

**Санкт-Петербургский политехнический университет  
Петра Великого  
Институт физики, нанотехнологий и телекоммуникаций**

На правах рукописи

**Колчина Нина Владимировна**

**Исследование селективности связывания дипептидов с глобулярными  
белками и днДНК**

Направление подготовки 03.06.01

---

*Код и наименование*

Направленность 03.06.01\_12

---

*Код и наименование*

**НАУЧНЫЙ ДОКЛАД**

об основных результатах научно-квалификационной работы (диссертации)

Автор работы: Колчина Нина Владимировна

Научный руководитель: доцент, в.н.с., доктор физико-математических наук,  
Петухов Михаил Геннадьевич

Санкт Петербург – 2020

Научно-квалификационная работа выполнена в ВШ/на кафедре Института физики, нанотехнологий и телекоммуникаций федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский политехнический университет Петра Великого», на базе «НИЦ Курчатовский институт» Федерального государственного бюджетного учреждения «Петербургский институт ядерной физики им. Б.П.Константинова»

Директор ВШ/Зав.каф. : *– Журихина Валентина Владимировна, доктор физико-математических наук, доцент*

Научный руководитель: *– Петухов Михаил Геннадьевич, доктор физико-математических наук, доцент*

Рецензент: *– Шабалин Константин Александрович, кандидат физико-математических наук, доцент, НИЦ «Курчатовский институт»  
– ПИЯФ, ученый секретарь  
ОМРБ*

С научным докладом можно ознакомиться в библиотеке ФГАОУ ВО «Санкт-Петербургский политехнический университет Петра Великого» и на сайте Электронной библиотеки СПбПУ по адресу: <http://elib.spbstu.ru>

## Общая характеристика работы

**Актуальность.** ДНК живых клеток в классической В-форме связывается с большим количеством разных белков, участвующих в важнейших процессах клеточного цикла – компактной упаковки и безопасного хранения генетической и эпигенетической информации, регуляции процессов репликации, транскрипции, рекомбинации и репарации ДНК. В базе данных PDB на данный момент имеется 4063 пространственных структур комплексов различных белков с ДНК. Из них только 97 структур включают в себя маленькие белки и другие короткие пептиды, имеющие длину аминокислотной последовательности менее 50 остатков. Только 4 структуры из них можно отнести к классу комплексов коротких пептидов с двунитевой ДНК (днДНК): 2EZF, 2EZD, 2EZE и 2EZG. Данные структуры являются фрагментами ДНК связывающего домена белка HMG-I(Y), связанного с участком промотора белка бета-интерферона. Центральным элементом этих 4 структур является один и тот же мотив последовательности трех аминокислотных остатков -Arg-Gly-Arg-, связывающийся в малой бороздке днДНК с последовательностью АТТС. Вероятно, существуют стабильные короткие пептиды, которые способны специфически и весьма эффективно взаимодействовать с днДНК или с белковыми молекулами. Систематическое исследование взаимодействия коротких пептидов имело бы большое значение для понимания механизмов работы живой клетки, а также для исследования возможности применения пептидов в качестве лигандов. Представляет особый интерес исследование селективности и вероятности побочного связывания для дипептидов и низкомолекулярных органических соединений при взаимодействии с днДНК и глобулярными белками. Исследование механизма и структурных особенностей пептидных и низкомолекулярных органических лигандов при взаимодействии с различными мишенями поможет комплексно изучить этот вопрос и впоследствии иметь более широкие представления о пептидных и органических лигандах для поиска более эффективных взаимодействий лиганд/мишень.

### Цели работы:

- исследование структурных закономерностей взаимодействий между дипептидами и днДНК в классической В-форме;
- исследование структурных закономерностей взаимодействий между дипептидами, низкомолекулярными органическими соединениями и глобулярными белками;

– исследование структурной организации активного центра немодифицированной рекомбинантной сульфатазы и предсказание эффективных и селективных лигандов пептидной природы.

Для достижения этих целей были поставлены следующие **задачи**:

1) Создание библиотек пространственных структур всех возможных дипептидов, состоящих из стандартных аминокислот, включая 400 вариантов со свободными N- и C-концами и 400 вариантов с нейтральными блокирующими терминальными группами, библиотек последовательностей днДНК длиной в 4 пары нуклеотидов (всего 136 вариантов, обладающих уникальной пространственной структурой) и приготовление набора комплексов глобулярных белков с нативными лигандами из базы данных PDBbind.

2) Докинг пептидных библиотек в центральные части всех возможных фрагментов днДНК и набора глобулярных белков, анализ структурных особенностей дипептидов, способствующих аффинному связыванию с мишенью и селективности взаимодействия.

3) Докинг низкомолекулярных органических соединений в нативные и ненативные белковые мишени, анализ структурных особенностей, способствующих селективному взаимодействию. Анализ вероятности побочного связывания и сравнение полученных результатов докинга пептидных и низкомолекулярных органических лигандов.

4) Построение пространственной структуры немодифицированной рекомбинантной сульфатазы. Методами молекулярного моделирования и молекулярной динамикой исследование взаимодействие активного центра с нативным лигандом и библиотекой коротких пептидов.

#### **Научная новизна работы.**

В работе мы впервые получили полную карту взаимодействий всех возможных дипептидов, состоящих из стандартных аминокислот, включая 400 вариантов со свободными N- и C-концами и 400 вариантов с блокирующими терминальными группами со всеми возможными последовательностями днДНК длиной в 4 пары нуклеотидов (всего 136 вариантов, обладающих уникальной пространственной структурой), карту взаимодействий всех возможных дипептидов и набора низкомолекулярных органических лигандов с набором глобулярных белков. Провели систематическое исследование структурных особенностей лигандов, обеспечивающих селективное взаимодействие с мишенью и высокую вероятность побочного связывания. В ходе работы впервые была построена теоретическая атомарная модель немодифицированного фермента F.p.Sulf-6His и выявлены ключевые аминокислотные остатки (C77, K327, K125, H127, R81). Произведен поиск селективных и эффективных лигандов пептидной природы для арилсульфатазы.

**Теоретическая и практическая значимость.** По результатам работы создана база данных предсказанных стабильных селективных комплексов дипептид/днДНК и дипептид/глобулярный белок, что может быть полезно при поиске селективного взаимодействия дипептидов. Предсказанные вероятности побочного связывания для всех подгрупп дипептидов с днДНК и белковыми мишенями и приведенные результаты структурного анализа дипептидов и низкомолекулярных органических соединений позволяет увидеть общую картину взаимодействия дипептид/днДНК и белок, а также использовать приведенные данные о селективности и эффективности для конструирования коротких пептидов или пептидометиков.

**Положения, выносимые на защиту.**

1. Аффинность дипептидов к днДНК зависит от аминокислотного состава дипептида и положения аминокислоты в дипептиде. За счёт увеличения главной цепи дипептиды с защитными группами на концах взаимодействуют с днДНК более селективно чем дипептиды со свободными концевыми группами.
2. Высокое значение вероятности побочного связывания дипептид/днДНК характерно для дипептидов, в составе которых есть положительно заряженные аминокислотные остатки. Селективное связывание наиболее вероятно среди групп с неполярными аминокислотами. Селективность и аффинность достигается сочетанием групп аминокислотных остатков.
3. Аффинность дипептидов к глобулярным белкам также зависит от аминокислотного состава дипептида, но не зависит от положения аминокислоты в дипептиде. Дипептиды с защитными группами на концах взаимодействуют с белками более селективно чем дипептиды со свободными концевыми группами.
4. Высокое значение вероятности побочного связывания дипептид/глобулярный белок характерно для дипептидов с заряженными остатками, селективное связывание для групп с неполярными аминокислотами.
5. Наибольший вклад в вероятность побочного связывания низкомолекулярных органических лигандов оказывает заряд и гибкость структуры при остальных равных параметрах.

**Апробация работы и публикации.** По материалам диссертации опубликовано 14 работ (в том числе 3 статьи в рецензируемых международных и отечественных журналах[1-3]).

