

**РОССИЙСКАЯ АКАДЕМИЯ НАУК  
СИБИРСКОЕ ОТДЕЛЕНИЕ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ  
ИНСТИТУТ ХИМИЧЕСКОЙ БИОЛОГИИ И ФУНДАМЕНТАЛЬНОЙ МЕДИЦИНЫ**

*на правах рукописи*

**ПОПОВ АЛЕКСАНДР ВИКТОРОВИЧ**

**МОЛЕКУЛЯРНО-ДИНАМИЧЕСКИЙ АНАЛИЗ СУБСТРАТНОЙ  
СПЕЦИФИЧНОСТИ 8-ОКСОГУАНИН-ДНК-ГЛИКОЗИЛАЗ  
БАКТЕРИЙ И ЧЕЛОВЕКА**

**03.01.04 — Биохимия**

**ДИССЕРТАЦИЯ**

**на соискание учёной степени  
кандидата биологических наук**

**Научные руководители:**

**доктор биологических наук, доцент, профессор РАН Жарков Д. О.  
доктор физико-математических наук, с.н.с. Воробьёв Ю. Н.**

**НОВОСИБИРСК — 2017**

## ОГЛАВЛЕНИЕ

1. ВВЕДЕНИЕ.....	6
2. ОБЗОР ЛИТЕРАТУРЫ.....	10
2.1. ОКИСЛИТЕЛЬНЫЕ ПОВРЕЖДЕНИЯ И ЭКСЦИЗИОННАЯ РЕПАРАЦИЯ ДНК.....	10
2.1.1. Окислители и антиоксиданты.....	10
2.1.2. Основные окислительные повреждения ДНК.....	12
2.1.3. Эксцизионная репарация оснований.....	15
2.1.4. ДНК-гликозилазы.....	19
2.1.4.1. Структурные суперсемейства.....	19
2.1.4.2. Формамидопиримидин-ДНК-N-гликозилаза.....	21
2.1.4.3. 8-оксогуанин-ДНК-N-гликозилаза эукариот.....	22
2.1.4.4. Другие ДНК-гликозилазы, исправляющие окислительные повреждения.....	23
2.1.5. GO-система.....	25
2.1.6. Структурные основы узнавания повреждённых оснований.....	26
2.1.7. Каталитический механизм ДНК-гликозилаз.....	31
2.2. КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ БИОПОЛИМЕРОВ.....	34
2.2.1. Молекулярная механика.....	34
2.2.2. Молекулярно-динамическая теория и методы.....	36
2.2.2.1. Силовые поля.....	36
2.2.2.2. Минимизация энергии.....	39
2.2.2.3. Адиабатический маппинг.....	41
2.2.2.4. Молекулярная динамика.....	42
2.2.2.5. Динамика Ланжевена.....	48
2.2.2.6. Броуновская динамика.....	49
2.2.2.7. Метод Монте-Карло.....	49
2.2.2.8. Симуляция отжига.....	50
2.2.2.9. Нединамические методы анализа.....	50
2.2.2.10. QM/MM.....	51
2.2.3. Молекулярная динамика и энергетические барьеры.....	52
2.2.3.1. Активированная молекулярная динамика.....	52
2.2.3.2. Направленная молекулярная динамика.....	52
2.2.4. Вычислительные аспекты молекулярной динамики.....	53
2.2.4.1. Аппаратное обеспечение.....	53

2.2.4.2.	Программное обеспечение .....	56
2.2.5.	Анализ молекулярно-динамических траекторий .....	58
2.2.6.	Развитие методов молекулярной динамики .....	58
2.3.	Исследование ДНК-гликозилаз методами компьютерного моделирования .....	59
2.3.1.	Урацил-ДНК-N-гликозилаза UNG.....	59
2.3.2.	Урацил-ДНК-N-гликозилаза SMUG1.....	61
2.3.3.	Формаидопиримидин-ДНК-N-гликозилаза бактерий.....	62
2.3.4.	8-оксогуанин-ДНК-N-гликозилаза человека .....	64
2.3.5.	ДНК-N-гликозилаза MutY .....	66
2.3.6.	Nei-подобная ДНК-N-гликозилаза NEIL1 .....	67
2.3.7.	Эндонуклеаза V бактериофага T4 (DenV) .....	68
2.3.8.	Гликозилазы алкилированных пуринов.....	68
3.	МАТЕРИАЛЫ И МЕТОДЫ.....	70
3.1.	Комплекс программ моделирования биополимеров BISON.....	70
3.1.1.	Программа BioPASED .....	70
3.1.2.	Силовое поле BioPASED .....	71
3.2.	Компьютерное моделирование .....	74
3.2.1.	Подготовка моделей белков и ДНК-белковых комплексов .....	74
3.2.2.	Оптимизация энергии и молекулярная динамика.....	77
3.2.3.	Анализ траекторий .....	78
3.3.	Экспериментальная часть.....	78
3.3.1.	Реактивы.....	78
3.3.2.	Дезоксирибоолигонуклеотиды.....	79
3.3.3.	Ферменты .....	79
3.3.4.	Штаммы бактерий и плазмиды .....	79
3.3.5.	Трансформация клеток .....	79
3.3.6.	Выделение мутантных форм OGG1 .....	80
3.3.7.	Гель-электрофорез и количественный обсчет результатов экспериментов.....	81
3.3.8.	Приготовление <sup>32</sup> P-меченых олигонуклеотидных субстратов .....	81
3.3.9.	Приготовление субстрата, содержащего АП-сайт .....	82
3.3.10.	Определение рН-зависимости активности ферментов .....	82
3.3.11.	Изучение зависимости расщепления субстрата от концентрации фермента.....	82
3.3.12.	Определение кинетических параметров реакций, катализируемых OGG1.....	83

4.	РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ.....	84
4.1.	Разработка программных средств.....	84
4.1.1.	Графический веб-интерфейс программы моделирования GUI-BioPASED.....	84
4.1.2.	Программа анализа молекулярно-динамических траекторий MDTRA.....	86
4.2.	Исследование структуры и активности Fpg .....	90
4.2.1.	Значимость водородных связей для поддержания структуры Fpg .....	91
4.2.2.	Состояния ионизации каталитически значимых аминокислотных остатков в активном центре Fpg.....	94
4.2.3.	Специфичность Fpg к основанию напротив повреждённого.....	101
4.2.4.	Роль молекул воды в механизме действия Fpg .....	108
4.3.	Исследование структуры и активности OGG1 .....	111
4.3.1.	Структурные основы снижения активности мутантных форм OGG1 на субстратах, содержащих 8-оксогуанин.....	111
4.3.2.	Структурные основы снижения активности мутантных форм OGG1 на субстратах, содержащих АП-сайт .....	118
4.3.3.	Специфичность OGG1 к основанию напротив повреждённого .....	122
4.3.4.	Кинетические параметры реакции мутантных форм OGG1 с субстратами, содержащими 8-оксогуанин или АП-сайт .....	124
5.	ЗАКЛЮЧЕНИЕ .....	128
6.	ВЫВОДЫ.....	129
7.	СПИСОК ЛИТЕРАТУРЫ.....	131



## СПИСОК ПРИНЯТЫХ СОКРАЩЕНИЙ

5-OH-Cyt	5-гидроксицитозин; 4-амино-5-гидроксипиримидин-2(1 <i>H</i> )-он
5-OH-Ura	5-гидроксиурацил; 5-гидроксипиримидин-2,4(1 <i>H</i> ,3 <i>H</i> )-дион
8-oxoAde	8-оксоаденин; 6-амино-7,9-дигидро-8 <i>H</i> -пурин-8-он
8-oxodGTP	8-оксо-2'-дезоксигуанозин-5'-трифосфат
8-oxoGua	8-оксогуанин; 2-амино-7,9-дигидро-1 <i>H</i> -пурин-6,8-дион
8-oxodGuo	8-оксо-2'-дезоксигуанозин
dRib	2'-дезоксирибоза
Fapy-Ade	4,6-диамино-5-формаидопиримидин
Fapy-Gua	2,6-диамино-4-оксо-5-формаидопиримидин
Fur	3-гидрокситетрагидрофуран
H2TH	мотив «спираль — два поворота — спираль» ( <b>h</b> elix — <b>2 t</b> urns — <b>h</b> elix)
HhH	мотив «спираль — шпилька — спираль» ( <b>h</b> elix — <b>h</b> airpin — <b>h</b> elix)
RMSD	среднеквадратичное отклонение ( <b>r</b> oot- <b>m</b> ean- <b>s</b> quare <b>d</b> eviation)
а. о.	аминокислотный остаток
АП	апуриновый-апиримидиновый
АП-в	2,3,5-тригидроксипентилфосфат (восстановленный АП-сайт)
АП-п	3-гидроксипропилфосфат (пропандиол-АП-сайт)
АФК	активные формы кислорода
БСА	бычий сывороточный альбумин
ДТТ	1,4-дителиотреитол
дцДНК	двухцепочечная дезоксирибонуклеиновая кислота
оцДНК	одноцепочечная дезоксирибонуклеиновая кислота
МД	молекулярная динамика
ОДН	олигодезоксирибонуклеотид
ПААГ	полиакриламидный гель
РСА	рентгеноструктурный анализ
ЦПД	циклобутановый пиримидиновый димер
ЭРО	эксцизионная репарация оснований
ЯМР	ядерный магнитный резонанс

## 1. ВВЕДЕНИЕ

Молекула ДНК, основной носитель наследственной информации, может повреждаться под действием различных агентов физической и химической природы. Накопление в большом количестве мутаций, вызванных этими повреждениями, приводит к старению, канцерогенезу и многим патологическим процессам. Один из наиболее распространённых типов повреждений ДНК — окисление, которое происходит при действии на макромолекулу активных форм кислорода. Они могут образоваться как под действием внешних агентов, например, ионизирующей радиации, так и в процессе нормального метаболизма и иммунного ответа аэробных организмов. Для исправления повреждений структуры ДНК в клетках существует система репарации, один из главных путей которой — эксцизионная репарация оснований (ЭРО). На ранних этапах ЭРО участвуют специфические ферменты двух классов: ДНК-*N*-гликозилазы (КФ 3.2.2, далее упоминаемые как ДНК-гликозилазы), узнающие повреждения в ДНК и выщепляющие повреждённые основания, и АП-эндонуклеазы, расщепляющие ДНК по вновь образованному апурин/апиримидиновому (АП) сайту. Каталитический механизм некоторых ДНК-гликозилаз может включать в себя АП-лиазную стадию, также приводящую к разрыву цепи.

Гликозилазы репарации окислительных повреждений ДНК присутствуют как в прокариотических, так и в эукариотических организмах. Обычно каждая из них специфична к достаточно широкому кругу окисленных пуринов или пиримидинов. В то же время эти ферменты могут быть высокоспецифичны к основанию, находящемуся напротив повреждённого, что является залогом эффективного функционирования системы предотвращения мутаций; однако природа такой специфичности пока не ясна.

Существующие в настоящий момент физические и химические методы исследования не позволяют в полной мере оценить вклад тех или иных структурных факторов ни в субстратную специфичность ферментов, ни в каталитический механизм их действия. Возросший в последнее время интерес к расчётным методам на фоне бурного роста вычислительных возможностей компьютеров способствует активному привлечению этого инструмента к задачам энзимологии. Методы молекулярной динамики позволяют изучать не только статичные структуры, но и, в рамках определённого рода приближений, процессы, в них происходящие. Поэтому детальное исследование каталитической активности и субстратной специфичности ДНК-гликозилаз на атомном уровне представляется современной и актуальной биохимической задачей.

Несмотря на то, что методы компьютерного моделирования развиваются уже не один десяток лет, имеющиеся программные средства по-прежнему несовершенны. Многие

инструменты не являются дружелюбными ни в отношении пользователя, ни в отношении современных аппаратных средств. На момент начала работы предпринимались лишь немногочисленные попытки создать программы анализа молекулярно-динамических траекторий, эффективные и удобные для широкого круга исследователей, не являющихся экспертами в области компьютерного моделирования. Поэтому разработка специализированных программных средств для нужд компьютерного моделирования — ещё одна актуальная задача, имеющая важнейшую практическую значимость.

**Цель настоящей работы** заключалась в выявлении структурных факторов активности и субстратной специфичности 8-оксогуанин-ДНК-гликозилаз бактерий и человека методом компьютерного моделирования и сопоставлении полученных данных с биохимическими показателями. В ходе исследования необходимо было решить следующие задачи:

1. разработать графический пользовательский интерфейс к программному пакету BISON/BioPASED для формирования заданий для молекулярного моделирования и автоматизированной первичной верификации моделей; разработать программу с графическим пользовательским интерфейсом для анализа протяжённых молекулярно-динамических траекторий на принципах эргономичности, расширяемости, эффективной визуализации результатов и с низкими системными требованиями; протестировать разработанное программное обеспечение на примере бактериальной формамидопиримидин ДНК-гликозилазы (Fpg);
2. изучить поведение каталитически значимых аминокислотных остатков в активном центре бактериальной ДНК-гликозилазы Fpg для получения репрезентативных моделей фермента; на основе полученных моделей исследовать факторы специфичности фермента к основанию напротив повреждённого;
3. исследовать структурные основы снижения активности мутантных форм 8-оксогуанин-ДНК-гликозилазы (OGG1) человека в отношении субстратов, содержащих 8-оксогуанин (8-охоGua) или AP-сайт; сопоставить полученные структурные данные с результатами кинетических экспериментов; исследовать факторы специфичности фермента к основанию напротив повреждённого.

**Научная новизна работы.** В рамках работы разработаны новые программные средства для подготовки, первичной верификации и анализа результатов моделирования методом молекулярной динамики. Впервые исследовано поведение каталитически важных аминокислот в активном центре фермента Fpg в зависимости от состояния их ионизации. Получены новые данные о структурных основах субстратной специфичности 8-оксогуанин-ДНК-гликозилаз.

Впервые показано влияние некоторых аминокислотных замен (C253I, C253L, Q315W) в активном центре OGG1 на активность фермента в отношении различных субстратов.

**Теоретическая и практическая значимость работы.** Полученные динамические данные о структурных детерминантах активности и субстратной специфичности 8-оксогуанин-ДНК-гликозилаз некоторых бактерий и человека позволяют углубить знания о различных стадиях ферментативного катализа, осуществляемого исследуемыми белками, для последующего использования их в таких практических областях, как рациональный дизайн ферментов и трансляционная медицина. Разработанные программные средства имеют большое прикладное значение и активно применяются в различных молекулярно-динамических исследованиях, в том числе вне темы настоящей работы.

**Методология и методы исследования.** В работе широко применялись компьютерные методы исследования биополимеров (метод молекулярной динамики), также были разработаны собственные вспомогательные алгоритмы и программы. Для экспериментальной проверки результатов расчётов применялись классические методы ферментативной кинетики.

**Положения, выносимые на защиту.**

1. Разработаны и протестированы новые программные средства: графический пользовательский интерфейс GUI-BioPASED к программному пакету BISON/BioPASED и программа анализа молекулярно-динамических траекторий MDTRA.
2. Каталитическая пара Pro1–Glu2 в белке Fpg при нейтральных значениях pH может существовать в виде смеси разных ионов, из которых структурно наиболее оптимальна форма с протонированными аминогруппой Pro1 и карбоксильной группой Glu2, а для протекания реакции требуется депротонирование только Pro1. Специфичность Fpg к основанию напротив повреждённого во многом определяется взаимодействиями вне активного центра. Молекулы воды, связанные в комплексе Fpg–ДНК, могут играть важную роль в конформационной динамике и катализе.
3. Причины снижения активности мутантных форм 8-оксогуанин-ДНК-гликозилазы OGG1 человека, содержащих объёмные аминокислотные замены в активном центре C253I, C253L и Q315W, заключаются в искажении структуры самого активного центра из-за его пластичности, а не в изменении конформации ДНК. Ориентация  $\epsilon$ -аминогруппы Lys249 служит определяющим структурным фактором активности фермента OGG1.

**Степень достоверности и апробация результатов.** Основные положения работы представлены и обсуждены на 8 российских и международных конференциях, в том числе: на международных конференциях «BGRS\SB'2010» (Новосибирск, Россия, 2010), «Постгеномные технологии — медицине» (Новосибирск, Россия, 2012), «Nucleic Acid – Protein Interactions for

Life Sciences» (Новосибирск, Россия, 2013); на международных конференциях «Albany 2013: Conversation 18» (Albany, США, 2013), «FEBS 2013» (Санкт-Петербург, Россия, 2013), «Albany 2015: Conversation 19» (Albany, США, 2015), материалы которых индексируются в Web of Science и Scopus. Разработанное программное обеспечение MDTRA применяется независимыми исследователями из Центра биотехнологии и геномики растений (Мадрид, Испания) и Института микробиологии АН Чешской республики (Прага, Чехия); с использованием MDTRA ими опубликовано 4 работы в зарубежных рецензируемых журналах. Кроме того, MDTRA активно применяется сотрудниками ИХБФМ СО РАН.

**Публикации.** По материалам диссертации опубликовано 4 научные статьи, из них 1 в российском рецензируемом журнале и 3 в международных рецензируемых журналах. Все журналы индексируются в международных базах Web of Science и Scopus.

**Структура и объём работы.** Диссертация состоит из введения, обзора литературы, экспериментальной части, результатов и их обсуждения, заключения, выводов и списка цитированной литературы. Работа изложена на 145 страницах, содержит 36 рисунков и 7 таблиц. Библиография включает 187 литературных источников.

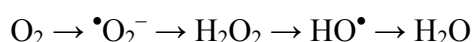
## 2. ОБЗОР ЛИТЕРАТУРЫ

### 2.1. ОКИСЛИТЕЛЬНЫЕ ПОВРЕЖДЕНИЯ И ЭКСЦИЗИОННАЯ РЕПАРАЦИЯ ДНК

Окисление клеточных макромолекул происходит со значительной частотой в аэробных организмах из-за побочных продуктов нормального метаболизма и иммунного ответа [1]. Окислительные повреждения митохондриальной и ядерной ДНК напрямую связаны с болезнями и старением. Они могут вызывать разрывы цепи, образование АП-сайтов и предмутагенное повреждение гетероциклических оснований [2]. Следует подробно рассмотреть различные типы окислительных повреждений, механизмы их возникновения и способы борьбы с ними.

#### 2.1.1. Окислители и антиоксиданты

Молекулярный кислород необходим для жизни всем эукариотам и многим прокариотам, так как их метаболизм основан на восстановлении  $O_2$  до  $H_2O$ . В то же время некоторые производные кислорода (т. н. активные формы кислорода, АФК), такие, как синглетный кислород, перекись водорода, супероксид, озон и гидроксильный радикал, являются высокотоксичными продуктами. Они образуются в процессе активирования или частичного восстановления кислорода. Внешняя электронная оболочка молекулы  $O_2$  обычно находится в триплетном состоянии и из-за спиновых ограничений не может одновременно принимать два или четыре электрона [3, 4]. При одноэлектронном восстановлении  $O_2$  до воды последовательно образуются три активные формы: супероксид-анион-радикал, пероксид водорода и гидроксильный радикал [2]:

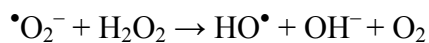
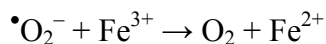


АФК обладают чрезвычайно высокой реакционной способностью, возрастающей от супероксида к гидроксилу [3, 4]. Однако относительно малореактивный супероксид может превращаться в более агрессивные формы спонтанно или в присутствии ионов переходных металлов. Среднее время существования АФК и, как следствие, расстояние, на которое они способны диффундировать от места возникновения, уменьшается с увеличением реактивности. Однако свободнорадикальные формы (супероксид-анион-радикал и гидроксильный радикал) способны выступать в роли инициаторов цепных реакций окисления, что позволяет неспаренным электронам перемещаться на большие расстояния. К АФК также относят особое метастабильное состояние молекулярного кислорода с более высокой энергией — так называемый «синглетный кислород» [4].

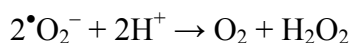
В эукариотических клетках АФК возникают главным образом в процессе передачи электрона по дыхательной цепи митохондрий, при этом около 0,1% кислорода, проходящего через дыхательную цепь, превращается в супероксид-анион-радикал [3]. В растительных организмах АФК также активно продуцирует фотосистема I [3]. АФК образуются в ходе многих внутриклеточных окислительно-восстановительных процессов, не связанных с митохондриями. Их источниками могут быть, например, цитохром P450 и NADPH:цитохром, мембранная NADPH-зависимая оксидаза и P450-редуктаза эндоплазматической сети [3].

Внешними источниками АФК являются ионизирующая радиация, приводящая к радиолизу воды [4], фотосенсибилизирующие красители (метиленовая синь, бенгальский розовый), ионы переходных металлов, некоторые лекарства и токсины, например, паракват (метилвиологен) и блеомицин [5].

Считается, что в повреждении биологических макромолекул АФК главную роль играют гидроксильный радикал и синглетный кислород в связи с их повышенной реакционной способностью. В присутствии двухвалентных катионов ионов железа или других низковалентных ионов переходных металлов гидроксильный радикал может образовываться из супероксид-анион-радикала и пероксида водорода в реакциях Фентона и Габера–Вейса [4]:



Последняя стадия несущественна из-за быстрой реакции дисмутации супероксид-анион-радикала [2]:



Особое значение для АФК-зависимого повреждения ДНК имеет то, что ионы металлов обычно образуют с ДНК комплексы [5]. АФК-зависимое повреждение ДНК может быть также связано с перекисным окислением липидов. Боковые группы полиненасыщенных жирных кислот очень чувствительны к свободнорадикальному окислению, а образующиеся при этом пероксильные радикалы распространяются по вышеупомянутому цепному механизму. Факт широкой распространённости мембранных структур внутри клетки подчёркивает важность вклада перекисного окисления липидов в общую токсичность АФК.

Предотвращение окисления биологических макромолекул осуществляется при помощи перехватчиков свободных радикалов, способных инактивировать АФК. Они включают как ферменты (супероксиддисмутазы, каталазы, пероксидазы), так и низкомолекулярные вещества, содержащие легко окисляемую группу (альдегиды, тиолы, гидрохиноны и т.п.). При этом физиологическая значимость перехватчика АФК определяется тем, способен ли он

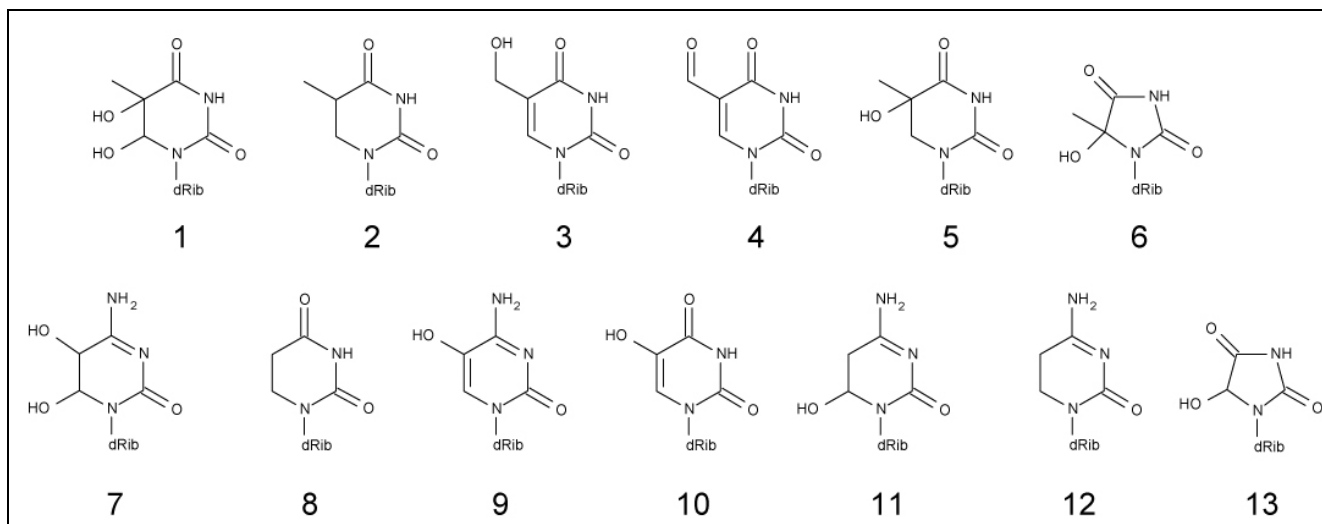
накапливаться в клетке в значительных количествах и концентрируется ли в чувствительных к окислительным повреждениям компартментах. В число таких перехватчиков входят как гидрофильные (аскорбиновая и мочевая кислоты, глутатион, цистеин, глюкоза), так и липофильные ( $\alpha$ -токоферол,  $\beta$ -каротин, билирубин) соединения.

### 2.1.2. Основные окислительные повреждения ДНК

АФК способны окислять множество биологически значимых соединений, и ДНК — не исключение. Известно несколько десятков модифицированных азотистых оснований, возникающих при её окислительном повреждении [4]. Окисление оснований может приводить к возникновению мутаций, гибели клетки из-за блокирования репликации и нарушению регуляторных процессов в ДНК. Помимо них, окислению могут также подвергаться моно-, ди- и трифосфаты нуклеозидов, что может привести к возникновению мутаций при включении в ДНК окисленных звеньев при репликации [6]. Окисление сахарофосфатного остова ДНК приводит к образованию окисленных по разным положениям АП-сайтов, а их окисление по атомам С4' и С5' — к разрыву цепи [7, 8].

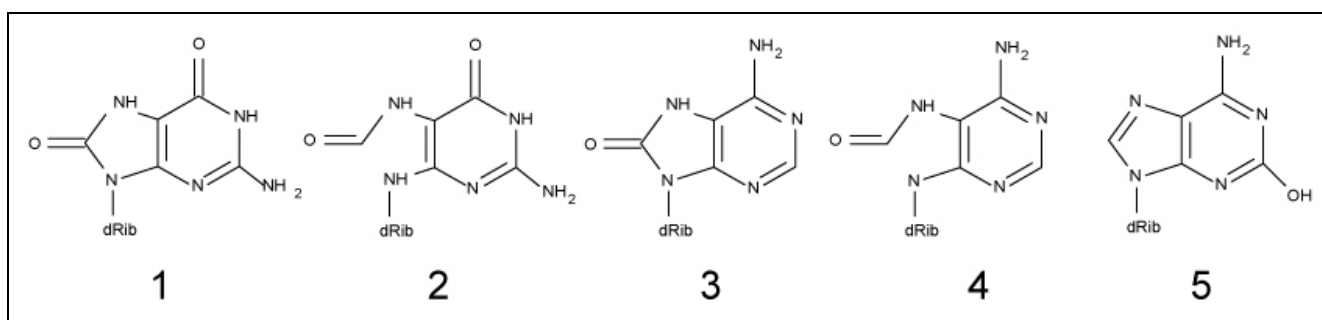
Окисление пиримидиновых оснований ДНК приводит к образованию нескольких основных и многих дополнительных продуктов (рисунок 1). Их спектр достаточно широк и различается в зависимости от окислительно-восстановительного потенциала раствора [4]. Основные продукты окисления тимина — тимингликоль, гидроксиметилурацил, 5-формилурацил, 5-гидрокси-5,6-дигидротимин и 5-гидрокси-5-метилгидантоин [9]. Основные продукты окисления цитозина — цитозингликоль, 5-гидроксицитозин (5-ОН-Cyt), 5-гидроксиурацил (5-ОН-Ura), 6-гидрокси-5,6-дигидроцитозин и 5-гидроксигидантоин [9]. Иногда к «окислительным повреждениям» пиримидинов относят также восстановленные основания 5,6-дигидроурацил, 5,6-дигидротимин и 5,6-дигидроцитозин, так как они наряду с окисленными основаниями образуются в водных растворах под действием ионизирующей радиации [4]. Биологические последствия возникновения окисленных пиримидиновых оснований зависят от конкретного вида повреждения. Например, достаточно подробно исследовано взаимодействие ДНК-полимераз с 5-гидроксипиримидинами. Они умеренно блокируются 5-ОН-Ura и 5-ОН-Cyt, но также способны включать различные dNMP в зависимости от контекста [10]. Транслезионные ДНК-полимеразы способны включать напротив 5-гидроксипиримидинов любой dNMP, но в случае 5-ОН-Ura преимущественно включают dGMP [11]. Это приводит к снижению мутагенного потенциала повреждения, так как 5-ОН-Ura образуется путём окислительного дезаминирования Cyt [11].





**Рисунок 1.** Окислительные повреждения пиримидиновых оснований: тимингликоль (1), 5,6-дигидротимин (2), 5-гидроксиметилурацил (3), 5-формулурацил (4), 5-гидрокси-6-гидротимин (5), 5-гидрокси-5-метилгидантоин (6), цитозингликоль (7), 5,6-дигидроурацил (8), 5-гидроксицитозин (9), 5-гидроксиурацил (10), 6-гидрокси-5,6-дигидроцитозин (11), 5,6-дигидроцитозин (12), 5-гидроксигидантоин (13).

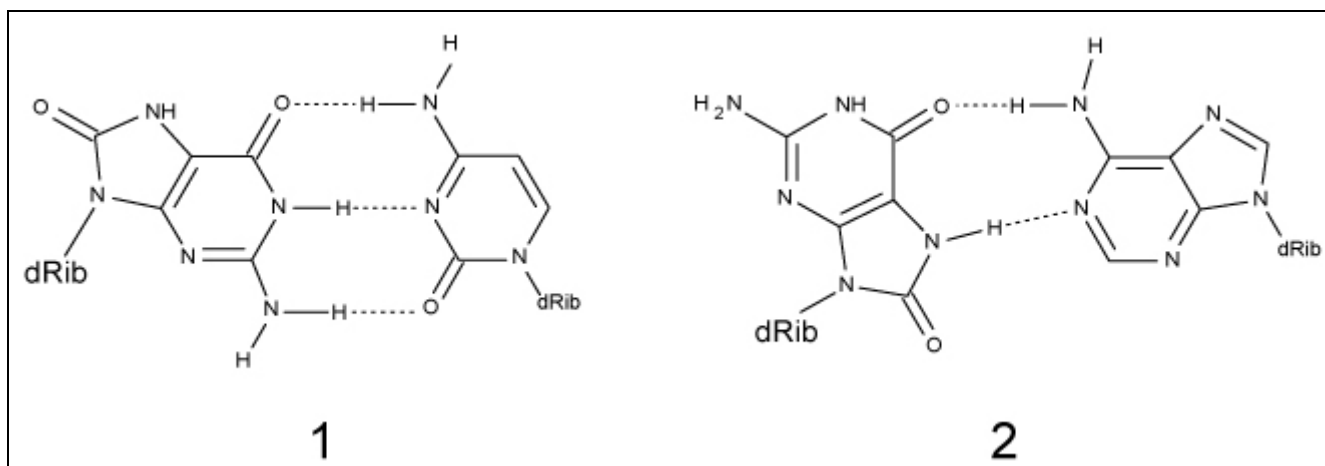
Основные продукты окисления оснований пуринового ряда — это 8-оксоаденин (8-охоAde) и, вероятно, наиболее изученное окисленное основание ДНК, 8-оксогуанин (8-охоGua). Они образуются при атаке гидроксильного радикала на атом С8 пуринового основания с образованием 8-гидроксипуринового радикала, который в дальнейшем окисляется до 8-оксопродукта [12]. Кроме того, радикал способен восстанавливаться с раскрытием пуринового кольца и образованием 2,6-диамино-4-оксо-5-формапидопиримидина (Fapy-Gua) или 4,6-диамино-5-формапидопиримидина (Fapy-Ade) [12]. Также после действия ионизирующего излучения или ряда оксидантов в клеточной ДНК выявляется 2-гидроксиаденин [13]. Структурные формулы основных продуктов окисления пуринов показаны на рисунке 2.



**Рисунок 2.** Окислительные повреждения пуриновых оснований: 8-оксогуанин (1), 2,6-диамино-4-оксо-5-формапидопиримидин (2), 8-оксоаденин (3), 4,6-диамино-5-формапидопиримидин (4), 2-гидроксиаденин (5).

8-охоGua — одно из наиболее часто встречающихся в природе повреждённых оснований. Первоначально оно было описано как 8-гидроксигуанин, но структурные исследования и результаты компьютерного моделирования *ab initio* показали, что в физиологических условиях кето-енольное равновесие при связи N7—C8 сдвинуто в сторону C8-кето-таутомера [14, 15]. Наличие этой кетогруппы приводит к тому, что свободный 8-

8-oxodGuo предпочтительно существует в *син*-конформации, в отличие от канонических пуриновых дезоксирибонуклеозидов. Эта конформация также стабилизируется при образовании хугстенновской пары с dAdo, и в меньшей степени — с dGuo и dThd. Пару с цитидином 8-oxodGuo образует в *анти*-конформации, и эта пара близка к канонической dGuo:dCyd (рисунок 3). Таким образом, потенциальное формирование устойчивой пары 8-oxodGuo:dAdo объясняет мутагенность этого повреждения, а именно — трансверсию G→T после прохождения второго цикла репликации в отсутствие репарации.



