# МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ имени М.В. ЛОМОНОСОВА

На правах рукописи

# Залевский Артур Олегович

# Атомистический механизм катион-зависимой активации тромбина

03.01.09 - Математическая биология, биоинформатика

ΑΒΤΟΡΕΦΕΡΑΤ

Диссертации на соискание ученой степени кандидата биологических наук

Москва – 2019

Работа выполнена на Факультете биоинженерии и биоинформатики Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В. Ломоносова»

Научный руководитель:	<b>Головин Андрей Викторович</b> доктор химических наук
<u>Официальные оппоненты:</u>	Спиридонова Вера Алексеевна доктор биологических наук, доцент, старший научный сотрудник отдела хроматографического анализа Научно-исследовательского института физико-химической биологии имени А.Н. Белозерского Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В. Ломоносова»
	Веселовский Александр Владимирович доктор биологических наук, заведующий лабораторией структурной биоинформатики отдела биоинформатики Федерального государственного бюджетного научного учреждения «Научно- исследовательский институт биомедицинской химии имени В.Н. Ореховича»
	<b>Миронов Андрей Александрович</b> доктор биологических наук, профессор Факультета биоинженерии и биоинформатики Федерального государственного бюджетного образовательного

биоинженерии и биоинформатики Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В. Ломоносова»

Защита состоится 24 декабря 2019 года в 15:00 на заседании диссертационного совета МГУ.03.04 Московского государственного университета имени М.В. Ломоносова по адресу: 119234, Москва, Ленинские горы, д. 1, стр. 73, Факультет биоинженерии и биоинформатики, ауд. 406. E-mail: dissovet@belozersky.msu.ru

С диссертацией можно ознакомиться в Научной библиотеке МГУ имени М.В. Ломоносова (Москва, Ломоносовский просп., д. 27, отдел диссертаций) и на сайте ИАС «ИСТИНА»: <u>https://istina.msu.ru/dissertations/253152327/</u> Автореферат разослан «\_\_» ноября 2019 г.

Ученый секретарь диссертационного совета, кандидат химических наук

M. Manof

И.В. Шаповалова

### СПИСОК СОКРАЩЕНИЙ

- АПФ ангиотензинпревращающий фермент
- ГАМКБ рецептор ү-аминомасляной кислоты класса Б
- КМ квантово-механические расчеты
- **КМ/ММ** гибридное квантово-механическое/молекулярно-механическое моделирование
- МД молекулярная динамика
- ПО программное обеспечение
- РСА рентгено-структурный анализ
- **ЯМР** ядерно-магнитный резонанс
- **API** программный интерфейс приложения (англ. application programming interface)
- DFT теория фукнционала плотности (англ. density functional theory)
- **МНР** молекулярный гидрофобный потенциал (англ. molecular hydrophobicity potential)
- **PDB** Protein Data Bank, банк данных пространственных структур биомолекул
- **RMSD** среднеквадратичное отклонение позиций атомов (англ. rootmean-square deviation of atomic positions)
- **MPI** интерфейс передачи сообщений между программами (англ. message passing interface)

#### введение

#### 1. Актуальность

Тромбин — ключевой фермент каскада свертывания крови, что делает его привлекательной фармакологической мишенью как для академических исследователей, так и для представителей фармацевтической индустрии. На сегодняшний день накоплен поражающий воображение объем биохимических, генетических и структурных данных, демонстрирующий связь тромбина не только с гемостазом, но и с регуляцией воспаления и регенерации. Все это делает тромбин, упомянутый более чем в 50 000 статей в базе данных PubMed, одним из самых исследованных ферментов. Что, тем не менее, не означает, что мы во всех деталях понимаем все аспекты его работы. Одним из таких белых пятен является механизм катион-зависимой активации тромбина. В отличие от большинства представителей семейства химотрипсиновых протеаз, активность тромбина зависит от присутствия ионов натрия — Na<sup>+</sup>. Причем появление катиона в натрий-связывающем сайте затрагивает как кинетику фермента, увеличивая скорость реакции расщепления белковых и пептидных субстратов почти на порядок, так и субстратную специфичность. Более того, из-за весьма умеренной степени сродства белка к натрию, при физиологических условиях только 60% белка находится в активной "быстрой" форме. Остальные же 40%, как считается, находятся в "медленной" форме, проявляющей антикоагулянтную активность. Таким образом, детальное понимание особенностей обеих форм и механизмов переключения между ними может быть использовано при создании антикоагулянтных препаратов нового поколения.

#### 2. Цели и задачи

Целью данной работы является выявление атомистических деталей механизма влияния катиона натрия на ферментативную активность тромбина.

Для достижения обозначенной цели необходимо решить следующие задачи:

- Проанализировать структуры человеческого тромбина, доступные в банке структур биомолекул PDB, и выделить структурный признак, характерный для быстрой и медленных форм фермента;
- С помощью методов молекулярного моделирования продемонстрировать влияние катиона натрия на поведение обнаруженного признака;
- Продемонстрировать, что переход между формами также вызывает изменение субстратной специфичности фермента.

