

# ОТЧЕТ О ПРОДЕЛАННОЙ РАБОТЕ С ИСПОЛЬЗОВАНИЕМ ОБОРУДОВАНИЯ ИВЦ НГУ

## 1. Аннотация

В рамках работы выполнялось моделирование структур мутантных вариантов двух белков системы эксцизионной репарации оснований ДНК – 8-оксогуанин-ДНК-гликозилазы OGG1 и урацил-ДНК-гликозилазы UNG – методом молекулярной динамики (МД), который до сих пор крайне редко использовался для предсказания эффекта мутаций. Проведено исследование влияния мутаций, встречающихся в клинических образцах в генах OGG1 и UNG, на конформацию активного центра этих белков. Для автоматизации молекулярно-динамического исследования создан конвейер задач от получения списка соматических мутаций и случайного выбора из него до итогового моделирования и анализа полученных траекторий.

## 2. Тема работы

Использование молекулярной динамики для предсказания функциональности встречающихся в раковых опухолях вариантов белков, отвечающих за устойчивость к химиотерапии

## 3. Состав коллектива

- 1) Попов Александр Викторович, к. б. н., м. н. с. ИХБФМ СО РАН, м. н. с. НГУ; руководитель проекта, программист, исследователь.
- 2) Жарков Дмитрий Олегович, чл.-корр. РАН, д. б. н., зав. лабораторией белковой инженерии НГУ; консультант.
- 3) Юдкина Анна Владимировна, ст. лаб. ИХБФМ СО РАН, ст. лаб. НГУ; исследователь.

## 4. Информация о гранте

РНФ 18-74-00052 «Использование молекулярной динамики для предсказания функциональности встречающихся в раковых опухолях вариантов белков, отвечающих за устойчивость к химиотерапии» (2018–2020 гг.), руководитель – Попов Александр Викторович.

## 5. Научное содержание работы

### 5.1. Постановка задачи

Прогресс в современных методах секвенирования сделал возможным персонализацию противоопухолевой терапии. Для расширения возможности персонализации значительный интерес представляют данные о функциональности тех или иных мутантных вариантов белков, связанных с ответом на повреждения ДНК. Особое внимание сегодня уделяется способам предсказания функциональности вариантов белков вычислительными методами. Однако подавляющее большинство таких методов сегодня основывается на анализе консервативности последовательностей и не привлекает всех современных возможностей компьютерного анализа пространственной структуры белков. В проекте предлагалось использовать метод молекулярной динамики (МД) для моделирования структур наблюдаемых в клинических образцах опухолей вариантов двух белков системы эксцизионной репарации оснований – 8-оксогуанин-ДНК-гликозилазы OGG1 и урацил-ДНК-гликозилазы UNG. На основе вычислительных данных и их экспериментальной проверки планировалось предложить общую процедуру использования молекулярной динамики для предсказания эффектов мутаций.

## **5.2. Современное состояние проблемы**

Моделирование методом МД в наши дни активно используется для наблюдения изменения во времени конформации биологических макромолекул, а также анализа их кинетических и термодинамических параметров. Современные методы позволяют достигать длительностей в микросекунды, считая обычными траектории в десятки и сотни наносекунд, а количество атомов при моделировании достигает миллиона (Krieger & Vriend, *J. Comput. Chem.* 36:996-1007, 2015). Рост популярности МД определяется, помимо всего прочего, совершенствованием программного обеспечения и растущей доступностью высокопроизводительных вычислительных систем.

Исследования методом МД и другими способами компьютерного моделирования проводились для нескольких ДНК-гликозилаз вирусов, бактерий и эукариот. Важным моментом здесь служит то, что известны пространственные структуры многих гликозилаз в комплексе с ДНК. Подробнее всего вычислительными методами исследована бактериальная форма мидопиримидин-ДНК-гликозилаза Fpg (MutM), для которой был изучен практически весь каталитический цикл от начального узнавания повреждения в составе ДНК до конформационных изменений фермент-субстратного комплекса и разрыва *N*-гликозидной и фосфодиэфирной связей и опубликовано множество работ, включая работы участников коллектива (Perlow-Poehnel et al., *Biochemistry* 43:16092-105, 2004; Popov et al., *J. Comput. Chem.* 34:319-325, 2013; Popov et al., *BMC Struct. Biol.* 17:5, 2017). Аналогичный цикл исследований был проведён для фермента OGG1 человека, который по субстратной специфичности идентичен Fpg, но обладает совершенно иной