**Рисунок 3.** Образование дезоксинуклеозидом 8-oxodGuo пар уотсон-криковского типа 8-oxodGuo(*анти*):dCyd(*анти*) (1) и пар хугстенновского типа 8-oxodGuo(*син*):dAdo(*анти*) (2).

В результате стандартизации методик определения количества этого повреждённого основания уровень эндогенного 8-oxoGua был оценен в 0,3—4,2 8-oxodGuo на  $10^6$  dGuo и показана возможность его повышения в несколько раз при генотоксическом стрессе [16], в том числе при старении, спонтанном развитии новообразований и ишемии–реперфузии [17]. Возрастание уровня 8-oxodGua в промоторных участках генов приводит к снижению их экспрессии, что наблюдается, например, в коре головного мозга при старении [18]. Упомянутые факты указывают на высокую биологическую и медицинскую значимость изучения этого повреждённого основания, а также ферментов, участвующих в его репарации.

Говоря о продуктах окислительного повреждения ДНК, нельзя не упомянуть АП-сайты, которые образуются в ДНК не только в результате окислительного стресса, но и в результате спонтанной апуринизации, т.е. гидролиза *N*-гликозидной связи в пуриновых нуклеозидах; в пиримидиновых она происходит на 1–2 порядка медленнее. Кроме того, они возникают как интермедиаты в процессе ЭРО (см. далее). Число АП-сайтов, присутствующих в любой момент времени в клетках млекопитающих, составляет ~5–20 на  $10^6$  оснований [19].

АП-сайты химически нестабильны, и на их месте под действием повышенной температуры, оснований, нуклеофильных реагентов и ионов  $Mg^{2+}$  могут возникать

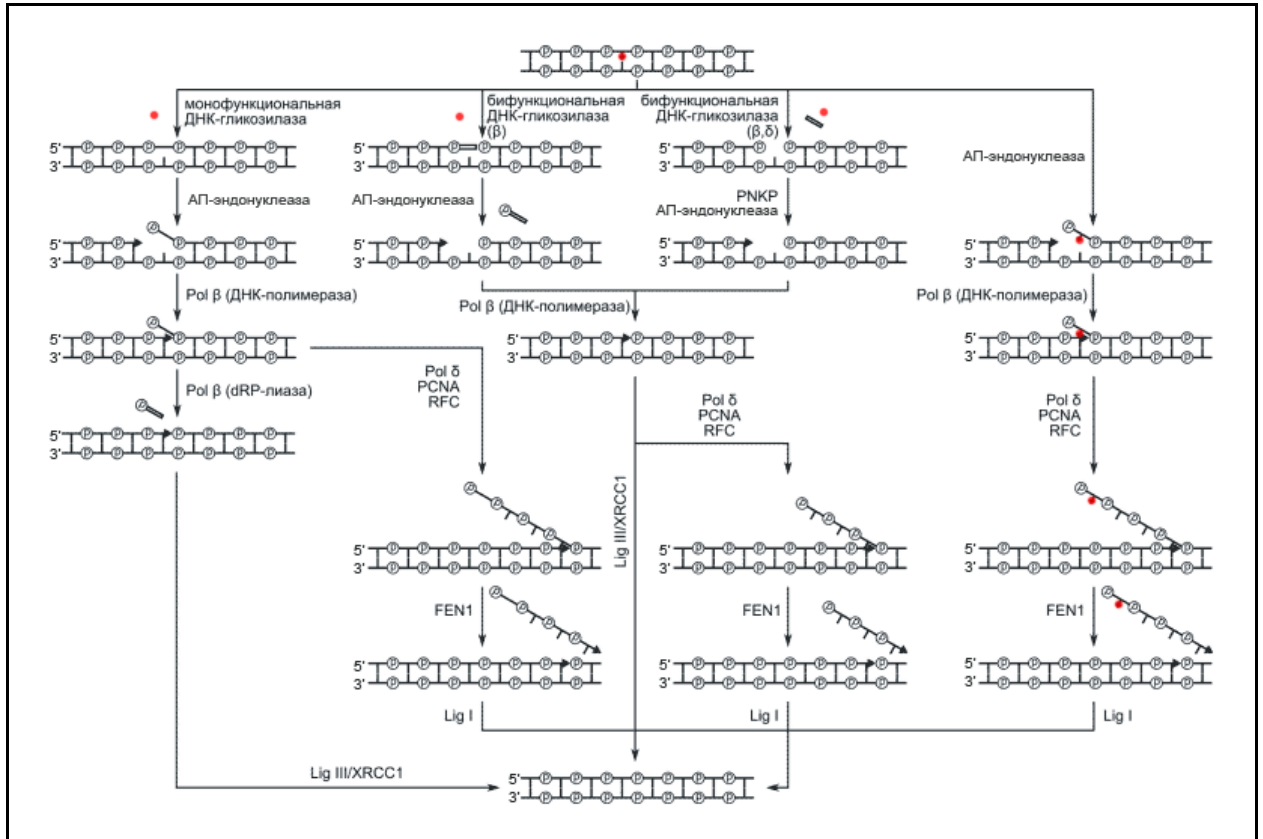
одноцепочечные разрывы ДНК. В связи с этим они обладают цитотоксическим и мутагенным действием [2, 20].

### 2.1.3. Эксцизионная репарация оснований

Объектами ЭРО служат в основном дезаминированные, алкилированные, окисленные и некоторые неправильно спаренные основания, а также АП-сайты [2, 20]. Эти типы повреждений не вызывают серьёзных нарушений в структуре двойной спирали ДНК. В целом механизм ЭРО, возможно, является наиболее древним и важным [21].

Механизм эксцизионной репарации с участием ДНК-гликозилаз выглядит следующим образом. Повреждённое основание узнаётся специфичной к нему ДНК-гликозилазой, которая гидролизует *N*-гликозидную связь между повреждённым основанием и сахарофосфатным остовом, что приводит к образованию АП-сайта и свободного основания. ДНК-гликозилазы различаются по своей субстратной специфичности и по механизмам действия. Большинство из них обладает достаточно широкой специфичностью, но некоторые — очень узкой. Ряд ДНК-гликозилаз обладает не только гликозилазной, но и АП-лиазной активностью, которая расщепляет ДНК по вновь образованным АП-сайтам по механизму  $\beta$ -элиминирования. Более подробно ДНК-гликозилазы описаны ниже.

АП-сайты репарируются системой, включающей ферменты АП-эндонуклеазы и, возможно, АП-лиазные активности ДНК-гликозилаз. АП-эндонуклеазы осуществляют гидролиз связи между атомом фосфора АП-сайта и 3'-О-атомом нуклеозида с 5'-стороны от АП-сайта; при этом образуются 5'-концевой остаток 2-дезоксирибо-5-фосфата и нуклеотид с 3'-гидроксильной группой. АП-эндонуклеазы ответственны за репарацию как спонтанно возникающих АП-сайтов, так и АП-сайтов, образующихся в ходе гидролиза *N*-гликозидной связи монофункциональными гликозилазами без лиазной активности. У *E. coli* имеются два негомологичных друг другу фермента, субстратами которых являются АП-сайты — эндонуклеаза IV (Nfo) и экзонуклеаза III (Xth). В дополнение к АП-эндонуклеазной активности, оба этих белка обладают 3'-диэстеразной активностью и способны удалять 3'-фосфат и 3'-фосфогликолатные аддукты; кроме того, ферменту Xth присуща 3'→5'-экзонуклеазная активность.



**Рисунок 4.** Общая схема различных путей эксцизионной репарации оснований. Приведены названия ферментов и белковых факторов эукариот. Стрелками в структуре ДНК показаны 3'-гидроксильные группы, прямоугольниками — ненасыщенные остатки сахара. Повреждённое основание ДНК выделено красным цветом (по [22]).

Главными АП-эндонуклеазами *E. coli* и клеток млекопитающих считаются Xth и гомологичный ей белок APEX1 соответственно. Отличительная их черта — наличие ярко выраженной 3'-фосфодиэстеразной активности, позволяющей удалять продукты атаки оснований ДНК различными свободными радикалами или активными формами кислорода. АП-эндонуклеаза человека обладает, помимо собственных эндонуклеазной и фосфодиэстеразной активностей, еще и окислительно-восстановительной функцией [23]. APEX1 действует как фактор, поддерживающий некоторые факторы транскрипции в активном (восстановленном) состоянии, стимулируя их ДНК-связывающую активность. У дрожжей, в отличие от высших эукариот, главной АП-эндонуклеазой является белок Ape1p, гомологичный не Xth, а Nfo, и практически идентичный последнему по своим ферментативным свойствам.

После удаления повреждённого основания в ДНК у млекопитающих возможны два механизма репарации (рисунок 4). В первом случае (short-patch repair, репарация с коротким ресинтезируемым участком) АП-эндонуклеаза и дезоксирибофосфатлиаза создают брешь размером в один нуклеотид, которая застраивается ДНК-полимеразой (Pol I прокариот или POL β эукариот), после чего разрыв ликвидируется ДНК-лигазой (LigA прокариот или LIG3α эукариот). Помимо упомянутых ферментов, у эукариот в репарации по этому механизму

участвует белок XRCC1 (X-Ray Cross-Complementation Protein 1), N-концевой участок которого взаимодействует с ДНК-полимеразой, а С-концевой участок — с ДНК-лигазой III, играя структурную роль. Второй путь репарации (long-patch repair; репарация с длинным ресинтезируемым участком) предполагает замену 2—13 нуклеотидов в 3'-направлении от повреждённого остатка. Ресинтез ДНК у прокариот продолжается той же Pol I, она же удаляет «свободно свешивающийся фрагмент» ДНК (флэп). У эукариот ресинтез ДНК осуществляется либо ДНК-полимеразой  $\beta$ , либо полимеразами  $\delta$  или  $\epsilon$ , функционирование которых зависит от белков PCNA (Proliferating Cell Nuclear Antigen), RFC (Replication Factor C) и, возможно, других, еще не определённых факторов; затем ДНКаза FEN1 (Flap Endonuclease 1) удаляет флэп, а последняя фосфодиэфирная связь восстанавливается ДНК-лигазой I или, в некоторых случаях, комплексом ДНК-лигаза III/XRCC1 [24]. Считается, что альдегидные АП-сайты репарируются по первому пути, а модифицированные (окисленные или восстановленные) — по второму.

Обнаруженный в 2002 году гликозилазозависимый механизм ЭРО дал основания выделить ещё одну её ветвь — инцизионную репарацию нуклеотидов [25]. Процесс начинается с расщепления цепи ДНК с 5'-стороны от повреждённого основания АП-эндонуклеазой и, так как дезоксирибофосфатлиазная активность POL  $\beta$  не способна удалить участок цепи, содержащий повреждённый нуклеотид, неизбежно идет по длиннозаплаточному пути [25].

Гликозилазы — это высокоспецифичные ферменты, узнающие, как правило, только определённый тип повреждённых оснований. Все известные ДНК-гликозилазы являются относительно небольшими мономерными белками, для функционирования которых не требуются кофакторы. Эти ферменты не зависят от ионов  $Mg^{2+}$ , что позволяет отличать их от магнийзависимых нуклеаз репарации [20]. ДНК-гликозилазы принято подразделять на два класса: бифункциональные, обладающие АП-лиазной активностью, и монофункциональные, ею не обладающие (см. раздел 2.1.7). Основные ДНК-гликозилазы *E. coli* и человека перечислены в таблице 1.

Поиск, узнавание и удаление повреждённых нуклеотидов ДНК-гликозилазами сопряжены со значительными структурными изменениями в молекулах ферментов и ДНК. Методом рентгеноструктурного анализа (РСА) установлено пространственное строение многих ДНК-гликозилаз. Большую роль в структурной биологии ДНК-гликозилаз сыграла разработка двух методов получения ковалентных сшивок этих ферментов с ДНК. Первый метод, борогидридная сшивка, применим только для ДНК-гликозилаз с АП-лиазной активностью и основан на восстановлении с помощью  $NaBH_4$  оснований Шиффа, образующихся в ходе катализа [27]. Эта реакция широко используется для доказательства АП-лиазной функции ДНК-

гликозилаз и для идентификации аминокислотных остатков (а. о.), играющих роль нуклеофилов. Восстановленные комплексы ДНК-гликозилаз легко кристаллизуются и дают структуры с высоким разрешением [28, 29]. Главным недостатком этого метода является выщепление повреждённого основания, что не дает возможности получить информацию о контактах, образуемых им в активном центре фермента.

**Таблица 1.** ДНК-гликозилазы *E. coli* и человека и их субстратная специфичность [26].

<i>E. coli</i>	Человек	Специфичность	Моно- или бифункциональная	Структурное суперсемейство
Ung	UNG	Ura в дц- и оцДНК, любой контекст	Моно	Ung
Mug	TDG	Ura в Ura:Gua, Thy в Thy:Gua (только TDG), C5-окисленные производные Thy и 5-метилцитозина, этеноцитозин	Моно	Ung
	SMUG1	Ura в оцДНК, гидроксиметилурацил, формилурацил	Моно	Ung
	MBD4	Thy и C5-окисленные производные Thy и 5-метилцитозина напротив Gua в CG-динуклеотидах	Моно	Ung
Tag		Алкилированные в кольцо пурины	Моно	Nth
AlkA	MPG	Алкилированные в кольцо пурины, гипоксантин, этеноаденин	Моно	Nth
Nth	NTHL1	Окисленные пиримидины	Би	Nth
Nei	NEIL1	Окисленные пиримидины или	Би	Fpg/Nei
	NEIL2	пурины	Би	Fpg/Nei
	NEIL3		Би	Fpg/Nei
Fpg		Окисленные пурины	Би	Fpg/Nei
	OGG1	Окисленные пурины	Би	Nth
MutY	MUTYH	Ade в Ade:Gua и Ade: 8-oxoGua	Моно	Nth

Второй подход — введение дисульфидной сшивки — в принципе, применим для анализа любых ДНК-связывающих белков. На основании уже полученных данных РСА делают выбор контакта ДНК с ферментом, по которому можно образовать дисульфидный мостик путем введения тиогруппы в соответствующее положение ДНК, и замены контактирующего

аминокислотного остатка на остаток цистеина [30]. Это позволяет легко установить структуру анализируемого фермента в комплексе с различными ДНК. В частности, метод дисульфидной сшивки использовали для того, чтобы с помощью известных структур комплексов ДНК-гликозилаз OGG1 и Fpg с повреждённой ДНК установить структуру этих же белков, связанных с неповреждённой ДНК. Без сшивки фермента с ДНК это представляется невозможным из-за их относительно низкого сродства к ДНК и статистической разнородности комплексов [31]. Однако такие структуры необходимо интерпретировать с осторожностью, поскольку введение сшивки может способствовать избирательной кристаллизации комплекса в той конформации, которая больше всего напоминает комплекс, использованный для определения мест модификации.

Сравнение данных РСА и аминокислотных последовательностей выявляет общие черты ДНК-гликозилаз [20]. У многих этих ферментов в активном центре обнаружен один и тот же повторяющийся мотив «спираль-шпилька-спираль» («helix-hairpin-helix», HhH) [20, 27]. Впервые мотив HhH обнаружен в ДНК-связывающем центре эндонуклеазы III *E. coli* [20]. С тех пор подобная структура обнаружена в ДНК-гликозилазах AlkA и Tag (*E. coli*) и Mag1p (дрожжи), удаляющих алкилированные основания; в 8-оксогуанин-ДНК-*N*-гликозилазе эукариот (OGG1); гликозилазах MutY (*E. coli*) и MUTYH (эукариоты), специфичных к неспаренному остатку аденина, и в некоторых других ферментах репарации. Помимо мотива HhH, в ДНК-связывающих центрах многих гликозилаз репарации обнаружены консервативный остаток Asp и Pro/Gly-богатый участок (GPD-петля) [20]. Эти наблюдения позволяют предполагать, что механизм узнавания повреждённых оснований моно- и бифункциональными ДНК-гликозилазами сходен [20]. Хотя каталитические механизмы выщепления повреждённых оснований могут отличаться, активные центры этих ферментов, по-видимому, могут связываться только с «вывернутым» из спирали ДНК основанием, как это показано на примере многих ДНК-гликозилаз [2].

## 2.1.4. ДНК-гликозилазы

### 2.1.4.1. Структурные суперсемейства

Известные ДНК-гликозилазы можно подразделить на три главных суперсемейства (таблица 1) [26]. Самое крупное и разнообразное из них — суперсемейство эндонуклеазы III (Nth), которое включает семейства собственно эндонуклеазы III, MutY, эндонуклеазы пиримидиновых димеров (Pdg2) и термофильной гликозилазы, выщепляющей неправильно спаренный тимин (MIG) [32, 33]. Все принадлежащие к этим семействам белки состоят как минимум из двух доменов (MutY — из трех), структурную основу одного из которых

составляет «бочонок» из 6  $\alpha$ -спиралей (6НВ-домен), а другой (FCL-домен) содержит 4  $\alpha$ -спирали и железо-серный кластер типа  $Fe_4S_4$  [26]. В 6НВ-домene расположены характеристические элементы данного суперсемейства — мотив HhH и GPD-петля. К этим центральным семействам примыкают архейная 8-оксогуанин-ДНК-гликозилаза AGOG [34] и атипичная 3-метиладенин-ДНК-гликозилаза MagIII [35], у которых сохранен 6НВ-домен, а FCL-домен в целом похож на имеющийся у белков центральных семейств, но не содержит железо-серного кластера [26]. В суперсемейство Nth входят также 8-оксогуанин-ДНК-гликозилаза эукариот (OGG1), алкиладенин-ДНК-гликозилазы бактерий AlkA и низших эукариот Mag и CpG-специфичная урацил-ДНК-гликозилаза MBD4 [1], у которых есть HhH-мотив и GPD-петля, но структуры доменов уже довольно сильно отличаются от прототипических белков и появляются другие структурные элементы [26]. Еще дальше от центральных семейств отстоят бактериальная 3-метиладенин-ДНК-гликозилаза Tag, обладающая HhH-мотивом, но лишённая GPD-петли, и эндонуклеаза V бактериофага T4 (DenV), у которой тоже нет GPD-петли, а вторая спираль HhH-мотива заменена неспиральной последовательностью [1].

Суперсемейство Fpg/Nei организовано вокруг центрального семейства Fpg/Nei [32, 33, 36]. Члены этого семейства — двухдоменные белки; основу их N-концевого домена составляет двухслойный антипараллельный  $\beta$ -сэндвич из 8  $\beta$ -складок, а C-концевой домен включает два консервативных структурных мотива — «спираль-два поворота-спираль» (H2TH) и цинковый палец типа Cys<sub>4</sub> [26]. К центральному семейству можно также отнести эукариотические ДНК-гликозилазы NEIL2 и NEIL3 (NEIL3 содержит ряд дополнительных доменов, но все основные структурные элементы в нем консервативны) [26]. Еще один эукариотический гомолог, NEIL1, вместо цинкового пальца содержит структуру из двух  $\beta$ -складок, не содержащую ионов металлов (так называемый «бесцинковый палец»), но в целом аналогичную цинковому пальцу Fpg/Nei. Кроме того, известен растительный белок MMH, где цинковый палец заменён положительно заряженным ДНК-связывающим мотивом. Оба этих фермента также относят к суперсемейству Fpg/Nei.

В суперсемейство урацил-ДНК-гликозилазы (Ung) входят шесть семейств [37]: 1) урацил-ДНК-гликозилазы (Ung); 2) гликозилазы неправильно спаренного урацила (Mug); 3) гликозилазы урацилов в оцДНК (SMUG1); 4) прокариотические урацил-ДНК-гликозилазы Unga, полный спектр субстратной специфичности которых пока не известен; 5) прокариотические урацил-ДНК-гликозилазы Ungb, удаляющие также ксантин и гипоксантин из дцДНК; 6) гипоксантин-ДНК-гликозилазы. Все эти белки объединены центральным ядром из

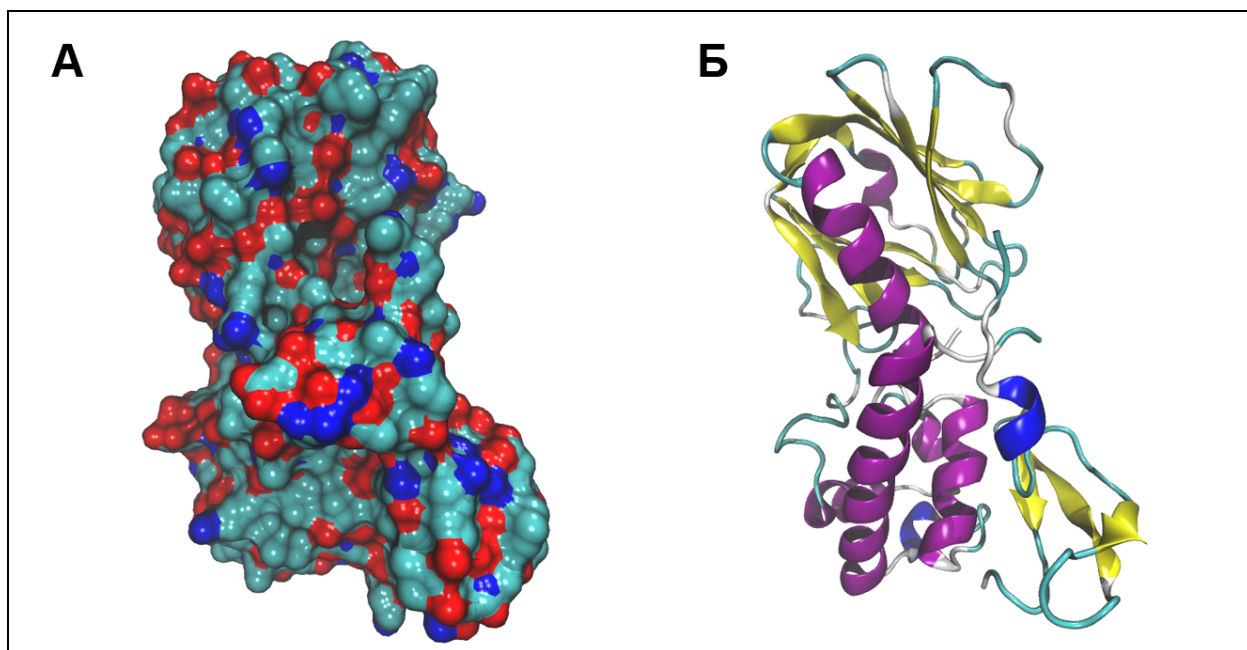


4  $\beta$ -складок, зажатых между 2  $\alpha$ -спиралями, однако гомология их последовательностей между отдельными семействами очень низка.

Ниже более подробно рассмотрены некоторые гликозилазы ЭРО, участвующие в удалении окисленных гетероциклических оснований.

#### 2.1.4.2. Формамидопиримидин-ДНК-*N*-гликозилаза

Формамидопиримидин-ДНК-*N*-гликозилаза *E. coli* (Fpg, или MutM) — фермент, узнающий и удаляющий из ДНК окисленные гетероциклические основания пуринового ряда: 8-охоGua, Fapy-Ade, Fapy-Gua [9, 38]. Это мономерный глобулярный белок (рисунок 5, А) с молекулярной массой 30,2 кДа, состоящий из 268 аминокислот [39]. Фермент, будучи металлопротеином, содержит один атом цинка на молекулу, который входит в мотив «цинкового пальца» (Cys-X<sub>2</sub>-Cys-X<sub>16</sub>-Cys-X<sub>2</sub>-Cys), локализованный на С-конце полипептидной цепи. Мотив третичной структуры «цинкового пальца», расположенный на С-конце Fpg, важен как для связывания, так и для катализа; он идентифицирован во многих ДНК-связывающих белках [39]. Последовательность Fpg не содержит мотивов HhH и GPD [36], однако С-концевой домен содержит четыре  $\alpha$ -спирали, из которых две ( $\alpha$ D и  $\alpha$ E) образуют родственный мотив «спираль-два поворота-спираль» (H2TH), также участвующий в связывании ДНК. N-концевой домен Fpg содержит  $\beta$ -сэндвич, состоящий из восьми  $\beta$ -складок, и две  $\alpha$ -спирали, параллельные его боковым сторонам (рисунок 5, Б).



**Рисунок 5.** Трёхмерная структура ДНК-гликозилазы Fpg *E. coli* (PDB ID: 1K82). (А) Поверхность белка, доступная растворителю. Бирюзовым цветом показаны атомы углерода, красным — кислорода, синим — азота. (Б) Вторичная структура белка. Жёлтым цветом показаны  $\beta$ -складки, фиолетовым и синим —  $\alpha$ -спирали, белым — неупорядоченные участки.

Наличие мотива H2TH и «цинкового пальца» определяет суперсемейство ДНК-гликозилаз, родственных Fpg.

Помимо *N*-гликозилазной, Fpg обладает еще двумя активностями: АП-лиазной и дезоксирибозофосфатлиазной [39]. АП-лиазная активность Fpg не ограничивается одним шагом  $\beta$ -элиминирования, а представляет собой две последовательные реакции элиминирования ( $\beta/\delta$ -элиминирование). В результате дезоксирибоза повреждённого нуклеотида высвобождается в виде 4-оксо-2-пентенала, а в ДНК образуется одноцепочечная брешь, фланкированная 3'- и 5'-концевыми фосфатами. При этом происходит последовательное расщепление фосфодиэфирных связей сначала с 3'-конца, а затем с 5'-конца от АП-участка [39].

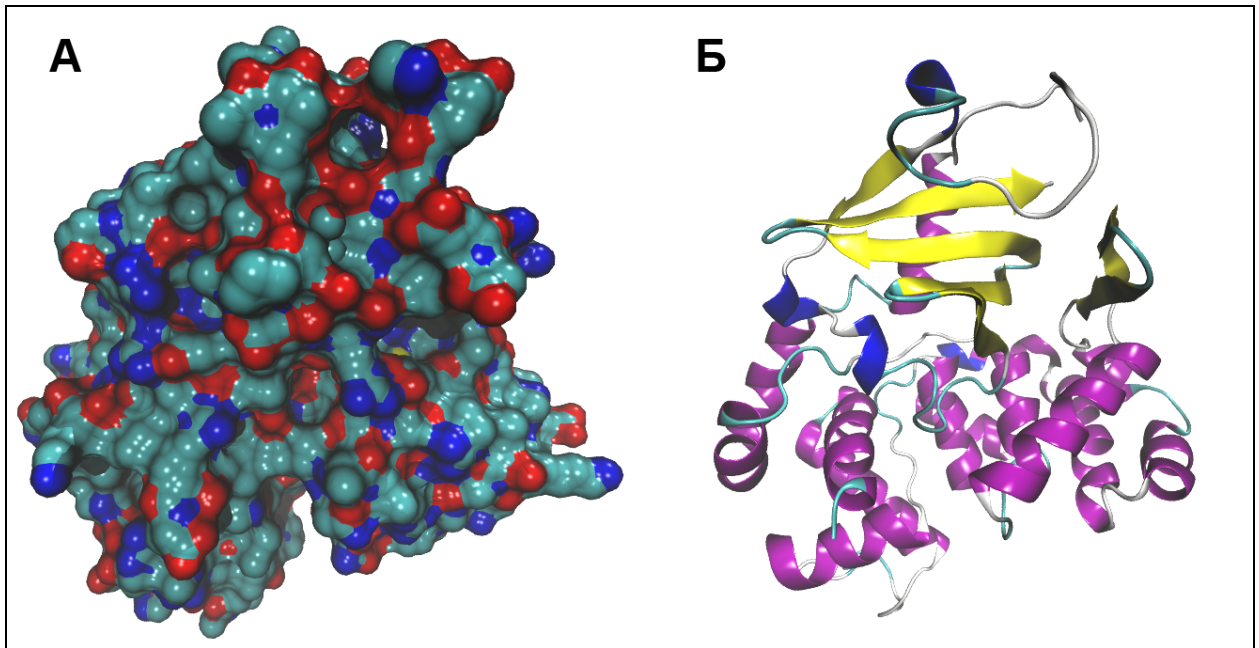
Активный центр фермента Fpg, определённый по данным РСА и аффинной модификации, содержит каталитически активный остаток N-концевого пролина, который образуется в ходе биосинтеза после отщепления иницирующего остатка формилметионина. Вторичная аминогруппа Pro1 депротонируется, возможно, с участием остатка Glu2 и вступает в реакцию с образованием основания Шиффа с атомом C1' повреждённого нуклеотида [26].

#### 2.1.4.3. 8-оксогуанин-ДНК-*N*-гликозилаза эукариот

Несмотря на то, что 8-оксогуанин-ДНК-*N*-гликозилаза эукариот (OGG1) перекрывается по субстратной специфичности с прокариотическим ферментом Fpg, эти белки не гомологичны друг другу [9]. Ген 8-оксогуанин-ДНК-*N*-гликозилазы (*OGG1*) из дрожжей был клонирован на основании его способности комплементировать *fpg*-дефицитные штаммы *E. coli*, затем по гомологии с *OGG1* из дрожжей были клонированы соответствующие гены высших эукариот.

Белок OGG1 из клеток человека — глобулярный (рисунок 6, А) и имеет молекулярную массу 38 кДа [39]. Анализ строения человеческого гена *ogg1* и его транскриптов показал, что в результате альтернативного сплайсинга образуются две группы изоформ мРНК *ogg1*, при этом в одной из групп транслируемый белок содержит сигнал ядерной локализации, а в другой — нет. В белке присутствуют два  $\alpha$ -спиральных домена, а также третий домен, организованный вокруг семи антипараллельных  $\beta$ -складок (рисунок 6, Б) [26].

В активном центре OGG1 содержит структурный мотив HhH-GPD, характерный для белков суперсемейства эндонуклеазы III. В этом мотиве, как правило, находится основная аминокислота, участвующая в образовании основания Шиффа [9]. Установлено, что в образовании основания Шиффа в случае OGG1 непосредственное участие принимает  $\epsilon$ -NH<sub>2</sub> группа остатка Lys249. OGG1 выщепляет из ДНК остатки 8-охоGua и Fapy-Gua, расположенные напротив Cyt. Фермент также способен расщеплять ДНК, содержащую АП-участки, по механизму  $\beta$ -элиминирования [9].



**Рисунок 6.** Трёхмерная структура ДНК-гликозилазы OGG1 человека (PDB ID: 1EBM). (А) Поверхность белка, доступная растворителю. Бирюзовым цветом показаны атомы углерода, красным — кислорода, синим — азота, жёлтым — серы. (Б) Вторичная структура белка. Жёлтым цветом показаны  $\beta$ -складки, фиолетовым и синим —  $\alpha$ -спирали, белым — неупорядоченные участки.

#### 2.1.4.4. Другие ДНК-гликозилазы, исправляющие окислительные повреждения

Следует кратко рассмотреть и другие ДНК-гликозилазы, субстратами которых являются окисленные нуклеозиды.

Монофункциональная ДНК-гликозилаза MutY принадлежит к структурному суперсемейству эндонуклеазы III и состоит из трёх доменов. Фермент удаляет аденин из неправильно образованных пар: с гуанином и 8-оксогуанином [40]. Структура полноразмерного белка MutY была определена для фермента из *Bacillus stearothermophilus* с использованием метода дисульфидной сшивки с повреждённой ДНК [41]. Структура свободного MutY известна только для белка из *E. coli* [40], при этом закристаллизован был лишь фрагмент фермента, состоящий из 6NB- и FCL-доменов, но лишённый третьего домена, сходного по строению с нуклеозиддифосфатгидролазами семейства NUDIX [41]. Аналогично, структура человеческого гомолога MUTYH в свободном виде была тоже получена лишь для каталитического фрагмента (а. о. 76—362) [42]. При секвенировании гена *mutY* обнаружено, что фермент имеет высокую степень гомологии с Nth (см. ниже) [39]. Ограниченный трипсинолиз приводит к отделению третьего С-концевого домена, не гомологичного Nth, с образованием фрагментов с молекулярными массами 26 и 13 кДа. Первый фрагмент содержит 15  $\alpha$ -спиралей и сохраняет все присущие MutY каталитические свойства. Детерминанты узнавания повреждённого основания, вероятно, находятся в С-концевом домене, состоящем из восьми  $\beta$ -складок,

обрамлённых одной спиралью с одной стороны и двумя  $\alpha$ -спиралями с другой [39]. Считается, что фермент MutY способен с очень малой эффективностью осуществлять разрыв цепи ДНК [43], но эта реакция протекает на порядки медленнее гликозилазной и, по всей вероятности, обусловлена присутствующими в активном центре нуклеофильными остатками, не приспособленными специально для этой роли. Число оборотов MutY ограничено ввиду прочного связывания с АП-продуктом [43].

Эндонуклеаза III (Nth) узнаёт и специфически выщепляет из ДНК окисленные гетероциклические основания пиримидинового ряда [9]. Nth из *E. coli* представляет собой мономерный глобулярный белок с молекулярной массой 23,4 кДа, состоящий из 211 а. о. [39]. Показано, что Nth выщепляет из двуцепочечной ДНК повреждённые пиримидиновые основания с насыщенным, открытым или фрагментированным гетероциклом, в том числе тимингликоль, 5,6-дигидротимин, 5-гидрокси-6-гидротимин, 5,6-цигидроурацил, аллоксан, 5-гидрокси-6-гидроурацил, урацилгликоль, 5-ОН-Cyt, 5-ОН-Ura,  $\beta$ -уреидоизомаляновую кислоту, 5-гидрокси-5-метилгидантоин, метилартронилмочевину, а также продукты фрагментации дигидротимидина и (5R)-тимидин-C5-гидрата. Помимо N-гликозилазной активности, фермент Nth обладает еще и АП-лиазной активностью и способен расщеплять фосфодиэфирные связи, непосредственно прилегающие к АП-участку, образуемому в результате расщепления, по механизму *син*- $\beta$ -элиминирования [39]. Гомолог Nth из клеток эукариот (NTHL1) очень близок к бактериальному ферменту по активности и субстратной специфичности.