#### 3. Научная новизна

В данной работе впервые произведен автоматический кластерный анализ всех структур человеческого тромбина, доступных в банке данных PDB. Полученный результат позволил сформулировать гипотезу о наличии ранее не наблюдавшейся водородной связи в активном центре фермента. Для проверки данной гипотезы были специально разработаны инструменты, использующие различные методы молекулярного моделирования: от мультиуровнего гибридного квантово-механического/молекулярномеханического, до пептидного докинга. Все инструменты исходно разрабатывались из соображений оптимизации для работы в массивнопараллельном суперкомпьютерном окружении.

С помощью разработанных инструментов удалось продемонстрировать наличие ингибирующей водородной связи Ser195 OG — Ser214 O в активном центре фермента, отсутствующей в системе с катионом натрия. Продемонстрирована роль сети молекул воды, соединенных водородными связями, в передаче структурирующего сигнала от натрий-связывающего кармана к активному центру фермента.

Наконец, с помощью моделирования молекулярного докинга комбинаторных пептидных библиотек, продемонстрировано, что появление ингибирующей водородной связи также кардинальным образом изменяет профиль специфичности фермента.

Таким образом, совокупность наблюдений, сделанных в ходе вычислительных экспериментов, позволяет расширить представление о медленной форме тромбина, добавив к нему новое состояние, обладающее характерной ингибирующей водородной связью.

#### 4. Практическая значимость

Данные о новой медленной форме тромбина могут быть использованы при разработке про- или антикоагулянтных препаратов, действие которых будет основано на целенаправленном воздействии на ингибирующую водородную связь. Такие молекулы должны будут избирательно связываться с медленной формой тромбина и фиксировать в ней белок, либо через аллостерические взаимодействия нарушать сеть водородных связей, между натрийсвязывающим сайтом и активным центром.

#### 5. Положения, выносимые на защиту

1. Новая реализация непараметрического алгоритма кластеризации Affinity Propagation, оптимизированная для структур биомолекул, позволяет выявлять биологически релевантные кластеры.

2. Сравнение кластеров, полученных из 400 записей тромбина в банке структур PDB, выявило возможность существования ранее не описанной ингибирующей водородной связи между О<sub>γ</sub>каталитического Ser195 и остовным кислородом Ser214.

3. С помощью нового пакета для гибридного КМ/ММ моделирования продемонстрировано влияние наличия иона Na<sup>+</sup> на образование ингибирующей водородной связи. Дополнительный энергетический барьер для переключения ингибирующей водородной связи в присутствии пептидного субстрата составляет ~0.8 ккал/моль, что соответствует экспериментальным данным о замедлении реакции в 3-10 раз.

4. Предложенный метод ранжирования аффинности библиотек пептидов к белковым мишеням позволяет идентифицировать новые биоактивные пептиды.

5. Образование ингибирующей водородной связи изменяет субстратную специфичность тромбина.

#### 6. Личный вклад автора

Личный вклад автора заключается в разработке инструментов для биоинформатического анализа пространственных структур и гибридного квантово-механического/молекулярно-механического моделирования тромбина, биологической интерпретации полученных результатов, представлении результатов на научных конференциях, публикации в рецензируемых научных журналах.

#### 7. Степень разработанности темы

С момента открытия феномена зависимости активности тромбина от концентрации катионов натрия, в этой области было напечатано большое экспериментальных и теоретических работ, получено много качественных структурных данных. Наибольший вклад в описание особенностей отличия быстрой и медленной форм тромбина внесли профессора Е. Di Cera и J. Huntington. Большое количество кристаллографических данных получено группами V. De Filippis, E. Di Cera, J. Huntington, G. Klebe.

В область разработки быстрых квантово-химических методов большой вклад внесли профессора J. Stewart и S. Grimme.

#### 8. Степень достоверности данных

Данные, представленные в работе, получены с использованием современных вычислительных методик. Результаты воспроизводимы, исходные коды разработанных инструментов опубликованы. Обзор литературы и обсуждение подготовлены с использованием актуальной литературы.

#### 9. Публикации по теме диссертации

По материалам диссертации опубликовано 5 статей в рецензируемых журналах, индексируемых в наукометрических базах данных Web of Science или Scopus.

#### 10. Апробация результатов

Результаты диссертационной работы были представлены на международных конференциях: 12th INTERNATIONAL CONFERENCE "BIOCATALYSIS. FUNDAMENTALS & APPLICATIONS" (Россия, теплоход Леонид Соболев, 2019), Russian Supercomputing Days (Россия, Москва, 2018), Информационные технологии и системы (ИТиС-18) (Россия, Казань, 2018), Bioinformatics of Genome Regulation and Structure/BGRS-2016 (Россия, Новосибирск, 2016).

#### 11. Структура диссертации

Диссертационная работа состоит из введения, обзора литературы, методов и материалов исследования, полученных результатов и их обсуждения, выводов и списка цитируемой литературы. Работа изложена на 125 страницах, иллюстрирована 32 рисунками и 5 таблицами. Список цитируемой литературы включает 145 наименований.

#### РЕЗУЛЬТАТЫ

#### 1. Кластерный анализ структур тромбина

В результате обработки массива из 400 записей человеческого тромбина, доступных в банке PDB (на 01.05.2019 г.), было получено 462 уникальных структур. Применение процедуры непараметрической кластеризации Affinity Propagation к этому набору структур приводит к выделению 20 кластеров (Таблица 1). При множестве преимуществ, которыми обладает алгоритм Affinity Propagation, нельзя не отметить, что он обладает и одним недостатком — из результатов кластеризации сложно оценить взаимоотношения полученных кластеров.