пространственной структурой (Fromme et al., Nat. Struct. Biol., 10:204-11, 2003; Banerjee et al., Nature, 434:612-8, 2005; Lukina et al., J. Biol. Chem., 288:28936-47, 2013; Li et al., J. Am. Chem. Soc., 139:2682-92, 2017). В результате было выяснено, что, несмотря на всю разницу в строении этих двух белков, для многостадийного узнавания повреждения используются практически идентичные ДНК-белковые взаимодействия вплоть до каталитической стадии, где механизмы реакции принципиально различны. Для обоих этих ферментов были изучены методом МД также эффекты некоторых мутаций, как существующих в природе, так и введённых с целью изучения механизма реакции. Для UNG расчётными методами был показан преимущественный вклад электростатических взаимодействий в активном центре в стабилизацию сильно искажённой предкаталитической конформации ДНК (Luo et al., Biochemistry, 38:9209-20, 1999; Ma et al., Biochemistry, 45:13687-96, 2006). Из других исследований ДНК-гликозилаз методом МД стоит отметить анализ влияния мутаций на субстратную специфичность и активность фермента SMUG1 из *Geobacter metallireducens* (Mi et al., J. Mol. Biol., 385:761-78, 2009) и изучение механизма ДНК-гликозилазы MutY, мутации в гене человеческого гомолога которой вызывают неполипозный рак толстого кишечника (Kellie et al., Biochemistry, 52:8753-65, 2013). Таким образом, молекулярно-динамические свойства ДНК-гликозилаз дикого типа исследованы достаточно хорошо, что создаёт предпосылки для успешного приложения метода МД к анализу последствий аминокислотных замен в этих белках.

### **5.3. Подробное описание работы, включая используемые алгоритмы**

Проводилось моделирование белков OGG1 и UNG с мутациями, встречающимися в раковых опухолях, методом молекулярной динамики. В качестве референтного PDB-файла для введения аминокислотных замен фермента OGG1 была использована структура каталитически неактивного варианта K249Q в комплексе с ДНК, содержащей остаток 8-оксогуанина (номер в БД PDB 1EBM); в ней проводили обратную замену Gln249 на Lys. Структуры комплексов ферментов с ДНК, полученные с использованием программного конвейера M3R-PDB, готовили для моделирования в программе LEaP: производилась верификация модели в рамках силового поля AMBER ff99SB+parmbsc0, сольватация в водном боксе (модель TIP3P) с минимальным расстоянием 10 Å от любого атома модели до стенки бокса, нейтрализация зарядов путем добавления противоионов Na<sup>+</sup> и сохранение топологии и исходных координат в форматы программы AMBER. В соответствии с литературными данными остаток Lys249 моделировали нейтральным, остаток Cys253 в виде тиолат-аниона, остаток Asp268 в протонированной форме. Параметры нестандартных остатков брались из литературы либо уже имелись в пакете AMBER. Оптимизацию энергии проводили в два этапа с использованием модуля

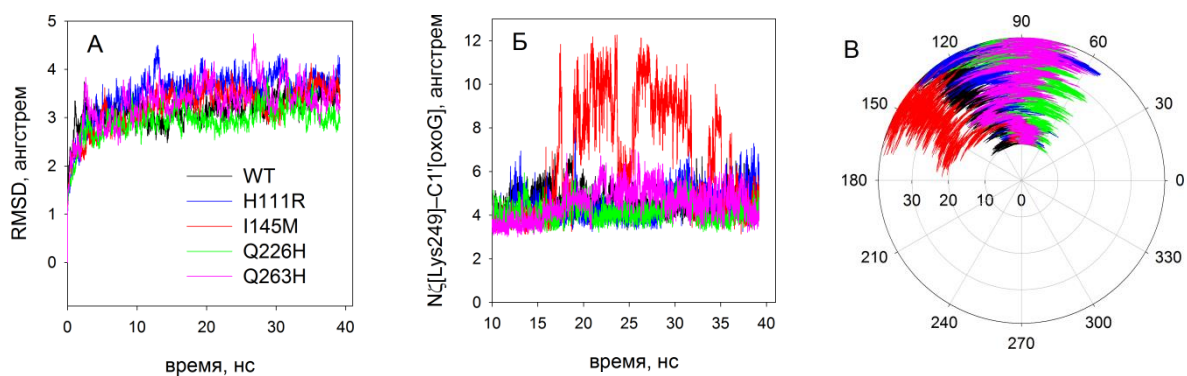
SANDER. На первом этапе выполняли релаксацию растворителя и противоионов путем оптимизации в течение 1000 шагов методом градиентного спуска и затем еще 1000 шагов методом сопряженных градиентов с наложенными на комплекс ДНК-белок сильными позиционными ограничениями  $500 \text{ ккал}/(\text{моль} \cdot \text{Å}^2)$ . На втором этапе выполняли релаксацию всей системы путем оптимизации в течение 2500 шагов методом градиентного спуска и затем еще 2500 шагов методом сопряженных градиентов без каких-либо ограничений. Затем осуществляли плавный разогрев системы с 0 К до 310 К в течение 200 пс с шагом 2 фс, в рамках периодических граничных условий, PME-электростатики и алгоритма SHAKE (модуль rmemd). Получение финальных траекторий для анализа проводили также с помощью модуля rmemd, в течение 40 нс (шаг 2 фс) в NTP-ансамбле. Мгновенные состояния системы (снимки) записывали каждые 8 пс, таким образом, длина одной траектории составила 5000 снимков, из которых для анализа использовали последние 2500 после прекращения роста среднеквадратичного отклонения координат атомов остова белка и ДНК. Траектории в формате MDCRD конвертировали в набор PDB-файлов с сохранением молекул растворителя, которые проводят 10% или более времени вблизи (не далее чем 5 Å) от поврежденного нуклеотида, при помощи разработанной в ходе проекта программы mdcrd2pdb, и затем анализировали в программе MDTRA. Помимо мутантных вариантов, для уточнения критических конформационных параметров активного центра OGG1 было проведено моделирование фермента дикого типа в комплексе с ДНК, содержащей апурин-апиримидиновый (АП-) сайт. Моделирование UNG проводилось аналогичным образом с использованием исходной структуры 1EMH и заменой содержащегося в ней остатка 2'-дезоксипсевдоуридина в ДНК на остаток 2'-деоксиуридина; аминокислотные замены в данном случае ограничивались изучаемыми мутациями.