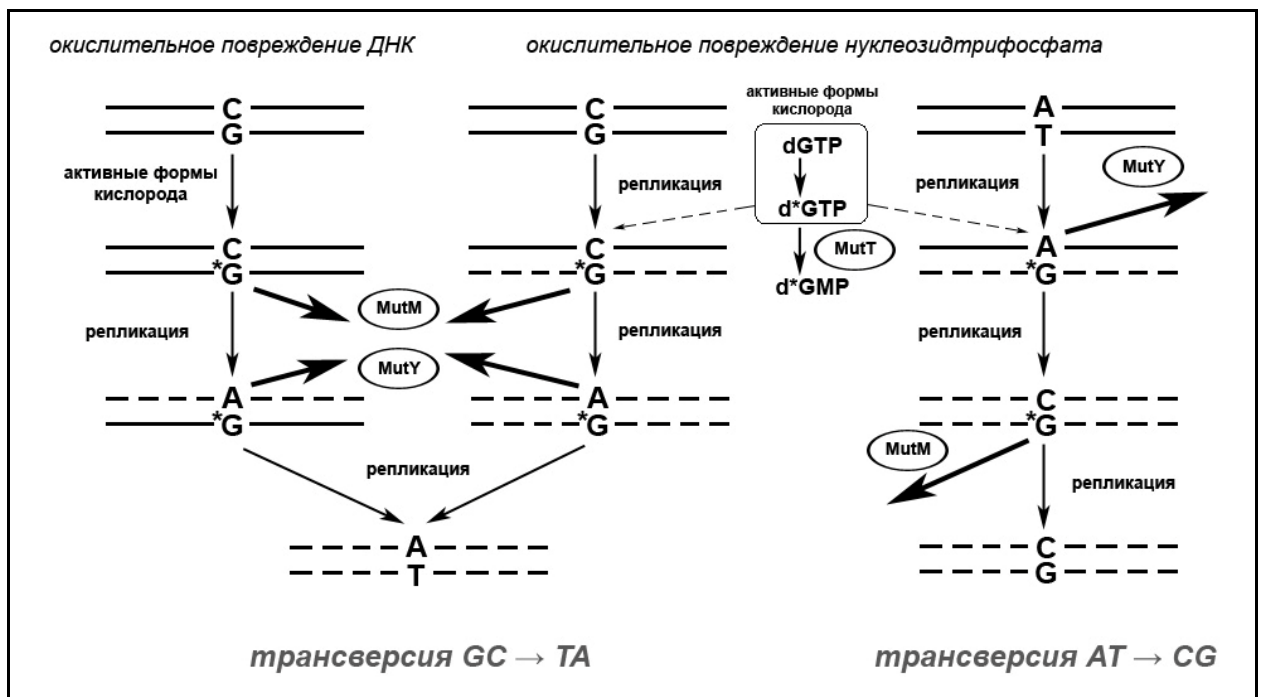
Эндонуклеаза VIII (Nei), узнаёт и выщепляет из ДНК широкий спектр окислительных повреждений: тимингликоли, дигидротимин,  $\beta$ -уреидоизомаляновую кислоту, окисленные производные цитозина, Fapy-Ade, а также остатки мочевины [44]. Таким образом, её субстратная специфичность во многом перекрывается с таковой Nth. Количество Nei, содержащееся в клетках *E. coli*, примерно в 20 раз меньше количества Nth (~5% Nei и 95% Nth) [39]. Ген *nei* кодирует белок, состоящий из 262 аминокислот, с молекулярной массой 29,7 кДа [39]. Несмотря на субстратную специфичность, очень сходную с ферментом Nth, аминокислотная последовательность Nei не имеет сходства с Nth, но обнаруживает высокую гомологию с бактериальной гликозилазой Fpg [45].

Продукты генов млекопитающих, имеющие сходство с Nei и Fpg, названы NEIL1, NEIL2 и NEIL3 [2, 46]. Все эти гомологи имеют ДНК-связывающий мотив H2TH, активный N-концевой нуклеофил (Pro1 у NEIL1 и NEIL2, Val1 у NEIL3) и необходимый для гликозилазной активности остаток Glu2 [46]. Предпочтительные субстраты для NEIL1 и NEIL2 — это окисленные пиримидины и остатки формамидопиримидинов. Для обоих этих ферментов хорошим субстратом служат АП-сайты [46]. Была также показана их

дезоксирибофосфатлиазная активность *in vitro* [46]. Функции белка NEIL3 долгое время оставались неизвестными, однако в итоге он был выделен [47] и охарактеризован как бифункциональная гликозилаза с субстратной специфичностью в отношении окисленных пуринов (спироиминодигидантоина, Fapy-Gua, Fapy-Ade, но не 8-охоGua) [48, 49]. Исследования *in vivo* на штаммах *E. coli*, дефицитных по различным комбинациям гликозилаз, свидетельствуют о важной роли NEIL3 в репарации повреждённого основания Fapy-Gua [49].

### 2.1.5. GO-система

Накопление 7,8-дигидро-8-оксогуанозинтрифосфата в нуклеотидном пуле, равно как и окислительные повреждения ДНК с образованием 8-охоGua, индуцируют как трансверсии G:C → T:A, так и трансверсии A:T → C:G (рисунок 7). В клетках бактерий, в частности, *E. coli*, существует механизм предотвращения таких мутаций. Этот механизм заключается в функциональном взаимодействии ферментов MutT, MutM и MutY [50] и называется GO-системой. Возможны два пути включения 8-охоGua в ДНК. В первом случае происходит окисление остатка Gua в цепи. Это повреждённое основание репарируется формамидопиримидин-ДНК-*N*-гликозилазой (MutM, или Fpg), удаляющей из ДНК остаток 8-охоGua напротив Cyt. В случае, если репарации не произошло, ДНК-полимераза может включить остаток Ade напротив 8-охоGua. Для коррекции этого повреждённого нуклеотида



**Рисунок 7.** Пути мутагенеза, связанные с 8-охоGua, и механизм их репарации (GO-система). dGTP в нуклеотидном пуле и остатки гуанина в ДНК подвергаются воздействию активных форм кислорода. Окислительные повреждения цепей ДНК ведут к G:C → T:A трансверсиям, что предотвращается белками MutM и MutY. Наличие 8-oxodGTP в нуклеотидном пуле приводит к трансверсиям A:T → C:G и G:C → T:A, которые предотвращаются белком MutT, гидролизующим 8-oxodGTP до монофосфата. \*G – 8-оксогуанин (по [50]).

существует белок MutY, выщепляющий, в частности, остаток аденина из пары 8-охоGua:Ade. Таким образом, удастся предотвратить трансверсию G:C → T:A, связанную с окислительным повреждением остатка Gua.

Второй путь включения 8-охоGua в ДНК – окисление dGTP до 8-оходGTP в нуклеотидном пуле. Остаток 8-охоGua может встраиваться при репликации как напротив Cyt (тогда в работу включаются белки MutM и MutY, как описано выше), так и напротив Ade. В последнем случае гликозилаза MutY, удаляя «правильный» остаток аденина, приводила бы к трансверсиям A:T → C:G. Поэтому существует третий белок, MutT, который эффективно расщепляет нуклеозидтрифосфат 8-оходGTP до монофосфата и препятствует его включению в ДНК. Клетки *E. coli*, в которых не экспрессируется MutT, имеют увеличенную в 200 раз частоту мутаций, а при отсутствии экспрессии всех трёх белков частота возрастает почти в 500 раз [50].

Существование GO-системы показано и для эукариот: клонированы и охарактеризованы гомологи MutT у мышей и человека (NUDT1), описан структурный гомолог и функциональный аналог MutY у человека — MUTYH. У эукариот обнаружено три структурных гомолога Fpg (белки NEIL), хотя они не являются в полной мере функциональными аналогами Fpg. Главным же функциональным аналогом Fpg в эукариотической GO-системе является белок OGG1 [51, 52].

### 2.1.6. Структурные основы узнавания повреждённых оснований

Как уже упоминалось выше, узнавание повреждённых оснований ДНК-гликозилазами сопряжено со значительными структурными изменениями в молекулах ферментов и ДНК, и важную роль в их понимании сыграл метод РСА. Прежде всего нужно отметить, что для всех гликозилаз, структура которых определена в комплексе с ДНК, характерен излом ДНК в месте повреждения на ~60–75° [26]. Повреждённый нуклеозид «вывернут» из двойной спирали в карман активного центра. Выворачивание осуществляется путем сближения межнуклеозидных фосфатных групп за счет сжатия повреждённой цепи в области повреждения и стабилизируется боковыми группами аминокислот фермента. Такое связывание гликозилазы с повреждённой ДНК не может происходить путём обычного комплексообразования. Для этого необходим ряд конформационных перестроек в молекулах фермента и ДНК-субстрата (изгиб молекулы ДНК, выворачивание повреждённого нуклеотида, внедрение аминокислотных остатков фермента в образовавшуюся полость и т.п.), которые могут происходить последовательно или частично одновременно. Например, узнавание 8-охоGua в составе ДНК белком Fpg *E. coli* состоит, по меньшей мере, из пяти последовательных равновесных стадий, предшествующих выщеплению повреждённого основания. Эти стадии можно отнести к следующим конформационным

изменениям: 1) неспецифическое связывание фермента с ДНК; 2) первичное узнавание повреждённого основания, в ходе которого происходит дестабилизация ДНК вокруг него и, возможно, интеркалирование остатка Phe110 между основаниями, находящимися в стэкинг-взаимодействии; 3) выворачивание повреждённого нуклеотида из двуцепочечной ДНК в активный центр фермента; 4) внедрение остатков Met73 и Arg108 в образовавшуюся полость; 5) переход активного центра фермента в каталитически компетентную форму. Наличие нескольких стадий в процессе узнавания повреждённого основания открывает возможность тонкой регуляции субстратной специфичности ДНК-гликозилаз. Весьма вероятно, что истинные субстраты этих ферментов эффективно проходят все стадии узнавания, в то время как канонические основания или повреждённые основания, не являющиеся субстратами для данного фермента, могут отвергаться на разных стадиях узнавания [53]. Для того, чтобы понять, как именно происходит узнавание субстратов, следует подробно рассмотреть структуру и конформационную подвижность отдельных гликозилаз.

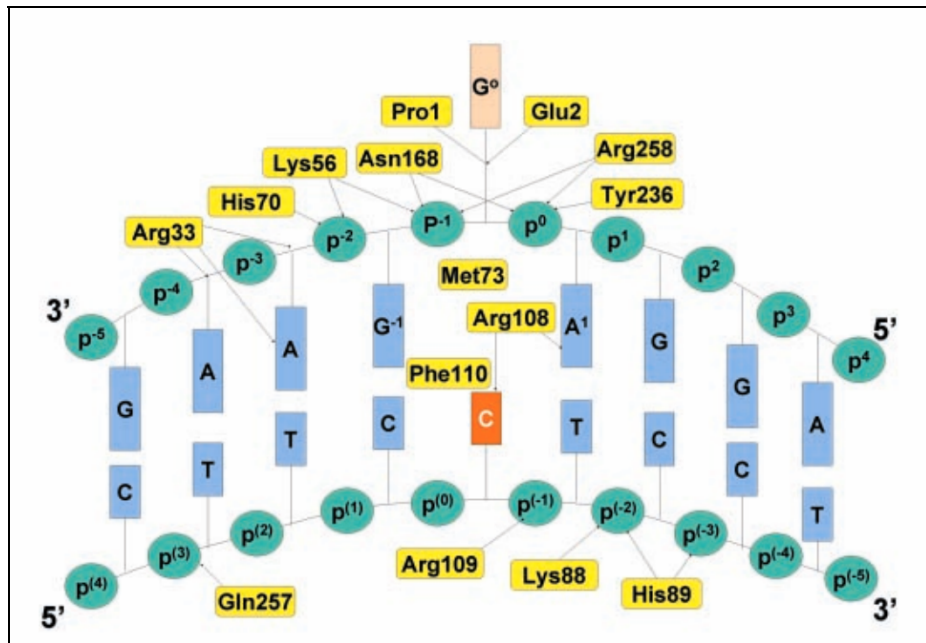
Гликозилаза Fpg состоит из N- и C-концевого доменов, соединенных гибким линкером. Граница между доменами образует положительно заряженную ДНК-связывающую бороздку, на её дне находится глубокий карман, в котором связывается повреждённое основание. На N-конце белка располагается описанный ранее каталитически активный дипептид Pro-Glu; мутации в нем приводят к полной потере каталитической активности фермента по отношению к субстратам, содержащим 8-охоGua. Как упоминалось в разделе 2.1.4.2, остаток Pro1 выступает в роли нуклеофила при атаке на атом C1' повреждённого нуклеотида и образует основание Шиффа в ходе реакции. Остаток Glu, по-видимому, участвует в активации Pro1 для нуклеофильной атаки и стабилизации открытой формы остатка дезоксирибозы [29, 45], а также в поддержании правильной конформации активного центра фермента [54]. На поворотах между отдельными складками  $\beta$ -сэндвича располагаются функционально важные аминокислотные остатки Met73, Arg108 и Phe110, которые внедряются внутрь спирали ДНК, фиксируя её искажённую конформацию, а Lys56 образует водородные связи с первым и вторым фосфатами с 3'-стороны от повреждённого нуклеотида.

Цинковый палец, расположенный в C-концевом домене Fpg, чрезвычайно важен для функционирования фермента. Обработка Fpg ионами  $Cd^{2+}$ ,  $Cu^{2+}$  или  $Hg^{2+}$  [26], а также мутации остатков цистеина в цинковом пальце (находящемся на C-конце полипептидной цепи), нарушающие его вторичную структуру, полностью инактивируют фермент [36], а мутации других консервативных аминокислот в этом мотиве значительно снижают активность [45]. Остаток Asn168 в мотиве H2TH и остаток Arg258 в цинковом пальце вместе с остатком Lys56 образуют жёсткую сеть водородных связей, фиксирующих ДНК вблизи повреждённого



нуклеозида [1, 26]. Петля Gly211–Phe229, находящаяся также в С-концевом домене, в структуре фермента Frg в свободном виде или связанного с ДНК, содержащей повреждённое основание, упорядочена [55, 56], а в структуре фермента, связанного с неповреждённой ДНК или с ДНК после выщепления повреждённого основания — разупорядочена [28, 29, 57]. Эти наблюдения свидетельствуют о значительной конформационной подвижности этой части молекулы белка.

Междоменная бороздка, на дне которой находится вывернутый из ДНК повреждённый нуклеозид, образована как высококонсервативными участками фермента (N-концевая спираль, мотив H2TH), так и остатками «неупорядоченной петли» Gly211–Phe229 [26]. Полость в ДНК, образовавшаяся в результате выворачивания нуклеозида, заполнена остатками Met73, Arg108 и Phe110 (рисунок 8). При этом ароматическое кольцо Phe110 интеркалирует между основанием Cyt напротив повреждённого нуклеотида и соседним с ним с 5'-стороны основанием ДНК, а гуанидиновая группа Arg108 образует специфические водородные связи с основанием Cyt напротив повреждённого нуклеотида, внося таким образом вклад в предпочтение Frg субстратов 8-охоGua:Cyt [26].



**Рисунок 8.** Схема зоны контакта Frg *E. coli* и ДНК. Центральные нуклеотиды модифицированной и комплементарной цепей содержат основания 8-охоGua и Cyt, соответственно, и имеют индекс 0. Индексы нуклеотидов комплементарной цепи взяты в скобки. Остатки Met73, Arg108 и Phe110 заполняют полость вывернутого основания и образуют ван-дер-ваальсовы контакты с ДНК [28].

Был проведен анализ комплексов ДНК, содержащей повреждённый нуклеотид, с ферментами Frg, полученными из *Bacillus stearotherophilus* [55] и *Lactococcus lactis* [58]. В первом случае ДНК содержала 8-охоGua, и нуклеозид связывается в *син*-конформации, а во втором случае в ДНК находился нерасщепляемый аналог Fary-Gua (карбо-Fary-Gua) и связывался в *анти*-конформации. Для изучения динамики возможных взаимодействий в

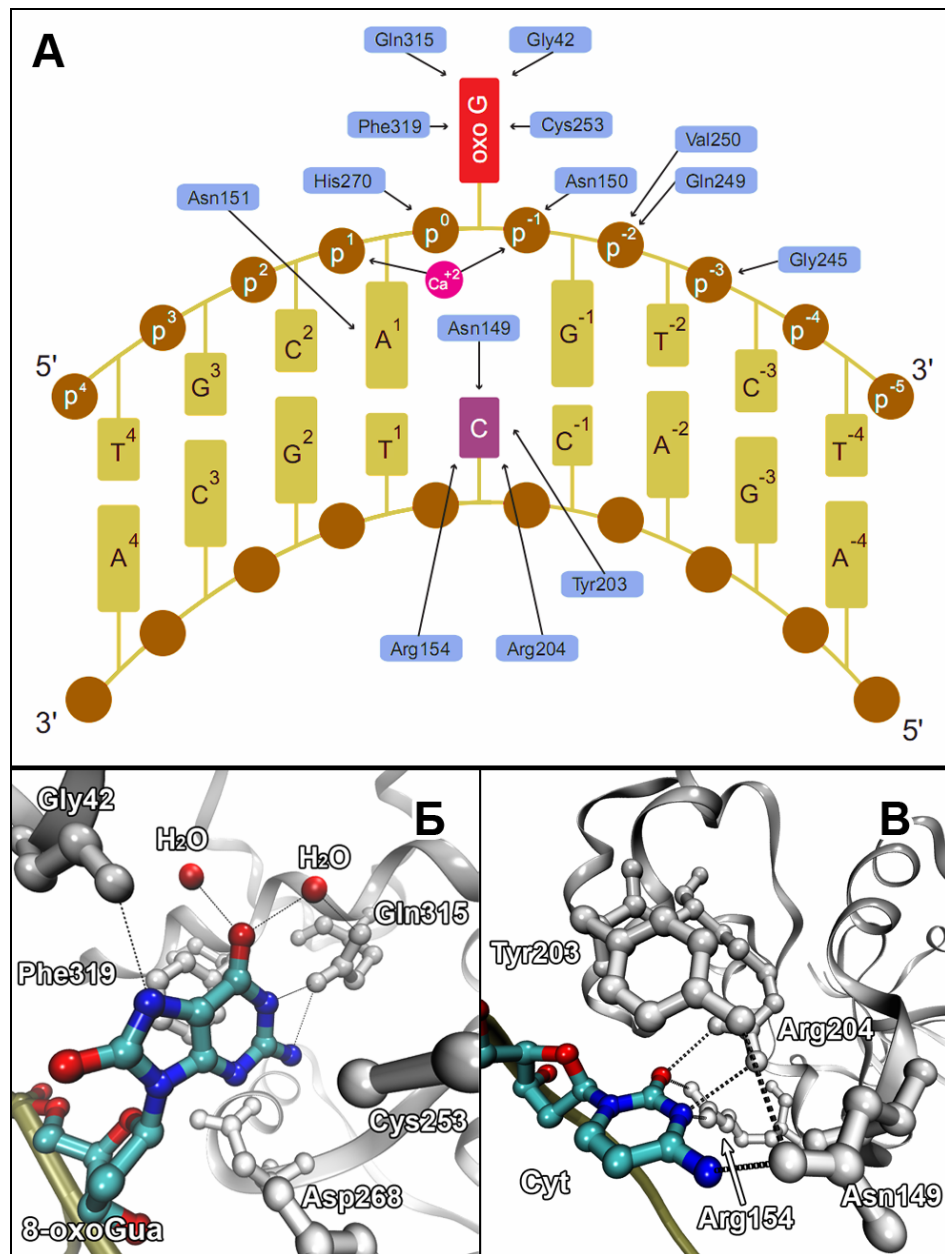


активном центре Fpg был использован метод компьютерного моделирования [54, 59]. Выяснилось, что мутация E2Q, использованная для инактивации *Bst*-Fpg, значительно изменяет электростатические взаимодействия в активном центре фермент-субстратного комплекса [1, 54], а наличие в молекуле *Bst*-Fpg остатка Glu77 сдвигает равновесие в сторону *син*-конформации остатка 8-*oxo*Guo. В полипептидной цепи *Eco*-Fpg на месте Glu77 находится остаток серина, и при моделировании структуры *Bst*-Fpg с заменой E77S более выгодной оказывается *анти*-конформация [26]. По всей видимости, в молекуле *Eco*-Fpg узнавание повреждённого основания происходит в *анти*-конформации [54, 59], для которой методом МД выявлены критические взаимодействия, объясняющие субстратную специфичность фермента (активность по отношению к 8-*oxo*Gua, Fapy-Gua и Fapy-Ade и неактивность по отношению к Gua и 8-*oxo*Ade). Важность некоторых из этих взаимодействий подтверждена методами сайт-направленного мутагенеза: при соответствующих аминокислотных заменах активность *Eco*-Fpg в отношении одних субстратов уменьшалась, а в отношении других не изменялась [59].

По данным ЯМР-исследования, проведённого с молекулами *Eco*-Fpg, многие аминокислотные остатки в междоменном линкере и в ДНК-связывающем центре фермента, а также некоторые остатки на поверхности доменов (до ~30% от общего числа остатков) высокомобільны в свободных молекулах Fpg [26]. После связывания повреждённой ДНК, содержащей нерасщепляемый аналог АП-сайта, конформация некоторых (но не всех) из них замораживается, при этом возникают новые мобильные участки (до ~40% от общего числа остатков), в том числе и в участках белковой глобулы, удалённых от места связывания ДНК. Участок «неупорядоченной петли» по данным ЯМР не обнаруживал значительной мобильности ни в свободном белке, ни в ДНК-белковом комплексе; причины такого расхождения между данными РСА и ЯМР не ясны [26].

В отличие от Fpg, конформационные изменения при связывании и расщеплении повреждённой ДНК белком OGG1 невелики [1, 26]. Связывание ДНК сопровождается поворотом петли Cys146–Asn151, в результате чего лежащий в этой петле остаток Asn149 оказывается внутри двуцепочечной ДНК, а повреждённое основание выворачивается в активный центр фермента [60]. Этот процесс сопровождается значительными перестройками боковых цепей аминокислотных остатков, выстилающих активный центр: например, в свободном ферменте он частично заполнен имидазольной группой остатка His270, входящей в стэкинг-взаимодействие с ароматическим кольцом Phe319 и закрывающей вход в активный центр, а при связывании ДНК остаток His270 отодвигается в сторону, освобождая вход и перемещая каталитический остаток Asp268 в нужное положение (рисунок 9, А, Б). При этом остаток Phe319 разворачивается и образует стэкинг-взаимодействие с повреждённым

основанием [60]. Еще один важный остаток, Cys253, может образовать водородную связь с  $\epsilon$ -аминогруппой Lys249, а также ван-дер-ваальсову связь с  $\pi$ -системой гетероциклического основания [61]. Как и в случае Fpg, фермент OGG1 высокоспецифичен к основанию Cyt напротив повреждённого основания [62, 63], и в комплексе с ДНК этот цитозин образует множественные водородные связи с остатками Asn149, Arg154 и Arg204, а остаток Tyr203, подобно Phe110 в Fpg, интеркалирует с 5'-стороны от него (рисунки 9, В).



**Рисунок 9.** Схема активного центра фермента OGG1 человека. (А) Диаграмма зоны контакта ДНК и белка. Центральные нуклеотиды модифицированной и комплементарной цепей содержат основания 8-охоGua и Cyt, соответственно; 8-охоGua имеет индекс 0. Стрелки указывают на взаимодействие определённых аминокислотных остатков с азотистыми основаниями и фосфатными группами нуклеотидов [63]. (Б) Остатки OGG1, участвующие в узнавании повреждённого основания — 8-охоGua. Показаны также две молекулы воды, потенциально участвующие в стабилизации паттерна водородных связей в активном центре. (В) Остатки OGG1, участвующие в узнавании основания напротив повреждённого — Cyt. Пояснения в тексте.

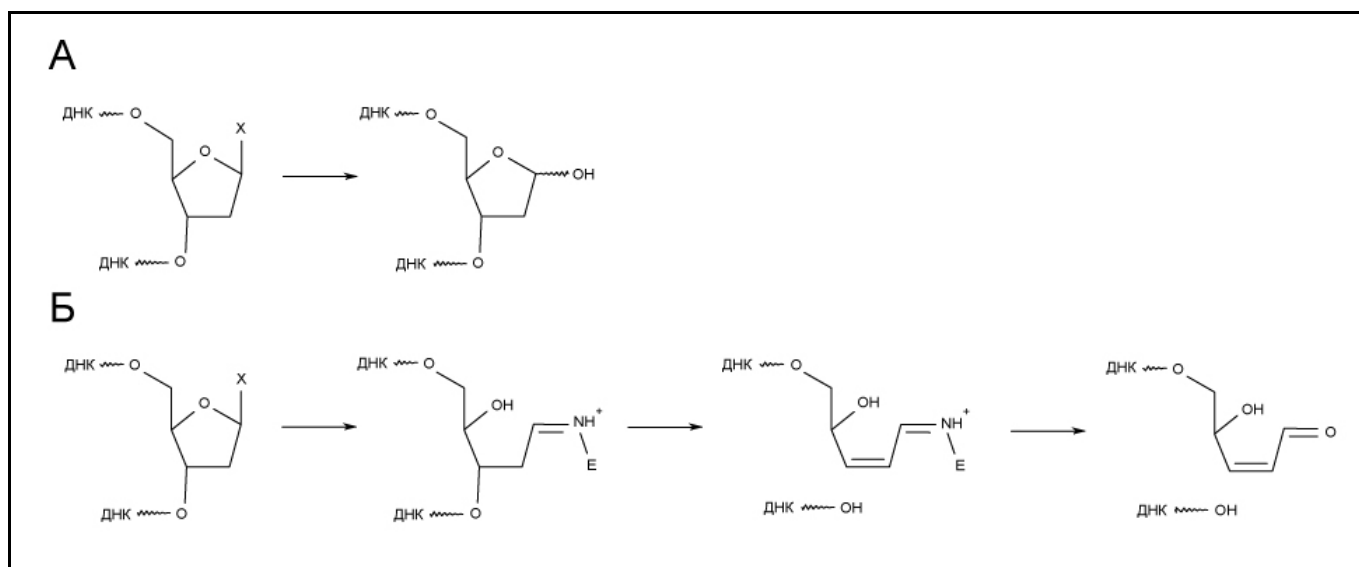
При связывании неповреждённой ДНК, в отличие от Fpg, основание, «читаемое» OGG1, вывернуто из спирали ДНК, но находится не в активном центре, а связано с участком белка («экзо-сайтом»), примыкающим к устью кармана, образующего активный центр [26]. Вероятно, именно в этом экзо-сайте происходит и первичное узнавание повреждённого основания, однако возможно и то, что выворачивание неповреждённого основания — артефакт, обусловленный расположением дисульфидной сшивки, использованной для кристаллизации неспецифического комплекса OGG1–ДНК [26], непосредственно между Cys149 (в мутанте OGG1 N149C) и основанием напротив повреждённого.

Структурные изменения белка MutY, происходящие при связывании субстрата и в ходе реакции, не исследованы в полной мере. При сравнении структур, отражающих состояние фермент-субстратного комплекса по завершении гликозилазной реакции и после диссоциации выщепленного основания из активного сайта, заметно движение спирали  $\alpha 5$  и поворота, примыкающего к ней с N-конца [26]. Этот участок фермента образует стенку кармана, связывающего вывернутый нуклеозид. Выщепленное основание быстро диссоциирует из комплекса фермент-продукт, а ДНК при этом остается связанной с MutY [64]. Интересно, что результаты исследований методом предстационарной кинетики с флюоресцентной детекцией свидетельствуют о том, что в процессе узнавания повреждённого остатка MutY выворачивает из ДНК как нуклеозид, содержащий удаляемое основание напротив повреждения (dAdo), так и комплементарный ему (dGuo или 8-oxodGuo) [65], но структурные данные этого не подтверждают [26].

### 2.1.7. Каталитический механизм ДНК-гликозилаз

Каталитический механизм действия различных ДНК-гликозилаз изучен достаточно подробно [66, 67]. Общий механизм их действия показан на рисунке 10. Как монофункциональные, так и бифункциональные гликозилазы расщепляют N-гликозидную связь по диссоциативному механизму с уходом повреждённого основания в форме аниона и образованием оксакарбениевого катиона дезоксирибозного кольца. Выщепленное основание чаще всего быстро диссоциирует из комплекса с ферментом, но иногда остается связанным с активным центром, способствуя катализу реакции  $\beta$ -элиминирования [26]. Далее в случае монофункциональных гликозилаз по атому C1' сахарного остатка присоединяется молекула воды. У бифункциональных гликозилаз в нуклеофильной атаке атома C1' участвует аминогруппа фермента (обычно  $\epsilon$ -аминогруппа остатка лизина или N-концевая  $\alpha$ -аминогруппа), в результате образуется ковалентный иминовый интермедиат (основание Шиффа) между ферментом и C1' повреждённого нуклеотида. Положительный заряд основания Шиффа

увеличивает кислотность протонов при С2', один из которых отрывается анионом повреждённого основания, инициируя серию электронных перегруппировок, приводящих к элиминированию 3'-фосфатной группы. Основание Шиффа затем гидролизуется, возвращая фермент в исходное состояние и оставляя в цепи разрыв с  $\alpha,\beta$ -ненасыщенным альдегидом на 3'-конце и фосфатной группой на 5'-конце ДНК. Некоторые ферменты, в частности, Frg [68], катализируют еще один шаг элиминирования ( $\delta$ -элиминирование), приводящий к выщеплению  $\alpha,\beta$ -ненасыщенного альдегида с образованием бреши размером в один нуклеозид с 3'- и 5'-фосфатными группами на концах.



**Рисунок 10.** Реакции, катализируемые ДНК-гликозилазами. (А) монофункциональные ДНК-гликозилазы катализируют выщепление повреждённого основания с образованием АП-сайта; (Б) бифункциональные ДНК-гликозилазы образуют основание Шиффа с повреждённым дезоксинуклеотидом и катализируют элиминирование 3'-фосфата. E — фермент (бифункциональная ДНК-гликозилаза) [26].

Несмотря на разные продукты реакции, единственное принципиальное различие между механизмами моно- и бифункциональных ДНК-гликозилаз заключается в природе нуклеофильной группы, что позволило предложить единый каталитический механизм ДНК-гликозилаз [69]. В случае монофункциональных гликозилаз в роли нуклеофила выступает активированная молекула воды; замещение ей азотистого основания приводит к образованию АП-сайта (рисунок 10, А). В случае бифункциональных гликозилаз нуклеофилом служит аминогруппа, принадлежащая одному из аминокислотных остатков фермента, которая замещает основание с образованием ковалентного интермедиата — протонированного основания Шиффа (рисунок 10, Б). Этот интермедиат затем претерпевает  $\beta$ -элиминирование с разрывом цепи (АП-лиазная активность бифункциональных ДНК-гликозилаз) и гидролизуется, восстанавливая активную форму фермента.

В случае Frg роль нуклеофильного остатка, образующего основание Шиффа, выполняет N-концевой остаток Pro1 [29], образующийся после отщепления иницирующего остатка

формилметионина. Остаток Glu2 взаимодействует с остовом ДНК за счет боковой карбоксильной группы и амидной карбонильной группы, стабилизируя структуру комплекса, а также образует водородные связи с О4'-гидроксильной группой разорванного рибозного кольца [29]. У бифункциональных ДНК-гликозилаз суперсемейства Nth в качестве нуклеофила выступает остаток Lys мотива HhH (например, Lys249 в hOGG1 [69]). В качестве кислого остатка, активирующего нуклеофил и стабилизирующего комплекс, может также выступать Asp (например, Asp268 в hOGG1 [61]).

В активном центре монофункциональной гликозилазы MutY *E. coli* важную роль играют остатки Asp138 и Glu37, причём остаток аспартата находится в депротонированном состоянии [70, 71]. Первый активирует молекулу воды, играющую роль нуклеофила, а второй – стабилизирует адениновый нуклеозид, образуя водородные связи между карбоксильной группой и атомами азота пуринового кольца [72].