В частности, расположение центров друг относительно друга. Для визуализации данной картины мы дополнительно применили метод полной связи (дальнего соседа) иерархической кластеризации. На полученной дендрограмме (Рис. 1) можно увидеть как отчетливо выделяются группы структур протромбина, а также медленных Е и Е<sup>\*</sup> форм. При этом большая группа структур, предположительно в E:Na<sup>+</sup> форме, соединена относительно короткими ребрами. Мы рассмотрели возможность объединения данных структур в один кластер. Объединение структур вплоть до ветки 1BCU\_HL существенно улучшало распределение силуэтов — метрики кластеризации, отражающей качество соответствия элемента кластеру (чем ближе к 1, тем лучше). Итоговое распределение силуэтов приведено на (Рис. 1).

PDB ID центра кластера	Разрешение (Å)	Размер кластера (шт)	Предполагаемая форма			
5JDU_CD 1NU9_A 5EDM_A	1.7 2.2 2.2	1 2 1	Е Претромбин-2 Протромбин			
Продолжение на следующей странице						

Таблица 1. Описание кластеров структур тромбина

PDB ID		Разрешение	Размер	Предполагаемая
центра	кла-	(Å)	кластера	форма
стера			(шт)	
3NXP_A		2.2	4	Претромбин-1
4NZQ_A		2.81	2	Протромбин
6C2W_A		4.12	3	Протромбин
2AFQ_CD		1.93	4	E
3BEI_AB		1.55	5	EE*
3K65_AB		1.85	7	Протромбин-2
4CH8_AB		1.75	46	?
3VXE_HL		1.25	6	?
1NT1_A		2.0	12	E:Na <sup>+</sup>
1A4W_HL		1.8	49	E:Na <sup>+</sup>
1BB0_AB		2.1	70	E:Na <sup>+</sup>
3EQ0_HL		1.53	55	E:Na <sup>+</sup>
5AFZ_HL		1.53	78	E:Na <sup>+</sup>
4YES_AB		1.5	39	E:Na <sup>+</sup>
5AFY_HL		1.12	45	E:Na <sup>+</sup>
2CN0_HL		1.3	20	E:Na <sup>+</sup>
1BCU_HL		2.0	7	E:Na <sup>+</sup>

Таблица 1 – Продолжение

Знаком ? отмечены структуры, которые, предположительно, могут также принадлежать к медленной Е форме.

Для оценки устойчивости результатов к небольшим изменениям в матрице расстояний, аналогичная процедура была повторена для структур, препроцессированных при помощи программы Reduce из пакета MolProbity [Chen et al., 2010]. Данная утилита использует набор геометрических критериев для оптимизации положения радикалов аминокислот, содержащих амидную связь: глутамин и аспарагин, а также гистидина. В силу особенностей метода PCA, нередко невозможно достоверно определить их ориентацию, однако, по косвенным данным, таким как наличие донор/акцепторов, заряженных групп и т.д. в окружении, можно сделать разумную догадку о корректной ориентации радикала. Таким образом, не внося существенных изменений в общую структуру, обработка при помощи программы Reduce внесет небольшой числовой шум в матрицу расстояний (медиана изменилась с -1.196 до -1.189). При этом результаты кластеризации были абсолютно идентичны полученным ранее.

Также мы использовали еще один вариант предобработки структур — проект PDB Redo использует стандартизированный протокол создания про-



Рисунок 1. Иерархическая кластеризация центров кластеров (слева), полученных методом непараметрической кластеризации Affinity Propagation. Структуры, отмеченные зеленым, являются протромбином. Отмеченные красным аннотированы как быстрая E:Na<sup>+</sup> форма, синим медленные E и E<sup>\*</sup> формы. Знаком ? отмечены структуры, которые, предположительно, могут также принадлежать к медленной E форме. Объединение (справа) E:Na<sup>+</sup> структур в один кластер улучшает метрику кластеризации.

странственных структур на основе данных электронной плотности, размещенных в исходном банке PDB [Joosten et al., 2009]. К сожалению, далеко не все исследователи размещают в PDB исходные данные, поэтому в данном случае в выборке оказалась только 301 уникальная структура (почти на треть меньше, чем в исходной). При этом медиана матрицы расстояний практически не изменилась (-1.191 против -1.196). Сохранились и ключевые особенности кластеризации: состав протромбиновых кластеров, состав кластеров E/E<sup>\*</sup>, наличие пограничных кластеров 4СН8 и 3EQ0.

Таким образом, можно заключить, что результаты кластеризации устойчивы к небольшим изменениям в структурах, индуцированных вращением радикалов или же полной реоптимизацией структуры. При этом, полученным кластерам можно сопоставить биологически релевантные метки: тип фермента (процессированный или профермент) или же форму. Последнее обстоятельство позволяет, во-первых, заключить, что вся процедура имеет биологический смысл, а во-вторых, позволяет предположить, что анализ промежуточных кластеров может привести к получению новой информации о структуре медленной Е формы.