#### **5.4. Полученные результаты**

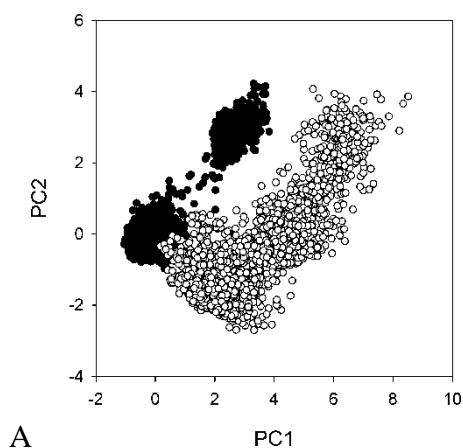
Были получены траектории МД для 21 варианта белка OGG1 и для 33 вариантов белка UNG (рис. 1). Поскольку МД сама по себе не выдает бинарного решения по категоризации мутаций на «нейтральные» и «вредные», способ принятия решений на основании анализа траекторий имеет первостепенное значение для ее применения в качестве предсказательного инструмента. Было предложено в качестве инструмента категоризации использовать близость траектории мутантного белка к траектории белка дикого типа, что позволяет достаточно просто вводить порог бинарного решения, либо количественный (некоторое пороговое расстояние в заданной метрике), либо качественный (кластеризация с белком дикого типа среди множества исследуемых вариантов). На основе литературных данных, моделирования и биохимических

исследований в ходе проекта для белка OGG1 был сформирован минимальный набор геометрических параметров, отражающих каталитически компетентную конформацию активного центра, куда входят расстояние от активного нуклеофила ( $\epsilon$ -аминогруппа остатка Lys-249), угол нуклеофильной атаки и расстояние O $\delta$ 1/O $\delta$ 2[Asp268]–O4'[oxodG]. Было проведено сравнение ряда способов выделения функционально важных параметров структуры белка, определения расстояния между траекториями и кластеризации мутантных вариантов (рис. 2). В результате первая и третья задачи оказались критическими для категоризации. При увеличении размеров функционально важного участка и соответственно числа учитываемых независимых переменных резко снижалась способность любых методов определения расстояния между траекториями давать кластеризованные результаты. Даже учет ближайшего окружения активного центра (5 Å от остатка 8-оксогуанина) выдавал результаты с низкой корреляцией с экспериментальными данными. Относительно методов кластеризации было показано, что категоризация мутаций как «нейтральных» при близкой кластеризации с белком дикого типа повышает количество ложнопозитивных предсказаний и дает меньшее согласие с экспериментальными данными, чем категоризация мутаций как «вредных» в наиболее удаленных кластерах. Таким образом, наилучшим методом категоризации, по крайней мере в нашем случае, является предсказание мутации как «вредной» при максимальном удалении от белка дикого типа, хотя при этом некоторые вредные мутации остаются пропущенными. Напротив, конкретный способ определения расстояния между траекториями оказался не так важен: по крайней мере качественно все они давали сходные результаты. Из соображений скорости расчета и минимальной потери информации в конечном итоге траектории анализировали разложением на главные компоненты, после чего для всех пар траекторий определяли коэффициент сходства Бхаттачарья (сумма квадратных корней из произведения количества точек каждой сравниваемой траектории в произвольно заданной решетке, в нашем случае 10×10) в плоскости первых двух главных компонент. После этого проводили кластеризацию методом попарного внутригруппового невзвешенного среднего. Аналогичным образом строился алгоритм категоризации мутаций для белка UNG. В этом случае на основании известной информации о механизме реакции в качестве критических геометрических параметров были выбраны расстояния O $\delta$ 1/O $\delta$ 2[Asp145]–O1'[U], N $\epsilon$ 2[His268]–O2[U] и расстояние от C5 основания урацила до плоскости фенильной группы Tyr148.