Основные результаты работы были представлены на 44м международном конгрессе Федерации Европейского Биохимического Общества FEBS (Польша, Краков, июль 2019) [4], XXV Российском национальном конгрессе «Человек и лекарство»

(Москва, Россия, апрель 2018) [5], XIX и XXI Всероссийской Зимней молодежной школе ПИЯФ по биофизики и молекулярной биологии (Ленинградская область, Россия, февраль 2018 и 2020) [6, 7], Всероссийской конференции с международным участием XLVII «Неделя науки СПбПУ» (Санкт-Петербург, Россия, ноябрь 2018) [8], V Пушинской школе-конференции «Биохимия, физиология и биосферная роль микроорганизмов» (Пушино, Россия, декабрь 2018) [9], Международной научной конференции по биоорганической химии «XII чтения памяти академика Ю.А. Овчинникова» и VIII Российском симпозиуме «Белки и пептиды» (Москва, Россия, сентябрь 2017) [10], Международной конференции Физика.СПб (Санкт-Петербург, Россия, октябрь 2017) [11], Международной конференции "Effective current approaches in anti-aging medicine and gerontology" International Symposium of Experts (Стокгольм, Швеция, апрель 2018) [12], научно-практической конференции «Инновационные Российские технологии в геронтологии и гериатрии 2017», посвященной 25-летию Санкт-Петербургского института биорегуляции и геронтологии (Санкт-Петербург, Россия, декабрь 2017) [13], Международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов» (Москва, Россия, апрель 2017) [14].

**Личный вклад автора.** Состоял в проведении компьютерного моделирования исследуемых молекулярных систем, в обработке, анализе и интерпретации полученных результатов, а также в подготовке докладов и публикаций.

**Структура и объем диссертации.** Диссертационная работа состоит из введения, четырех глав, выводов, списка цитируемой литературы (121 наименование). Работа изложена на 131 странице, включает 38 рисунка, 19 таблиц и 2 приложения.

## **Содержание работы**

Во **введение** обоснована актуальность диссертационной работы, сформулирована цель и аргументирована научная новизна исследования, показана практическая значимость полученных результатов, представлены выносимые на защиту научные положения.

В **первой главе** подробно описан метод виртуального скрининга, который мы применяли для исследования взаимодействия дипептидов и низкомолекулярных органических соединений с днДНК и глобулярными белками. Данный подход стал использоваться чуть более 20 лет назад. На сегодняшний день это стандартный подход при поиске лекарственно подобных веществ, поскольку помогает сократить исследуемое химическое пространство соединений.

Поскольку в молекулярном моделировании используется упрощенная модель взаимодействия лиганд/мишень без учета молекулярного окружения, то такие свойства, как стабильность лиганда, возможность взаимодействия с нецелевыми мишенями, возможность прохождения через биологические барьеры к целевой мишени необходимо исследовать на начальном этапе, чтобы исключить из рассмотрения заведомо неподходящие лиганды и сократить химическое пространство для поиска. Одним из таких критериев ограничения физико-химических свойств лигандов при поиске лекарственно подобных веществ является правило Липински, также известное как правило пяти.

Также необходимо понимать, что метод виртуального скрининга имеет ограничения, они относятся к качеству кристаллических структур, проблемам жесткого докинга и к оценочной функции взаимодействия лиганд/мишень. В одной из работ было показано, что с помощью значения оценочной функции удалось предсказать только 57% взаимодействий лиганд/мишень. Вероятно, это связано с завышением или занижением значения оценочной функции для комплексов, которые схожи или наоборот, сильно отличаются от комплексов из тестового набора для параметризации оценочной функции.

Тем не менее, метод молекулярного моделирования незаменим и хорошо себя зарекомендовал при исследовании больших наборов данных лигандов, мишеней, хотя для более точных результатов необходима статистическая обработка данных, более глубокое исследование конкретных взаимодействий методами молекулярной динамики, квантово-химическими методами и, конечно, поскольку любая модель не может отразить всех свойств биологической системы, то экспериментальные анализы полученных результатов *in vitro* и *in vivo* являются самым надежным показателем взаимодействия лиганд/мишень.

В нашем исследовании рассматривался вопрос о возможности взаимодействия таких коротких пептидов, как дипептиды с днДНК и глобулярными белками, а также оценивалась разница во взаимодействии с глобулярными белками для дипептидов и низкомолекулярных органических соединений. Поднимались такие важные вопросы, как селективность и вероятность побочного связывания. Именно эти вопросы легли в основу структурного анализа дипептидов и низкомолекулярных органических лигандов для выявления закономерностей свойств селективности и вероятности побочного связывания со структурой лиганда. Поскольку в настоящее время взаимодействие коротких пептидов с днДНК недостаточно изучено в литературном обзоре была выделена отдельная глава, посвященная этому вопросу. В этой главе описаны особенности днДНК в В-форме в качестве рассматриваемой мишени, также приведены примеры низкомолекулярных органических лигандов днДНК и структура комплекса самого короткого пептида с днДНК (PDB: 2EZG). Данный пептид состоит из 10 аминокислотных остатков и не обладает

вторичной формой, но во взаимодействии с азотистыми основаниями днДНК участвуют всего 3 центральных аминокислотных остатка –Arg-Gly-Arg-. Также в литературном обзоре описаны примеры известных биологических активностей коротких пептидов, состоящих из 2-20 аминокислотных остатков.

При упоминании коротких пептидов в качестве лигандов, в научной литературе нередко упоминается низкая стабильность и высокая вероятность побочного связывания, поэтому несмотря на высокую аффинность, низкую токсичность, доступность при исследовании методом молекулярного моделирования, доступность синтеза, на практике применение коротких пептидов в качестве лигандов уступает место низкомолекулярным органическим соединениям. Тем не менее, последние данные говорят о недооцененности этого направления. Все больше информации появляется о биологических активностях коротких пептидов, появилось даже отдельное направление – пептидомика. Также изменяются представления о селективности и вероятности побочного связывания. Было показано, что в среднем лекарственный препарат может образовывать устойчивые комплексы с 329 белковыми мишенями. Высокая вероятность образования побочного комплекса, не всегда может быть критерием побочного действия лекарственных средств.

Во **второй главе** описана методическая часть и результаты исследования взаимодействия дипептидов с днДНК.

На первом этапе работы был проведен структурный анализ дипептидов для оценки получаемых результатов докинга. Все дипептиды были разделены на 5 групп и 25 подгрупп в зависимости от физико-химических свойств аминокислотных остатков в их составе (Таблица 1.).

Таблица 1. Группы аминокислот

№	Группа	Аминокислота
1	Полярные заряженные отрицательно при pH=7	D, E
2	Неполярные	G, A, V, I, L, P
3	Полярные незаряженные (заряды скомпенсированы) при pH=7	S, T, C, M, N, Q
4	Ароматические	F, W, Y
5	Полярные заряженные положительно при pH=7	K, R, H

Группы дипептидов были условно обозначены R1R2, где R1, R2 первый и второй аминокислотный остаток, соответственно, номером группы указан в таблице 1. Таким образом, в нашем исследовании рассматривались 5 групп дипептидов: 1R2 (1), 2R2 (2),



3R2 (3), 4R2 (4), 5R2 (5). Также 25 подгрупп дипептидов: 11, 12, 13, 14, 15, 21, 22, 23, 24, 25, 31, 32, 33, 34, 35, 41, 42, 43, 44, 45, 51, 52, 53, 54, 55. Подгруппы представлены разным количеством дипептидов, в зависимости от числа аминокислот в представленных группах (Таблица 1).

После процедуры виртуального скрининга 400 дипептидов со свободными N-, C-концами и с защитными группами на концах в 136 рассматриваемых участков днДНК были получены их комплексы (по 54400 для дипептидов со свободными и защитными концами). Для графической интерпретации полученных результатов была составлена полная карта взаимодействия днДНК/дипептид для пептидов со свободными группами (верхняя карта) и с защитными (нижняя) (рис. 1).