#### 1.1. Анализ и аннотация кластеров

В рамках темы данной работы наибольший интерес представляют промежуточные кластеры 3VXE\_HL и 4CH8\_AB, расположенные между объединенным кластером структуры в активной форме E:Na<sup>+</sup> и кластерами в E/E<sup>\*</sup> форме. Для детального сравнения структур мы также выбрали референсную модель — это такая структура из всего набора, что сумма расстояний от нее до остальных является минимальной. Референсной моделью оказалась 5AFZ\_HL, обладающая разрешением 1.53 Å, что является пусть и не лучшим показателем в выборке, однако все равно достаточно высоко и входит в топ-50 структур с лучшим разрешением. Также данная структура представляет собой один из 20 исходных центров кластеров и принадлежит подмножеству активной E:Na<sup>+</sup> формы (Puc. 1).

Можно предположить, что наиболее вероятным местом наличия небольших, но критичных для работы фермента, изменений будет окружение активного центра. Тем более, что мажорные изменения, характеризующиеся, например, разрушением геометрии Na<sup>+</sup>-связывающего центра, как следует из результатов кластеризации, выделяются в отдельные хорошо различимые кластеры на границе со структурами профермента.



Рисунок 2. Каталитическая триада (Ser195, His57, Asp102) и соединяющие водородные связи в активном центре референсной структуры 5AFZ\_HL (слева). Наложение (справа) референсной структуры 5AFZ\_HL (серый), структуры в медленной Е форме 3BEI\_AB (голубой) и структуры из промежуточного кластера 3VXE\_HL (фиолетовый). Атомы азота, кислорода и серы окрашены в синий, красный и желтый соответственно.

На (Рис. 2) приведена геометрия активного центра быстрой E:Na<sup>+</sup> формы из структуры 5AFZ\_HL. Ключевым элементом является каталитическая триада Ser195, His57, Asp102, остатки которой связаны цепочкой водородных связей.

При сравнении референсной структуры 5AFZ\_HL, структуры в медленной E форме 3BEI\_AB и структуры из промежуточного кластера 3VXE\_HL (Puc. 2), в глаза бросается разница в положении остовного кислорода остатка Ser214. Причем наиболее выраженное смещение в наблюдается структуре 3BEI\_AB. При этом, в результате смещения, расстояние между кислородом OG каталитического Ser195 до остовного кислорода Ser214 уменьшается с 4.2 Å до 3.3 Å, что попадает в эмпирические границы расстояния между донором и акцептором водородной связи, верхняя граница для которых общепринято считается на расстоянии до 3.5 Å [Arunan et al., 2011]. При этом, промежуточный кластер 3VXE\_HL также занимает некоторое промежуточное состояние с точки зрения положения треугольника Ser195 OG, His57 NE2, Ser214 O.



Рисунок 3. Распределение расстояний Ser195 OG — His57 NE2 и Ser195 OG — Ser214 в структурах тромбина.

Чтобы проверить, насколько это смещение — воспроизводимое явление, мы провели анализ расстояний Ser195 OG — His57 NE2 и Ser195 OG — Ser214 O во всех структурах тромбина, использованных для кластеризации. Как видно, существует значимое количество структур (29), в которых расстояние Ser195 OG — Ser214 O меньше эмпирической границы 3.5 Å (Рис. 3) . Более того, выделяется группа из трех структур, у которых при этом расстояние Ser195 OG — His57 NE2 больше данной границы, что может свидетельствовать об формировании новой водородной связи. Отметим, что данные структуры относятся к группе медленных. Таким образом, возможно, смещение Ser195 и Ser214 может быть следствием формирования новой водородной связи. При этом, может разрушаться ключевая каталитическая водородная связь Ser195 OG — His57 NE2. Возможность ослабления или полного разрушения каталитической водородной связи ранее допускалась исследователями [Vogt et al., 2010].

#### 2. QM/MM моделирование GROMACS/GFN2-xTB

Так как GROMACS/MOPAC2012 оказался непригоден для задачи описания иона в составе натрий-связывающего сайта в тромбине, нам пришлось разработать еще один инструмент для гибридного KM/MM моделирования. На этот раз мы использовали метод GFN-xTB2, входящий в новое семейство полуэмпирических DFTB методов общего назначения, оптимизированных для корректного воспроизведения геометрий, колебательных свойств, а также энергий нековалентных взаимодействий [Bannwarth et al., 2019, Grimme et al., 2017].

При оптимизации геометрии системы, состоящей из кубика воды и катиона натрия, методом GFN-xTB2, наблюдалось координационное число 6.0 (Рис. 4), характерное для этого иона [Bucher et al., 2010]. Однако расстояния в статичной системе составляют 2.2-2.4Å, при характерном диапазоне 2-3Å с пиком в области 2.45-2.55 [Bucher et al., 2010].



Рисунок 4. Координационная сфера Na<sup>+</sup> в кубе воды. Слева — направо: стартовая структура, оптимизация полуэмпирическим методом PM6-D3, оптимизация методом GFN-xTB2.