Исследования неферментативной апуринизации рибо- и дезоксирибонуклеозидов показывают, что расщепление *N*-гликозидной связи идет с образованием оксокарбениевого переходного состояния, в котором связь с азотистым основанием уже практически исчезает, связь с замещающей молекулой воды только начинает образовываться, а в области C1'–O4' накапливается положительный заряд. Такое диссоциативное переходное состояние может быть стабилизировано электростатическими взаимодействиями с положительно заряженным фуранозным остатком или протонированием уходящего основания [2]. Показано, что, по крайней мере, в случае урацил-ДНК-гликозилаз Ung и UNG реализуется именно такой механизм [66]. Значения кинетического изотопного эффекта по атомам Н и С в положениях 1' и 2' дезоксирибозного цикла согласуются с диссоциативным механизмом  $D_N^*A_N$ , в котором сначала связь C1'–N1 разрывается практически полностью, а затем оксокарбениевый ион атакуется молекулой воды.

На сегодняшний момент неясно, протекает ли катализ с участием других ДНК-гликозилаз, в особенности бифункциональных, с образованием того же типа переходного состояния, что и в случае Ung/UNG. С одной стороны, многие ДНК-гликозилазы обладают повышенным сродством к ДНК, содержащей положительно заряженные аналоги АП-сайта или пирролидиновые гомодезоксинуклеозиды, которые в некоторой степени имитируют переходное состояние диссоциативной реакции нуклеофильного замещения. С другой стороны, высокоэффективная АП-лиазная реакция, катализируемая бифункциональными ДНК-гликозилазами, проходит по хорошо изученному механизму нуклеофильного замещения карбонильного атома кислорода аминогруппой с образованием тетраэдрического интермедиата. Трудно представить, что активные центры нескольких групп ферментов, имеющих разное

пространственное строение, будут одновременно оптимизированы для стабилизации двух переходных состояний разной природы — диссоциативного для ДНК-гликозилазной реакции и ассоциативного для АП-лиазной.

Существующие в настоящий момент методы исследования не позволяют в полной мере оценить вклад тех или иных структурных факторов в каталитический механизм действия ферментов. Получаемые методом РСА структуры статичны и зачастую содержат инактивирующие мутации, необходимые для кристаллизации комплекса, поэтому не могут однозначно характеризовать протяжённый процесс *in vivo*. Выявление каталитически важных аминокислотных остатков методом сайт-специфического мутагенеза осложняется тем, что такие мутации могут приводить к не очевидному на первый взгляд нарушению вторичной и третичной структуры биополимера. Такие мощные методы, как, например, исследование кинетического изотопного эффекта, являются сложными процедурами для высокомолекулярных соединений, так как требуют введения изотопных меток в строго определённые позиции. Поэтому в последнее время возрастает интерес к расчётным методам, позволяющим аппроксимировать динамические процессы в биополимерах с разумной точностью. С появлением доступных компьютеров и бурным ростом их вычислительных возможностей такие методы становятся незаменимым инструментом структурной биологии и энзимологии.

## **2.2. КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ БИОПОЛИМЕРОВ**

### **2.2.1. Молекулярная механика**

Молекулярная механика подразумевает моделирование молекулярных систем с использованием простых функций потенциальной энергии, например, гармонических осцилляторов или потенциалов Кулона. Этот подход широко применяется в различных методах компьютерного моделирования — молекулярной динамике (МД), моделировании методом Монте-Карло, расчёте докинга и т.д. [73] Атомы обычно представляются в виде сфер, а связи между ними – в виде пружин. В этом случае внутренние силы могут быть описаны несложными математическими функциями. Например, жёсткость ковалентных связей обычно описывается законом Гука, а взаимодействие ковалентно не связанных атомов — потенциалом Леннарда–Джонса. Используя эти уравнения для вычисления полной потенциальной энергии, можно произвести численное интегрирование уравнения Ньютона классической механики, таким образом наблюдая структурные флуктуации во времени.

Моделирование динамики часто используется для наблюдения изменения во времени конформации белков и других биологических макромолекул, а также кинетических и

термодинамических параметров системы [74–79]. Моделирование, вообще говоря, позволяет наблюдать такие движения частиц и на таких временных промежутках, которые недоступны другим методам анализа. Так как метод опирается на расчёты, есть возможность внести изменения в модель, которые не являются физически реалистичными, но, тем не менее, позволяют оценить молекулярные характеристики в этих условиях. Таким образом, МД вкупе с некоторыми другими дополнительными расчётными методами становится ценным инструментом в выяснении основ строения и функционирования биологических макромолекул.

Первое моделирование белка методом МД было проведено в 1976 году и длительность полученной траектории составила всего 9,2 пс [73]. В качестве объекта исследования был выбран панкреатический ингибитор трипсина крупного рогатого скота размером около 500 атомов. Современные методы позволяют достичь длительностей в микросекунды [80], считая обычными траектории в десятки и даже сотни наносекунд, а количество атомов при моделировании составляет  $10^4$ – $10^6$ . Однако динамические процессы в белках протекают с различными скоростями в различных временных интервалах. Например, быстрые локальные перемещения играют важную роль в ферментативных реакциях. Более медленные движения в масштабе молекулы целиком включают в себя аллостерическое соединение и конформационные свертывания. Временные интервалы для некоторых событий в белках приведены в таблице 2.

**Таблица 2.** Характеристические временные интервалы динамических событий в белке [73].

<i>Событие в белке</i>	<i>Временной интервал, с</i>
Колебания длины связей	$10^{-14} - 10^{-13}$
Эластичная вибрация глобулярного домена	$10^{-12} - 10^{-11}$
Вращение наружных боковых цепей в растворителе	$10^{-11} - 10^{-10}$
Вращательные колебания внутренних групп	$10^{-11} - 10^{-9}$
Шарнирный изгиб (относительное движение глобулярных доменов)	$10^{-11} - 10^{-7}$
Вращение внутренних боковых цепей	$10^{-4} - 1$
Аллостерические переходы	$10^{-5} - 1$
Локальная денатурация	$10^{-5} - 10$
Движение петель	$10^{-9} - 10^{-5}$
Движение жёстких элементов (спиралей)	$10^{-9} - 10^{-6}$
Перемещение витков спирали	$10^{-7} - 10^4$

Новейшие достижения в разработке программного обеспечения для МД делают возможным моделирование до 160 нс в день на современном персональном компьютере [81] и даже больше, с использованием графических ускорителей.

МД широко применяется для исследования большого количества динамических свойств и процессов в таких областях, как структурная биохимия, биофизика, энзимология, молекулярная биология, фармацевтическая химия и биотехнология [82–87]. Хотя использование метода МД самого по себе ограничивает исследователя данными о координатах, скоростях и энергиях атомов, использование методов статистической механики позволяет связать макроскопические параметры (давление, теплоёмкость, свободная энергия) с микроскопическими расчётными значениями. Статистическая механика заняла прочное место среди методов анализа молекулярно-динамических траекторий и позволяет предсказывать по данным моделирования, например, изменение свободной энергии связывания потенциального лекарства или энергетические последствия конформационного перехода в белке. МД, кроме того, применяется в молекулярном дизайне (конструировании лекарств и белков с заданными свойствами) и улучшении кристаллических структур [88].

## **2.2.2. Молекулярно-динамическая теория и методы**

### **2.2.2.1. Силовые поля**

Для получения достоверных результатов компьютерного моделирования требуется выбор правильной функции, описывающей энергию внутримолекулярных и межмолекулярных взаимодействий. В традиционной МД энергетическая функция для нековалентных взаимодействий является довольно простой, парной и аддитивной функцией ядерных координат. Использование координат ядра в качестве координат атома оправдано в рамках приближения Борна–Оппенгеймера [73]. Для описания ковалентных связей используются уравнения для двух (связи), трёх (углы) и четырёх (двугранные углы) тел.

Энергетические функции обычно состоят из большого числа параметризованных выражений [89]. Значения параметров берутся главным образом из экспериментальных исследований и квантовомеханических расчётов малых молекул и фрагментов. Предполагается, что полученные данные можно обобщить и на случай больших молекул. Набор этих функций вместе с набором соответствующих параметров называют силовым полем. Для моделирования белков было разработано несколько специальных силовых полей.

Хотя это не является повсеместным, но одним из важных условий для силового поля является полная аддитивность [73]. Это означает, что длины связей не зависят от углов, а



значения частичных зарядов атомов фиксированы. Такой подход даёт приемлемый, хотя и не безупречный, профиль потенциальной энергии.

Некоторые силовые поля, называемые силовыми полями второго класса, включают в себя парные зависимости или зависимости более высоких порядков. Обычно они используются для точного воспроизведения колебательного спектра или обработки структур с геометрией, сильно отличающейся от равновесной.

Одним из широко распространённых силовых полей является силовое поле CHARMM22 (Chemistry at HARvard Macromolecular Mechanics) [90]. Как и подобные ему распространённые поля, оно состоит из набора дискретных слагаемых. Каждое слагаемое представлено в относительно простой функциональной форме и описывает внутримолекулярные и межмолекулярные силы в системе с известными относительными атомными координатами:

$$U(\vec{\mathbf{R}}) = \sum_{\text{bonds}} K_d (d - d_0)^2 + \sum_{\text{Urey-Bradley}} K_{UB} (S - S_0)^2 + \sum_{\text{angle}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\chi (1 + \cos(n\chi - \delta)) + \\ + \sum_{\text{impropers}} K_\phi (\phi - \phi_0)^2 + \sum_{\text{nonbond}} \left\{ \varepsilon_{ij} \left[ \left( \frac{R_{ij}^{\min}}{r_{ij}} \right)^{12} - \left( \frac{R_{ij}^{\min}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\varepsilon_1 r_{ij}} \right\} \quad (1)$$

где  $U$  — потенциальная энергия как функция радиус-векторов атомов  $\vec{\mathbf{R}}$ ;  $K_d$ ,  $K_{UB}$ ,  $K_\theta$ ,  $K_\chi$ ,  $K_\phi$  — параметрические константы длин связей, дистанции Юри–Брэдли (длины связей между атомами через один), валентных углов, двугранных углов и двугранных импроперов (специальных корректировок для, например, заведомо плоских структур, наподобие бензольных колец) соответственно;  $d$ ,  $s$ ,  $\theta$ ,  $\chi$ ,  $\phi$  — значения длин связей, дистанций Юри–Брэдли, валентных углов, двугранных углов и двугранных импроперов соответственно, индекс 0 указывает на равновесное значение [73]. Это слагаемые для ковалентных взаимодействий. Последнее слагаемое представляет собой описание нековалентных взаимодействий, а именно взаимодействия Леннарда–Джонса и Кулона. Так,  $\varepsilon_{ij}$  — это глубина потенциальной ямы Леннарда–Джонса,  $R_{ij}^{\min}$  — дистанция, на которой потенциал становится нулевым,  $q_i$  — частичный атомный заряд атома  $i$ ,  $\varepsilon_1$  — эффективная диэлектрическая константа, а  $r_{ij}$  — расстояние между атомами  $i$  и  $j$ .

Для получения параметров потенциала Леннарда–Джонса используются правила Лоренца–Бертло [91]. Значения  $\varepsilon_{ij}$  вычисляются как средние арифметические значений  $\varepsilon_{ii}$  и  $\varepsilon_{jj}$ , а  $R_{ij}^{\min}$  — средние арифметические значений  $R_{ii}^{\min}$  и  $R_{jj}^{\min}$ . Значения частичных атомных зарядов получают методами варьирования заряда для воспроизводства эффектов диэлектрического экранирования. Константа  $\varepsilon_1$  обычно принимается равной единице для моделей, содержащих растворитель в виде явных молекул воды.

Используя уравнение (1), можно легко получить суммарную потенциальную энергию биополимера как функцию атомных координат в его конкретном состоянии, называемом снапшотом системы.

Силовые поля CHARMM были независимо параметризованы для белков, нуклеиновых кислот, липидов и углеводов [73] с целью соответствия между собой, что позволяет моделировать гетерогенные системы. Кроме того, существуют силовые поля для малых органических молекул и небιологических макромолекул, таких, как цеолиты [73].

Силовые поля для биологических макромолекул во многом похожи между собой. Гармонические члены описывают длины связей и валентные углы, ряды Фурье описывают вращение, а парные межатомные взаимодействия описываются потенциалом Леннарда–Джонса и законом Кулона [89]. Обычно, хотя и не всегда, параметры силовых полей определяются сначала для белков, а потом распространяются на нуклеиновые кислоты, липиды и другие биологические макромолекулы, чтобы получить совместимость с белковыми полями. Основная разница между силовыми полями состоит в том, откуда берутся параметры. Иногда они содержат взаимосвязи и поправочные коэффициенты для того, чтобы расчёты лучше соответствовали наблюдаемым явлениям. Малозначимая на первый взгляд разница между тем, как различные программные пакеты обрабатывают технические детали моделирования, может привести к совершенно иным значениям энергии. Таким образом, сравнивать параметры одних и тех же атомов, но разных силовых полей не представляется возможным, а прямой перенос параметров от одного поля к другому, вообще говоря, невозможен [73].

Силовое поле AMBER (Assisted Model Building with Energy Refinement) для белков — пример одного из широко распространённых силовых полей, разрабатываемых параллельно с программой молекулярно-динамических расчётов [92]. Точность расчёта частичных зарядов молекулы белка в различных её частях очень важна, и для поля AMBER они были определены методом ограниченного электростатического потенциала (Restrained ElectroStatic Potential, RESP) [93, 94]. Этот метод вычисляет электростатический потенциал молекулярных поверхностей квантовомеханическими методами в рамках модели точечного заряда в центре атома. Были проведены исследования, которые показали, что этот метод даёт наилучшие результаты в вычислении конформационной энергии.

Обобщённое силовое поле AMBER (General AMBER Force Field, GAFF) для органических молекул было разработано в рамках совместимости с существующими силовыми полями AMBER [95]. Оно содержит параметры для большого числа фармакологически значимых агентов и может использоваться в рациональном дизайне лекарств с

автоматизированной методологией, где задание параметров вручную для каждого исследуемого соединения весьма непрактично.

Большинство параметров длин связей и валентных углов силового поля OPLS-AA (Optimized Potentials for Liquid Simulations — All Atoms) было получено напрямую из силового поля AMBER, а торсионные параметры и параметры нековалентных взаимодействий были получены методом комбинации расчётов молекулярных орбиталей *ab initio* и моделирования Монте-Карло [96]. Впоследствии торсионные параметры этого поля были улучшены путем пересчёта коэффициентов.

Хотя традиционные силовые поля не могут моделировать химические реакции из-за того, что для этого требуется разрыв и формирование связей, которые считаются явно заданными, существуют разработки так называемых реакционных силовых полей, которые позволяют обойти эти ограничения. Например, силовое поле ReaxFF [97] вместо явно заданных валентных связей оперирует понятиями кратности связи, что позволяет непрерывно осуществлять их формирование и разрыв. ReaxFF разрабатывалось как максимально обобщённое силовое поле. Оно было параметризовано и протестировано для реакций углеводородов, образования нанотрубок, катализируемого переходными металлами и высокоэнергетических материалов.

#### 2.2.2.2. Минимизация энергии

Хотя метод МД — это действенный метод оптимизации молекулы, позволяющий преодолевать энергетические барьеры локальных минимумов, её запуск для напряжённой структуры весьма не желателен [73]. Это может привести к тому, что потенциальная энергия окажется очень высокой, произойдёт резкий скачок температуры моделируемой системы и результаты могут оказаться непредсказуемыми. Поэтому предварительно обычно осуществляется минимизация энергии исходной структуры. Оптимизация методом минимизации энергии — одна из фундаментальных теоретических концепций молекулярного моделирования.

Задача начальной оптимизации энергии сводится к поиску минимума функции многих переменных  $U(\vec{r})$ , где  $\vec{r}$  — набор векторов, определяющих позиции атомов, а функция  $U$  — это функция суммарной потенциальной энергии системы [89]. Таким образом, при наличии  $N$  атомов и трёх пространственных степеней свободы у каждого функция  $U$  является функцией  $3N$  переменных. Существует множество алгоритмов для решения этой задачи, но далеко не все применимы к задачам молекулярной механики [73].

Поиск глобального минимума функции тысяч переменных — задача, вообще говоря, непростая. Кроме того, если методы оптимизации энергии справляются с задачей улучшения структуры, то совершенно не подходят для обхода конформационного пространства. Но для таких задач, как исправление неверных длин связей или искажённых углов, минимизация энергии является рутинной процедурой. Наиболее часто применяемыми методами являются методы первого порядка (т.е. оперирующие с первой производной функции), такие, как метод градиентного спуска или метод сопряжённых градиентов, и методы второго порядка (т.е. оперирующие со второй производной функции), такой, как метод Ньютона [73].

Метод градиентного спуска — это итеративный метод первого порядка, основная идея которого заключается в том, чтобы идти в направлении наискорейшего спуска, задаваемом антиградиентом  $-\nabla U$ , являющимся, по сути, равнодействующей силой системы. Это означает, что атомы сдвигаются в направлении, задаваемом силой, что имеет под собой физическую основу. На шаге  $i+1$  выполняется расчёт:

$$\vec{\mathbf{r}}^{[i+1]} = \vec{\mathbf{r}}^{[i]} - \lambda^{[i]} \nabla U(\vec{\mathbf{r}}^{[i]}),$$

где вектор  $\vec{\mathbf{r}}$  —  $3N$ -мерный аргумент функции,  $\lambda$  — размер шага,  $U$  — функция потенциальной энергии. Размер шага может быть константным или дробным (т.е. меняться в процессе расчёта). Кроме того, он может рассчитываться как наименьшее значение аргумента  $\lambda$  функции  $F(\vec{\mathbf{r}}^{[i]} - \lambda^{[i]} \nabla U(\vec{\mathbf{r}}^{[i]}))$  — в этом случае метод носит название «метод наискорейшего спуска» [98]. Обычно размер шага является дробным. В том случае, если потенциальная энергия системы после очередной итерации увеличилась, считается, что локальный минимум был пройден, и размер шага умножается на некоторый уменьшающий фактор, обычно 0,5. Если же энергия уменьшилась, то шаг слегка увеличивается, обычно умножением на 1,2 [73]. Хотя градиентный метод неэффективен для нерегулярных поверхностей потенциальной энергии с множественными локальными минимумами, он достаточно хорошо подходит для поиска ближайшего локального минимума.

Метод Ньютона является популярным методом второго порядка, хотя требует некоторой доработки для использования в типичных молекулярно-механических системах. Суть метода заключается в предположении, что справедливо приближённое равенство:

$$U(\vec{\mathbf{r}} + \Delta\vec{\mathbf{r}}) \approx U(\vec{\mathbf{r}}) + U'(\vec{\mathbf{r}})\Delta\vec{\mathbf{r}} + \frac{1}{2}U''(\vec{\mathbf{r}})\Delta\vec{\mathbf{r}}^2,$$

что, по сути, является суммой первых трёх членов тейлоровского разложения  $U(\vec{\mathbf{r}})$  в окрестности точки  $\vec{\mathbf{r}}$  [98]. Минимум правой части, если он существует, достигается при приращении  $\Delta\vec{\mathbf{r}}$ , удовлетворяющем равенству:

$$U''(\bar{\mathbf{r}})\Delta\bar{\mathbf{r}} = -U'(\bar{\mathbf{r}}),$$

откуда следует, что если функция дважды дифференцируема и хорошо аппроксимируется тейлоровским разложением второго порядка, а также удачно выбрана начальная точка  $\bar{\mathbf{r}}$ , то:

$$\bar{\mathbf{r}}_{\min} = \bar{\mathbf{r}} - \frac{U'(\bar{\mathbf{r}})}{U''(\bar{\mathbf{r}})}.$$

Для квадратичных поверхностей не требуется вычислять итерации, т.к. точный минимум находится на основе текущих точек и производных в этих точках. Биомолекулярные системы являются далеко не квадратичными и часто имеют много локальных минимумов, что делает метод Ньютона мало применимым [73]. Однако существуют его модификации, например, метод адаптированного базиса Ньютона–Рафсона (Adopted Basis Newton–Raphson, ABNR), где схема Ньютона применяется к малому подпространству молекулы, а полное смещение молекулы является суперпозицией градиентного метода и метода Ньютона [90]. Метод ABNR показал свою эффективность для больших биологических молекулярных систем [90].

### 2.2.2.3. Адиабатический маппинг

Простейший способ изучения перемещения групп в белках — это характеристика низкоэнергетических профилей отдельных перемещений. Этот подход носит название «адиабатический маппинг» [73]. Конкретные атомы направляются по заранее определённом пути, вызывая интересующие структурные изменения. Оставшимся атомам позволяется двигаться свободно, подчиняясь профилю потенциальной энергии системы, для уменьшения общей потенциальной энергии в каждой точке пути. Предполагается, что сдвиги атомных координат примерно соответствуют структурным флуктуациям, требующимся для осуществления движения, и энергии аппроксимируют реальную энергию, наблюдаемую во время спонтанного движения.

Адиабатический маппинг является малозатратным в вычислительном плане и поэтому широко применяется для изучения структурных перестроек разнообразного масштаба. Информацию о временных интервалах исследуемых процессов получить нельзя, хотя некоторые приблизительные результаты можно получить из релаксированных энергий в аналитических моделях динамики (например, динамика Ланжевена).

Основной недостаток метода — зависимость результатов от выбранного пути движения. Если на самом деле движение происходит по другому пути, результаты будут неверными и вводящими в заблуждение. Существуют также количественные ошибки из-за неполной релаксации конформации, главным образом переоценка энтальпийных барьеров, связанная с неполной релаксацией делокализованной цепи, а методы минимизации энергии обычно

оказываются неэффективными. Кроме того, подход игнорирует некоторые важные термодинамические эффекты, такие, как энтропия и зависимость энтальпии от температуры.

#### 2.2.2.4. Молекулярная динамика

##### 2.2.2.4.1. Уравнение движения

По своей сути оптимизация методом МД — это итеративное численное интегрирование уравнений движения частиц, основанных на мгновенных силах, действующих в молекулярно-механической системе, с последующим движением системы. Уравнения движения соответствуют классическим ньютоновским уравнениям механики. Классическая МД достаточно эффективна благодаря определённому набору приближений. Например, квантовые эффекты обычно полностью игнорируются. Вместо этого атомы (а иногда и группы жёстко связанных атомов) представляются точечными массами согласно модели Борна–Оппенгеймера, т.е. учитываются только перемещения ядер атомов.

Рассмотрим уравнение движения отдельного атома  $i$  массой  $m_i$  и позицией, заданной радиус-вектором  $\mathbf{r}_i$ . Суммарная сила  $\mathbf{F}_i$ , действующая на атом, выражается через отрицательный градиент:

$$\vec{\mathbf{F}}_i = -\frac{d\vec{U}}{d\vec{\mathbf{r}}_i},$$

где  $U$  — функция потенциальной энергии системы. Согласно второму закону Ньютона:

$$\vec{\mathbf{F}}_i = m_i \frac{d^2 \vec{\mathbf{r}}_i}{dt^2} = \vec{\mathbf{p}}_i \frac{d\vec{\mathbf{r}}_i}{dt},$$

где  $\mathbf{p}_i$  — импульс атома  $i$ . Каждая из компонент вектора  $\mathbf{r}_i$  (например,  $x$ ) за короткий конечный интервал  $\Delta t$  может быть описана обычным тейлоровским разложением:

$$x(t + \Delta t) = x(t) + \frac{dx(t)}{dt} \Delta t + \frac{d^2 x(t)}{dt^2} \frac{\Delta t^2}{2} + \dots$$

Этих функций — позиции  $x(t)$ , скорости  $dx(t)/dt$  и ускорения  $d^2x(t)/dt^2$  — вполне достаточно для численного решения уравнений движения (иногда к ним добавляется рывок  $d^3x(t)/dt^3$ ) при условии, что члены ряда Тейлора более высокого порядка каким-либо образом аппроксимируются. Например, их сумма часто принимается равной нулю и ряд, таким образом, обрывается на ускорении, хотя надо отметить, что в общем случае это плохое приближение, так как оно приводит к заметным колебаниям потенциальной энергии системы в процессе расчётов.

Для численного интегрирования уравнений движения существует множество алгоритмов. Большинство из них основано на методе конечных разностей, когда временной отрезок разбивается на маленькие конечные шаги  $\Delta t$ . Не все они являются стабильными при

интегрировании движения осцилляторов (а молекулярно-динамическое движение именно такое), поэтому такие достаточно точные методы, как метод Рунге–Кутты, не подходят. Для нужд МД существуют осцилляторно-стабильные алгоритмы. Например, популярный алгоритм Верле (Verlet) [73] использует позицию и ускорение в момент времени  $t$ , а также предыдущую позицию в момент времени  $t - \Delta t$ :

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \frac{d^2x(t)}{dt^2} \Delta t^2 + O(\Delta t^4).$$

Этот алгоритм не вычисляет скорости, так что их приходится искать другим методом. Но при расчёте МД скорости — важный параметр, поэтому гораздо удобнее применять алгоритм интегрирования, оперирующий скоростями. Одна из таких модификаций, называемая «метод с перешагиваниями», или leapfrog [99], вычисляет скорость на полуинтервале  $t + \Delta t/2$  и использует её в расчёте позиции:

$$\begin{aligned} x(t + \Delta t) &= x(t) + \frac{dx(t)}{dt} \left(t + \frac{\Delta t}{2}\right) \Delta t + O(\Delta t^4) \\ \frac{dx(t)}{dt} \left(t + \frac{\Delta t}{2}\right) &= \frac{dx(t)}{dt} \left(t - \frac{\Delta t}{2}\right) + \frac{d^2x(t)}{dt^2} \Delta t + O(\Delta t^2) \end{aligned}$$

Недостаток такого подхода в том, что скорости получаются не на заданных интервалах, а на полуинтервалах, т.е. рассчитываются в разные моменты времени; этим и объясняется название метода [99]. Скорость, впрочем, несложно посчитать отдельно:

$$\frac{dx(t)}{dt} \left(t\right) = \frac{dx(t)}{dt} \left(t - \frac{\Delta t}{2}\right) + \frac{d^2x(t)}{dt^2} \frac{\Delta t}{2} + O(\Delta t^2).$$

Другая модификация метода Верле, называемая Velocity-Verlet [73], оперирует уже скоростями в тех же точках, в которых рассчитываются координаты:

$$\begin{aligned} x(t + \Delta t) &= x(t) + \frac{dx(t)}{dt} (t) \Delta t + \frac{1}{2} \frac{d^2x(t)}{dt^2} (t) \Delta t^2 + O(\Delta t^4) \\ \frac{dx(t)}{dt} (t + \Delta t) &= \frac{dx(t)}{dt} (t) + \frac{\frac{d^2x(t)}{dt^2} (t) + \frac{d^2x(t)}{dt^2} (t + \Delta t)}{2} \Delta t + O(\Delta t^2) \end{aligned}$$

Наиболее точной модификацией алгоритма Верле, обладающей увеличенной стабильностью и разработанной специально для нужд МД, является алгоритм Бимана [100]:

$$\begin{aligned} x(t + \Delta t) &= x(t) + \frac{dx(t)}{dt} (t) \Delta t + \frac{1}{6} \left( 4 \frac{d^2x(t)}{dt^2} (t) - \frac{d^2x(t)}{dt^2} (t - \Delta t) \right) \Delta t^2 + O(\Delta t^4) \\ \frac{dx(t)}{dt} (t + \Delta t) &= \frac{dx(t)}{dt} (t) + \frac{1}{6} \left( 2 \frac{d^2x(t)}{dt^2} (t + \Delta t) + 5 \frac{d^2x(t)}{dt^2} (t) - \frac{d^2x(t)}{dt^2} (t - \Delta t) \right) \Delta t + O(\Delta t^3) \end{aligned}$$

Все эти часто используемые методы интегрирования являются обратимыми, что означает, что выбор направления моделирования во времени произволен. Иными словами, если знаки скоростей будут изменены на противоположные, моделирование пойдёт в обратном направлении.

Вычислительная сложность алгоритма интегрирования, безусловно, имеет значение, однако, как правило, время, затрачиваемое на него, намного меньше времени расчёта сил. Существует естественное стремление минимизировать количество таких расчётов — например, увеличивая шаг интегрирования. Поэтому точность алгоритма интегрирования имеет первостепенное значение. Она зависит от того, какое количество членов ряда Тейлора учитывается: номер первого неучтённого члена даёт так называемый порядок ошибки алгоритма. Алгоритмы, корректируемые к выбранному порядку ошибки, носят общее название «методы прогноза и коррекции» [99]. Например, таким методом является вышеупомянутый алгоритм Бимана [100]. Он даёт порядок ошибки  $O(\Delta t^4)$  для позиций и  $O(\Delta t^3)$  для скоростей (метод Верле даёт порядок ошибки  $O(\Delta t^4)$  для позиций и  $O(\Delta t^2)$  для скоростей).

#### **2.2.2.4.2. Молекулярное окружение**

При помощи метода МД можно моделировать различные экспериментальные условия. В ранних работах молекулы представлялись изолированными структурами в вакууме [73]. Позднее в расчёты были вовлечены явные молекулы воды и соседние белковые молекулы в кристаллическом окружении. Сейчас широко распространён подход, называемый «периодические граничные условия». Система повторяется периодически по всем направлениям, формируя трёхмерную пространственную решётку, где каждой ячейке соответствует копия центральной ячейки. При таком подходе можно быть уверенным, что все атомы окружены атомами-соседями, неважно, оригинальные это атомы или их копии. Атом, покидающий область моделирования сквозь определённую граничную поверхность, немедленно входит в область через противоположную грань («циклический возврат») [99]. Для того чтобы исключить дублирующие взаимодействия между атомами и их копиями, применяется специальный критерий: для определения взаимодействия между атомами  $i$  и  $j$  используется сам атом  $i$  и либо атом  $j$ , либо его копия — в зависимости от того, что ближе к атому  $i$ . Периодические граничные условия не ограничены кубическими решётками — можно использовать додекаэдры или усечённые октаэдры [73]. Это позволяет существенно уменьшить число атомов растворителя в системе и, таким образом, увеличить скорость расчётов. Количество геометрических примитивов, совместимых с периодическими граничными условиями, ограничено, однако в отсутствие требования строгой периодичности



(стохастические граничные условия) можно использовать любые примитивы [73]. Такие граничные условия находят применение в моделировании небольших регионов, например, лиганд-связывающего сайта при исследовании докинга или активного сайта фермента [101], и позволяют исключить большую часть молекулы из процесса расчётов.

#### 2.2.2.4.3. Увеличенный шаг интегрирования

Для увеличения длины траектории моделирования при неизменной вычислительной сложности представляется наиболее очевидным увеличение длины шага интегрирования. Однако это может привести к нестабильности системы и снижению эффективности моделирования. Существуют алгоритмы повышения стабильности при увеличенной длине шага, которые стабилизируют длины водородных связей и таким образом приводят к снижению наиболее быстрых колебаний системы. Одним из таких алгоритмов является SHAKE, или ограниченный метод Верле — модификация алгоритма Верле для фиксирования таких внутренних координат, как длины связей и углы [73]. Величина шага симуляции ограничена требованием, чтобы  $\Delta t$  было достаточно мало в сравнении с периодом наиболее высокочастотных движений системы. В случае биомолекулярных систем это растягивающие колебания связей, которые обычно не интересуют исследователей. Поэтому такие алгоритмы, как SHAKE, ограничивающие длины связей их равновесными значениями, оказываются весьма полезными. Для этого используются фиксированные межатомные расстояния для задания ограничений на длины связей, и три дистанции между тремя атомами для задания ограничения на величины валентных углов. Практически ограничение длины связи между атомами  $i$  и  $j$  задаётся таким образом:

$$(\mathbf{r}_j - \mathbf{r}_i)^2 - d_{ij}^2 < \frac{\varepsilon}{d_k^2},$$

где  $\mathbf{r}_i$  и  $\mathbf{r}_j$  — радиус-векторы атомов,  $d_{ij}$  — требуемое расстояние,  $\varepsilon$  — константа,  $d_k$  — равновесная длина связи. Такие ограничения рекомендуется применять лишь для атомов, не участвующих в образовании важных двугранных углов. Такому требованию отвечают обычно атомы водорода. Применение этих ограничений позволяет увеличить длину шага в три раза по сравнению с оригинальным алгоритмом Верле [73].

Предложенный вначале вариант алгоритма SHAKE был итеративным и решал систему нелинейных уравнений методом Ньютона [102]. Позднее был разработан неитеративный метод [103], а также усовершенствование алгоритма для работы с замкнутыми структурами — QSHAKE, оперирующий кватернионами [104]. Последняя модификация алгоритма — квадратично-сходящийся P-SHAKE [105], но в то же время алгоритмически более сложный.

Существуют также альтернативные алгоритмы, такие, как LINCS (Linear Constraint Solver), оперирующий непосредственно дистанциями, а не их производными, т.е. скоростями, и дающий дополнительный четырёхкратный прирост скорости вычислений [106].