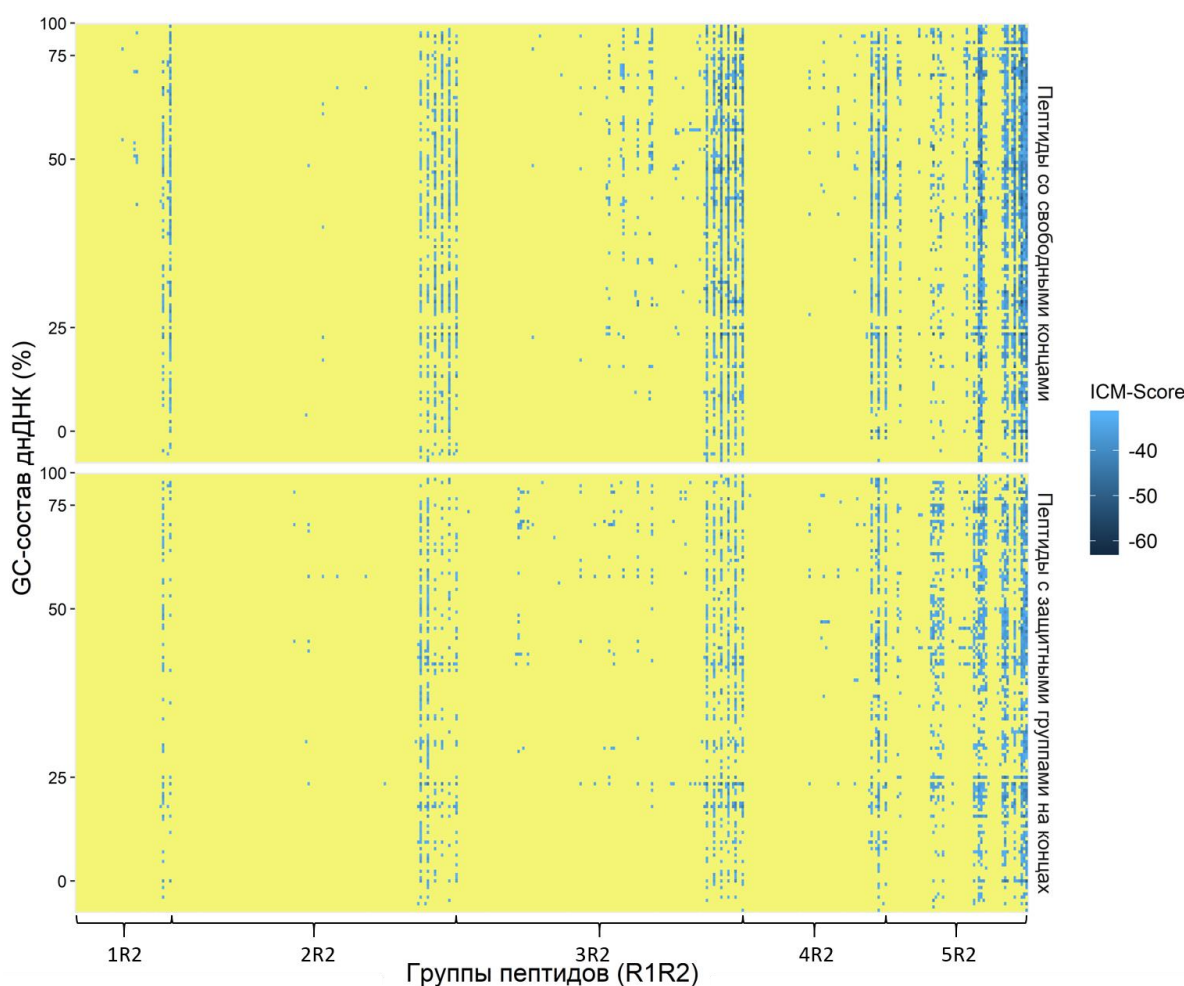


Рис. 1. Карта взаимодействия днДНК/дипептид в распределении по 25 подгруппам пептидов

По вертикальной оси на графике представлены кодовые обозначения всех подгрупп дипептидов. По горизонтальной оси представлены все исследуемые участки днДНК, отсортированные в порядке увеличения процентного отношения GC-состава. Синим

цветом обозначено значение ICM-Score: более темный цвет соответствует меньшей величине ICM-Score, более устойчивым комплексам днднк с дипептидами. Жёлтым цветом обозначены комплексы со значением  $ICM-Score > 32.0$ , в этой области возможность связывания дипептида маловероятна.

Дипептиды с защитными группами на N-, C-конце по сравнению с дипептидами со свободными концами образуют меньше устойчивых комплексов (синие полосы становятся более разряженные), что говорит о повышении селективности при увеличении длины пептида.

Для дипептидов со свободными концами статистически значимо отличаются вероятности побочного связывания групп: 5R2 (имеет наибольшую вероятность побочного связывания 20.58%), 2R2 (имеет наименьшую вероятность побочного связывания 2.58%). Для дипептидов с защитными группами на концах достоверно отличаются вероятности побочного связывания групп: 5R2 (имеет наибольшую вероятность побочного связывания 20.58%), 1R2 и 2R2 (имеют наименьшую вероятность побочного связывания 1.31% и 1.86%, соответственно).

Подгруппа дипептидов 55 характеризуется высокой вероятностью побочного связывания: для дипептидов со свободными концами – 62.3%, для дипептидов с защитными группами на концах – 41.60%.

Было проведено исследование влияния аминокислотного остатка и его положения на возможность образования устойчивого комплекса с днднк. На рисунке 2 показана диаграмма размаха значений ICM-Score в зависимости от положения аминокислот в дипептиде.

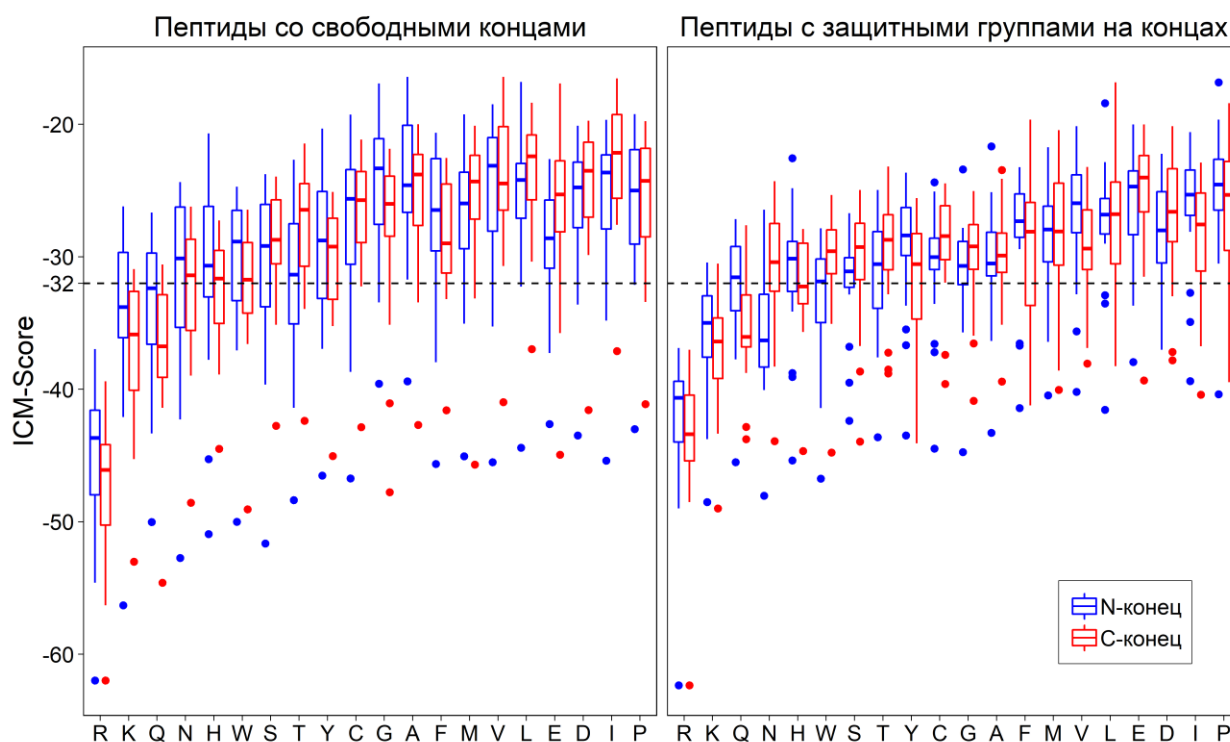


Рис. 2. Диаграмма размаха значений функции ICM-Score от положения аминокислотных остатков в дипептидах со свободными и защитными группами на N-, C-концах

В результате статистического анализа было выявлено, что аффинность дипептидов (способность образовывать стабильный комплекс,  $ICM-Score \leq -32.0$ ) к днднк достоверно зависит от аминокислотного состава дипептида и положения аминокислоты в дипептиде. Таким образом, аминокислотные остатки R, K, N, T, Q, H, W, S, C, Y вносят положительный вклад в связывание с днднк, а аминокислоты V, L, D – отрицательный вклад. Аминокислоты M, T с N-конца дипептида и R, Q, K, Y, H, W, N, F, G с C-конца оказывают положительное влияние на аффинность дипептидов к днднк.