До недавнего времени GFN2-хТВ распространяется в виде бинарного исполняемого файла (исходные коды и программный интерфейс API были открыты 02.10.2019 г.), поэтому просто перенести на него использованный ранее с MOPAC2012 подход невозможно. Однако, используя ту же идею и архитектуру, мы реализовали интерфейс общего назначения с подключаемыми плагинами для различных квантово-механических пакетов. Данный интерфейс устроен аналогично интерфейсу для GROMACS/MOPAC2012. На языке программирования Cython – диалекте языка программирования Python, конвертируемом в код на языке С, пишется сценарий-плагин, отвечающий за а) запись входного файла для квантово-механического пакета; б) вызов исполняемого файла с необходимыми параметрами; в) чтение выходного файла. Для удобства управления и тонкой настройки квантово-механического пакета, можно опционально добавлять переменные окружения, которые можно изменять "на лету". Так как плагин компилируется в разделяемую библиотеку, то, благодаря механизмам управления поиска путей к библиотекам (с помощью переменной окружения LD\_LIBRARY\_PATH), можно быстро переключать используемый инструмент без пересборки всего пакета GROMACS. Это чрезвычайно удобная возможность, когда необходимо валидировать полученный результат с помощью квантово-химических методов различного уровня точности. Чтобы минимизировать задержки, связанные с использованием сетевых файлов систем, особенно на кластерах и суперкомпьютерах (как мы продемонстрировали на примере GROMACS/ORCA/MOPAC2012, такие задержки могут замедлять расчет практически на порядок), обмен данными происходит через временную локальную файловую систему, размещенную в оперативной памяти (TMPFS).

Также такая реализация интерфейса не накладывает никаких ограничений на исходную кодовою базу, поэтому можно использовать дополнительные плагины для пакета GROMACS. В частности, набор инструментов PLUMED2 для проведения расчетов с использованием методов повышения эффективности сканирования фазового пространства [Bonomi et al., 2019, Tribello et al., 2014].

В данном случае мы использовали модификацию метода метадинамики (well-tempered metadynamics). На основе статистики распределения расстояний Ser195 OG — His57 NE2 и Ser195 OG — Ser214 в PDB, а также предварительных данных моделирования в GROMACS/DFTB3, мы дополнительно ограничили фазовое пространство при помощи гармонических потенциалов.

Как видно на (Рис. 5), системы с катионом и без продолжают демонстрировать различное поведение. В случае системы без иона, присутствует несколько локальных минимумов ((3.4, 3.7), (3.5, 2.8)) в области, находящейся за пределами границ водородной связи О-N (Ser195 OG — His57 NE2). Причем глубина последнего на (3.5, 2.8) составляет около ~0.8 ккал/моль. Непосредственное наличие в данном минимуме водородной связи с Ser214 О подтверждается визуальным анализом (Рис. 6).

Важной характеристикой моделирования с использованием метадинамики является сходимость полученных результатов. В данном случае мы отслеживали величину барьера (Рис. 6) на пути между двумя локальными минимумами, обнаруженными в системе без иона: (2.7, 5.3) (характерный для исходной водородной связи) и (3.5, 2.8) (связь с Ser214 O). Барьер вычислялся как разница максимального и минимального значений на данном пути. Помимо процедуры реконструкции поверхности свободной энергии на основе суммирования наложенных в ходе метадинамики потенциалов,



Рисунок 5. Ландшафт свободной энергии для расстояний Ser195 OG — His57 NE2 и Ser195 OG — Ser214 O в зависимости от наличия (слева) или отсутствия (справа) иона Na<sup>+</sup> в натрий-связывающем сайте. Наличие катиона существенно ограничивает доступную область конформационного пространства.

мы также применили реконструкцию с использованием взвешенного гистограммного анализа.

Анализ водородных связей внутри квантовой системы между молекулами воды и белковой частью системы также демонстрирует, что общее пространство водородных связей сужено: в системах без катиона, в среднем, 26 водородных связей, против 23 в системе с катионом (Рис. 7). Амплитуды изменений в системе без иона тоже больше. Все вместе это можно интерпретировать следующим образом: в системе с катионом поддерживается стабильная структура водородных связей, которые гораздо реже обмениваются.

#### 3. Моделирование взаимодействия с субстратом

Учитывая наличие альтернативных конфигураций активного центра, нам было интересно узнать, могут ли они оказывать эффект на взаимодействие с субстратом. Для этой задачи мы провели классическое молекулярномеханическое моделирование комплекса тромбина с фрагментом нативного пептидного субстрата Ace-PRSFL-Nme. Нами были созданы две стартовых конфигурации: исходная, соответствующая E:Na<sup>+</sup> форме из кристалла, и альтернативная — со связью Ser195 OG – Ser214 О. Последняя была получена путем применения к исходной гармонического потенциала, ограничивающего расстояние между протоном каталитического серина и остовным кислородом Ser214 на расстоянии 1.8 Å.

14



Рисунок 6. Величина энергетического барьера на пути в локальный минимум (слева), характерный для системы без катиона. Система с ионом Na<sup>+</sup> в натрий-связывающем сайте изображена красным, без катиона синим. Прерывистая линия отображает анализ на основе реконструкции поверхности свободной энергии с использование взвешенного гистограммного метода. Водородная связь (справа) Ser195 OG — Ser214 O в системе без катиона Na<sup>+</sup>.