#### 2.2.2.4.4. Экспериментальные условия

Набор всех возможных систем, имеющих различные микроскопические состояния, но принадлежащих к одному макроскопическому, или термодинамическому, состоянию, называют ансамблем [73]. Существуют различные формальные ансамбли с различными характеристиками. Наиболее широко распространены следующие:

- 1) канонический ансамбль (NVT): термодинамическое состояние систем характеризуется фиксированным набором атомов  $N$ , постоянным объёмом  $V$  и постоянной температурой  $T$ ;
- 2) изобарно-изоэнтальпийный ансамбль (NPH): термодинамическое состояние систем характеризуется фиксированным набором атомов  $N$ , постоянным давлением  $P$  и постоянной энтальпией  $H$ ;
- 3) изобарно-изотермический ансамбль (NPT): термодинамическое состояние систем характеризуется фиксированным набором атомов  $N$ , постоянным давлением  $P$  и постоянной температурой  $T$ ;
- 4) большой канонический ансамбль ( $\mu VT$ ): термодинамическое состояние систем характеризуется постоянным химическим потенциалом  $\mu$ , постоянным объёмом  $V$  и постоянной температурой  $T$ ;
- 5) микроканонический ансамбль (NVE): термодинамическое состояние систем характеризуется фиксированным набором атомов  $N$ , постоянным объёмом  $V$  и постоянной энергией  $E$ ; это соответствует закрытой системе, т.к. энергия сохраняется [73].

Большинство ранних симуляций осуществлялось в рамках микроканонического ансамбля — так называемая «свободная динамика». Однако реальные эксперименты проводятся, как правило, при постоянном объёме и постоянной температуре (т.е. в рамках канонического ансамбля) либо при постоянных давлении и температуре (в рамках изобарно-изотермического ансамбля), поэтому использование этих ансамблей оказывается предпочтительным. Во время симуляции в условиях постоянной энергии температура колеблется из-за спонтанных превращений между кинетической и потенциальной компонентами общей энергии. Мгновенная температура  $T$  может быть рассчитана из скоростей атомов, используя уравнение

$$3k_B T = \frac{\sum_{i=1}^N m_i v_i^2}{N},$$

где  $k_B$  — константа Больцмана,  $m_i$  и  $v_i$  — масса и величина скорости атома  $i$  соответственно,  $N$  — общее число атомов в системе. Если нужно, скорости атомов могут масштабироваться или модифицироваться ещё каким-либо способом для того, чтобы сохранить температуру постоянной на время симуляции. Нужно отметить, что это уравнение должно быть модифицировано соответствующим образом, если используются ограничивающие алгоритмы, такие, как SHAKE.

Для сохранения постоянного давления объём должен быть переменным. Для этого масштабируются размеры периодического объёма. Координаты атомов масштабируются соответственно.

#### 2.2.2.4.5. Растворители

Чтобы достоверно моделировать динамику макромолекул, нужно правильно учитывать окружение. На современном уровне развития методов и вычислительных мощностей невозможно полностью взять в рассмотрение обобщённый случай физиологического окружения типичного белка. Для большинства исследований используется вода в качестве моделируемого растворителя, и тут возможны два варианта. Можно представлять растворитель в виде набора молекул, каждая из которых — полноправный участник ансамбля. Это модель «эксплицитного растворителя» (явного). В этом случае много времени будет тратиться на расчёт взаимодействий между молекулами растворителя, поэтому явного моделирования молекул растворителя хотелось бы избежать. Это реализуемо с помощью подхода «имплицитного растворителя» (неявного) [107]. Модель неявного растворителя не только снижает аппаратные требования для расчёта модели, но и позволяет лучше оценивать свободную энергию растворения [108, 109], а также облегчает конформационную динамику биополимера [110]. Одна из наиболее популярных моделей неявного растворителя — обобщённая модель Борна [111]. Существует и гибридный подход: статическое поле растворителя считается методом конечных разностей Пуассона–Больцмана, а в заданном интересующем регионе используются явные молекулы воды.

Для явного моделирования молекул воды также существует несколько алгоритмов, в числе которых — TIP3P, TIP4P, TIP5P, SPC и SPC/E [73]. Параметры водной модели настраиваются таким образом, что энтальпия испарения и плотность растворителя воспроизводятся между моделированиями. Все перечисленные модели имеют дипольный момент 2,3 Д вместо экспериментального значения 1,85 Д. Зависимость плотности воды от температуры не описывается хорошо ни одной моделью, кроме, быть может, TIP5P. Все

популярные модели явного растворителя хорошо сочетаются с алгоритмами SHAKE, т.к. считают молекулы воды полностью жёсткими.

Некоторые проблемы неявного растворителя можно преодолеть, используя тонкий слой воды вокруг белка — так называемая молекулярно-континуальная модель. Для удерживания этого слоя используется слабый гармонический потенциал. Также часто оказывается разумным сохранить так называемую «кристаллографическую» воду, т.е. молекулы воды, обнаруженные в кристаллических структурах, зачастую находящиеся на интерфейсе биополимер-лиганд, т.к. такие молекулы воды могут играть важную роль в формировании комплексов [112, 113].

Дальние электростатические взаимодействия играют главную роль в стабильности структуры биополимера, а также представляются основными инициаторами процессов ассоциации [73]. Обычно расчёт таких взаимодействий — наиболее вычислительно сложная задача в процессе моделирования методом МД. Количество вычисляемых кулоновских взаимодействий пропорционально квадрату количества зарядов в системе. В ранних моделях использовался подход под названием «сферическое отсечение». Все взаимодействия за пределами заданной сферы, например,  $12 \text{ \AA}$ , не принимались в расчёт. Однако простое отсечение приводит к нарушению непрерывности функции энергии и, в конечном итоге, к нестабильности системы. Поэтому используются различные сглаживающие функции на границах сферы. Кроме того, был предложен метод отсечения по двум дистанциям — ближние электростатические взаимодействия пересчитываются на каждом шаге, а дальние — только после обновления списка нековалентно связанных соседей. Для эффективного вычисления энергии электростатического взаимодействия в больших периодических боксах может также применяться метод суммирования по Эвальду, доработанный специально для использования в методах МД с алгоритмической сложностью  $O(N \log N)$ . Также для дальних электростатических взаимодействий может применяться быстрый мультипольный метод с алгоритмической сложностью  $O(N)$  [114].

#### 2.2.2.5. Динамика Ланжевена

Динамика Ланжевена включает стохастические элементы, аппроксимирующие эффекты степеней свободы, которыми пренебрегают при моделировании [73, 89]. Она основана на уравнении Ланжевена как альтернативе второму закону Ньютона и включает в себя два дополнительных члена. Первый — функция вязкого трения, отражающая сопротивление движению биополимера за счёт наличия молекул растворителя. Второй — шумовой член, позволяющий учесть случайные силы, возникающие за счёт соударений частиц биополимера с молекулами растворителя. Уравнение движения Ланжевена для  $i$ -го атома:

$$m_i \frac{d^2 \vec{r}}{dt^2} = F_i(\vec{r}) - \zeta_i \frac{d\vec{r}}{dt} + R_i(t),$$

где  $F_i(\vec{r})$  — обычный член ньютоновского уравнения МД,  $\zeta_i$  — коэффициент вязкого трения, а  $R_i(t)$  — случайные силы, действующие на атом. Температура системы поддерживается взаимосвязью между  $\zeta_i$  и  $R_i(t)$  (флуктуационно-диссипационная теорема). При  $\zeta_i = 0$  динамика Ланжевена эквивалентна обычной МД. При  $\zeta_i > 0$  случайные импульсы в системе облегчают осуществление движений, связанные с преодолением барьеров, и поэтому динамика Ланжевена может предложить улучшенные характеристики исследования конформационного пространства в сравнении с классической МД.

#### 2.2.2.6. Броуновская динамика

Броуновская динамика — диффузионный аналог МД, рассчитываемый численным интегрированием уравнения Ланжевена [73, 89]. При достаточно высокой эффективной вязкости растворителя движение молекулы биополимера может быть описано в рамках случайных перемещений, так как тормозящий эффект превысит инерционный. Метод броуновской динамики стремится смоделировать случайные перемещения, укладываемые в репрезентативную диффузную траекторию. Это достигается использованием высоких значений коэффициента вязкого трения  $\zeta_i$  в уравнении Ланжевена. В случае, если интересующий исследователя процесс диффузионный по своей сути, броуновская динамика оказывается полезным и широко применяющимся методом в дополнение к МД. Обычно, хотя и не всегда, в броуновских расчётах белки считаются твёрдыми телами. Поэтому в сравнении с классической МД задача оказывается вычислительно менее сложной, и траектории длиной в микросекунды или даже миллисекунды оказываются практически достижимыми. Примеры биологических процессов, которые можно изучать методом броуновской динамики: диффузионно-контролируемые реакции, диффузионные столкновения, ионная диффузия под действием электростатического поля.

Выбор между классической МД, динамикой Ланжевена и броуновской динамикой должен определяться тем, какие факторы дают наиболее весомый вклад в интересующее исследователя движение молекул.

#### 2.2.2.7. Метод Монте-Карло

Структурные и термодинамические свойства системы можно оценить, используя моделирование методом Монте-Карло, важной альтернативой моделированию методом МД [73, 89]. Метод Монте-Карло — это стохастический подход к созданию набора репрезентативных конфигураций в данных термодинамических условиях, таких, как температура и объём. Одна из

привлекательных особенностей метода Монте-Карло — использование только потенциальной энергии при переходе между конфигурациями; расчёт сил не требуется, поэтому вычисления оказываются более простыми. Впрочем, некоторые варианты метода так требуют расчёта сил.

В самом общем случае метод Монте-Карло — это метод численного интегрирования. Выбирается набор случайных аргументов, от которых вычисляются значения функции. Результаты большого количества таких шагов объединяются, и при наличии достаточно широкой выборки вероятность любого полученного результата может быть легко оценена.

Существуют определённые проблемы, препятствующие использованию метода Монте-Карло для больших биологических молекул. Для них сложно определить эффективные движения и, таким образом, сложно продумать простые структурные изменения с достаточно большим отклонением, но в то же время не приводящие к энергетически невыгодной конформации.

Обычные методы Монте-Карло оказываются неэффективными для исследования пространства возможных конфигураций больших молекул в сравнении с методами МД. Кроме того, они не позволяют увидеть эволюцию структуры во времени. Поэтому были разработаны гибридные методы, совмещение методов Монте-Карло и МД, решающие эти проблемы [73].

#### **2.2.2.8. Симуляция отжига**

Алгоритм симуляции отжига связан с методом Монте-Карло и представляет собой эффективный способ поиска конфигурации системы с минимальной энергией. Обычный метод Монте-Карло оказывается неэффективным для поиска конфигураций, лежащих за барьерами потенциальной энергии и полезен лишь тогда, когда исходная конфигурация лежит вблизи ямы потенциального минимума. Симуляция отжига позволяет решить эту проблему путём осуществления начальных шагов методом Монте-Карло при очень высокой температуре. В соответствии с периодическим протоколом температура уменьшается логарифмически (иногда — линейно), пока не достигает нуля. Этот подход часто используется при моделировании белков и в программах оптимизации структуры. Аналитический эквивалент симуляции отжига — метод диффузного уравнения сглаживания потенциала [73].

#### **2.2.2.9. Нединамические методы анализа**

##### **2.2.2.9.1. Конформационная выборка**

Помимо методов Монте-Карло и симуляции отжига, существуют и другие нединамические методы для исследования конформационного пространства биополимеров. Метод CONCOORD (CONstraints to COORDinates) [115] не создает серии конформаций во

времени, но создает набор конформаций, удовлетворяющих набору дистанционных ограничений. В отличие от вышеописанных алгоритмов, он не работает с поверхностью потенциальной энергии. Структуры получаются из исключительно геометрических соображений. Это позволяет получить конформации, которые не могли бы возникнуть в процессе моделирования методом МД или другим методом, основанным на энергии, и поэтому хорошо дополняет динамические методы.

#### **2.2.2.9.2. Метод главных компонент**

Метод главных компонент (Principal Component Analysis) часто используется для снижения размерности набора данных [73]. Для произвольного набора данных, где существует заметная корреляция между размерностями или индивидуальными переменными, первая главная компонента будет линейной комбинацией этих переменных, максимально близко аппроксимирующей весь набор данных. Вторая и последующие главные компоненты строятся таким же образом на наборе данных, оставшемся после исключения более значимых главных компонент. Все главные компоненты будут ортогональными.

Отделение функционально важных движений от термальных флуктуаций биополимера представляется наиболее важной задачей при анализе траекторий. Анализ ковариационной матрицы координат атомов по методу главных компонент называется «эссенциальной динамикой» [73]. Это мощный метод извлечения значимых, масштабных и коррелированных движений, происходящих в процессе моделирования. В этом смысле главные компоненты — это набор ортогональных базисов для атомных координат траектории. Главные компоненты, соответствующие наибольшему отклонению, могут применяться как к структуре биополимера в целом, так и индивидуально, так и в сочетании. Остальные движения, включая небольшие термальные флуктуации, могут быть отброшены. Это упрощает визуализацию и оценку значительных движений, которые могут оказаться биологически важными.

#### **2.2.2.10. QM/MM**

Гибридные квантовомеханические/молекулярно-механические (QM/MM) подходы стали вполне жизнеспособными и быстро приобретают популярность [73]. Они оказываются полезными, так как позволяют изучать механизмы реакций. Обычная МД не подходит для таких исследований ввиду допущения, что связи никогда не разрушаются и не образуются заново. Классические квантовомеханические методы, в которых электронные степени свободы описываются волновыми функциями, например, алгоритм Кара–Паринелло [116], также не подходят для этих целей из-за их чрезвычайной вычислительной сложности, не позволяющей получать сколько-нибудь значимые траектории для биополимера в растворе. Совмещение этих

методов, например, моделирование интересующего фрагмента системы методом квантовой механики, а остальной части — методом молекулярной механики, позволяет получить траектории приемлемой длины на современном аппаратном обеспечении.

### **2.2.3. Молекулярная динамика и энергетические барьеры**

#### **2.2.3.1. Активированная молекулярная динамика**

Множество биологических процессов протекает быстро, но относительно редко, на протяжении длительных временных интервалов. Например, многие реакции и конформационные переходы проявляются на длительных интервалах из-за того, что они состоят из множества активируемых процессов [73]. Вообще говоря, активируемые процессы широко распространены в биологии. Типичный пример такого процесса — локальный конформационный переход в ходе связывания лиганда. Активируемый процесс — это такой процесс, между начальным и конечным состояниями которого существует энергетический барьер, который нужно преодолеть. Само по себе преодоление энергетического барьера — процесс быстрый, однако время, затрачиваемое на придание атомам необходимого момента за счет произвольных температурных флуктуаций, может быть очень велико. Классический молекулярно-динамический подход нельзя использовать для исследования активируемых процессов, т.к. его временные интервалы имеют порядок наносекунд, в то время как временные интервалы активируемых процессов составляют миллисекунды или больше. Но если основные структурные изменения в процессе заранее известны, можно использовать такой метод моделирования, как активированная МД. Это процесс, состоящий из двух стадий. Сначала проводится серия моделирований последовательных процессов по пути конформационного перехода. Цель этих моделирований — определение энергетических барьеров, иными словами, точек активации. Далее в районе этих барьеров проводится расчёт методом обычной МД, причем его можно проводить как в прямую, так и в обратную сторону, чтобы получить репрезентативную траекторию процесса пересечения барьера. Анализ полученных траекторий даёт информацию о механизме активируемого процесса [117].

#### **2.2.3.2. Направленная молекулярная динамика**

В процессе направленной МД (Steered Molecular Dynamics, SMD) система подвергается действию дополнительной силы, зависящей либо от времени, либо от координат, для того, чтобы направить систему вдоль каких-либо конкретных степеней свободы. Это позволяет сфокусироваться на интересующих динамических событиях и при этом снизить вычислительную ёмкость задачи моделирования [73]. Внешняя сила может, например,



направлять событие ассоциации или диссоциации комплекса. SMD оказывается полезной там, где происходят крупные структурные изменения и, соответственно, значительные отклонения от равновесия: диссоциация лиганда, развёртывание белка, инициируемое растягиванием его конца. Существует также схожий метод «нацеленной» МД, где внешняя сила определяется разностью между текущей и целевой конформациями. Задача таких расчётов — привести систему в целевую конформацию. Этот подход может быть использован для предсказания путей, по которым осуществляются переходы между определёнными белковыми конформациями, а также процесс свёртывания белка.

#### **2.2.4. Вычислительные аспекты молекулярной динамики**

Практическое применение методов МД определяется, помимо всего прочего, широким распространением соответствующего программного обеспечения и растущей доступностью высокопроизводительных вычислительных систем.

##### **2.2.4.1. Аппаратное обеспечение**

Огромные успехи, достигнутые в области молекулярного моделирования биополимеров, во многом обязаны стремительному развитию вычислительной техники. Расчёт динамики белков, нуклеиновых кислот и их комплексов считается одной из самых вычислительно сложных задач биологии. Длина и точность моделирования зависит от производительности процессора, объёма оперативной и дисковой памяти. Кратко рассмотрим наиболее значимые разработки в области вычислительной инфраструктуры, которые можно применить к молекулярно-механическим расчётам.

##### **2.2.4.1.1. Параллельные вычисления**

Современные высокопроизводительные вычисления немыслимы без множества процессоров, работающих в одном компьютере. Такие системы называют кластерными вычислительными системами. Однако мощь таких систем может быть использована лишь в том случае, когда задача легко разделяется на множество подзадач, которые могут вычисляться независимо, т.е. распараллеливаться. Такое распараллеливание возможно как на уровне процессов (т.е. запуск независимых молекулярно-динамических расчётов — разных моделей, в разных условиях моделирования и т.д.) или на уровне потоков (параллельный расчёт одной модели). Также требуется минимизировать обмен данными между потоками, чтобы избежать необходимости синхронизации, которая способна полностью уничтожить выигрыш от использования нескольких процессоров или процессорных ядер. Потоки начинают ждать друг друга, чтобы обменяться какой-либо информацией, выстраиваясь в последовательную очередь,

к которой добавляются также задержки, связанные с выполнением синхронизации, и в итоге это может работать медленнее, чем однопоточный расчёт. К сожалению, алгоритмы МД не всегда легко распараллеливаются и, как правило, требуют интенсивного обмена данными между потоками. Однако существуют приёмы, позволяющие частично преодолеть это ограничение и запускать многопоточные процессы моделирования. Поэтому современные суперкомпьютерные проекты представляют интерес и для специалистов, работающих в области молекулярного моделирования. Один из таких проектов — Blue Gene, проект компьютерной архитектуры, разработанный для создания нескольких суперкомпьютеров и направленный на достижение скорости обработки данных, превышающей 1 петафлопс (флопс — внесистемная единица производительности компьютеров, соответствующая одной операции с плавающей запятой в секунду). Было предусмотрено четыре этапа проекта: Blue Gene/L, Blue Gene/C, Blue Gene/P и Blue Gene/Q. На последнем этапе была успешно достигнута скорость в 17 петафлопс (при заявленных 20), и в 2015 году разработка Blue Gene была неофициально прекращена. Существуют также реализованные проекты кластеров из недорогого аппаратного обеспечения пользовательского сектора, например, кластеры серии *Beowulf*. Такой способ соединения обычных компьютеров в кластер иногда называют «виртуальным суперкомпьютером».

#### 2.2.4.1.2. SIMD

Помимо распределения расчётной нагрузки на несколько процессорных ядер существует технология, позволяющая выполнять распараллеливание на уровне одного процессорного ядра. Речь идёт о технологии SIMD — Single Instruction Multiple Data (одна инструкция, много данных). Как следует из названия, SIMD — это набор дополнительных процессорных инструкций, которые позволяют выполнять несколько операций за один процессорный такт. Простой пример — сложение двух четырёхмерных векторов:

$$(x_1, y_1, z_1, w_1) + (x_2, y_2, z_2, w_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2, w_1 + w_2).$$

Как видно, такая операция требует четырёх сложений чисел с плавающей запятой, т.е. четырёх операций. Однако использование соответствующей команды SIMD позволяет выполнить сложение всех четырёх чисел одновременно. Поскольку векторы активно используются в молекулярно-динамических расчётах, выигрыш от таких операций может оказаться значительным.

Вместе с развитием процессоров развиваются и инструкции SIMD, в частности — реализация SIMD от компании Intel под названием SSE (Streaming SIMD Extensions; потоковые SIMD-расширения). Появились инструкции для работы с числами двойной (SSE2) и четверной (AVX) точности (оригинальные инструкции работают с числами одинарной точности),

инструкции для вычисления скалярного произведения и т.д. Набор инструкций, AVX2, позволяет выполнять операции одновременно над четырьмя 64-битными целыми числами, что может использоваться, например, для расчёта индексов табличного поиска [81]. В 2016 году компания Intel запустила на рынок первую модель процессора с поддержкой новейшего набора инструкций AVX-512, работающих с 512-битными регистрами.

Использование SIMD-расширений процессоров сопряжено с некоторыми трудностями. Во-первых, для достижения высокой производительности требуется выравнивание данных по 16-битной границе (что, впрочем, легко реализуемо во всех современных компиляторах). Во-вторых, эффективное использование SIMD возможно лишь на низком уровне программирования — применение ассемблерных вставок или специальных процедур-интринсиков, что подразумевает высокую квалификацию программиста. В-третьих, до недавнего времени разные модели процессоров реализовали разные модели SIMD (например, SSE от компании Intel, 3DNow! от компании AMD, Altivec от компаний Apple, IBM и Motorola, NEON от компании ARM Holdings). Впрочем, сейчас ситуация несколько улучшилась — расширение SSE завоёвывает всё большую популярность, и им оснащаются не только процессоры Intel, но и многие процессоры AMD. Появляются также руководства по портированию кода между различными SIMD-моделями (например, SSE→NEON).

#### **2.2.4.1.3. GPGPU**

Ещё одним перспективным направлением развития аппаратных средств является эволюция видеокарт. Изначально разработанные для задач ускорения отрисовки компьютерной (и в первую очередь — игровой) графики [118], графические процессоры (GPU) стали в определённой степени универсальными программируемыми процессорами с крайне высокой степенью параллельности. Например, видеокарта GeForce GTX 275 содержит 240 процессорных ядер, суммарная производительность которых достигает 1 терафлопс. Существуют также специализированные вычислительные устройства, оптимизированные под вычисления, например, Tesla от компании NVIDIA. Использование возможностей параллельной обработки таких устройств позволяет достичь высокой производительности вычислений, в том числе не связанных с отрисовкой — вычислений общего назначения (General-purpose computations on GPU; GPGPU). Для этого были разработаны различные программные средства — AMD FireStream, CUDA, DirectCompute, OpenCL.

Одной из самых популярных технологий является технология CUDA (Compute Unified Device Architecture), разработанная компанией NVIDIA и поддерживаемая её видеокартами и специализированными вычислительными модулями Tesla [118, 119]. Она позволяет

программистам реализовывать на языке программирования, похожем на C, алгоритмы, выполнимые на графических процессорах, а главное — совмещать фрагменты программ, выполняемых на центральном процессоре (параллельные или последовательные) и функций, выполняемых на GPU. Многие популярные пакеты моделирования методом МД уже реализуют те или иные возможности с использованием графических процессоров.

#### **2.2.4.2. Программное обеспечение**

На сегодняшний день доступно большое количество разнообразных программных решений. Хотя все они стараются охватить как можно больший спектр возможностей, у каждой программы существуют свои преимущества, выделяющие её среди аналогов [73]. Не все исследователи ограничиваются в своей работе каким-либо одним программным пакетом, поэтому большое значение приобретают вопросы совместимости и приведения к единому стандарту. Многие популярные молекулярно-динамические пакеты могут использовать форматы параметров силового поля, топологии и траекторий, изначально разработанные для других программ. Это в определённой степени упрощает воспроизведение опубликованных результатов даже без использования оригинального программного обеспечения.

Множество пакетов моделирования общего назначения содержат какое-либо средство проведения моделирования методом МД, однако часто оно является лишь интерфейсом к одной из специализированных программ. Такие интерфейсы являются полезными, т.к. позволяют запускать задания без глубинного понимания процессов, лежащих в основе расчётов. Кроме того, существуют веб-интерфейсы для таких программ, упрощающих процесс формирования заданий, традиционно являющийся одним из источников ошибок. Примером может служить веб-интерфейс CHARMM-GUI для программы CHARMM [120].

Нужно отметить, что наличие конкуренции в области программного обеспечения ведёт к непрерывному его усовершенствованию, что в целом отражает процесс роста качества и функциональности программ метода МД.

##### **2.2.4.2.1. GROMACS**

GROMACS (GRONingen MACHine for Chemical Simulations) позиционируется как многоцелевой пакет, в первую очередь предназначенный для моделирования таких биологических молекул, как белки и липиды [121]. Утверждается, что программа оптимизирована для быстрого вычисления нековалентных взаимодействий, поэтому используется также для моделирования небологических систем, например, полимеров. GROMACS — это по сути переписанная программа GROMOS [122], которая, как и AMBER, была разработана на основе ранней версии CHARMM. Основные преимущества GROMACS —

простота использования и высокая производительность на персональных компьютерах. Кроме того, для использования программы не требуется изучать какой-либо скриптовый язык. Пакет GROMACS представлен в виде набора исполняемых файлов, принимающих небольшое число параметров. Форматы файлов программы — текстовые, что позволяет избежать проблемы платформенной зависимости. Недостатком является их относительно большой размер, однако программа эффективно использует сжатие данных. Кроме того, для траекторий может быть применено сжатие с потерями, значительно снижающее их размер ценой небольшой потери точности. Программа и её исходный код распространяются по лицензии GNU GPL, позволяющей его использование и доработку в своих целях.

#### **2.2.4.2.2. *NAMD***

В отличие от GROMACS, пакет NAMD (Not Another Molecular Dynamics program) оптимизирован под современные параллельные платформы с большим числом процессорных ядер [123]. Он позволяет моделировать системы огромного размера, недостижимого многими пакетами молекулярного моделирования. Расчёты эффективно распараллеливаются на тысячи процессоров, а последние версии программы используют также GPGPU-вычисления. Ещё одним преимуществом является интеграция с программой визуализации структур VMD [124], облегчающей проведение интерактивного анализа результатов. Впрочем, набор инструментов пакета NAMD скромный по сравнению с другими пакетами. Однако благодаря совместимости форматов файлов существует возможность использовать сторонние инструменты для подготовки, постобработки и анализа данных. Программа NAMD также бесплатная и распространяется с полным исходным кодом, хотя и с некоторыми ограничениями на его использование.

#### **2.2.4.2.3. *AMBER***

Пакет AMBER состоит из набора программ, каждая из которых предназначена для выполнения определённого задания [125]. Разработку ведёт большое число соавторов, и часто новые алгоритмы и методы впервые реализуются именно в этом пакете. Существует возможность использования AMBER как на персональном компьютере, так и на массивно-параллельных платформах (PMEMD — Particle Mesh Ewald Molecular Dynamics). Программа моделирования пакета AMBER использует собственное силовое поле, разрабатываемое параллельно с самим пакетом. В отличие от GROMACS и NAMD, пакет является коммерческим; распространяется также вместе с исходным кодом.

#### **2.2.4.2.4. CHARMM**

Программный пакет CHARMM является одним из старейших пакетов молекулярного моделирования [126]. В отличие от AMBER, он демонстрирует высокую степень интеграции — одна программа выполняет все операции, начиная с подготовки модели и заканчивая анализом результата моделирования. Как и AMBER, для программы моделирования была разработана собственная функциональная форма силового поля. Пакет CHARMM является платным программным продуктом. Существует также специальная коммерческая версия CHARMM, лишённая самых современных возможностей, но более стабильная.

#### **2.2.5. Анализ молекулярно-динамических траекторий**

Результаты молекулярного моделирования зачастую представлены в виде набора состояний моделируемой системы в отдельные моменты времени — снапшотов. Это могут быть как специальные файлы, содержащие координаты атомов системы, так и набор файлов, каждый из которых представляет собой законченную модель. Многие описанные программные пакеты предлагают собственные инструменты анализа: как консольные программы (ptraj из пакета AMBER Tools [125]), так и программы с графическим пользовательским интерфейсом (VMD [124], разработанная в связке с пакетом NAMD). Однако зачастую эти инструменты сводятся к набору аналитических функций, отвергая логическую организацию данных в задачах анализа и возможность многократно возвращаться к систематизированным ранее посчитанным результатам в интерактивном режиме. Некоторые из существующих решений (в частности, консольные инструменты) требуют написания специальных подпрограмм на внутреннем языке (Python, Tcl, собственные скриптовые языки), что требует высокой квалификации пользователя, хорошего владения этим языком, навыков отладки исходных текстов программ, и в целом усложняет процесс анализа. Имеются и ограничения по аппаратным требованиям. В частности, программа VMD, обеспечивающая интерактивный анализ, требует загрузки всей траектории в оперативную память компьютера, что приводит к высоким требованиям к её количеству, так и длительному времени загрузки программы до того момента, когда можно будет приступить к работе. Поэтому разработка высококачественных программных средств анализа траекторий представляет собой одну из важнейших задач МД.

#### **2.2.6. Развитие методов молекулярной динамики**

Стандартные методы МД часто не могут адекватно покрыть конформационное пространство для вычисления термодинамических и кинетических параметров биополимеров. Это связано с наличием высоких энтальпийных и энтропийных барьеров при относительно

низкой тепловой энергии в физиологических условиях [73]. Попадание модели в определённые регионы, ограниченные барьерами, во временном интервале моделирования, даёт неэргодичную траекторию и, таким образом, наблюдаемые средние характеристики оказываются далёкими от средних характеристик ансамбля. Существующие методы улучшенной выборки далеки от идеала, поэтому интерес представляет усовершенствование таких методов. Кроме того, важно повышение эффективности собственно расчётов — улучшение алгоритмов интегрирования для увеличения величины шага, параллелизм в вычислении дистанционных сил, использование более строгих моделей учёта окружения (например, модели Пуассона–Больцмана вместо обобщённой модели Борна) и доработанных моделей неявного растворителя и т.д. [73]. Вместе с этим большое значение имеет развитие аппаратных средств — увеличение числа процессорных ядер, объёма оперативной и дисковой памяти. Разумеется, полученные траектории должны быть своевременно обработаны и проанализированы, поэтому разработка удобных программных средств анализа также оказывается задачей первостепенной важности.

## **2.3. Исследование ДНК-гликозилаз методами компьютерного моделирования**

### **2.3.1. Урацил-ДНК-*N*-гликозилаза UNG**

Одним из первых молекулярно-динамических исследований урацил-ДНК-*N*-гликозилазы (UNG) было моделирование двух структур фермента в комплексе с 9-звенным дуплексом ДНК [127]. Структура такого комплекса была определена экспериментально для комплекса фермент–продукт, а начальная структура для комплекса фермент–субстрат была реконструирована присоединением урацила заново к дезоксирибозе. Выяснилось, что гидролиз гликозидной связи вызывает значительное изменение в позициях элементов остова ДНК. Сайт узнавания урацила в целом сохраняется одинаковым в обеих моделях. Однако водородная связь между амидной группой Asn204, которая отвечает за различие урацила и цитозина, и O<sup>4</sup> урацила нарушается при флуктуациях боковой цепи в комплексе фермент–субстрат и заменена водородной связью между молекулами воды, находящимися внутри белковых полостей позади кармана. Моделирование прояснило роль двух остатков, His268 и Asp145, которые были квалифицированы как важные для катализа в экспериментах по мутагенезу. В комплексе фермент–субстрат His268 находится на расстоянии  $3,45 \pm 0,35 \text{ \AA}$  от урацила, позволяя молекуле воды формировать мостик с O<sup>2</sup>. Окружение увеличивает pK<sub>a</sub> His268 до 7,1, переводя его в протонированное состояние, что облегчает гидролиз гликозидной связи. В соответствии с кристаллической структурой, остов ДНК втягивается после реакции и позволяет His268 приблизиться к O<sup>2</sup> урацила с сопутствующим высвобождением молекулы воды и снижением

$pK_a$  до 5,5, в результате чего протон переходит к продукту. Боковая цепь Asp145 в комплексе фермент–субстрат полностью сольватирована и связана с 3'-фосфатом дезоксиуридина через молекулу воды. Близость Asp145 к отрицательно заряженному фосфату и её  $pK_a$  4,4 указывают на то, что она не может выступать в роли общего основного катализатора. Предложен механизм, в котором мостиковая молекула воды между Asp145 и 3'-фосфатом принимает протон от другой молекулы воды и стабилизирует мостик с помощью иона гидроксония, а также образует гидроксил-анион, нужный на стадии гидролиза [127].

Моделирование показало, что UNG использует электростатическое взаимодействие с фосфатными группами остова субстрата для катализа [128, 129]. Хотя результаты вычислений были проверены экспериментами, подтвердившими предсказанную важность фосфатов  $-2$ ,  $-1$  и  $+1$ , существовало явное несогласование для группы  $+2$ . Расчёты показали, что она даёт важный вклад, но в экспериментах её отсутствие или нейтрализация оказывали слабое влияние на катализ. Фосфатная группа  $+2$  напрямую взаимодействует с гистидином активного центра (His148 у человека) в кристаллической структуре UNG в комплексе с двуцепочечной ДНК. Было предположено, что His148 даёт сильный антикаталитический эффект из-за протонирования, и путём расчёта свободной энергии был оценен его  $pK_a$  в активном центре. Обнаружено, что His148 заряжен положительно во всём экспериментальном диапазоне pH (4–10), так что его депротонирование не может компенсировать делецию или нейтрализацию фосфатной группы  $+2$ . Расчёт свободной энергии был облегчен эффективной процедурой масштабирования заряда, позволяющей осуществить количественную коррекцию системы с неявным растворителем далеко от активного центра; метод был улучшен для учёта разницы отсечения электростатических взаимодействий в молекулярно-механическом подходе и электростатическом подходе в сплошных средах. Для того чтобы показать, что фосфатная группа  $+2$  полностью окружена растворителем в комплексах с оцДНК, похожих на использованные в экспериментах, были проведены дополнительные циклы моделирования. Однако в оригинальном моделировании с дцДНК эта группа хорошо структурирована и окружена. Разница в экранировании растворителем, по-видимому, отвечает за наблюдаемое отсутствие эффекта в экспериментах по нейтрализации и делеции этой группы [128].