Исследование дипептидов на селективность показало наличие среди полученных комплексов 28 и 29 селективных комплексов дипептид/днднк со свободными концами и с защитными группами на концах, соответственно. Селективными мы считали комплексы, в которых дипептид связывается только с одним участком из всех исследуемых участков днднк со значением  $ICM-Score \leq -32.0$ . Большинство селективных пептидов содержат аминокислотные остатки (в порядке уменьшения частот встречаемости): полярные незаряженные (3), ароматические (4), полярные заряженные положительно (5) и неполярные (2). Среди селективных комплексов не было обнаружено ни одного с аминокислотным остатком R (5).

В результате анализа полученных данных были предложены следующие утверждения для проверки их методом молекулярной динамики:

1. Положительно заряженные аминокислоты в дипептидах увеличивают конформационную стабильность комплексов дипептид/днДНК, при этом аминокислотный остаток R будет обеспечивать наибольшую устойчивость связывания дипептида с днДНК;
2. N- и C-концевые положения в дипептидах не являются одинаково предпочтительными для аминокислотных остатков (M, T с N-конца дипептида и R, Q, K, Y, H, W, N, F, G с C-конца);
3. Поскольку положительно заряженные дипептиды имеют высокую аффинность к днДНК, но для них характерна низкая селективность, при этом незаряженные дипептиды имеют относительно низкое сродство и высокую селективность по отношению к днДНК, то комбинация двух типов аминокислот может привести к балансу между аффинностью и селективностью к днДНК в коротких пептидах;
4. Часть предсказанных селективных комплексов является стабильными.

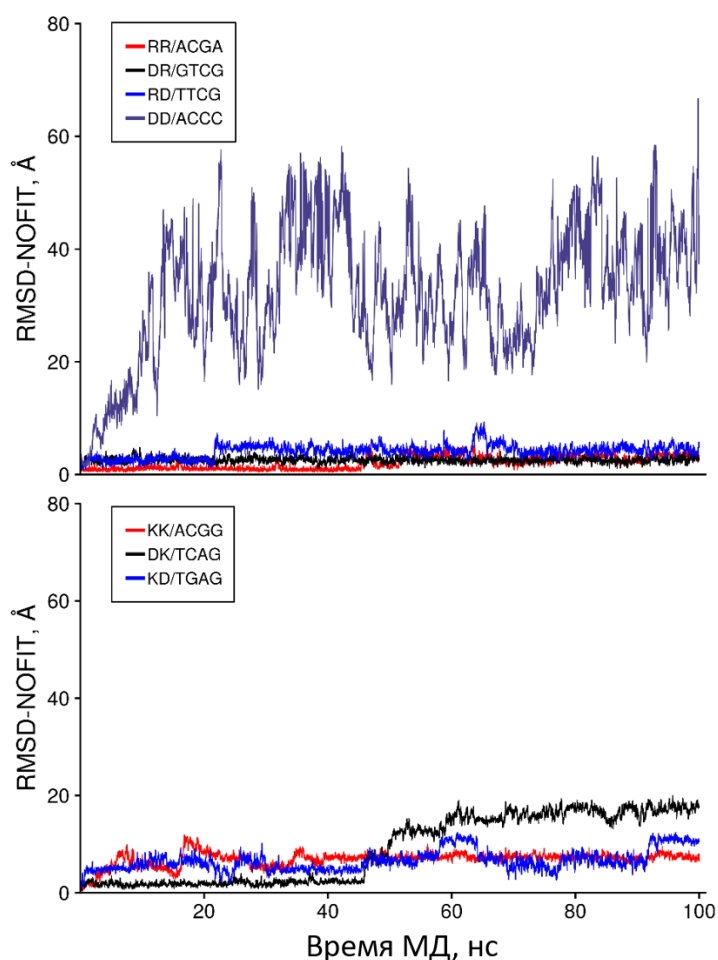


Рис.3. Среднеквадратичное отклонение (RMSD) атомов Ca, C, N, O дипептида от времени для комплексов RR/ACGA, DR/GTCG, RD/TTCG, DD/ACCC, KK/ACGG, DK/TCAG и KD/TGAG, цветовая кодировка комплексов указана в легенде

Для проверки первых двух предположений был проведен запуск молекулярной динамики для комплексов RR/ACGA, DR/GTCG, RD/TTCG, DD/ACCC, KK/ACGG, DK/TCAG и KD/TGAG (рис. 3). Как видно из представленного графика наиболее стабильное взаимодействие для комплексов RR/ACGA, DR/GTCG, RD/TTCG (дипептиды в составе которых есть R образуют более устойчивые комплексы с днДНК). При этом комплекс RD/TTCG является чуть менее стабильным, чем DR/GTCG (для R предпочтительное положение с С-конца), предпочтение положения аминокислотного остатка можно наблюдать для комплексов DK/TCAG и KD/TGAG, где комплекс с KD является менее устойчивым (для К предпочтительное положение с С-конца). Таким образом, первое и второе предположение можно считать верными.

Для проверки третьего предположения о возможности повышения стабильности селективных комплексов дипептид/днДНК с помощью добавления положительно заряженных аминокислотных остатков были сконструированы трипептиды QQR и MQR, которые в точности повторяли взаимодействие дипептидов QQ и MQ в их комплексах с днДНК, а дополнительная группа R взаимодействовала с фосфатным остовом днДНК (неселективное взаимодействие), обеспечивая стабильность связывания. Таким образом, был проведен запуск молекулярной динамики для комплексов QQ/GTCG, QQR/GTCG, MQ/GTCG, MQR/GTCG (рис. 4).

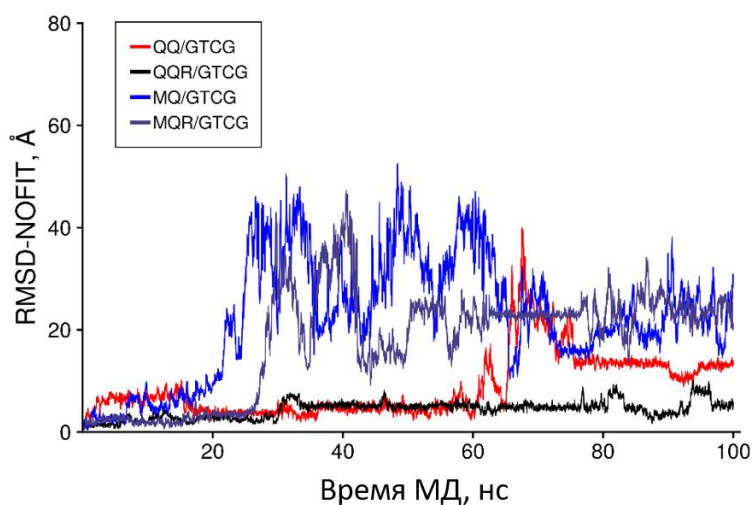


Рис.4. Среднеквадратичное отклонение (RMSD) атомов C $\alpha$ , C, N, O дипептида от времени для комплексов QQ/GTCG, QQR/GTCG, MQ/GTCG, MQR/GTCG, цветовая кодировка комплексов указана в легенде

В результате молекулярной динамики на графике 4 показано изменение траекторий дипептидов при добавлении аминокислотного остатка R на C-конец. Стабильность комплексов увеличивается, следовательно, повышение стабильности селективных комплексов с помощью добавления положительных остатков в дипептид является возможным.

Для проверки последнего предположения о стабильности предсказанных селективных комплексов были выбраны комплексы с дипептидами, которые содержат положительные аминокислотные остатки: KE/TCGA, KA/GGGG, HS/TCGA, FK/CACG, HW/TCGA (рис. 5).

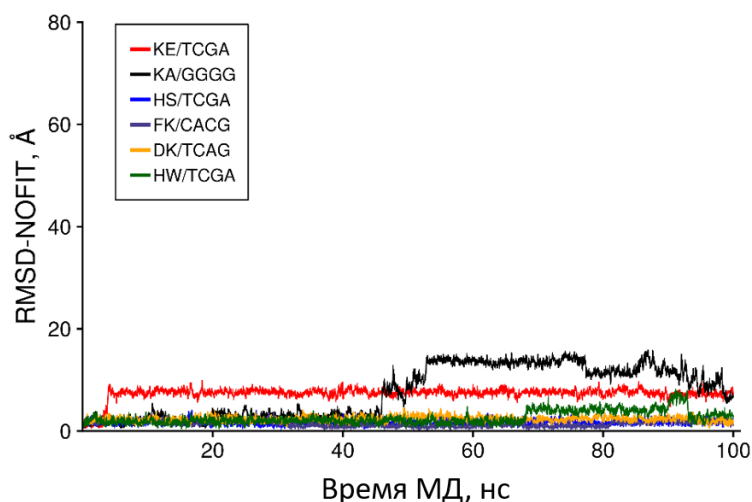


Рис.5. Среднеквадратичное отклонение (RMSD) атомов C $\alpha$ , C, N, O дипептида от времени для комплексов KE/TCGA, KA/GGGG, HS/TCGA, FK/CACG, HW/TCGA, цветовая кодировка комплексов указана в легенде

Результаты молекулярной динамики показали, что комплексы HS/TCGA, FK/CACG, HW/TCGA были стабильны в течение 100нс, комплекс KE/TCGA также был стабилен в течение этого времени, но на 5нс произошла переориентация боковых групп. Комплекс KA/GGGG был стабилен в течение 45нс молекулярной динамики.