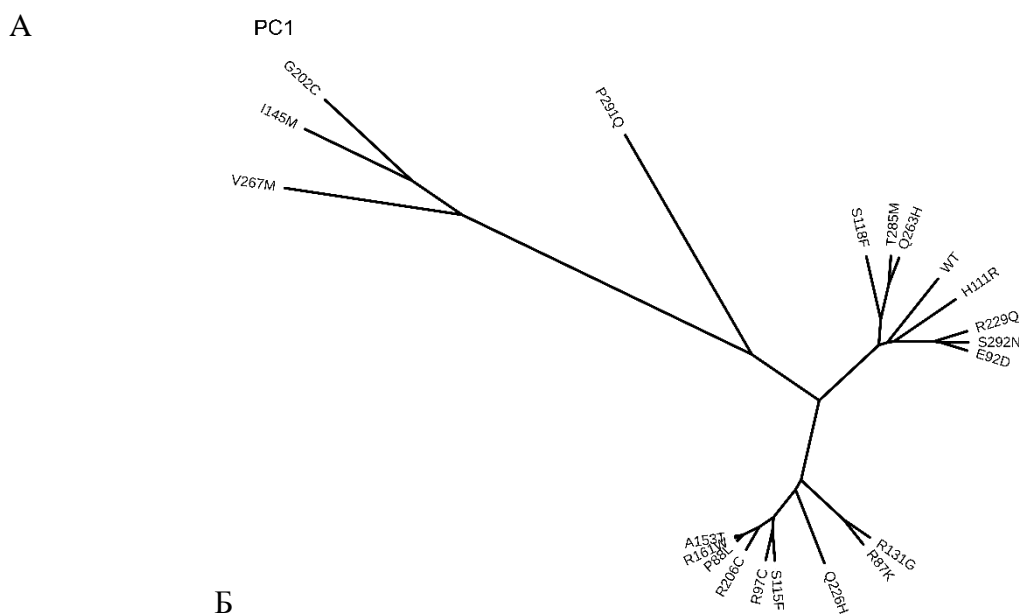
## 5.5. Иллюстрации, визуализация результатов



**Рис. 1.** Данные МД для OGG1 дикого типа и нескольких мутантных форм. (А), среднее квадратичное отклонение от начальной структуры. (Б), расстояние  $N\zeta[\text{Lys249}] - C1'[\text{oxoG}]$ . (В), угол  $N\zeta[\text{Lys249}] - C1'[\text{oxoG}] - N9[\text{oxoG}]$ . На всех графиках черным цветом обозначен OGG1 дикого типа, синим — мутант H111R, красным — I145M, зеленым — Q226H, розовым — Q263H.



**Рис. 2.** Кластеризация траекторий МД для белка OGG1. А – проекция траекторий OGG1 дикого типа (черные кружки) и G202C (белые кружки) на плоскость первых двух главных компонент. Б – кластеризация 21 траектории для белка OGG1 дикого типа и его мутантных вариантов. Во всех случаях траектории построены по трем ключевым геометрическим параметрам активного центра. Варианты I145M, G202C, V267M и P291Q предсказываются как «вредные».



## 6. Эффект от использования кластера в достижении целей работы

Метод МД является довольно ресурсоемким, и решение современных задач этим методом возможно лишь с применением высокопроизводительных вычислительных систем, обеспечивающих высокую скорость расчета и распараллеливание вычислений. Таким образом, использование ресурсов кластера критично для выполняемой работы.

## **7. Перечень публикаций, содержащих результаты работы**

Popov A.V., Endutkin A.V., Yatsenko D.D., Yudkina A.V., Barmatov A.E., Makasheva K.A., Rasporova D.Y., Diatlova E.A., Zharkov D.O. Molecular dynamics approach to identification of new OGG1 cancer-associated somatic variants with impaired activity. Статья на ревью после рецензирования в журнале *J. Biol. Chem.* (JCR IF = 4.106)