Для определения механизма действия UNG использовалось моделирование методом QM/MM [129]. В отличие от предложенного изначально согласованного ассоциативного механизма показано, что реакция идет поэтапно по диссоциативному механизму. Расщепление гликозидной связи приводит к интермедиату, состоящему из оксокарбениевого катиона и урацилат-аниона. Дальнейшая атака молекулой воды и перенос протона на Asp145 приводит к образованию продуктов. Основной вклад в снижение энергии активации принадлежит не



ферменту, а субстрату. Такой «автокатализ» происходит благодаря положению четырёх фосфатных групп, стабилизирующих переходное состояние скоростьлимитирующей стадии. Важность этих фосфатов объясняет остаточную активность, наблюдаемую у мутантов, лишённых ключевых аминокислотных остатков. [129].

Помимо мезофильного фермента UNG человека, были проведены экспериментальные и расчётные структурные исследования для психрофильного фермента UNG из атлантической трески *Gadus morhua*. Адаптированный к низкой температуре фермент каталитически более эффективен, но в то же время менее устойчив к нагреванию по сравнению с термостабильным ферментом, работающим при более высокой температуре [130]. Расчёты методом МД при трёх разных температурах (375 К, 400 К и 425 К) показывают, что существуют отчётливые структурные отличия в пути денатурации между двумя гомологами, наиболее заметные на N- и C-концах. Искажение мезофильного фермента начинается одновременно на N- и C-концах, в то время как C-конец играет ключевую роль в стабильности психрофильного фермента. Моделирование также демонстрирует, что при определённых температурах адаптированный к холоду фермент денатурирует быстрее, чем термостабильный гомолог, что согласуется с экспериментальными данными [130]. Расчёты электростатического континуума показывают, что поверхностные потенциалы адаптированного к холоду фермента более комплементарны потенциалам ДНК в активном центре и рядом с ним в сравнении с термостабильной формой, что свидетельствует о лучшем связывании субстрата [131, 132]. Таким образом, компьютерное моделирование помогает определить структурные детерминанты адаптации ферментов к разным температурам.

### 2.3.2. Урацил-ДНК-N-гликозилаза SMUG1

Специфичная к оцДНК монофункциональная урацил-ДНК-гликозилаза SMUG1 принадлежит к отдельному семейству суперсемейства урацил-ДНК-гликозилазы. Бактериальный ортолог SMUG1 из *Geobacter metallireducens* (Gme) и SMUG1 человека являются не только урацил-ДНК-гликозилазами, но и ксантин-ДНК-гликозилазами [133]. Анализ мутаций и моделирование методом МД Gme-SMUG1 выявили важные структурные детерминанты в консервативных мотивах 1 и 2 для обеих активностей. Мутации аминокислотных остатков Met57 (M57L) и His210 (H210G, H210M, H210N), которые вовлечены во взаимодействие с карбонильным углеродом C2 урацила или ксантина, вызывают существенное снижение активностей, удаляющих эти основания из ДНК. Увеличенная селективность достигается в мутантной форме фермента Gme-SMUG1 A214R, причем остаток Ala214 является высококонсервативным во всех бактериальных гомологах и находится в

участке, взаимодействующем с основанием в процессе выворачивания. При этой замене специфичность фермента напоминает специфичность SMUG1 человека, что иллюстрируется сохранением способности удалять Ura из неканонических пар и слабой активностью по отношению к ксантину. Моделирование методом МД также показало, что замена M57L увеличивает гибкость региона петли мотива 2 и особенно Ala214, что может отвечать за снижение каталитической активности. Замена G60Y полностью уничтожает урацил- и ксантин-ДНК-гликозилазную активность, что согласуется с моделью, в которой G60Y блокирует попадание повреждённого основания в активный центр фермента. Интересно также то, что замена G63P превращает фермент Gme SMUG1 в универсальную урацил-ДНК-гликозилазу, выщепляющую Ura из оцДНК и из любых пар в дцДНК, а также приводит к полной потере способности удалять ксантин. Расчёты МД показали, что в основе такого серьёзного влияния замены G63P на профиль активности SMUG1 может лежать комбинация снижения свободного объёма и изменённой гибкости петель активного центра. Конформационная гибкость, таким образом, играет важную роль в определении субстратной специфичности и каталитической эффективности [133].

### 2.3.3. Формамидопиримидин-ДНК-N-гликозилаза бактерий

Методы МД и биоинформатики были успешно использованы для установления аминокислотных остатков, отвечающих за узнавание повреждённого основания ДНК-гликозилазой Fpg [59]. Для проверки предсказаний использовался сайт-специфический мутагенез с последующими экспериментами по ферментативной кинетике. На основе данных моделирования методом МД предсказано, что Lys217 взаимодействует с O8 внецепочечного 8-охоGua, находящегося в центре связывания. Действительно, мутация Lys217 (K217T) селективно снижает способность Fpg выщеплять 8-охоGua из ДНК. Дигидроурацил, будучи также субстратом Fpg, выступал в роли неспецифического контроля. Другие остатки, вовлечённые в процесс узнавания повреждённого основания (His89, Arg108, Arg109), были определены методом совмещённого анализа консервативности и структуры. Arg108 образует две водородные связи с цитозином ДНК и является основным фактором специфичности к основанию напротив повреждённого. Мутация этого остатка (R108A) снижает выщепление 8-охоGua из температурно-нестабильных неправильных пар с Gua и Thy, хотя на удаление из стабильных пар с Cyt и Ade влияет мало. Мутация His89 (H89A) селективно уменьшает частоту выщепления 8-охоGua, в то время как мутация Arg109 (R109A) практически полностью нарушает связывание фермента с повреждённой ДНК. Все эти факты свидетельствуют о том, что His89 и Arg109 образуют часть «считывающей головки» — элемента структуры,

используемый ферментами для сканирования ДНК на предмет наличия повреждённых оснований. His89 и Lys217 участвуют в формировании специфичности Fpg к окислительным повреждениям, а Arg108 отвечает за специфичность к основанию напротив повреждённого [59].

Субстратная специфичность Fpg представляет особенный интерес с точки зрения МД. Так, повреждённые основания 8-охоGua, Fapy-Gua и Fapy-Ade связываются и удаляются Fpg, а структурно близкое к ним основание 8-охоAde удаляется Fpg очень неэффективно [54]. С помощью метода МД было выяснено, что Lys217 и Met73 в активном центре фермента играют важную роль в узнавании окисленного имидазольного кольца в основаниях-субстратах, и уотсон-криковская сторона повреждённого основания играет роль в его оптимальном позиционировании в активном центре. Узнавание и выщепление Fapy-Ade, по-видимому, происходит из-за открытого имидазольного кольца, в то время как недостаток гибкости 8-охоAde и закрытое имидазольное кольцо могут играть определяющую роль в том, что Fpg не удаляет 8-охоAde [54].

Для исследования высокой степени селективности Fpg к 8-охоGua по сравнению с Gua были получены две модели Fpg *Bacillus stearothermophilus* (неактивный мутант E2Q) в комплексе с ДНК, различающиеся по положению повреждённого основания в активном центре; в одной 8-оходGuo связан в *син*-конформации, в другой — в *анти* [134]. Было высказано предположение, что *син*-конформация является более благоприятной для 8-оходGuo в случае Fpg *Lactococcus lactis*, что связывают с возможной ролью С-концевой петли этого фермента (так называемая  $\alpha$ F- $\beta$ 10 петля) в катализе [135]. Однако моделирование методом МД показало, что *син*-конформация, обнаруженная в кристаллической структуре этого комплекса, стабилизирована взаимодействием с неконсервативным остатком Glu77; замена его на Ser — остаток, находящийся в гомологичной позиции в Fpg *E. coli* — ведёт к предпочтительному связыванию 8-оходGuo в *анти*-конформации [134]. Гомологичная петля присутствует у Fpg *Bacillus stearothermophilus* и тесно взаимодействует с 8-охоGua (атомом N7), будучи неупорядоченной в комплексах фермента с ДНК, содержащей АП-сайт [136]. Однако следует заметить, что в случае *Lla*-Fpg экспериментально определенная структура комплекса с ДНК, содержащей нерасщепляемый карбоциклический аналог Fapy-Gua, содержит вывернутый поврежденный нуклеотид в анти-конформации, и достаточно странным было бы то, что фермент связывает в активном центре два структурно близких повреждения в принципиально разных конформациях. Сравнительные данные МД комплексов с ДНК, содержащей 8-охоGua и Gua, показывают, что перемещения петли  $\alpha$ F- $\beta$ 10 требуются для различения ферментом Fpg повреждённого и неповреждённого нуклеотидов [136]. В дальнейшем компьютерное моделирование комплексов предсказало путь, по которому повреждённое основание при

встрече с ферментом выворачивается из ДНК-дуплекса; оно также продемонстрировало серьёзные различия в свободной энергии между Gua и 8-охоGua вдоль этого пути [137]. В частности, разница в протонировании атома N7 основания ведет к тому, что амидный карбонил Ser220 образует водородную связь с N7 8-охоGua, но отталкивается от N7 Gua, имеющего неподелённую электронную пару [138]. Таким образом, исследование структуры таких комплексов методом МД позволяет понять механизм различения вывернутых нормальных и повреждённых оснований, а также процесс выворачивания в целом.

#### **2.3.4. 8-оксогуанин-ДНК-N-гликозилаза человека**

Гликозилаза эукариот OGG1, удаляющая из ДНК окисленные пурины, не гомологична бактериальному ферменту Fpg и значительно отличается от него по механизму действия. Так, на примере OGG1 человека впервые продемонстрирован катализ этапа ферментативной реакции с участием продукта предыдущего этапа [139]. Методом борогидридной сшивки получен комплекс каталитического интермедиата 8-охоGua с ферментом. Структурные данные, моделирование методом МД и последующее биохимическое исследование кинетики продемонстрировали, что фермент удерживает выщепленное повреждённое основание и использует его как кофактор, участвующий в катализе [139].

Использование метода МД не ограничено анализом известных структур фермент-субстратных комплексов. Описан динамический процесс формирования комплекса OGG1 человека и ДНК, содержащей повреждённое основание 8-охоGua [140]. Моделирования длительностью 1 нс было достаточно для регистрации формирования ДНК-белкового комплекса — он получился через 900 пс динамики и существовал до окончания расчётов. Комплекс был в основном представлен перекрывающимися ван-дер-ваальсовыми поверхностями молекул ДНК и белка. Гуанидиновая группа Arg324 находилась близко к фосфодиэфирному мостику нуклеотида, содержащего 8-охоGua, так что было предположено, что она участвует в катализе. В области повреждения были разрушены водородные связи, что приводило к локальному нарушению структуры В-ДНК. Фосфодиэфирная связь при С5' нуклеотида с 8-охоGua была смещена по направлению к Arg324. Во всех контактах ДНК с ферментом была сформирована сеть водородных связей через молекулы воды, увеличивая стабильность комплекса. В контрольном моделировании такой же системы, но с нативной ДНК, не наблюдалось формирование ни комплекса, ни сети водородных связей через молекулы воды [140].

Для узнавания повреждённого основания OGG1 использует стратегию захвата во время сканирования ДНК [141]. Кристаллическая структура комплекса OGG1 с неповреждённой ДНК

показывает, что основание Gua выворачивается из ДНК, но не попадает в центр узнавания фермента. Расчёты разницы в свободной энергии демонстрируют, что как силы притяжения, так и силы отталкивания играют роль в предпочтительном связывании 8-охоGua в активном центре. Происходит как бы «защита входа» в активный центр при узнавании повреждённого остатка [141].

При расщеплении гликозидной связи ферментом OGG1 каталитически активным является остаток Lys249. Изначально считалось, что он играет роль нуклеофила по механизму  $S_N2$  после активации при отдаче протона строго консервативному остатку Asp268, также находящемуся в активном центре. Однако данные как кристаллографии, так и биохимии ставили это предположение под вопрос, в основном из-за того, что лизин не идеально расположен для такой атаки. Кроме того, каталитическая активность сохраняется после мутации Asp268 в остаток, не являющийся акцептором протона, но сохраняющий основные взаимодействия в активном центре [142]. Исследовалось несколько механизмов реакции для того, чтобы определить правдоподобные способы расщепления гликозидной связи лизином, используя гибридную теорию функциональной плотности для описания ассоциативного и диссоциативного механизмов. Обнаружено, что наименьший энергетический барьер оказывается при механизме  $S_N1$ , где лизин электростатически стабилизирует уходящее основание, затем отдаёт протон с очень малым энергетическим барьером и, наконец, атакует рибозное кольцо с образованием ковалентного интермедиата белок-ДНК. Механизм  $S_N2$  тоже имеет энергетический барьер ниже, чем «спонтанное» разрушение связи, но значительно выше, чем у механизма  $S_N1$ . Однако реагенты, находящиеся в конформации, подходящей для реакции по механизму  $S_N2$ , оказались существенно более стабильны, чем для механизма  $S_N1$ , что указывает на то, что в последнем случае активный центр должен сильнее их связывать для достижения полного каталитического эффекта. Анализ поляризации переходных состояний показал, что наиболее выгодным для стабилизации переходного состояния является распределение зарядов в реакции по механизму для  $S_N1$ , и этот путь будет наилучшим в случае, если в активном центре фермента будет заранее существовать сильный диполь. Эти исследования поддерживают гипотезу о том, для ДНК-гликозилаз предпочтительным является диссоциативный механизм [142]. Однако в дальнейшем диссоциативный механизм был поставлен под сомнение. Для каталитического цикла, ответственного за гликозилазную активность OGG1 — активация фермента, расщепление гликозидной связи и освобождение повреждённого основания — проведено теоретическое исследование, основанное на теории функциональной плотности [143]. Использована беспрецедентно большая квантовомеханическая система, включающая основание 8-охоGua целиком, кольцо

дезоксирибозы, связанное с фосфатными группами, и большинство окружающих аминокислотных остатков, формирующих активный центр фермента. Подтверждено, что Asp268 не играет никакой роли в активации Lys249. В то же время показано, что 8-охоGua играет роль кофермента, запускающего нуклеофильную активацию путём депротонирования Lys249. Затем следует нуклеофильная атака Lys249 по аномерному атому углерода по механизму  $S_N2$ . Это скоростьлимитирующая стадия процесса с активационным барьером 16,7 ккал/моль, что хорошо согласуется с экспериментально полученным значением 17,1 ккал/моль. Выщепленное основание 8-охоGua снова играет роль кофактора фермента в конце процесса путём активации (переносом протона) открытия рибозного кольца и формирования основания Шиффа. Исследование предлагает рекуррентную модель ферментативного расщепления пуринового нуклеозида: активация уходящей группы путём протонирования основания нуклеозида (общий кислотный катализ) активирует расщепление гликозидной связи [143].

Молекулярно-динамическое исследование мутантных форм OGG1 человека имеет большое значение, так как известно, что некоторые полиморфные формы фермента, включающие одиночные аминокислотные замены, связаны с разнообразными формами рака [144]. Некоторые из таких мутаций — R46Q, R131Q и R154H — изначально были охарактеризованы экспериментально, а в дальнейшем исследованы методом МД для предсказания структурных и динамических эффектов замен по отношению ко всему белку и, в особенности, в активном центре и центре узнавания субстрата. Ни одна из замен не вызывает разворачивания или каких-либо глобальных структурных изменений, их влияние ограничено активным центром и центром узнавания. Выяснилось, что хотя активный центр фермента расположен на удалении 18-21 Å от трёх точек мутации, их структурные эффекты распространяются пространственно и вызывают крупное изменение ориентации и химического окружения боковых цепей аминокислот активного центра. Это изменение, по-видимому, мешает боковой цепи Lys249 претерпеть необходимый шаг: депротонирование перед нуклеофильной атакой. Мутации также вызывают расширение полости активного центра, что может объяснить экспериментально наблюдаемое снижение субстратной специфичности [144].

### 2.3.5. ДНК-*N*-гликозилаза MutY

ДНК-гликозилаза MutY — важный участник GO-системы, удаляющий аденин из неправильных пар Gua:Ade или 8-охоGua:Ade. Протекание *N*-гликозилазной реакции, катализируемой этим ферментом, требует определённой организации молекул растворителя [145]. Для исследования влияния растворителя и окружения на катализ были использованы классическое моделирование методом МД, а также QM/MM. Полученные

результаты говорят о том, что вся реакция — процесс энергетически нейтральный, в котором скоростьюлимитирующей стадией оказывается первая, а именно: протонирование атома N7 аденина для инициации расщепления. Вторая стадия — экзотермическая, она включает в себя гидролиз оксокарбениевого интермедиата. Хотя молекулы воды участвуют в катализе на обоих этапах, первая стадия требует совершенно иной организации растворителя, нежели вторая. Для обеспечения протонирования N7 формируется долговременный паттерн растворителя, облегчающий участие трёх из пяти структурных молекул воды в активном центре. Это устойчивое распределение координирует каталитически активные молекулы воды в основные позиции для обеспечения переноса протона: молекула воды, создающая мостики между каталитическими остатками, и одна-две молекулы воды, «поддерживающие» вышеупомянутую. Для поддержания такой конфигурации MutY использует комбинацию гидрофобных и гидрофильных остатков, превращая окружение в своеобразную «ловушку для воды». Гидрофильные остатки увеличивают время существования нужного паттерна растворителя путём формирования сети водородных связей, в то время как гидрофобные остатки ограничивают расположение каталитически важных молекул воды, участвующих в переносе протона [145]. Таким образом, катализируемая MutY реакция, вероятно, регулируется как энтропийными, так и энтальпийными факторами.

### 2.3.6. Nei-подобная ДНК-N-гликозилаза NEIL1

Гликозилаза NEIL1 человека — многофункциональный фермент ЭРО, исправляющий разнообразный набор окислительных повреждений; наиболее распространённый, 8-охоGua, выщепляется слабо [146]. Для установления структурных свойств, определяющих субстратную специфичность NEIL1, исследовались модели фермента в комплексе с двумя парами типичных хорошо репарируемых субстратов: *R*- и *S*-стереоизомерами спироиминодигидантоина, непланарными продуктами дальнейшего окисления гуанина, и *5R,6S*- и *5S,6R*-стереоизомерами тимингликоля, преобладающими продуктами окисления тимина. Также исследовался и слабо репарируемый 8-охоGua. Проводилось моделирование методом МД в течение 10 нс. Объяснение способности NEIL1 удалять многие продукты окислительных повреждений видится в том, что у них есть такие общие химические черты, как пиримидиноподобное кольцо и общие доноры и акцепторы водородных связей, позволяющие повреждённому основанию хорошо связываться в активном центре, который в определённой степени гибок. Однако плоский 8-охоGua не так хорошо подгоняется под неглубокий и сравнительно тесный карман узнавания: образует меньше водородных связей с ферментом, его шестичленное кольцо экспонировано в растворитель. Всё это приводит к слабому взаимодействию его с NEIL1 [146].

### 2.3.7. Эндонуклеаза V бактериофага T4 (DenV)

Для выяснения роли аминокислотных остатков в активном центре DenV рассчитаны их значения  $pK_a$ , используя полумикроскопическую версию белковых диполей — диполей Ланжевена (PDL/D/S) [147]. Проанализированы различные вклады в  $pK_a$ , такие, как межрядовое взаимодействие, конформационные перегруппировки, релаксация белка и связывание с ДНК. Зарядовое состояние аминокислотных остатков в активном центре менее благоприятно в случае комплекса, нежели свободного фермента.  $pK_a$  каталитической N-концевой аминокислотной группы уменьшается с 8,01 в свободном ферменте до 6,52 в комплексе, в то же время  $pK_a$  Glu23 увеличивается с 1,52 до 7,82, что указывает на нейтральность ключевых остатков в комплексе фермент–субстрат гликозилазного этапа реакции. Полученные значения  $pK_a$  согласуются с оптимальным диапазоном pH реакции и допускают, что N-конец играет роль нуклеофила. Glu23 в протонированной форме образует водородную связь с O4' сахара 5'-тимина димера и может способствовать увеличению положительного заряда на C1', таким образом, ускоряя нуклеофильное замещение. Более того, нейтральный Glu23 является подходящим кандидатом для протонирования O4' для индуцирования открытия кольца, что требуется для завершения гликозилазного этапа реакции, проводимой DenV. Положительно заряженные Arg22 и Arg26 создают электростатически благоприятное окружение для уходящей группы основания. Для различения механизмов  $S_N1$  и  $S_N2$  рассчитана энергия протонирования O<sup>2</sup> 5'-тимина димера. Фермент стабилизирует нейтральный тимин с энергией ~3,6 ккал/моль и дестабилизирует протонированный тимин на ~6,6 ккал/моль с учетом водного окружения. Следовательно, формирование протонированного тиминового интермедиата маловероятно, и это свидетельствует в пользу механизма  $S_N2$  [147].

### 2.3.8. Гликозилазы алкилированных пуринов

Экзоциклические этеноаддукты – распространённые повреждённые остатки в ДНК. Один из них, 1,N<sup>6</sup>-этенoadенин, является субстратом для многих гликозилаз, удаляющих алкилированные пурины. Это 3-метиладенин-ДНК-N-гликозилаза *E. coli* (AlkA), метилпурин-ДНК-N-гликозилаза человека (MPG) и некоторые другие. Помимо этеноаддуктов, существуют насыщенные аналоги этенопроизводных, формируемые, например, терапевтическими хлорэтилнитрозомочевинами [148, 149]. Это мутагены и канцерогены, и зачастую тоже являются субстратами гликозилаз алкилированных пуринов, хотя активность в их отношении оказывается сниженной: в 20 раз в случае AlkA [148] и в 65 раз в случае MPG [149]. Моделирование методом МД показало общие структурные черты этено- и этанопроизводных в ДНК-дуплеках, и в то же время — различия во взаимодействиях повреждённого основания и



аминокислот активного центра фермента [148, 149]. Моделирование 1,N<sup>6</sup>-этанoadенина в активном центре MPG (на основе кристаллографической структуры комплекса MPG с ДНК, содержащей 1,N<sup>6</sup>-этанoadенин) в течение 2 нс показало заметное ослабление стекинг-взаимодействий между повреждённым основанием и ароматическими боковыми группами ключевых аминокислотных остатков в активном центре. В то же время плоский 1,N<sup>6</sup>-этанoadенин лучше устраивается в активном центре за счёт стекинг-взаимодействий [149]. Таким образом, данные МД позволили внести ясность в вопрос различения ферментами структурно похожих субстратов.

Показано, что ионы металлов ингибируют те ферменты репарации ДНК, что используют их для собственных нужд, т.е. для функционирования и/или построения структуры. Однако фермент MPG не содержит ионов металлов и не требует их наличия для функционирования, но при этом его активность в расщеплении субстратов, содержащих 1,N<sup>6</sup>-этанoadенин, ингибируется ионами Cd<sup>2+</sup>, Ni<sup>2+</sup> и Zn<sup>2+</sup> [150]. Моделирование методом МД с ионом цинка показывает, что активный центр MPG содержит потенциальный сайт связывания Zn<sup>2+</sup>, сформированный несколькими каталитически важными и консервативными аминокислотными остатками. Связывание металла с этим сайтом оказывается несовместимым с механизмом ферментативного катализа [150]. Ингибирование активности MPG может вносить вклад в генотоксичность металлов и сниженную репарацию продуктов алкилирующих повреждений *in vivo*, поэтому структурное объяснение этого феномена с привлечением данных компьютерного моделирования представляет значительный интерес.

Таким образом, компьютерный анализ гликозилаз в последнее время успешно используется для углубленного исследования их структуры и механизма действия. Однако многие моменты, в частности, структурные детерминанты специфичности 8-оксогуанин-ДНК-гликозилаз бактерий и человека в отношении различных субстратов, равно как и влияние объёмных аминокислотных заместителей на конформационную стабильность активного центра, не исследованы в полной мере. В настоящей диссертации предпринята попытка смоделировать белково-нуклеиновые системы *in silico*, разработать комплексный подход к анализу их молекулярно-динамических траекторий и получить новые данные о структуре и механизме действия этих ферментов.

### 3. МАТЕРИАЛЫ И МЕТОДЫ

#### 3.1. Комплекс программ моделирования биополимеров BISON

Практика применения коммерчески доступных программных комплексов, показывает, что их применение к нестандартным биополимерам, т.е. биополимерам, содержащим неканонические аминокислотные остатки или нуклеотиды, затруднительно. Во-первых, это обусловлено ограниченностью библиотек топологий аминокислотных остатков и нуклеотидов в распространяемых пакетах. Во-вторых, приближенность известных силовых полей и моделей сольватации оставляют возможность для поиска более точной модели описания энергии биополимера в водном растворе. Это явилось причиной разработки нового комплекса программ для моделирования биополимеров методом МД BISON (BioSimulator Origin Novosibirsk). Комплекс программ BISON состоит из программы BioPASED (BioPolymer Analysis of Structure, Energy and Dynamics), реализующей метод МД, и новых версий программ: SIMS — для расчёта гладкой молекулярной поверхности [151]; FAMBE — для расчёта энергии поляризации водного раствора решением уравнения Пуассона методом пограничных элементов и энергии сольватации [152, 153]; FAMBEpH — для расчёта энергии молекулы белка в водном растворе для заданного pH с учетом ионизационного равновесия и сольватации [154].

##### 3.1.1. Программа BioPASED

Программа BioPASED предназначена для работы с трехмерными структурами белков и нуклеиновых кислот с целью их анализа, оптимизации по конформационной энергии, статистического анализа конформационной динамики и расчёта свободной энергии конформации биополимера. Используется математическая модель молекулы в рамках классического метода молекулярной механики [89]. Программа предназначена для работы с молекулами белков, состоящими из 20 типов природных и нескольких типов минорных аминокислотных остатков, молекулами нуклеиновых кислот, ДНК и РНК, состоящих из 4-х типов природных нуклеотидов и нескольких типов минорных нуклеотидов, наблюдающихся в природной тРНК. Основными функциями программы BioPASED являются: 1) анализ пространственной исходной структуры молекулы биополимера и её соответствия библиотеке топологий, восстановление отсутствующих координат атомов боковых цепей аминокислотных остатков; 2) восстановление атомов водорода в структуре из тяжёлых атомов; 3) расчёт конформационной энергии заданной структуры молекулы; 4) оптимизация структуры путем минимизации конформационной энергии молекулы, учёт дополнительных позиционных и дистанционных ограничений на атомы молекулы при оптимизации и МД структуры

биополимера; 5) расчёт сольватации биополимера по модели гауссовских гидратных оболочек; 6) расчёт водной оболочки заданной толщины из явных молекул воды; 7) расчёт сильно связанных молекул воды — водных мостиков между атомами биополимера; 8) расчёт позиций противоположно заряженных ионов вокруг биополимера, соответствующих глобальному минимуму электростатического потенциала; 9) расчёт молекулярной поверхности, доступной водному раствору вокруг биополимера; 10) расчёт молекулярно-динамической траектории движения атомов биополимера под действием микроскопических сил атом-атомных взаимодействий и тепловых флуктуаций; 11) квазиглобальная оптимизация конформации биополимера путем симуляции отжига методом МД; 12) расчёт существенных конформационных движений на основе анализа матрицы позиционных флуктуаций атомов биополимера; 13) расчёт конформационной энтропии метастабильной конформации; 14) расчёт свободной энергии макромолекулы в метастабильной конформации.

### 3.1.2. Силовое поле BioPASED

Программа BioPASED рассчитывает молекулярно-динамическую фазовую траекторию молекулы биополимера в NVT ансамбле, интегрируя классические уравнения движения атомов с тремя временными шагами. Шаги задаются для быстро, умеренно и медленно меняющихся сил, действующих на атомы, соответственно, вследствие: 1) деформации валентного остова и внутреннего вращения; 2) ван-дер-ваальсовых и электростатических взаимодействий; 3) дальнедействующих электростатических взаимодействий и взаимодействий с растворителем. Список взаимодействующих пар атомов рассчитывается для двух радиусов взаимодействий, 8 и 15 Å для ван-дер-ваальсовых и электростатических взаимодействий соответственно, и обновляющихся с разными частотами. Взаимодействие с растворителем учитывается моделью неявного растворителя (гауссовских сольватных оболочек). Используется силовое поле AMBER99 [92, 93] с несколькими модификациями функциональной формы.

Энергия биополимера в растворе  $U$  описывается выражением:

$$U = E_{\text{bond}} + E_{\text{angle}} + E_{\text{tors}} + E_{\text{vdw}}^{14} + E_{\text{elect}}^{14} + E_{\text{vdw}} + E_{\text{elect}} + E_{\text{H-bond}} + E_{\text{solv}}$$

Слагаемые  $E_{\text{bond}}$ ,  $E_{\text{angle}}$ ,  $E_{\text{tors}}$  описывают энергию деформации валентных связей, валентных углов и энергию внутреннего вращения соответственно. Функциональная форма и параметры соответствуют силовому полю AMBER99. Слагаемые  $E_{\text{vdw}}^{14}$  и  $E_{\text{vdw}}$  описывают парные ван-дер-ваальсовы взаимодействия между атомами, разделёнными тремя и более чем тремя валентными связями соответственно. Ван-дер-ваальсов потенциал  $v_{ij}$  имеет функциональную форму с регулируемой жёсткостью потенциала на малых расстояниях:

$$v_{ij}(r) = \begin{cases} \varepsilon_{ij} \left[ \left( \frac{Rm_{ij}}{r} \right)^{12} - 2 \left( \frac{Rm_{ij}}{r} \right)^6 \right] & \text{при } r > R_{sij} \\ v_{ij}(R_{sij}) + \left( \frac{\partial v_{ij}(r)}{\partial r} \right)_{r=R_{sij}} (r - R_{sij}) & \text{при } r \leq R_{sij} \end{cases},$$

где  $r$  — расстояние между атомами,  $\varepsilon_{ij}$  — глубина потенциальной ямы,  $Rm_{ij}$  — точка минимума потенциала,  $R_{sij}$  — радиус «мягкого» (soft-core) потенциала, который в свою очередь задаётся как:

$$R_{sij} = \alpha_s Rm_{ij}.$$

Варьирование параметра  $\alpha_s$  от 0 до 1 позволяет получить управляемую деформацию поверхности потенциальной энергии и используется при квазиглобальной оптимизации структуры молекулы методом отжига. Параметры  $\varepsilon_{ij}$  и  $Rm_{ij}$  ван-дер-ваальсового потенциала в формуле взяты из AMBER FF99. При значении параметра  $\alpha_s = 0$  ван-дер-ваальсов потенциал эквивалентен стандартному. Энергии ван-дер-ваальсовых взаимодействий рассчитываются в радиусе  $R_{Cvdw} = 8 \text{ \AA}$ .

Энергии  $E_{\text{elect}}^{14}$  и  $E_{\text{elect}}$  представляют собой энергии электростатических взаимодействий атомов разделённых тремя и более чем тремя, валентными связями. Электростатическая энергия  $E_{\text{elect}}$  биополимера, погружённого в водный раствор, является суммой межатомных взаимодействий  $u_{ij}$ , которые рассчитываются методом обобщённого поля реакции:

$$u_{ij}(r) = \begin{cases} 0 & \text{при } r \geq R_c \\ \frac{q_i q_j}{D(r)} \left[ \frac{1}{r} + \frac{r^2}{2R_c^3} - \frac{3}{2R_c} \right] & \text{при } r < R_c \end{cases},$$

где  $q_i, q_j$  — атомные заряды,  $R_c$  — радиус сферы взаимодействия, равный  $15 \text{ \AA}$ . Величина  $D(r)$  — диэлектрическая константа, зависящая от межатомного расстояния:

$$D(r) = K_D r,$$

где  $r$  измеряется в единицах  $\text{ \AA}$ , а константа  $K_D = 3,3 \text{ \AA}^{-1}$ . Значение константы  $K_D$  выбрано так, чтобы аппроксимировать эффективную константу электростатических взаимодействий, рассчитанную решением уравнения Пуассона [154, 155]. Электростатические взаимодействия атомов на малых расстояниях  $r$  модифицированы так, чтобы иметь мягкое притяжение или отталкивание:

$$u_{ij}(r) = \begin{cases} 0 & \text{при } r \geq R_C \\ \frac{q_i q_j}{D(r)} \left[ \frac{1}{r} + \frac{r^2}{2R_C^3} - \frac{3}{2R_C} \right] & \text{при } R_{Sij} < r < R_C, \\ u_{ij}(R_{Sij}) + \left( \frac{\partial u_{ij}(r)}{\partial r} \right)_{r=R_{Sij}} (r - R_{Sij}) & \text{при } r \leq R_{Sij} \end{cases}$$

где  $R_{Sij}$  — радиус «мягкого» потенциала, определённый выше.