Было проведено экспериментальное исследование связывания дипептидов с днДНК с использованием анализа сдвига электрофоретической подвижности для дипептидов DR,

КЕ, в результате которого было качественно показано взаимодействие дипептидов DR, КЕ с участком днднк.

В **третьей главе** описана методическая часть и результаты исследования взаимодействия с глобулярными белками дипептидов, а также низкомолекулярных органических лигандов.

На первом этапе работы был проведен структурный анализ набора низкомолекулярных органических лигандов для структурирования и оценки получаемых результатов докинга. Лиганды были разделены на кластеры с помощью метода полной связи. В результате было получено 22 кластера и только пять кластеров из этого набора содержали 10 и более лигандов (3, 4, 10, 17, 20). Поскольку химические структуры внутри кластера достаточно разнообразны были рассчитаны наиболее часто встречающиеся фрагменты внутри кластера и рассчитаны средние значения физико-химических характеристик по кластерам: молекулярная масса, количество циклических фрагментов, количество двугранных углов, количество доноров водородной связи, количество акцепторов водородной связи и заряд.

После процедуры виртуального скрининга 400 дипептидов и 136 низкомолекулярных органических лигандов в набор из 136 глобулярных белков были получены их комплексы (54400 для дипептидов со свободными концами и с защитными группами на концах и 18496 для низкомолекулярных органических лигандов).

Для наглядной графической интерпретации полученных результатов была составлена полная карта взаимодействия глобулярный белок/низкомолекулярный органический лиганд (рис. 6).

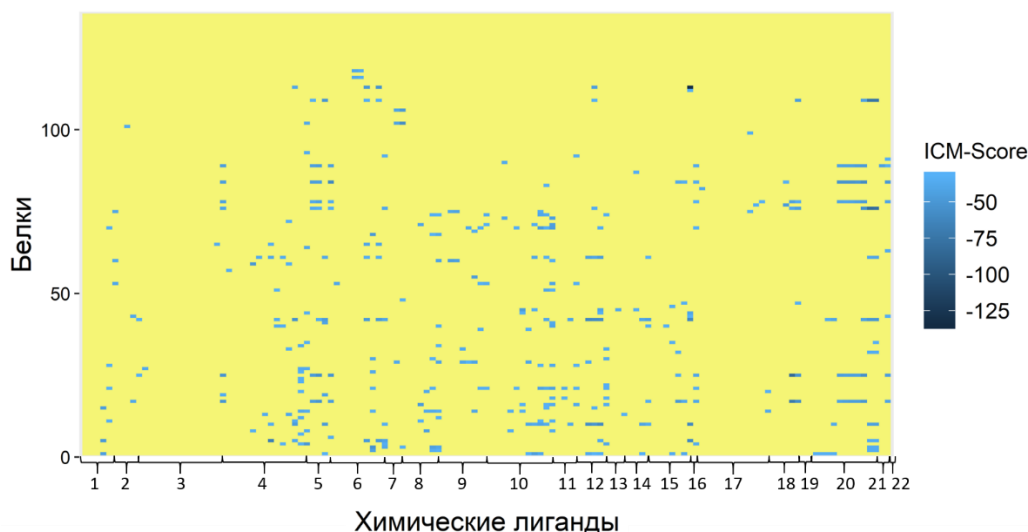


Рис. 6. Карта взаимодействия белок/низкомолекулярный органический лиганд в распределении по 22 кластерам

На рисунке 6 по вертикальной оси на графике представлены кластеры низкомолекулярных органических лигандов. По горизонтальной оси представлены все исследуемые глобулярные белки, отсортированные в порядке увеличения среднего значения ICM-Score. Синим цветом обозначено значение ICM-Score. Анализ значений ICM-Score для органических лигандов показал, что статистически значимо отличаются значения вероятностей связывания у кластеров: 20 (имеет наибольшую вероятность побочного связывания 3.37%), 3 и 17 (имеют наименьшую вероятность побочного связывания по сравнению с другими группами 0.42% и 0.31%, соответственно). Было показано, что наибольший вклад в вероятность побочного связывания органических лигандов оказывает заряд, при более отрицательном заряде повышается вероятность побочного связывания. Кроме того, структуры кластеров 3 и 17 имеют меньшее количество двугранных углов по сравнению с другими кластерами, возможно, это также ограничивает их конформационную лабильность к ненативным мишеням и приводит к более селективному связыванию.

Для наглядной графической интерпретации полученных результатов была также составлена полная карта взаимодействия глобулярный белок/дипептид для пептидов с заряженными группами (верхняя карта) и с защитными группами на концах (нижняя) (рис. 7).



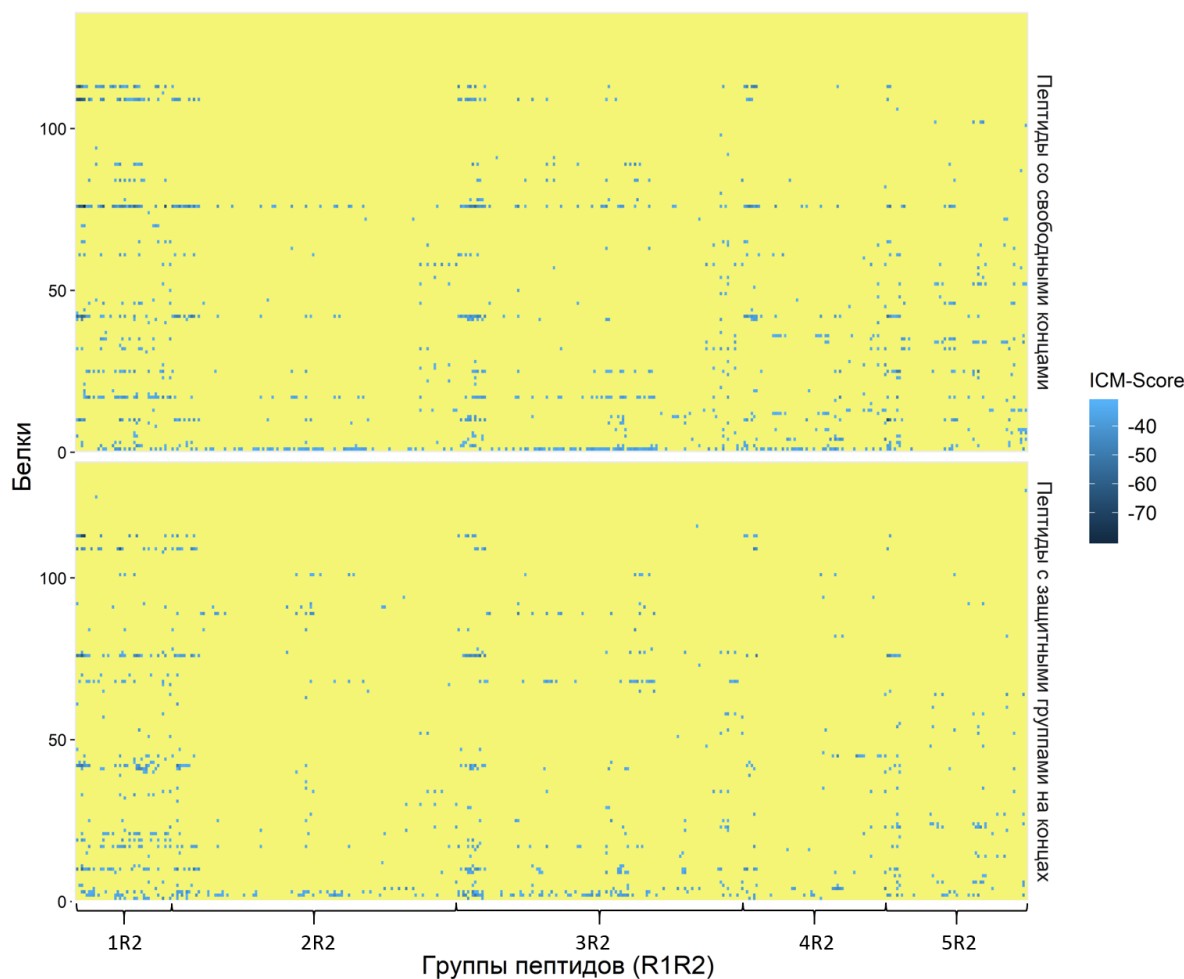


Рис. 7. Карта взаимодействия глобулярный белок/дипептид в распределении по 25 подгруппам пептидов