Рисунок 7. Количество водородных связей в система с Na<sup>+</sup> (красный) и без (синий). Слева приведено изменение количества связей во времени, справа — в виде гистограммы.

В результате моделирования классической (без применения методов повышения эффективности сканирования) молекулярной динамики, мы получили двумерный ландшафт свободной энергии в уже многократно использованных ранее координатах: Ser195 OG – His57 NE2, Ser195 OG – Ser214 О (Рис. 8). Данный ландшафт качественно воспроизводит картину KM/MM моделирования: минимум, соответствующй каталитически активной E:Na<sup>+</sup> форме существенно глубже, нежели Е форме, образованной за счет появления ингибирующей водородной связи с остовом Ser214. Отдельно можно отметить появление еще одного состояния – вытесненного (или Expelled).

Более детальный анализ фрагмента ландшафта, соединяющего области N-O и O-O (Рис. 8), демонстрирует, что если фермент-субстратный ком-



Рисунок 8. Поверхность свободной энергии комплекса тромбина с пептидным субстратом. Выделенные области соответствуют: N-O – каталитическая конфигурация, O-O – ингибирующая водородная связь, Expelled – вытесненная конфигурация (слева). Одномерный профиль энергии вдоль пути между минимумами зон O-N и O-O (справа).

плекс образовался с конформацией фермента, имеющей водородную связь с остовом Ser214, то для выхода из этого состояния необходимо преодолеть барьер ~0.8 ккал/моль. А это, очевидным образом, будет вносить вклад в общую кинетику процесса и соответствует уменьшению наблюдаемой скорости приблизительно на порядок.

#### 4. Предсказание субстратной специфичности

#### 4.1. Разработка и оптимизация алгоритма PeptoGrid

Одно из экспериментальных наблюдений, касающихся медленной формы, состоит в том, что она имеет измененную субстратную специфичность. При этом, по всей видимости, происходит не смена профиля к другому набору субстратов, а скорее падение специфичности к фибриногену. В пользу этой гипотезы говорит тот факт, что взаимодействие медленной формы тромбина с протеином С (лиганд, характерный для медленной формы) улучшается в 1 700 раз за счет образования тройного комплекса с тромбомодулином [Pozzi et al., 2012].

Взаимодействие белка с пептидными лигандами может быть изучено с помощью моделирования молекулярного докинга. Одной из лучших оценочных функций, специально сконструированных для пептидных библиотек, на сегодняшний день является разработанный нами алгоритм PeptoGrid [Zalevsky et al., 2019]. Он использует набор результатов докинга, полученных с помощью любого из популярных пакетов: AutoDock Vina [Trott and Olson, 2010], PLANTS [Korb et al., 2009] или LeDock [Wang et al., 2016]. Процедура основана на анализе частот встречаемости атомов определенного типа в различных областях пространства и последующем ранжировании

структур в соответствии с этими пространственными частотными картами. Структуры, лучше соответствующие такому консенсусному облаку контактов, получают более высокие значения оценочной функции (Рис. 9).



Рисунок 9. Блок-схема алгоритма оценочной функции PeptoGrid [Zalevsky et al., 2019]

Процедура была валидирована на библиотеке трипептидов, для которых известны экспериментальные константы связывания с ангтионтензинпревращающим ферментом (АПФ). А ее эффективность подтверждена результатами виртуального скрининга биогенных пептидов, способных взаимодействовать с внеклеточный доменом ГАМКБ рецептора — фармакологической мишени для лечения зависимостей, депрессии и тревожных расстройств [Froestl, 2010]. Отобранные в результате вычислительного предсказания пептиды продемонстрировали выраженное влияние на поведение рыбок *Danio rerio in vivo* [Zalevsky et al., 2019].

Также аналогичная процедура может использоваться и для олигомеров нуклеиновых кислот. Например, с помощью докинга комбинаторной библиотеки тринуклеотидов в структуру тромбина и используя для построения не индивидуальные типа атомов, а все атомы оснований, можно выделить области наиболее предпочтительного расположения азотистых оснований на его поверхности (Рис. 10). Причем оказывается, что в структурах тромбина с ДНК-аптамером 15ТВА и его производными, в первом локальном максимуме (слева) действительно располагается один из тиминов ТТ-петли. А во второй области может располагаться ароматическое кольцо из химической модификации по 12 нуклеотиду — 5-(3-(2-(1H-индол-3-ил)ацетамид-Nил)-1-пропен-1-ил)-2'-дезоксиуридин. Это означает, что информация, получаемая при помощи подхода PeptoGrid, может использоваться не только для дизайна новых молекул, но также для подбора мест введения химических модификаций.



Рисунок 10. Совмещение карты пространственной плотности и структур лигандов. (А) Карта для ароматических углеродов и структура пептидомиметика в активном центре ангиотензинпревращающего фермента (PDB ID 2XY9). (Б) Карта для тимина и структура химически модифицированного ДНК-аптамера к тромбину (PDB ID 6EO6). Увеличение частоты встречаемости атомов происходит от синего, через желтый, к пурпурному.