Энергия водородных связей  $E_{H\text{-bond}}$  рассчитывается между атомом водорода  $H$ , присоединённым к электроотрицательному атому  $X$ , и электроотрицательным атомом акцептором  $Y$ . Энергия водородной связи зависит от координат атомов  $X\text{-}H\text{...}Y$ , расстояния  $r_{HY}$  и угла  $\Theta_{XHY}$  между связями  $X\text{-}H$  и  $H\text{...}Y$ . Потенциал водородной связи аппроксимирован функцией с мягким взаимодействием на малых расстояниях  $r_{HY}$ :

$$u_H(r_{HY}, \Theta_{XHY}) = \begin{cases} \epsilon_H \left[ \left( \frac{R_{mHY}}{r_{HY}} \right)^{12} - 2 \left( \frac{R_{mHY}}{r_{HY}} \right)^6 \right] e^{-\frac{(1-\cos \Theta_{XHY})^2}{\sigma_H^2}} & \text{при } r_{HY} > R_{sHY} \\ u_H(R_{sHY}, \Theta_{XHY}) + \left( \frac{\partial u_H(r)}{\partial r} \right)_{r=R_{sHY}} (r - R_{sHY}) & \text{при } r_{HY} \leq R_{sHY} \end{cases}$$

$$R_{sHY} = \alpha_H R_{mHY}, \text{ где } 0 < \alpha_H < 1$$

где параметры  $\epsilon_H$ ,  $R_{mHY}$  представляют собой максимальную энергию водородной связи и оптимальное расстояние между атомами водорода  $H$  и акцептора  $Y$ . Параметр  $\alpha_H$  уменьшает отталкивание атомов  $H$  и  $Y$  на малых расстояниях  $r_{HY}$ . Параметры  $\epsilon_H$ ,  $R_{mHY}$ ,  $\sigma_H$  подобраны так, чтобы воспроизводить геометрию и энергии водородных связей, полученных анализом структур белков и комплексов белок-белок [156].

Энергия сольватации  $E_{solv}$  вычисляется в рамках модели неявного растворителя EEF1 (гауссовских гидратных оболочек) [107], которая представляет энергию сольватации в виде суммы парных межатомных слагаемых, зависящих от межатомного расстояния  $r$ . Модель гауссовских гидратных оболочек показывает хорошую вычислительную эффективность и достаточную точность. Вклад противоионов также учитывается неявно методом масштабирования зарядов нуклеотидов согласно модели Маннинга [157, 158].

Качество описанного силового поля было протестировано расчётом молекулярно-динамических траекторий белков, которые демонстрируют стабильность нативных структур, а также расчётом стабильности неправильно сложенных структур белков для набора Парка–Левитта [159].

## 3.2. Компьютерное моделирование

Компьютерное моделирование белков и ДНК-белковых комплексов проводили в несколько этапов: 1) подготовка моделей для моделирования и проверка их правильности; 2) оптимизация методом минимизации энергии и затем методом МД; 3) анализ полученных траекторий.

### 3.2.1. Подготовка моделей белков и ДНК-белковых комплексов

Трёхмерная структура комплекса ДНК-гликозилазы OGG1 человека с ДНК (PDB ID: 1EBM) включала в себя дцДНК длиной 15 п. н., содержащую 8-охоGua в *анти*-конформации, вывернутый в активный центр фермента [63], и комплементарный ему Cyt. В структуре отсутствовали N-концевой фрагмент 1—8 а. о., а также внутренний участок 80—82. Внутренние участки брали из другой структуры OGG1 человека (PDB ID: 1KO9), затем выравнивали со аминокислотными остатками на краях разрыва восстанавливаемой структуры с использованием алгоритма Кабша [160] и встраивали выравненные фрагменты в модель редактированием PDB-файла. Таким образом была получена модель WT-C (таблица 3, 1). Для получения модели WT-A проводили замену Cyt напротив 8-охоGua на Ade по следующему алгоритму. Атомную структуру нуклеозида dAdo брали из этой же цепи, затем выравнивали со структурой заменяемого нуклеозида dCyd по атомам O4', C1' и N1/N9 с использованием алгоритма Кабша [160]. Выравненный фрагмент нуклеозида — основание Ade — встраивали в модель вместо Cyt редактированием PDB-файла (модель WT-A). Инактивирующую мутацию K149Q отменяли редактированием PDB-файла: удалением атомов боковой группы аминокислоты и переименованием остатка. Этим же способом осуществляли другие замены: C253I, C253L и Q315W (см. одноимённые модели). В соответствии с квантовомеханическими данными, некоторые аминокислотные остатки моделировали в нестандартных зарядовых состояниях: Cys253 в форме тиолат-аниона, Lys249 с нейтральной  $\epsilon$ -аминогруппой, Asp268 с нейтральной карбоксильной группой. Эти модификации вводили редактированием PDB-файла: заменой названий остатков (CYS→CYM; LYS→LYN; ASP→ASH). Замену 8-охоGua на АП-сайт (модели 20—23; таблица 3) проводили редактированием PDB-файла: переименованием остатка и удалением атомов основания, за исключением атома азота, образующего гликозидную связь: его переименовывали в атом кислорода. Восстановление недостающих атомов водорода и боковых цепей аминокислотных остатков проводили в автоматическом режиме в программе BioPASED.

Таблица 3. Сводная таблица моделей, использованных в компьютерном моделировании.

№ n/n	Условное обозначение	Фермент	Описание
1	Eco-Fpg-1	Eco-Fpg	Дикий тип; моделирование с учётом водородных связей
2	Eco-Fpg-2	Eco-Fpg	Дикий тип; моделирование без учёта водородных связей
3	PRN-GLH-C	Lla-Fpg	Дикий тип; Pro1, Glu[H]2; дцДНК, 8-охоGua:Cyt ( <i>анти</i> )
4	PRN-GLU-C	Lla-Fpg	Дикий тип; Pro1, Glu2; дцДНК, 8-охоGua:Cyt ( <i>анти</i> )
5	PRO-GLH-C	Lla-Fpg	Дикий тип; Pro[H]1, Glu[H]2; дцДНК, 8-охоGua:Cyt ( <i>анти</i> )
6	PRO-GLU-C	Lla-Fpg	Дикий тип; Pro[H]1, Glu2; дцДНК, 8-охоGua:Cyt ( <i>анти</i> )
7	PRN-GLH-Aa	Lla-Fpg	Дикий тип; Pro1, Glu[H]2; дцДНК, 8-охоGua:Ade ( <i>анти</i> )
8	PRN-GLU-Aa	Lla-Fpg	Дикий тип; Pro1, Glu2; дцДНК, 8-охоGua:Ade ( <i>анти</i> )
9	PRO-GLH-Aa	Lla-Fpg	Дикий тип; Pro[H]1, Glu[H]2; дцДНК, 8-охоGua:Ade ( <i>анти</i> )
10	PRO-GLU-Aa	Lla-Fpg	Дикий тип; Pro[H]1, Glu2; дцДНК, 8-охоGua:Ade ( <i>анти</i> )
11	PRN-GLH-As	Lla-Fpg	Дикий тип; Pro1, Glu[H]2; дцДНК, 8-охоGua:Ade ( <i>син</i> )
12	PRN-GLU-As	Lla-Fpg	Дикий тип; Pro1, Glu2; дцДНК, 8-охоGua:Ade ( <i>син</i> )
13	PRO-GLH-As	Lla-Fpg	Дикий тип; Pro[H]1, Glu[H]2; дцДНК, 8-охоGua:Ade ( <i>син</i> )
14	PRO-GLU-As	Lla-Fpg	Дикий тип; Pro[H]1, Glu2; дцДНК, 8-охоGua:Ade ( <i>син</i> )
15	WT-C	hOGG1	Дикий тип; дцДНК, 8-охоGua:Cyt
16	WT-A	hOGG1	Дикий тип; дцДНК, 8-охоGua:Ade
17	C253I	hOGG1	Мутация C253I; дцДНК, 8-охоGua:Cyt
18	C253L	hOGG1	Мутация C253L; дцДНК, 8-охоGua:Cyt
19	Q315W	hOGG1	Мутация Q315W; дцДНК, 8-охоGua:Cyt
20	WT-AP	hOGG1	Дикий тип; дцДНК, AP-сайт:Cyt
21	C253I-AP	hOGG1	C253I; дцДНК, AP-сайт:Cyt
22	C253L-AP	hOGG1	C253L; дцДНК, AP-сайт:Cyt
23	Q315W-AP	hOGG1	Q315W; дцДНК, AP-сайт:Cyt

В качестве лиганда для моделирования OGG1 использовали 15-звенный ОДН, состав и нумерация оснований приводятся ниже:

	+7	+6	+5	+4	+3	+2	+1	0	-1	-2	-3	-4	-5	-6	-7
5'	G	C	G	T	C	C	A	OG	G	T	C	T	A	C	C 3'
3'	C	G	C	A	G	G	T	C/A	C	A	G	A	T	G	G 5'
	(-7)	(-6)	(-5)	(-4)	(-3)	(-2)	(-1)	(0)	(+1)	(+2)	(+3)	(+4)	(+5)	(+6)	(+7)

В моделях, содержащих AP-сайт, OG<sup>0</sup> был, соответственно, заменён остатком AP-сайта.

Трёхмерная структура ДНК-гликозилазы Fpg *E. coli* (PDB ID: 1K82; цепь A) с восстановленными а. о. 217—224, отсутствовавшими в оригинальном PDB-файле, была предоставлена Воробьёвым Ю. Н. На основе структуры были построены две модели

(таблица 3, 10, 11), различавшиеся протоколами моделирования (см. раздел 3.2.2): модель *Eco-Fpg-1* моделировалась с учётом водородных связей, модель *Eco-Fpg-2* — без учёта водородных связей.

При выборе стартовой структуры для МД формамидопиримидин-ДНК-*N*-гликозилазы из Protein Data Bank, содержащего в базе данных 56 кристаллических структур Fpg, принадлежащих различным видам бактерий и описывающих различные стадии реакции [28, 29, 55, 57, 58, 137], были приняты во внимание следующие соображения. Во-первых, структура должна содержать ДНК с повреждённым основанием, расположенным в активном центре белка. Во-вторых, структура должна максимально соответствовать ферменту дикого типа. В-третьих, модель должна быть высококачественной в хорошем разрешении (не более 2 Å) и минимумом отсутствующих остатков. В-четвёртых, выбранный фермент Fpg должен быть хорошо охарактеризован биохимически, по меньшей мере, должна быть известна его специфичность к субстратам *охоGua:Cyt* и *охоGua:Ade*. В итоге выбрали структуру PDB ID 1XC8 (разрешение 1,95 Å) дикого типа белка Fpg *Lactococcus lactis* (*Lla-Fpg*), связанного с дцДНК длиной 14 п. н., содержащей нерасщепляемый карбоциклический аналог *Фару-Gua* в *анти*-конформации, вывернутый в активный центр фермента [58], и *Cyt* напротив него. *Lla-Fpg* практически идентичен *Eco-Fpg* в плане селективности в отношении основания напротив повреждённого и спектра субстратов [161]. Повреждённое основание меняли на *анти-8-охоGua* по следующему алгоритму. Атомную структуру основания *8-охоGua* брали из трёхмерной структуры комплекса Fpg *E. coli* с ДНК (PDB ID: 1R2Y), затем выравнивали со структурой основания *Фару-Gua* с использованием алгоритма Кабша [160] и встраивали полученное выравненное основание в модель вместо *Фару-Gua*. Метиленовую группу циклопентанового кольца, изостерическую O4', заменяли атомом кислорода (C6'→O4') путём редактирования названия атома в файле модели. Из 397 молекул воды, присутствовавших в кристаллической структуре, оставляли 7 молекул, находящихся в непосредственной близости от активного центра фермента. Помимо модели *анти-8-охоGua:анти-Cyt* (таблица 3, 12—15), конструировали также модели *анти-8-охоGua:анти-Ade* (таблица 3, 16—19) и *анти-8-охоGua:син-Ade* (табл. Таблица 3, 20—23). Замены нуклеозидов осуществляли по алгоритму, описанному выше для OGG1. Каждый описанный комплекс моделировали в четырёх различных зарядовых состояниях остатков Pro1 и Glu2: депротонированный Pro1 и протонированный Glu2 (модели PRN-GLH-C, PRN-GLH-Aa и PRN-GLH-As); депротонированные Pro1 и Glu2 (модели PRN-GLU-C, PRN-GLU-Aa и PRN-GLU-As); протонированные Pro1 и Glu2 (модели PRO-GLH-C, PRO-GLH-Aa и PRO-GLH-As); протонированный Pro1 и депротонированный Glu2 (модели PRO-GLU-C, PRO-GLU-Aa и PRO-GLU-As). Эти модификации вводили редактированием PDB-



файла: заменой названий остатков, где это было необходимо (PRO→PRN; GLU→GLH). Так как N-концевые остатки в силовом поле Amber99 всегда считаются протонированными, то название остатка PRN означает *депротонированный* пролин; литера «N» в данном случае подразумевает нейтральный статус протонирования.

В качестве лиганда для моделирования Fpg использовали 14-звенный ОДН, состав и нумерация оснований приводятся ниже:

		+5	+4	+3	+2	+1	0	-1	-2	-3	-4	-5	-6	-7	
5'	C	T	C	T	T	T	OG	T	T	T	C	T	C	G	3'
3'		A	G	A	A	A	C/A	A	A	A	G	A	G	C	G
		(-5)	(-4)	(-3)	(-2)	(-1)	(0)	(+1)	(+2)	(+3)	(+4)	(+5)	(+6)	(+7)	

Все описанные модели корректировали на предмет нумерации атомов и остатков, а также проверяли на правильность с использованием модуля проверки PDB-файлов программы GUI-BioPASED (см. раздел 4.1.1).

### 3.2.2. Оптимизация энергии и молекулярная динамика

Параметры силового поля для неканонических аминокислотных остатков (заряженные остатки пролина и цистеина, нейтральные остатки лизина, глутаминовой и аспарагиновой кислоты) и иона цинка были предоставлены Воробьёвым Ю. Н. Параметры силового поля для 8-охоGua были взяты из статьи [54]. Параметры силового поля для АП-сайта были предоставлены К. Симмерлингом (Университет штата Нью-Йорк, Стони-Брук, США) в личном сообщении.

Моделирование «цинкового пальца», присутствующего в моделях фермента Fpg, проводили в присутствии дополнительных ограничений потенциала силового поля для нековалентных связей между парами атомов: ионом  $Zn^{2+}$  и атомами серы четырёх цистеинов, формирующих «цинковый палец», для поддержания геометрии структурного мотива. Величина константы гармонического потенциала составляла  $2,1 \text{ ккал/\AA}^2$ .

Все подготовленные структуры оптимизировали в течение 500 шагов методом сопряжённых градиентов (алгоритм Флетчера), и затем в течение 150 пс методом МД с постепенным разогревом от 50 К до 300 К в программе BioPASED. Шаг интегрирования составлял 1 фс. Моделирование проводилось в рамках канонического NVT-ансамбля. В качестве модели сольватации выбрали модель неявного растворителя, позволяющую реализовать более эффективную выборку в конформационном пространстве благодаря отсутствию динамического трения явных молекул воды [111]. Явно моделировались лишь те молекулы воды, что были оставлены в составе моделей.

Повторные моделирования PRO-GLH проводили независимо с разными временами начального разогрева системы от 50 К до 300 К: 150, 200 и 250 пс. Таким образом, создавали

стартовые структуры с различными локальными энергетическими минимумами для дальнейшей оптимизации.

Окончательную оптимизацию методом МД проводили следующим образом. Структуру Frg *E. coli* моделировали в двух режимах: с учётом водородных связей (множитель 1,0) и без учёта таковых в течение 2,03 нс. При этом не использовали гармонические ограничения потенциала силового поля (за исключением «цинкового пальца», как описано выше). Остальные структуры моделировали в течение 10 нс с учётом водородных связей (множитель 2,5) и искусственными ограничениями потенциала силового поля: 0,001 ккал/Å<sup>2</sup> для атомов белка, 0,25 ккал/Å<sup>2</sup> для атомов крайних оснований ДНК и 0,0025 ккал/Å<sup>2</sup> для остальных атомов ДНК. Запись координат атомов вдоль траектории осуществляли каждые 2 пс. Таким образом, длина траектории составила 1015 PDB-файлов для модели Frg *E. coli* и 5000 PDB-файлов для остальных моделей.

Протоколы моделирования были подготовлены с помощью разработанной в рамках настоящей работы программы GUI-BioPASED (см. раздел 4.1.1). Расчёты выполняли на суперкомпьютере «NKS-30T» Сибирского Суперкомпьютерного Центра СО РАН.

### 3.2.3. Анализ траекторий

Анализ полученных траекторий и построение графиков осуществляли с помощью разработанной в рамках настоящей работы программы Molecular Dynamics Trajectory Reader and Analyzer (MDTRA; см. раздел 4.1.2). Для анализа угла изгиба ДНК использовался собственный алгоритм, основанный на аппроксимации опорных точек ДНК (срединных точек между противоположными фосфатами) плоскостью методом главных компонент [162, 163], регрессии полученного 2D-набора полиномом четвёртой степени, аналитическом вычислении производных и вычислении угла пересечения касательных в краевых областях. Для визуализации траекторий использовали программы VMD [124], RasMol [164] и PyMol (The PyMOL Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC).

## 3.3. Экспериментальная часть

### 3.3.1. Реактивы

В работе были использованы трис(гидроксиметил)аминометан, мочеви́на («ICN Biomedicals», США), бычий сывороточный альбумин («New England Biolabs», США), дитиотреитол, N,N,N',N'-этилендиаминтетрауксусная кислота, NiSO<sub>4</sub>, α-D-глюкоза, α-D-лактоза, ампициллин, АТФ («Sigma», США), N,N'-метиленабисакриламид, акриламид («Reanal», Венгрия), ксиленианол FF, бромфеноловый синий, персульфат аммония, MgCl<sub>2</sub> («Serva»,

Швейцария), бакто-агар, бакто-триптон, бакто-дрожжевой экстракт («DIFCO Laboratories», США),  $\gamma$ -[ $^{32}\text{P}$ ]АТФ с удельной активностью  $1,8 \cdot 10^{14}$  Бк/ммоль («БИОСАН», Новосибирск), остальные реактивы и растворители квалификации х. ч. и ос. ч. отечественного производства. Для всех процедур, кроме электрофореза, использовалась бидистиллированная вода. При выделении белков для хроматографии использовались сорбенты HiTrap Ni-иминодиацетат-сефароза и HiTrap гепарин-сефароза («GE Healthcare», США).

### 3.3.2. Дезоксирибоолигонуклеотиды

Для исследования активности ферментов OGG1 (дикого типа и мутантных форм) и Eco-Frg были использованы ОДН следующих последовательностей:

OG23	5' -CTCTCCCTTCXCTCCTTTCCCTCT-3' (X = 8-oxoGua)
U23	5' -CTCTCCCTTCXCTCCTTTCCCTCT-3' (X = Ura)
C23'	5' -AGAGGAAAGGAGCGAAGGGAGAG-3'

Олигонуклеотиды были синтезированы в Лаборатории медицинской химии ИХБФМ СО РАН.

### 3.3.3. Ферменты

В работе использовали полинуклеотидкиназу фага T4 (10 000 ед. акт./мл, «БИОСАН», Новосибирск), урацил-ДНК-гликозилазу из *E. coli* (1 000 ед. акт./мл; «New England Biolabs», США). Выделение мутантных форм белков OGG1 человека описано ниже. Ферменты OGG1 человека дикого типа и Frg *E. coli* были предоставлены Жарковым Д. О.

### 3.3.4. Штаммы бактерий и плазмиды

В работе использовали штамм *E. coli* BL21(DE3)RIL ( $F^-$  *ompT hsdS<sub>B</sub>(r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup>) gal dcm* ( $\lambda$ cIts857 *ind1 Sam7 nin5 lacUV5-T7 gene 1*) pRIL) («Stratagene», США). Плазмиды pET-15b, несущие вставки вариантов изоформы OGG1-1a, были предоставлены Жарковым Д. О.

### 3.3.5. Трансформация клеток

Для получения электрокомпетентных клеток *E. coli* клетки ночной культуры выращивали в 1 л среды L (10 г/л бакто-триптон, 5 г/л бакто-дрожжевой экстракт, 5 г/л NaCl) при 37° С на качалке при 250 об/мин до оптической плотности 0,7, отделяли от среды центрифугированием (15 мин, 4 000 g, 4° С), дважды промывали 10%-ным глицерином (v/v) при 4° С и замораживали в 10%-ном глицерине (v/v) при -70° С.

Плазмиды трансформировали в полученные клетки методом электропорации с использованием аппарата *E. coli* Pulser («Bio-Rad», США) в 0,1-см кюветах при 1,8 кВ, после чего инкубировали в среде SOC (20 г/л бакто-триптон, 5 г/л бакто-дрожжевой экстракт, 10 мМ

NaCl, 2,5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub>, 20 mM α-D-глюкоза) при 37° С в течение часа и высевались на селективную среду, содержащую 100 мкг/мл ампициллин.

### 3.3.6. Выделение мутантных форм OGG1

Клетки *E. coli* BL21(DE3)RIL с плазмидой pET-15b, несущей вставку одного из вариантов изоформы OGG1-1a, выращивали в 2 мл среды ZYP-0,8G (10 г/л бакто-триптон, 5 г/л бакто-дрожжевой экстракт, 1 mM MgSO<sub>4</sub>, 0,8% w/w α-D-глюкоза, 25 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 50 mM KH<sub>2</sub>PO<sub>4</sub>, 50 mM Na<sub>2</sub>HPO<sub>4</sub>), содержащей 100 мкг/мл ампициллин, при 37° С на качалке при 250 об/мин в течение ночи.

Для проведения автоиндукции ночную культуру инокулировали в 330 мл среды ZYP-5052 (10 г/л бакто-триптон, 5 г/л бакто-дрожжевой экстракт, 1 mM MgSO<sub>4</sub>, 25 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 50 mM KH<sub>2</sub>PO<sub>4</sub>, 50 mM Na<sub>2</sub>HPO<sub>4</sub>, 0,05% w/v α-D-глюкоза, 0,2% w/v α-D-лактоза), содержащей 100 мкг/мл ампициллин, и инкубировали при 37° С на качалке при 250 об/мин также в течение ночи. Клетки отделяли от среды центрифугированием (5 мин, 10 000 g, 4° С). Полученную клеточную массу ресуспендировали в 37 мл буфера TE (10 mM Трис-НСl, pH 8,0, 1 mM ЭДТА) и замораживали при –70° С.

Для осаждения белка клетки размораживали и разрушали с использованием генератора ультразвука УЗДН-2Т («SELMI») 10 импульсами по 30 с на максимальной мощности с интервалом 1,5 мин; суспензию всё время охлаждали на льду. Затем супернатант отделяли от осадка центрифугированием (20 мин, 14 000 g, 4° С) и добавляли кристаллический (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> при перемешивании на льду до 80% насыщения. Суспензию инкубировали в течение ночи при 4° С, после чего осадок, содержащий белок, отделяли центрифугированием (20 мин, 14 000 g, 4° С) и замораживали при –70° С.

Сульфатаммонийные осадки белков растворяли в 25 мл буфера С (25 mM фосфат калия, pH 7,5, 1 M NaCl), затем добавляли 2 мл сорбента HiTrap Chelating, заряженного ионами Ni<sup>2+</sup>, и инкубировали при 4° С в течение 1 ч. Сорбент осаждали центрифугированием (5 мин, 5 000 g, 4° С), осадок переносили в 5-мл колонку Econo-Pac («Bio-Rad», США) и промывали 10 мл буфера С, затем 5 мл буфера С, содержащего 50 mM имидазол. Белок элюировали в две фракции по 2 мл буфера С, содержащего 500 mM имидазол. Фракции анализировали электрофорезом в 12% ПААГ (система Лэммли [165]) с окрашиванием кумасси голубым. Наиболее чистую фракцию затем очищали на колонке HiTrap Heparin («Pharmacia», Швеция) с элюцией градиентом NaCl (0–1 M) в буфере А (25 mM фосфат калия, pH 7,5, 1 mM ЭДТА) на хроматографе BioCAD («Applied Biosystems», США). Полученные фракции вновь анализировали электрофорезом. Фракции белка с чистотой не менее 95% диализовали в течение

20 ч против двух смен буфера для хранения (25 мМ фосфат калия, рН 7,5, 400 мМ NaCl, 1 мМ ЭДТА, 1 мМ ДТТ, 50% v/v глицерин). Белок хранили при  $-20^{\circ}\text{C}$ .

Молярную концентрацию очищенных белков определяли спектрофотометрически по поглощению при 280 нМ с использованием коэффициентов молярной экстинкции, рассчитанных по формуле Гилла–фон Хиппеля [166] (OGG1 C253I, C253L:  $\varepsilon = 68410 \text{ см}^{-1}\cdot\text{М}^{-1}$ ; OGG1 Q315W:  $\varepsilon = 73910 \text{ см}^{-1}\cdot\text{М}^{-1}$ ).

### 3.3.7. Гель-электрофорез и количественный обсчет результатов экспериментов

Во всех описанных ниже экспериментах, если не указано иное, при проведении гель-электрофореза в денатурирующих условиях использовали 20% ПААГ (соотношение акриламид:бис-акриламид 19:1), содержащий 8 М мочевины и 1× буфер TBE (90 мМ Трис, 90 мМ борная кислота, 2 мМ ЭДТА); электрофорез проводили в полиакриламидной пластине толщиной 0,4–0,6 мм и длиной 20–30 см в 1× буфере TBE при напряжении 800–1000 В без специального охлаждения. После разделения продуктов реакции в геле радиоактивность полос определяли радиолуминесцентным сканированием экрана Image Screen K («Kodak», США) с использованием системы Molecular Imager FX («Bio-Rad», США). При необходимости результаты обсчитывали с использованием программы Quantity One v4.6.5 («Bio-Rad», США), а кинетические параметры определяли с помощью процедуры линейной либо нелинейной регрессии, используя программу SigmaPlot v8.0 («SPSS Inc.», США).

### 3.3.8. Приготовление $^{32}\text{P}$ -меченых олигонуклеотидных субстратов

Фосфорилирование ОДН проводили в 70 мМ Трис-НСl (рН 7,6), содержащем 10 мМ  $\text{MgCl}_2$ , 100 мМ KCl и 1 мМ 2-меркаптоэтанол. На 100–200 пмоль ОДН в объеме 20–50 мкл добавляли 20 ед. акт. полинуклеотидкиназы фага Т4 и 2 МБк  $\gamma$ [ $^{32}\text{P}$ ]АТР (примерно 1 пмоль). После инкубации в течение 40 мин. при  $37^{\circ}\text{C}$  отбирали 0,5 мкл реакционной смеси для последующего определения выхода продукта, а затем отделяли меченые ОДН от невключившегося [ $^{32}\text{P}$ ]АТР обращённофазовой хроматографией на сорбенте  $\text{C}_{18}$  NenSorb. К реакционной смеси добавляли 500 мкл 100 мМ Трис-НСl, рН 7,5, 1 мМ ЭДТА. Перед нанесением данного образца колонку  $\text{C}_{18}$  объемом 200 мкл смачивали 1 мл метанола, а затем уравнивали 2 мл того же буфера. После нанесения образца колонку промывали последовательно 2 мл буфера и 2 мл воды. Меченый ОДН смывали с колонки 600 мкл 50% метанола. Затем очищенный ОДН упаривали досуха под вакуумом и растворяли в 30 мкл 0,1×TE буфера (10 мМ Трис-НСl, рН 7,5, 0,1 мМ ЭДТА). Для определения выхода отбирали 0,5 мкл очищенного ОДН и анализировали его и аликвоту, отобранную на предыдущем этапе,

гель-электрофорезом в денатурирующих условиях. Выход меченого продукта после очистки составлял обычно 40–80% при эффективности включения метки ~20%.

Отжиг комплементарных олигонуклеотида вели с двукратным молярным избытком немеченой цепи в следующем режиме: 1 мин при 95° С, 5 мин при 37° С, 30 мин при комнатной температуре. Полученный дуплекс хранили при –20° С.

### **3.3.9. Приготовление субстрата, содержащего АП-сайт**

Для анализа расщепления ДНК ферментом OGG1 использовали двуцепочечный ОДН, содержащий АП-сайт. Так как АП-сайт недостаточно стабилен при высокой температуре и в присутствии нуклеофильных соединений в растворе, субстрат готовили непосредственно перед реакцией. Для его приготовления брали 5'-меченый двуцепочечный ОДН, содержащий остаток урацила (см. раздел 3.3.2), и обрабатывали урацил-ДНК-*N*-гликозилазой из *E. coli*. Реакцию вели в буфере R (50 мМ Трис-НСl, рН 7,5, 100 мМ NaCl, 1 мМ ЭДТА, 1 мМ ДТТ) при 37° С в течение 30 мин.

### **3.3.10. Определение рН-зависимости активности ферментов**

Для определения рН-зависимости активности фермента использовали 25 мМ натрий-фосфатный буфер, рН которого варьировал в интервале 4,0–9,0 (использовали пары кислота/1-замещённый фосфат, 1-/2-замещённый фосфаты и 2-/3-замещённый фосфаты). Реакционная смесь содержала 25 мМ буфер с соответствующим рН, 100 нМ субстрат (5'-меченый двуцепочечный ОДН, содержащий остаток оксогуанина; см. раздел 3.3.2) и либо 2 нМ, либо 500 нМ фермент. Реакционную смесь инкубировали при 37° С (2 нМ фермент) или на льду (500 нМ фермент) в течение 1 мин. Реакцию останавливали добавлением равного объёма буфера для нанесения, содержащего 95% формамид (v/v), после чего пробы прогревали при 95° С в течение 3 мин. Продукты реакции разделяли гель-электрофорезом в ПААГ и анализировали (см. раздел 3.3.7).

### **3.3.11. Изучение зависимости расщепления субстрата от концентрации фермента**

Для определения зависимости расщепления субстрата от концентрации фермента брали следующие концентрации фермента: 1000 нМ, 500 нМ, 200 нМ, 100 нМ, 50 нМ, 20 нМ, 10 нМ, 5 нМ, 2 нМ, 1 нМ. Фермент разводили в 0,5× буфере R (рН 7,5; 25 мМ Трис-НСl, 50 мМ NaCl, 0,5 мМ ЭДТА, 0,5 мМ ДТТ), содержащем 0,5 мг/мл БСА, и добавляли в пробирки с 50 нМ раствором дцОДН, содержащего повреждение 8-охоGua или АП-сайт (приготовленный по методике, описанной в разделе 3.3.9). Реакцию проводили в течение 10 мин при 37° С, после

чего продукты реакции разделяли гель-электрофорезом в ПААГ и анализировали (см. раздел 3.3.7).

### 3.3.12. Определение кинетических параметров реакций, катализируемых OGG1

Стандартная реакционная смесь содержала 1× буфер R (pH 7,5; 50 mM Трис-НCl, 100 mM NaCl, 1 mM ЭДТА, 1 mM ДТТ). Фермент разводили в 0,5× буфере R (pH 7,5; 25 mM Трис-НCl, 50 mM NaCl, 0,5 mM ЭДТА, 0,5 mM ДТТ), содержащем 0,5 мг/мл БСА.

Чтобы определить константы скорости  $k_{ex}$  для реакции выщепления повреждённого основания на субстрате, содержащем 8-охоGua, и  $k_{cl}$  для реакции  $\beta$ -элиминирования на субстрате, содержащем АП-сайт, 10 нМ субстрат инкубировали с 200 нМ ферментом при 37° С или при 13° С соответственно, в течение 0,5–30 мин. В случае 8-охоGua реакцию останавливали добавлением NaOH до концентрации 100 mM, после чего пробы прогревали в течение 2 мин при 95° С, затем нейтрализовали раствор добавлением эквивалентного количества HCl и добавляли равный объём буфера для нанесения, содержащий 95% формамид (v/v). Перед анализом электрофорезом в денатурирующих условиях пробы прогревали при 95° С в течение 2 мин. В случае АП-сайта реакцию останавливали добавлением равного объёма буфера для нанесения, содержащего 95% формамид (v/v), после чего пробы прогревали в течение 1 мин при 95° С и охлаждали на льду. Значение  $k_{ex}$  рассчитывали методом нелинейной регрессии из кривой зависимости накопления продукта от времени по формуле:

$$[P] = [S]_0 \left(1 - e^{-k_{ex}t}\right),$$

где [P] — концентрация продукта,  $[S]_0$  — начальная концентрация субстрата,  $t$  — время реакции. Для определения константы расщепления  $k_{cl}$  субстрата, содержащего 8-охоGua, по механизму  $\beta$ -элиминирования, 100 нМ субстрат (двухцепочечный ОДН, содержащий повреждённое основание 8-охоGua или АП-сайт) инкубировали с 5 нМ ферментом при 37° С в течение 0,5–10 мин. В случае 8-охоGua реакцию останавливали добавлением NaOH до концентрации 100 mM, после чего пробы прогревали в течение 2 мин при 95° С, затем нейтрализовали раствор добавлением эквивалентного количества HCl и добавляли равный объём буфера для нанесения, содержащий 95% формамид (v/v). Перед анализом электрофорезом в денатурирующих условиях пробы прогревали при 95° С в течение 2 мин. В случае АП-сайта реакцию останавливали добавлением равного объёма буфера для нанесения, содержащего 95% формамид (v/v), после чего пробы прогревали в течение 30 секунд при 95° С и охлаждали на льду. Значение  $k_{cl}$  рассчитывали методом линейной регрессии из линейной части кривой зависимости накопления продукта от времени по формуле:

$$[P] = [S]_0 (1 + k_{cl}t),$$

где  $[P]$  — концентрация продукта,  $[S]_0$  — начальная концентрация субстрата,  $t$  — время реакции.