На рисунке 7 по вертикальной оси на графике представлены кодовые значения всех дипептидов. По горизонтальной оси представлены все исследуемые глобулярные белки, отсортированные в порядке увеличения среднего значения ICM-Score. Синим цветом обозначено значение ICM-Score. В отличие от взаимодействий дипептид/днДНК связывание с глобулярными белками происходит без выраженных аффинных взаимодействий определенных групп дипептидов и кластеров низкомолекулярных органических лигандов. Комплексы достаточно равномерно распределены по всем группам. Сравнивая взаимодействия дипептид/днДНК и дипептид/глобулярный белок, необходимо отметить, что около 31.63% из 800 дипептидов (400 дипептидов со свободными концами и 400 с защитными группами на концах) имеют возможность связывания с днДНК и 71.00% дипептидов имеют возможность связывания с белками. Дипептиды с защитными группами на N-, C-конце по сравнению с дипептидами со свободными концами проявляют большую селективность к глобулярным белкам, как и в случае взаимодействия дипептид/днДНК.

Анализ вероятности побочного связывания с глобулярными белками для всех наборов, рассматриваемых лигандов, показал, что распределения вероятности побочного связывания низкомолекулярных органических соединений не отличается от распределения в дипептидах с защитными группами на концах. Что означает, с дипептидов с защитными группами на концах короткие пептиды начинают проявлять селективность в связывании с глобулярными белками на уровне низкомолекулярных органических лигандов. Вероятно, при увеличении главной цепи дипептида селективность будет расти, однако активный центр глобулярных белков имеет ограниченный размер, таким образом, эффективность и селективность связывания будет определяться характеристиками как пептида, так и активного центра белка.

Статистически значимо отличаются значения вероятностей связывания у групп дипептидов со свободными концами: 1R2 (имеет наибольшую вероятность побочного связывания 4.54%), 2R2 (имеет наименьшую вероятность побочного связывания 1.04%). Для дипептидов с защитными группами на концах статистически значимо отличаются группы: 1R2 (имеет наибольшую вероятность побочного связывания 4.28%), 2R2 и 4R2 (имеют наименьшую вероятность побочного связывания по сравнению с другими группами 1.06% и 0.86%, соответственно).

Анализ вероятности побочного связывания показал, что для подгруппы 11 характерно самая высокая из всех подгрупп дипептидов вероятность побочного связывания: со свободными концами 7.17% и для дипептидов с защитными группами на концах 6.99%.

Было проведено исследование влияния аминокислотного остатка и его положения на возможность образования устойчивого комплекса с глобулярным белком. На рисунке 8 показана диаграмма размаха значений ICM-Score в зависимости от положения аминокислот в дипептиде.

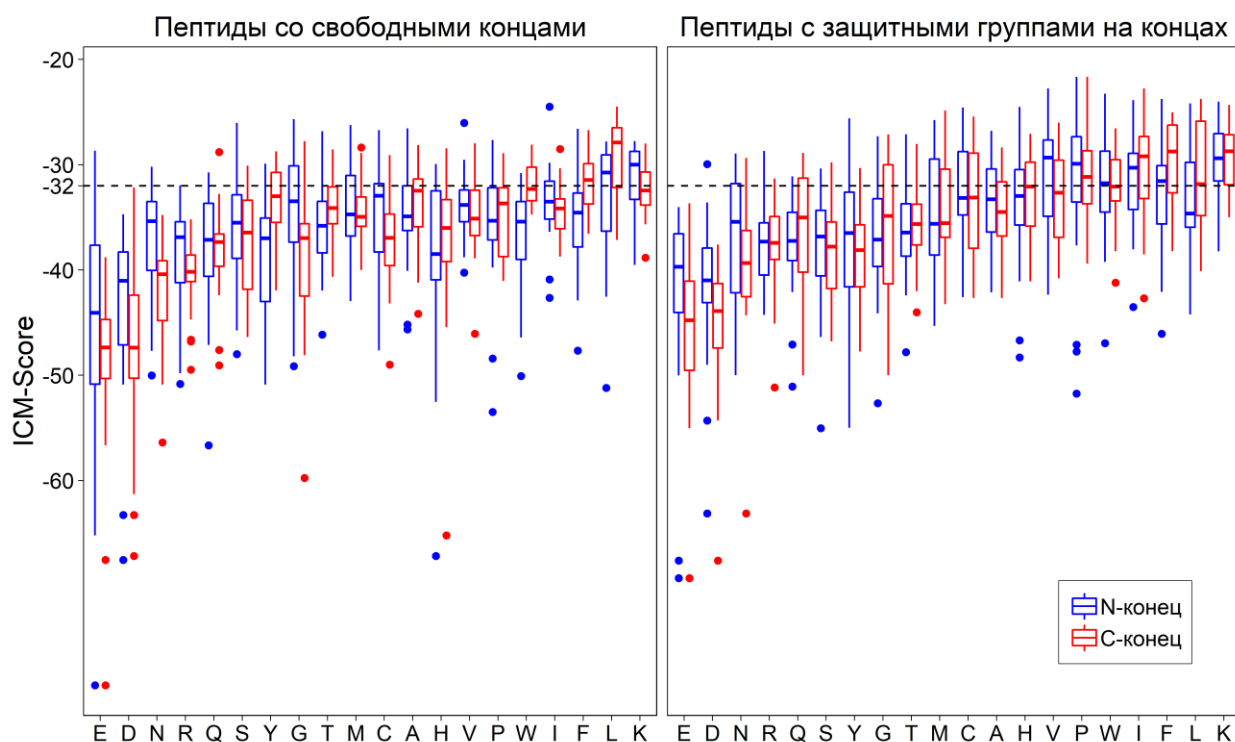


Рис. 8. Диаграмма размаха наименьших значений функции ICM-Score для каждого дипептида в зависимости от положения аминокислотных остатков в дипептидах со свободными и защитными группами на N-, C-концах

В результате регрессионного анализа было статистически выявлено, что аффинность дипептидов к белкам достоверно зависит от аминокислотного состава дипептида, но не зависит от положения аминокислоты в дипептиде. Положительный вклад во взаимодействие дипептид/глобулярный белок вносят аминокислоты E, D, R, N, Q, S, G, Y, H, T, а аминокислоты V, F, L, I, K – отрицательный вклад.

Исследование селективности взаимодействия дипептид/глобулярный белок показало 105 и 81 комплекс для дипептидов со свободными концами и с защитными группами на концах, соответственно, что значительно превышает количество селективных комплексов дипептид/днДНК. Большинство селективных комплексов глобулярный белок/дипептид содержит группы неполярные (индекс 2) и полярные незаряженные (индекс 3), что составляет около 70% от общего числа аминокислотных остатков в дипептидах селективных комплексов.

На основании результатов исследования взаимодействия дипептидов с днДНК и дипептидов с глобулярными белками, была составлена таблица со значениями вероятности побочного связывания дипептидов с изученными мишенями (Таблица 2). На

основании табличных данных можно осуществлять предварительную оценку побочного связывания для исследуемой подгруппы дипептида.