Таким образом, набор инструментов PeptoGrid может оказаться полезным для изучения взаимодействия тромбина с пептидными субстратами. Однако для предсказания субстратной специфичности необходимо иметь возможность многократно производить массовый параллельный докинг пептидных библиотек. Для решения данной задачи мы оптимизировали разработанный ранее подход по следующим направлениям:

- Оптимизация структуры хранения данных;
- Оптимизация для высокопараллельных окружений.
- Увеличение скорости моделирования докинга;

Нативные пакеты для моделирования молекулярного докинга, такие какие как AutoDock Vina, PLANTS и LeDock, в процессе работы генерируют большое количество промежуточных и итоговых файлов. В случае AutoDock Vina, для полной комбинаторной библиотеки трипептидов из 8 000, общее количество поз будет более 160 000 (при 20 индивидуальных позах для каждого пептида). При работе в суперкомпьютерных окружениях, использующих распределенные сетевые файловые системы, такое большое количество файлов, расположенных, к тому же, в небольшом количестве директорий, способно существенно понизить производительность [Kovatch et al., 2015]. Для обхода данной проблемы мы реализовали хранение полученных координат в виде единой матрицы формата HDF5. Помимо уменьшения количества файлов, HDF5 также позволяет работать в нескольких режимах: один писатель — множество читателей и множество писателей — множество читателей. Таким образом, при использовании технологии параллельного программирования MPI мы получаем возможность запускать до нескольких сотен параллельных процессов, независимо осуществляющих молекулярный докинг и запись результатов в один общий файл, без необходимости дополнительного постпроцессинга по объединению результатов.

Скорость молекулярного докинга лимитирована выбранным пакетом. При этом лучший результат по качеству предсказаний ранее продемонстрировала AutoDock Vina [Zalevsky et al., 2019], которая является самой медленной из тройки AutoDock Vina, PLANTS, LeDock.

К счастью, некоторое время назад была разработан пакет Quick Vina 2, имплементирующий логику и скоринг функцию AutoDock Vina, но с несколькими эвристиками, значительно ускоряющими моделирование (до 20.5 раз) [Alhossary et al., 2015]. При этом, так как Quick Vina 2 имеет полностью совместимый синтаксис с Autodock Vina, то ее внедрение в PeptoGrid элементарно. Из-за особенностей модификаций Quick Vina 2, зависимость качества результатов от параметра exhaustivness, влияющего на эффективность сканирования конформационного пространства существенно отличается от исходной. Поэтому мы сравнили эффективность предсказания при автоматически выбираемом значении данного параметра с полученными ранее результатами. Сравнение производилось путем сопоставления коэффициентов корреляции с экспериментальными данными. Каждый раз случайно выбиралось 5 точек из разных диапазонов константы связывания, процедура была повторена 1000 раз. Как видно из (Рис. 11), даже исходные результаты Quick Vina 2 оказываются лучше, чем AutoDock Vina. Применение же рескоринг-функции PeptoGrid дополнительно улучшает медиану коэффициента корреляции Спирмена на 6.4 процентных пункта. Исходный код обновленного пакета PeptoGrid доступен по ссылке https://github.com/aozalevsky/ peptogrid2/



Рисунок 11. Качество предсказаний различных скоринг функций для пептид-белкового докинга. Синим и красным отмечены результаты применения рескоринг функции PeptoGrid к Vina и Qvina 2, соответственно.

С использованием Quick Vina 2 и 15 узлов суперкомпьютера "Ломоносов-2", сканирование полной комбинаторной библиотеки трипептидов занимает порядка 100 минут.

#### 4.2. Докинг пептидных библиотек

Для анализа субстратной специфичности с помощью разработанного нами инструмента PeptoGrid [Zalevsky et al., 2019] мы провели массовый параллельный докинг библиотеки трипептидов в структуры тромбина, полученные в предыдущей главе.

Реконструированный профиль специфичности, основанный на пептидах, чьи позы имеют конфигурацию, пригодную для гидролиза (всего 11 153



Рисунок 12. Профиль специфичности тромбина в положения P2, P1, P1<sup>°</sup> быстрой формы тромбина

из 160 000 поз) демонстрирует, что наиболее частое сочетание в положениях P2-P1 — GR. В целом же, в P2 встречаются аминокислоты GPST, в P1 — R, в P1 <sup>'</sup> — AGS, что хорошо соотносится с экспериментальными данными о субстратной специфичности [Gallwitz et al., 2012]. Отдельно интересно отметить наличие H, Y и F в P1 положении. Для фактора свертывания VIII мутация A372H являет-ся одним из природных вариантов, который вызывает гемофилию из-за того, что скорость активации такого белка тромбина падает в 80 раз [Nogami et al., 2005]. Расположение же тирозина и фенилаланина (и его аналогов) в P1 положении является характерной особенностью целого ряда высокоселективных ингибиторов тромбина [Lee et al., 1997, Riester et al., 2005].



Рисунок 13. Профиль специфичности тромбина в положения P2, P1, P1<sup>'</sup> тромбина с ингибирующей водородной связью

Для аналогичного профиля тромбина с ингибирующей водородной свя-

зью подходят всего 53 позы из 160 000, что скорее можно трактовать как отсутствие пригодных для катализа поз вообще. Такое изменение геометрии активного центра вызывает общий сдвиг специфичности (коэффициент корреляции Пирсона между значениями оценочной функции PeptoGrid для двух форм тромбина равен 0.3). После чего данный фермент-субстратный комплекс все равно должен перейти в каталитическую конфигурацию. На кривой накопление продукта в ходе реакции данный переход может быть детектирован в виде дополнительной лаг-фазы при старте из условий, в которых заведомо превалирует медленная форма фермента [Zhang and Kovach, 2005].

