T. 37, № 3. – C. 191-319.
46. Hogben H. J., Krzystyniak M., Charnock G. T. P., Hore P. J., Kuprov I. Spinach – A software library for simulation of spin dynamics in large spin systems // *Journal of Magnetic Resonance*. – 2011. – T. 208, № 2. – C. 179-194.