Таблица. 2. Значения вероятности побочного связывания с днДНК и глобулярными белками для подгрупп дипептидов

Подгруппы дипептидов (R1R2)	Мишень днДНК				Мишень глобулярный белок			
	Дипептиды со свободными группами на концах		Дипептиды с защитными группами на концах		Дипептиды со свободными группами на концах		Дипептиды с защитными группами на концах	
	$\mu$ Score	$\mu$ P	$\mu$ Score	$\mu$ P	$\mu$ Score	$\mu$ P	$\mu$ Score	$\mu$ P
11	-13	0.00	-13	0.00	-7.5	7.17	-6.4	6.99
12	-15	0.00	-14	0.00	-6.4	3.19	-5.3	3.13
13	-19	0.74	-17	0.00	-7.9	6.13	-8.0	4.66
14	-20	0.00	-17	0.00	-7.1	3.31	-6.2	4.29
15	-26	19.10	-23	8.70	-8.6	3.55	-8.7	4.04
21	-12	0.00	-14	0.00	-4.4	3.13	-4.9	4.11
22	-13	0.00	-15	0.00	-3.8	0.59	-1.9	0.47
23	-16	0.18	-18	0.27	-5.4	1.12	-4.6	1.08
24	-18	0.00	-19	0.04	-3.7	0.20	-2.8	0.53
25	-25	18.60	-25	11.80	-5.5	1.23	-5.2	0.69
31	-16	0.00	-16	0.06	-6.5	5.70	-6.6	4.84
32	-16	0.10	-18	0.69	-5.5	1.02	-3.6	0.96
33	-20	3.23	-21	1.29	-6.7	2.02	-5.6	1.67
34	-23	1.35	-21	0.57	-5.0	0.57	-4.2	0.74
35	-29	32.70	-27	17.70	-7.6	2.17	-6.2	1.39
41	-15	0.00	-17	0.12	-4.5	5.15	-4.9	2.94
42	-16	0.00	-18	0.00	-3.7	1.43	-1.5	0.16
43	-19	1.06	-20	0.78	-5.3	1.92	-3.7	0.94
44	-22	1.14	-20	0.57	-3.3	0.82	-1.0	0.49
45	-28	27.30	-26	18.60	-5.3	2.29	-3.5	1.06
51	-22	6.99	-20	3.55	-6.9	5.27	-6.8	4.41
52	-20	6.05	-23	9.44	-5.2	0.86	-4.1	0.69
53	-25	18.60	-25	15.30	-6.9	2.45	-5.5	1.43
54	-26	20.80	-26	19.00	-5.5	0.82	-5.6	0.82
55	-35	62.30	-31	41.60	-7.5	4.08	-5.1	1.39

$\mu$ Score – средние значения ICM-Score,  $\mu$ P – средние значения вероятности побочного связывания

В четвертой главе исследовался активный центр арилсульфатазы из мицелиального гриба *F. proliferatum* LE1 (F.p.Sulf-6His). Целью данной работы являлось исследование активного центра F.p.Sulf-6His и определение ключевых аминокислотных остатков, которые участвуют в процессе гидролиза, а также поиск эффективных и селективных лигандов среди коротких пептидов с защитными группами на концах, для повышения селективности.

В базе данных PDB отсутствует кристаллическая структура белка F.p.Sulf-6His. Таким образом, на первом этапе работы была построена модель пространственной структуры арилсульфатазы F.p.Sulf-6His по гомологии с известной кристаллической структурой сульфатазы из *S. melliloti* (SmCS) (PDB: 4UG4), являющейся близким гомологом исследуемого фермента (идентичность последовательностей 44%).

Стабильность полученной структуры исследовали методом молекулярной динамики. Релаксация сконструированной модели белка F.p.Sulf-6His в течение 150 нс показала стабильность полученной структуры белка. Среди наиболее заселённых конформаций белка, найденных методами кластеризации, была выбрана конформация с открытым входом в активный центр, объём которого был достаточен для размещения нативного субстрата pNPS (пара-нитрофенил сульфата).

На следующем этапе исследования был проведен докинг лиганда pNPS в активный центр белка F.p.Sulf-6His (рис. 9), с последующей оценкой стабильности полученного комплекса методом молекулярной динамики. Таким образом, согласно положению лиганда в активном центре арилсульфатазы F.p.Sulf-6His были предложены ключевые аминокислотные остатки (C77, K327, K125, H127, R81), участвующие в ферментативном гидролизе сульфозэфиров. Анализ гидролитической активности мутантных форм F.p.Sulf-6His показал, что данные аминокислотные остатки участвуют в процессе гидролиза сульфозэфиров и активный центр предсказан верно.

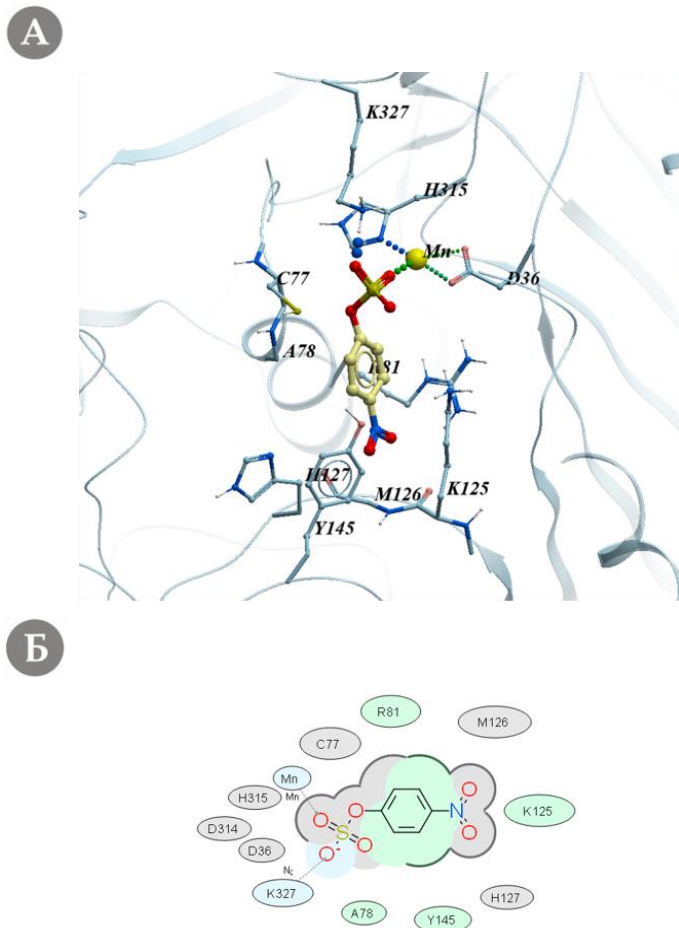


Рис.9. (А) Расположение субстрата pNPS в активном центре F.p.Sulf-6His. (Б) Диаграмма взаимодействия аминокислотных остатков F.p.Sulf-6His с pNPS. Зелёным цветом показана гидрофобная область, голубым затенением показаны акцепторы водородной связи, серыми дугами показана поверхность, доступная растворителю

После получения структуры активного центра арилсульфатазы F.p.Sulf-6His и определения ключевых аминокислотных остатков был проведен поиск ингибитора арилсульфатазы среди коротких пептидов. Для этого был проведен докинг 20 аминокислотных остатков, 400 дипептидов и 8000 трипептидов с защитными группами на N- и C-концах в активный центр белка F.p.Sulf-6His.

Анализ результатов докинга показал, что аминокислоты оказались малы для эффективного связывания, а трипептиды превышали размер активного центра исследуемого белка, при взаимодействии наблюдались стерические напряжения. Наиболее эффективным в качестве пептидного лиганда для сульфатазы оказались дипептиды с защитными группами на концах.

Основываясь на исследовании взаимодействий дипептид/белок (Таблица 2) и результатов докинга, в качестве селективных и эффективных ингибиторов

арилсульфатазы были предложены дипептиды с защитными группами на концах (вероятность побочного связывания менее 1,5%): NG, NH, PH, GN, PT, NW.

Наилучший результат показал дипептид NG, образуя комплекс дипептид NG/F.p.Sulf-6His со значением ICM-Score=-49.3 (рис. 10).