#### выводы

1. Новая реализация непараметрического алгоритма кластеризации Affinity Propagation позволяет выявлять биологически релевантные кластеры.

2. Сравнение кластеров, полученных из 400 записей тромбина в банке структур PDB, выявило возможность существования ранее не описанной ингибирующей водородной связи между О<sub>γ</sub>каталитического Ser195 и остовным кислородом Ser214.

3. С помощью нового пакета для гибридного КМ/ММ моделирования продемонстрировано влияние наличия иона Na+ на образование ингибирующей водородной связи. Дополнительный энергетический барьер для переключения ингибирующей водородной связи в присутствии пептидного субстрата составляет 0.8 ккал/моль, что соответствует экспериментальным данным о замедлении реакции в 3-10 раз.

4. Новый метод ранжирования аффинности библиотек пептидов к белковым мишеням позволяет идентифицировать новые биоактивные пептиды.

5. Образование ингибирующей водородной связи изменяет субстратную специфичность тромбина.

# НАУЧНЫЕ СТАТЬИ ПО ТЕМЕ ДИССЕРТАЦИИ, ОПУБЛИКОВАННЫЕ В ЖУРНАЛАХ SCOPUS, WOS, RSCI

 Arthur Zalevsky, Alexander Zlobin, Vasilina Gedzun, Roman Reshetnikov, Maxim Lovat, Anton Malyshev, Igor Doronin, Gennady Babkin, and Andrey Golovin. Peptogrid—rescoring function for autodock vina to identify new bioactive molecules from short peptide libraries. Molecules, 24(2):277, 2019. IF=3.060

- Elena G. Zavyalova, Valeriia A. Legatova, Rugiya Sh Alieva, Arthur O. Zalevsky, Vadim N. Tashlitsky, Alexander M. Arutyunyan, and Alexey M. Kopylov. Putative mechanisms underlying high inhibitory activities of bimodular dna aptamers to thrombin. Biomolecules, 9(2):41–41, 2019. IF=4.694
- 3. А. Злобин<sup>\*</sup>, **А. О. Залевский**<sup>\*</sup>, Ю. А. Мокрушина, О. В. Карцева, А. В. Головин, И. В. Смирнов. Предпочтительная конформация связывания канонических субстратов бутирилхолинэстеразы непродуктивна для экотиофата. Acta Naturae, 10(4):121–124, 2018. **IF=1.657**
- 4. Alexander Zlobin, Yuliana Mokrushina, Stanislav Terekhov, **Arthur Zalevsky**, Tatiana Bobik, Anastasiya Stepanova, Maria Aliseychik, Olga Kartseva, Sergey Panteleev, Andrey Golovin, Alexey Belogurov, Alexander Gabibov, and Ivan Smirnov. Qm/mm description of newly selected catalytic bioscavengers against organophosphorus compounds revealed reactivation stimulus mediated by histidine residue in the acyl-binding loop. Frontiers in pharmacology, 9:e834, 2018. **IF=3.845**
- R. Reshetnikov\*, A. Stolyarova\*, A. Zalevsky\*, D. Panteleev, G. Pavlova, D. Klinov, A. Golovin, and A. Protopopova. A coarse-grained model for DNA origami. Nucleic Acids Research, 46(3):1102–1112, 2018. IF=11.147

# ДРУГИЕ НАУЧНЫЕ РАБОТЫ, ОПУБЛИКОВАННЫЕ ПО ТЕМЕ ДИССЕРТАЦИИ:

- 1. Arthur Zalevsky, Alexander Zlobin, Ivan Smirnov, Andrey Golovin, and Roman Reshetnikov. Unraveling na+ dependent thrombin activation with large-scale QM/MM simulations. In Biocatalysis-2019: Abstracts of 12th International Conference "Biocatalysis: Fundamentals and Applications", pages 56–56. Moscow, 2019.
- Arthur O. Zalevsky, Roman V. Reshetnikov, and Andrey V. Golovin. New QM/MM implementation of the MOPAC2012 in the GROMACS. In RuSCDays 2018: Supercomputing, Communications in Computer and Information Science (CCIS), pages 279–288. Springer Nature Switzerland AG, 2018.
- Евгения Елизарова и Артур Залевский. Предсказание субстратной специфичности протеаз методами молекулярного моделирования. В Сборник трудов 42-й междисциплинарной школы-конференции ИППИ РАН "Информационные технологии и системы 2018", страницы 573–577. Институт проблем передачи информации им. А.А. Харкевича РАН Москва, 2018.
- 4. A. O. Zalevsky, A. O. Demkiv, and A. V. Golovin. In silico design of

aptamers containing g-quadruplexes. In The tenth international conference on bioinformatics of genome regulation and structure\systems biology (BGRS\SB-2016), pages 347–347. Novosibirsk, Russia, 2016.