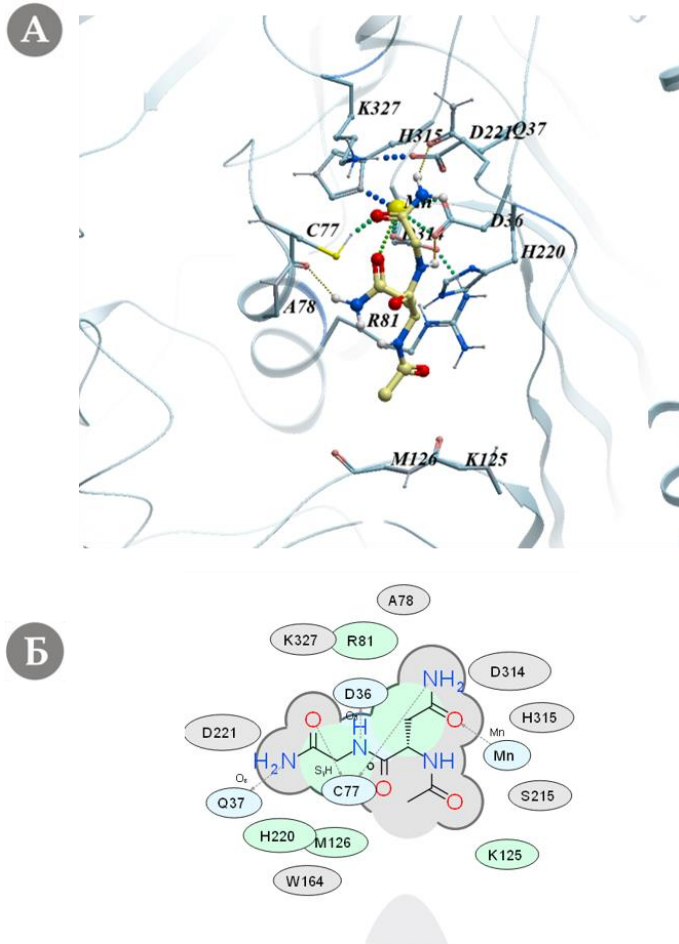


Рис.10. (А) Расположение субстрата NG в активном центре F.p.Sulf-6His. (Б) Диаграмма взаимодействия аминокислотных остатков F.p.Sulf-6His с NG. Зелёным цветом показана гидрофобная область, голубым затенением показаны акцепторы водородной связи, серыми дугами показана поверхность, доступная растворителю

Дипептид NG взаимодействует с ключевыми аминокислотными остатками активного центра сульфатазы: C77, K327, K125, R81. Дипептид NG принадлежит к 32 подгруппе дипептидов, селективность которой 0.96%.

## Заключение

1. Проведенный анализ структурных особенностей комплексов дипептид/днДНК показал, что аффинность дипептидов к днДНК зависит от аминокислотного состава дипептида и положения аминокислоты в дипептиде. Таким образом, аминокислотные остатки R, K, N, T, Q, H, W, S, C, Y вносят положительный вклад в связывание с днДНК, а аминокислоты V, L, D – отрицательный вклад. Аминокислоты M, T с N-конца дипептида и R, Q, K, Y, H, W, N, F, G с C-конца оказывают положительное влияние на аффинность дипептидов к днДНК. Также было подтверждено, что N- и C-концевые защитные группы способствуют к образованию устойчивого комплекса с днДНК.
2. При взаимодействии дипептид/днДНК было показано, что вероятность побочного связывания характерно для дипептидов, в составе которых есть положительно заряженные аминокислотные остатки, особенно R. Селективное связывание наиболее вероятно среди групп с неполярными аминокислотами, в случае с дипептидов с защитными группами на концах селективное связывание возможно также у дипептидов с отрицательно заряженными группами.
3. Структурный анализ особенностей дипептидов, показал, что аффинность дипептидов к белкам достоверно зависит от аминокислотного состава дипептида, но не зависит от положения аминокислоты в дипептиде. Положительный вклад во взаимодействие дипептид/глобулярный белок вносят аминокислоты E, D, R, N, Q, S, G, Y, H, T, а аминокислоты V, F, L, I, K – отрицательный вклад. Также был подтвержден отрицательный вклад N-, C-концевых защитных групп на способность образования устойчивого комплекса с глобулярными белками.
4. При взаимодействии дипептид/глобулярный белок было показано, что вероятность побочного связывания характерно для дипептидов с отрицательными остатками, селективное связывание для групп с неполярными аминокислотами. Для дипептидов с защитными группами на концах также селективное связывание возможно для групп с ароматическими аминокислотами.
5. Анализ низкомолекулярных органических лигандов показал, что вероятность побочного связывания у отрицательно заряженных низкомолекулярных соединений имеет наибольшее значение среди органических лигандов, а жесткие структуры со средней молекулярной массой, заряженные положительно имеют наименьшую вероятность побочного связывания по сравнению с другими группами.



## Благодарности

Я хочу поблагодарить моего научного руководителя Михаила Геннадьевича Петухова за предоставленную возможность работы над интересным проектом, а также за проявленное терпение и понимание. А также Георгия Николаевича Рычкова, Арину Сергеевну Афанасьеву, Александра Павловича Якимова и Андрея Владимировича Илатовского за поддержку и оказание неоценимой помощи вначале моей исследовательской деятельности. Анну Николаевну Кульминскую и Кирилла Сергеевича Боброва за возможность применить вычислительные методы в проекте по арилсульфатазе. Владимира Хацкелевича Хавинсона и Наталью Сергеевну Линькову за проявленный интерес к нашей работе. Сотрудников кафедры Биофизики и сотрудников Лаборатории биофизики макромолекул ОМРБ ПИЯФ за поддержку и интерес к проекту.

## Список публикаций по теме диссертации

1. Кузник Б.И., Линькова Н.С., Колчина Н.В., Куканова Е.О., Хавинсон ВХ. Семейство молекул JAM и их роль в регуляции физиологических и патологических процессов // Успехи физиологических наук. 2016;47(4):76-97.
2. Kolchina N., Khavinson V., Linkova N., Yakimov A., Baitin D., Afanasyeva A., et al. Systematic search for structural motifs of peptide binding to double-stranded DNA // Nucleic Acids Res. 2019;47(20):10553-63.
3. Колчина Н.В., Рычков Г.Н., Кульминская А.А., Ибатуллин Ф.М., Петухов М.Г., Бобров К.С. Исследование структурной организации активного центра немодифицированной рекомбинантной сульфатазы из мицелиального гриба *Fusarium proliferatum* LE1 // Биоорганическая химия. 2020;46(4):425-34.
4. Kolchina N., Afanasyeva A., Linkova N., Khavinson V., Petukhov M. Investigation of dipeptides binding to doublestranded DNA using molecular modeling and docking // FEBS Open Bio. 2019;9:215.
5. Колчина Н.В., Линькова Н.С., Хавинсон В.Х., Петухов М.Г. Конструирование коротких пептидов для связывания с двунитевой ДНК в классической В-форме // Всероссийский 13 Симпозиум "Биоинформатика и компьютерное конструирование лекарств". 2018:73.
6. Колчина Н.В., Линькова Н.С., Хавинсон В.Х., Петухов М.Г. Систематический поиск структурных мотивов связывания дипептидов с днДНК // Всероссийская Зимняя молодежная школа ПИЯФ по биофизике и молекулярной биологии. 2018:137.

7. Колчина Н.В., Якимов А.П., Петухов М.Г.. Разработка нового подхода конструирования лигандов белков с помощью модифицированного метода молекулярных фрагментов // Всероссийская Зимняя молодежная школа ПИЯФ по биофизике и молекулярной биологии. 2020:109-10.
8. Колчина Н.В., Бобров К.С., Петухов М.Г., Рычков Г.Н., Кульминская А.А. Исследование структуры активного центра немодифицированной рекомбинантной сульфатазы из мицелиального гриба *Fusarium proliferatum* LE1 // Всероссийская конференция с международным участием XLVII «Неделя науки СПбПУ». 2018:235.
9. Бобров К.С., Колчина Н.В., Петухов М.Г., Рычков Г.Н., Кульминская А.А. Структурно-функциональные исследования рекомбинантной сульфатазы из *Fusarium proliferatum* LE1 // V Пушкинская школа-конференция «Биохимия, физиология и биосферная роль микроорганизмов». 2018:11-3.
10. Колчина Н.В., Афанасьева А.С., Салль Т.С., Линькова Н.С., Петухов М.Г., Хавинсон В.Х. Анализ взаимодействия дипептидов с двунитевой ДНК. // Международная научная конференция по биоорганической химии «XII чтения памяти академика Ю.А. Овчинникова» и VIII Российский симпозиум «Белки и пептиды» 2017:61.
11. Колчина Н.В., Петухов М.Г. Поиск оптимальных пар входных-выходных атомов при конструировании лекарств с помощью метода фрагментов в активных центрах белков // Международная конференция ФизикАСПб. 2017:42.
12. Kolchina N.V., Linkova N.S., Petukhov M.G. Systematic search for structural motifs of short peptides binding to double-stranded DNA // "Effective current approaches in anti-aging medicine and gerontology" International Symposium of Experts. 2018:53.
13. Колчина Н.В., Петухов М.Г. Анализ способов связывания дипептидов с двунитевой ДНК в классической В-форме // Успехи геронтологии. 2017;6(30):33.
14. Колчина Н.В., Афанасьева А.С. Вычислительный метод поиска термочувствительных участков структуры глобулярных белков // Ломоносов. 2017.