## 4. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

В настоящей работе предложены новые программные средства для улучшения существующего метода компьютерного анализа структуры и динамики биополимеров и их комплексов. На их основе проведено моделирование методом МД некоторых 8-оксогуанин-ДНК-гликозилаз бактерий и человека, а полученные расчётные данные подвергнуты экспериментальной проверке биохимическими методами. Исследование ферментов ЭРО дикого типа и мутантных форм на различных субстратах позволяет выявить новые структурные основы узнавания ими повреждённых остатков.

### 4.1. Разработка программных средств

В рамках настоящей работы были разработаны две программы: 1) графический веб-интерфейс GUI-BioPASED к программе BioPASED; 2) программа анализа траекторий МД и построения графиков — MDTRA.

#### 4.1.1. Графический веб-интерфейс программы моделирования GUI-BioPASED

Пакет BISON/BioPASED предназначен для моделирования трехмерных структур белков и нуклеиновых кислот методом МД в рамках классических молекулярно-механических подходов. В его основе лежит программа BioPASED, выполняющая основные задачи по восстановлению, оптимизации энергии и интегрированию траекторий. Работа с программными пакетами, реализующими метод МД, достаточно сложна, особенно для экспериментаторов, не являющихся экспертами в области использования программных средств вычислительной структурной биологии, и BioPASED — не исключение. Формирование задания, выбор нужных параметров, правильное их форматирование, проверка моделей на корректность и правильный запуск расчётов — все эти операции традиционно являются источником ошибок. Разработанная веб-программа GUI-BioPASED для моделирования МД с использованием программы BioPASED позволяет сформировать задание в виде одного исполняемого файла, экспортировать этот файл на компьютер пользователя, обеспечить его исполнение и, таким образом, частично автоматизировать эту часть работы. Создан дружественный межплатформенный интерфейс, который осуществляет проверку исходных данных, сообщает об ошибках и дает подсказки по возможным способам их решения.

Главная цель программы GUI-BioPASED — позволить пользователю ввести данные, необходимые для расчётов, и сформировать файлы с данными для последующего запуска



расчётов. Программа GUI-BioPASED проверяет как тип введённых значений (строка, число), так и принадлежность их диапазону разумных значений. Если обнаруживаются ошибки, программа выдает об этом сообщение, а также подсказку об их исправлении. Некоторые параметры программы BioPASED являются взаимоисключающими, и программа GUI-BioPASED учитывает это, предлагая для ввода согласованный набор параметров и делая один параметр недоступным при выборе другого. При создании управляющих файлов программа строго придерживается требуемого форматирования, что позволяет гарантировать, что программа BioPASED сможет прочитать эти файлы без ошибок. Проверка PDB-файлов позволяет заранее убедиться, что нужный файл может быть безошибочно прочитан программой BioPASED. Интерфейс GUI-BioPASED поддерживает русский и английский языки.

Программа BioPASED работает как в среде операционной системы Microsoft Windows, так и в Linux. Пользователю, в зависимости от его операционной системы, предлагается загрузить архив с подходящей версией программы. Программа GUI-BioPASED написана на языке PHP с использованием СУБД MySQL и полностью поддерживает все распространённые веб-браузеры: Microsoft Internet Explorer (версии не ниже 6.0) и все браузеры, основанные на его ядре (например, Maxthon), Mozilla Firefox (версии 2 и 3, в том числе версии для Linux), Opera, Google Chrome. Могут иметься некоторые различия в разметке страницы и внешнем виде элементов управления, которые, однако, не будут мешать работе с программой.

Основные модули программы: 1) модуль создания управляющих файлов (позволяет сформировать архив с файлами задания, загрузить его на компьютер пользователя и запустить процесс расчётов с помощью командного файла); 2) модуль проверки PDB-файлов (выполняет проверку соответствия модели библиотеке топологии аминокислотных и нуклеотидных остатков, целостности цепей, а также исправляет некоторые незначительные ошибки); 3) модуль администрирования (предназначен для настройки различных аспектов работы самой программы, доступ предоставляется только уполномоченным сотрудникам). Для идентификации пользователей используется собственный механизм сессий, основанный на IP-адресах и уникальном идентификаторе, сохраняющемся в cookies.

Программа свободно доступна для работы на веб-сайте, расположенном на сервере ИХБФМ СО РАН (<http://biopased.niboch.nsc.ru>). Там же находятся подробное справочное руководство и пример составления типичного задания, на русском и английском языках.

GUI-BioPASED активно использовалась в работе для формирования управляющих файлов и для проверки PDB-файлов моделей на совместимость с программой моделирования BioPASED.

#### 4.1.2. Программа анализа молекулярно-динамических траекторий MDTRA

Анализ молекулярно-динамических траекторий большого размера вручную представляет собой очень сложную задачу. Большинство имеющихся решений [124, 125, 167—171] имеют существенные недостатки. Одним из значительных недостатков многих из них является отсутствие логической организации обработки данных в виде проектного файла, что зачастую исключает возможность быстро вернуться к ранее посчитанным данным в интерактивном режиме. Далее, некоторые из существующих решений требуют написания специальных подпрограмм на встроенном скриптовом языке с большим количеством команд при отсутствии графического интерфейса, что усложняет процесс анализа и, как любая программа, требует отладки исходных текстов заданий. Существуют также определённые ограничения по аппаратным требованиям (например, минимальному количеству оперативной памяти) и скорости работы.

Разработанная программа Molecular Dynamics Trajectory Reader and Analyzer (MDTRA) является частью интегрированного пакета моделирования BISON и способна работать как с PDB-траекториями вообще, так и с траекториями, создаваемыми программой BioPASED. Дизайн MDTRA основан на следующих принципах: 1) эргономичный и расширяемый графический интерфейс, не требующий в большинстве случаев написания подпрограмм, хотя и позволяющий это делать; 2) возможность быстро представить данные в графическом виде, пригодном не только для анализа, но и для публикации; 3) низкие требования к количеству оперативной памяти и интенсивное использование современных аппаратных возможностей.

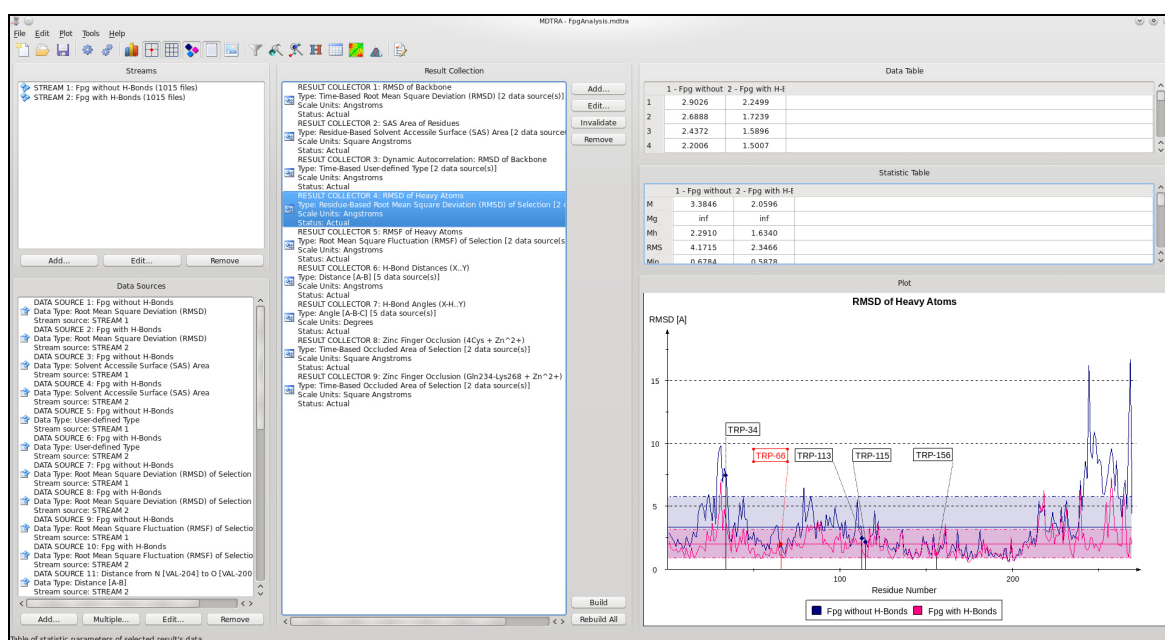


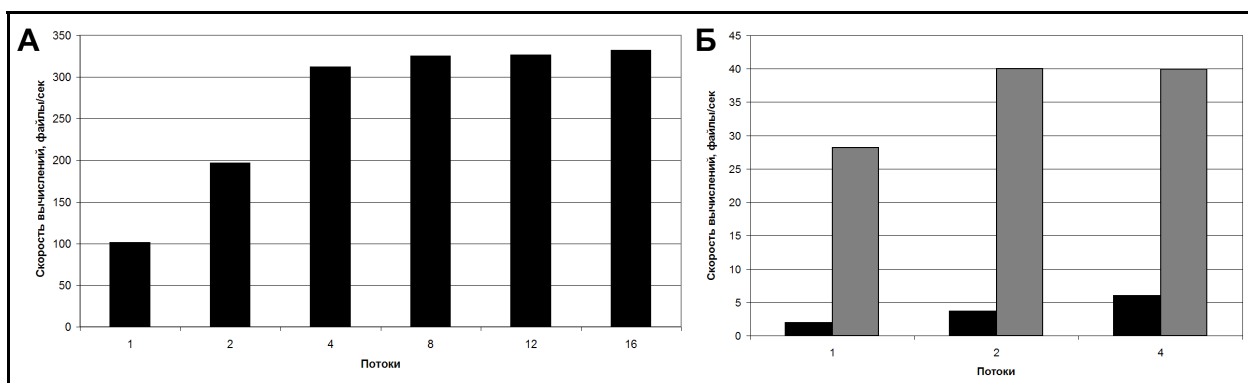
Рисунок 11. Общий вид графического пользовательского интерфейса программы MDTRA.

MDTRA написана на объектно-ориентированном языке программирования C++ с использованием межплатформенных библиотек Qt 4 (The Qt Company). Выше приводится общий вид графического пользовательского интерфейса (рисунок 11). Программа работает с траекториями, представляющими собой набор PDB-файлов для конкретного запуска расчётов МД. Возможен анализ траекторий и других форматов (например, DCD и MDCRD) после преобразования. В качестве источников данных MDTRA выступают: 1) среднеквадратичное отклонение (RMSD) или колебание остова биополимера ( $N-C\alpha-C-O$  для белков,  $O1P-O2P-P-O5'-C5'-C4'-O4'-C3'-O3'-C2'-O2'-C1'$  для нуклеиновых кислот); 2) RMSD или среднеквадратичное колебание по выбранным фрагментам; 3) среднеквадратичный радиус молекулы; 4) расстояние между двумя атомами; 5) угол между тремя атомами; 6) угол между отрезками, каждый из которых задан парой атомов; 7) торсионный или двугранный угол, заданный четырьмя атомами; 8) угол между плоскостями, каждая из которых задана тремя атомами; 9) сила, действующая на атом, или равнодействующая сил, действующих на два атома — только для формата траекторий BioPASED; 10) площадь поверхности биополимера, доступная растворителю (всей или выбранного фрагмента); 11) площадь поверхности биополимера (всей или выбранного фрагмента), закрытой другим фрагментом; 12) пользовательские типы данных.

Обработанные источники данных представляют собой наборы чисел с плавающей точкой, которые можно объединить в результаты. Результаты могут представлять данные как в виде функции от времени, так и от номера остатка. Перед расчётами каждый элемент траектории выравнивается относительно первого элемента с использованием алгоритма Кабша [160]. Производится также статистический анализ данных с вычислением следующих параметров: 1) среднее значение (арифметическое, геометрическое, гармоническое и квадратичное); 2) максимум и минимум; 3) амплитуда; 4) медиана; 5) дисперсия; 6) стандартное отклонение и стандартная ошибка. Возможен расчёт до шести дополнительных (пользовательских) статистических параметров. Если в результат собрано несколько источников данных, для них попарно вычисляется линейный коэффициент корреляции Пирсона [172].

В отличие от некоторых других программ анализа [124], MDTRA не загружает предварительно все снимки траектории в оперативную память, т.к. их количество может быть очень велико, достигая  $10^6$  [80] и занимая десятки и сотни гигабайт. Обработка траекторий осуществляется последовательно по одному файлу в один проход, что особенно эффективно при наличии быстрых дисковых накопителей (SSD).

Программа в расчётах активно использует многопоточность (до 16 потоков) и SIMD-оптимизацию (SSE-инструкции), что позволяет добиться высокой производительности на современных компьютерах. Например, при использовании четырёх потоков на четырёхъядерном CPU наблюдается трехкратный прирост производительности (рисунок 12, А); это значение учитывает не только время вычислений, но и время обращения к файлам на жёстком диске.



**Рисунок 12.** Многопоточная оптимизация программы MDTRA. (А) Прирост производительности при использовании нескольких потоков CPU. (Б) Прирост производительности при использовании массивно-параллельных вычислений на графическом процессоре. По оси абсцисс отложено число потоков CPU. Чёрным цветом показаны результаты профилирования расчётов без использования GPGPU, серым – с его использованием. Данные получены в системе с CPU Intel Core i5 2,66 ГГц (4 ядра), GPU NVIDIA GeForce GTX275 1,4 ГГц (30 потоковых мультипроцессоров).

Площадь поверхности, доступной растворителю, вычисляется по алгоритму Шрейка–Рапли [173] на основе ван-дер-ваальсовых радиусов атомов, расширенных на величину радиуса растворителя (для воды это 1,4 Å). Каждый атом описывается как набор точек, равномерно распределённый на сфере заданного радиуса. Плотность точек зависит от выбранной точности и соответствует числу разбиений начального икосаэдра по алгоритму геосферы, создающему высококачественное нормальное распределение. По умолчанию, используется 162 точки на атом. MDTRA вычисляет число точек, находящихся внутри других сфер; тогда площадь поверхности атома, доступной растворителю, будет равна:

$$S = 4\pi R^2 \left( 1 - \frac{N_{\text{внутр.}}}{N_{\text{общее}}} \right),$$

где  $S$  — площадь поверхности атома, доступной растворителю (в квадратных ангстремах),  $R$  — расширенный радиус атома,  $N_{\text{внутр.}}$  — число внутренних точек,  $N_{\text{общее}}$  — общее число точек. Общая площадь поверхности биополимера, доступная растворителю, вычисляется как сумма этих значений для всех его атомов.

Можно также посчитать площадь поверхности, закрытую заданным фрагментом. Например, в ДНК-белковых комплексах некоторые аминокислотные остатки плотно контактируют с ДНК, а в отсутствие ДНК доступны растворителю. Их суммарная площадь

поверхности есть площадь поверхности белка, закрытой ДНК. В MDTRA реализован эффективный массивно-параллельный алгоритм расчёта площади поверхности с использованием графических процессоров (GPGPU) на основе NVIDIA CUDA, что может приводить к 14-кратному ускорению расчётов по сравнению с однопоточной версией, исполняемой на центральном процессоре (рисунок 12, Б).

Некоторые инструменты MDTRA предоставляют возможность поиска значимых данных вдоль траектории или реализуют возможности, выходящие за рамки описанной архитектуры. Это такие модули, как поиск по дистанциям, поиск по торсионным углам, поиск различий в действующих силах, поиск водородных связей, быстрое создание проекта анализа ДНК и расчёт двумерного среднеквадратичного отклонения (2D-RMSD). Поиск по дистанциям — это инструмент анализа статистически значимых отклонений в расстояниях между атомами в двух траекториях. Он может быть использован для поиска источников данных и снижения их размерности. Поиск по торсионным углам во многом аналогичен поиску по дистанциям, но сравнивает разницу в торсионных углах белкового и нуклеинового остова, а так же боковых групп а. о. Модуль поиска водородных связей используется для нахождения водородных связей, стабилизирующих структуру на значимых временных промежутках. Поиск осуществляется в несколько этапов: 1) построение списка всех возможных триплетов X–H··Y (где X — донор, Y — акцептор, H — атом водорода, сплошная черта означает ковалентную связь, пунктирная – водородную); 2) расчёт энергии каждого триплета для каждого элемента траектории; 3) сведение полученных результатов в таблицу, выбирая только значимые на основе заданных пороговой энергии и частоты встречаемости в траектории. Энергия водородной связи вычисляется по формуле (соответствующей используемой в BioPASED):

$$E = E_m \left[ \left( \frac{R_m}{r} \right)^{12} - 2 \left( \frac{R_m}{r} \right)^6 \right] e^{-\left( \frac{1+\cos\theta}{s} \right)^2},$$

где E — энергия водородной связи, r — расстояние между атомами Y и H (в ангстремах),  $\theta$  — угол в триплете,  $R_m$  и  $E_m$  — параметры силового поля AMBER для конкретной пары атомов Y и H, и s — константа мягкости потенциала.

Модуль быстрого создания проекта анализа ДНК позволяет в несколько простых действий добавить в проект десятки и сотни источников данных для отдельных оснований ДНК и комплементарных пар, что ускоряет анализ этих биополимеров, как в свободном виде, так и в комплексах. Среди параметров, которые создает этот инструмент: торсионные углы нуклеотидов, фазовый угол сахара, параметры Дикерсона [174] для отдельных пар оснований и для соседних пар.

В MDTRA также присутствует инструмент построения специальных 2D-RMSD графиков, отражающих относительное отклонение между парами элементов траектории. Каждая точка на таком графике с координатами  $(n, m)$  показывает среднее значение RMSD по выбранному набору атомов между элементами траектории  $n$  и  $m$ :

$$\text{RMSD}(n, m) = \sqrt{\frac{1}{N} \sum_{i=1}^N |\bar{\mathbf{r}}_i(n) - \bar{\mathbf{r}}_i(m)|^2},$$

где  $\mathbf{r}_i(n)$  и  $\mathbf{r}_i(m)$  — радиус-векторы атома  $i$  в элементах траектории  $n$  и  $m$ , соответственно, а  $N$  — общее количество атомов в выборке. Таким образом, становится очевидной диагональная симметрия графика. Инструмент 2D-RMSD требует загрузки в память всей траектории и поэтому требователен к количеству оперативной памяти в системе.

Некоторые инструменты MDTRA оперируют с фрагментом молекулы. Синтаксис языка выбора максимально приближен к языку RasMol [164], хотя есть и заметные отличия. Интерпретатор языка выборки написан с использованием инструментов Bison (генератор грамматики; не путать с пакетом моделирования BISON) и Flex [175].

Хотя в MDTRA присутствует большое количество заранее заданных источников данных и статистических параметров, реализован также механизм расширения: программируемые источники данных. Собственные подпрограммы анализа могут быть написаны на слегка модифицированном языке Lua [176] в рамках интерфейса программирования (API), и сразу же выполнены.

Таким образом, программа MDTRA — это межплатформенное, свободное программное решение с открытым исходным кодом и графическим пользовательским интерфейсом, предназначенное для детального анализа МД траекторий. Подробная документация по программе на английском языке, а также сама программа и её исходный код доступны на веб-сайте (<http://bison.niboch.nsc.ru/mdtra.html>).

Программа активно использовалась в работе для анализа молекулярно-динамических траекторий и построения графиков.

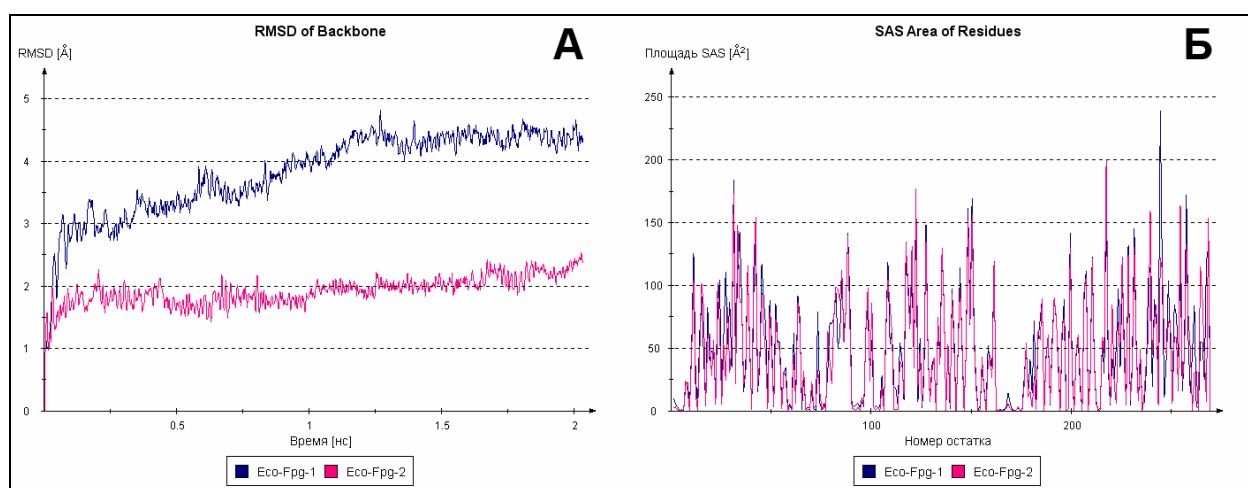
## 4.2. Исследование структуры и активности Frg

В работе проведено исследование структуры и динамики некоторых формамидопиримидин-ДНК-*N*-гликозилаз бактерий: подтверждены данные о мобильности отдельных фрагментов, исследовано поведение каталитической пары Pro-Glu в активном центре в зависимости от состояния ионизации аминокислотных остатков с оценкой их  $pK_a$ , получены новые данные о специфичности ферментов к основанию напротив повреждённого, изучен вклад молекул растворителя в формирование и поддержание стабильности фермент-субстратных

комплексов. Выявленные новые структурные детерминанты в целом согласуются с концепцией пластичности активного центра, которая хорошо объясняет широкую субстратную специфичность Fpg.

#### 4.2.1. Значимость водородных связей для поддержания структуры Fpg

Ранее было показано, что отдельные сегменты молекулы фермента Fpg имеют повышенную мобильность [177]. В частности, интерес представляли участки, которые могли давать вклад во флюоресценцию белка в процессе реакции. Поэтому особое внимание уделялось остаткам триптофана, которые могут делать вклад в изменение флюоресценции во время экспериментов по предстационарной кинетике [177]. Однако ранее проводилась оценка мобильности сегментов по результатам МД без учёта водородных связей. Для окончательного решения вопроса, а также с целью протестировать разработанное программное обеспечение, исследовали конформационную мобильность Fpg *E. coli* в течение 2 нс как с учётом водородных связей (модель Eco-Fpg-1), так и без учёта их (Eco-Fpg-2), для проверки соответствия результатов ранее полученным данным.

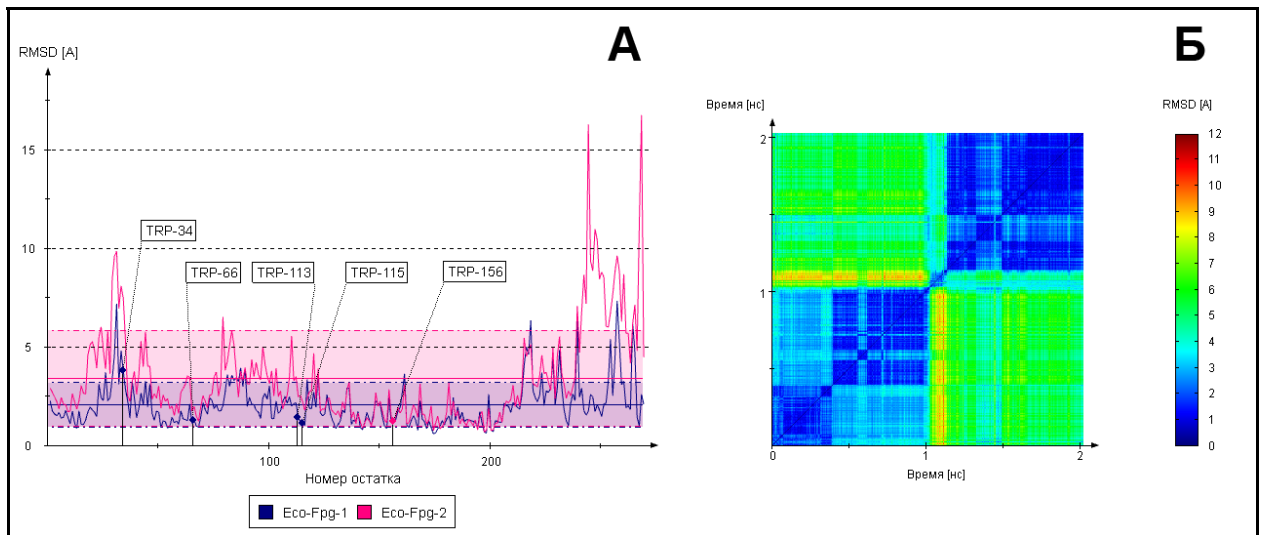


**Рисунок 13.** Общий анализ траекторий моделей Fpg *E. coli*. (А) RMSD атомов остова вдоль траекторий. (Б) Поверхность белка, доступная молекулам воды, для каждого аминокислотного остатка. Пояснения в тексте.

Анализ RMSD атомов остова моделей вдоль траектории моделирования показал, что в целом они оказались стабильными (рисунок 13, А). Как и ожидалось, модель Eco-Fpg-2 демонстрировала более высокое значение колебаний со средним значением RMSD 3,81 Å против 1,92 Å у модели Eco-Fpg-1. Площадь поверхности биополимера, доступной молекулам воды, с высокой степенью коррелирует между обеими моделями (рисунок 13, Б); коэффициент корреляции Пирсона составляет 0,90. Среднеквадратичный радиус, хотя и несколько возрастает на начальных этапах МД у модели Eco-Fpg-2, в итоге стабилизируется и составляет 19,6 Å и 19,9 Å для первой и второй моделей, соответственно. Это означает, что в обеих моделях

конформация молекулы белка в целом оставалась сходной; однако некоторые остатки модели Eсо-Fpg-2 оказались более доступными растворителю.

Для того, чтобы выяснить, насколько подвижен тот или иной аминокислотный остаток в моделях, проводили анализ RMSD тяжёлых атомов (т.е. всех атомов, кроме атомов водорода) для каждого остатка (рисунок 14, А). Критерием повышенной подвижности было превышение значением RMSD величины среднего значения по всем остаткам плюс одно стандартное отклонение. В целом данные по мобильности модели Eсо-Fpg-2 согласуются с ранее опубликованными [177]; в модели Eсо-Fpg-1 обнаружили следующие аминокислотные остатки и сегменты с повышенной подвижностью: 26, 29–34, 81, 83–87, 117, 122, 161, 215–218, 223, 228, 231, 239, 253–254, 256–258. Некоторые из них отсутствуют в модели Eсо-Fpg-2, что свидетельствует о том, что учёт водородных связей не приводит к равномерному снижению конформационной мобильности.



**Рисунок 14.** Анализ конформационной подвижности моделей Fpg *E. coli*. (А) RMSD тяжёлых атомов для каждого аминокислотного остатка. Позиции остатков триптофана отмечены маркерами и подписями. Горизонтальные сплошные линии показывают средние значения RMSD по всем остаткам; цветная заливка между ними и штрих-пунктирными линиями показывает диапазон одного стандартного отклонения. (Б) 2D-RMSD тяжёлых атомов остатка Trp34 вдоль траектории, модель Eсо-Fpg-1. Пояснения в тексте.

Повышенную мобильность остатка Trp34 обнаружили как в модели Eсо-Fpg-1, так и Eсо-Fpg-2, с амплитудами 3,77 Å и 7,47 Å, соответственно. Любопытно, что большинство высокоподвижных остатков сосредоточено в двух областях белка, цинковом пальце и петле  $\beta 1/\beta 2$  (последняя область включает Trp34), которые входят в большую и малую бороздки ДНК, соответственно, когда фермент сканирует ДНК на предмет повреждённых нуклеотидов [28]. Такая подвижность может быть важна для наилучшей адаптации к изменениям геометрии ДНК во время поиска повреждённых остатков и катализа. Ещё один сегмент с высоким значением RMSD, петля  $\beta 5/\beta 6$ , лежит на поверхности белка далеко от ДНК-связывающей бороздки, и её подвижность вряд ли имеет какую-либо функциональную значимость. Необходимо отметить,



что обе модели не содержат протяжённых высококомобильных участков. Подвижность отдельных аминокислотных остатков, например, Lys217, играющего важную роль в узнавании повреждённого основания [54, 59], оказывается сравнимой в обеих моделях (4,6 Å в Eсо-Fpg-1, 4,8 Å в Eсо-Fpg-2). Это не удивительно с учётом того факта, что Lys217 формирует водородные связи с ДНК [54], которой в моделях нет; однако это наблюдение оказывается важным свидетельством в пользу того, что в свободном белке этот аминокислотный остаток не участвует в формировании внутренней сети водородных связей.

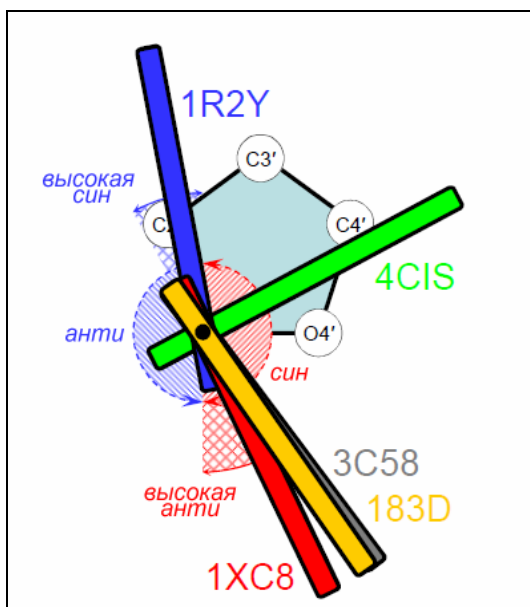
Остальные остатки триптофана не продемонстрировали высокую подвижность в обеих моделях, что, впрочем, не означает, что они не дают вклад в наблюдаемое изменение флуоресценции. Проводили анализ 2D-RMSD тяжёлых атомов для каждого остатка триптофана (Trp34, Trp66, Trp113, Trp115 и Trp156) в модели Eсо-Fpg-1. У остатков Trp66, Trp113, Trp115 и Trp156 не наблюдалось значительной подвижности, хотя присутствуют небольшие конформационные сдвиги порядка 3.0 Å. Остаток Trp34, напротив, оказался высококомобильным (рисунок 14, Б). Он претерпевает два заметных конформационных перехода (широкие голубая и зелёная полосы на графике) и один кратковременный скачок (узкая красная полоса). Таким образом, все остатки триптофана могут давать вклад во флуоресценцию, но остаток Trp34 оказывается наиболее значимым.

Поиск водородных связей в модели Eсо-Fpg-1 выявил 72 значимых водородных связи. Значимой считали связь с энергией больше 0,4 ккал/моль и присутствующую, по крайней мере, в 95% записанных состояний системы (снапшотов). Большая часть этих связей — связи, стабилизирующие  $\alpha$ -спирали и  $\beta$ -складки. Многие из них находятся в зоне линкера, связывающего домены белка, или рядом с этой зоной, а также вокруг цинкового пальца. Связи оказались довольно стабильными со средней дистанцией 3,0 Å и углом триплета X–H $\cdots$ Y порядка 166°; это однозначно свидетельствует об участии найденных связей в формировании и поддержании вторичной структуры Fpg.

Для уточнения влияния водородных связей на стабильность «цинкового пальца» вычисляли заслонённую площадь белка его фрагментом: либо собственно остатками цистеина «цинкового пальца» (Cys243–Cys266) и ионом Zn<sup>2+</sup>, либо ионом Zn<sup>2+</sup> и участком молекулы белка, участвующим в формировании  $\beta$ -шпильки и неупорядоченных регионов, окружающих «цинковый палец» (Gln234–Lys268). Зона контакта в Eсо-Fpg-1 увеличивалась и стабилизировалась быстрее для обоих способов задания фрагмента, заслоняющего белок. Это согласуется с ранее полученными из статических кристаллических структур данными о протяжённой сети водородных связей, привязывающих «цинковый палец» к мотиву H2TH в C-концевом домене Fpg [28], и демонстрирует эту сеть в динамике.

#### 4.2.2. Состояния ионизации каталитически значимых аминокислотных остатков в активном центре Fpg

Прежде, чем исследовать активный центр Fpg методом МД, необходимо ответить на вопрос, имеющий большое значение для катализа: в каких зарядовых состояниях находятся каталитически значимые аминокислотные остатки? Протонирование этих аминокислотных остатков (Pro1 и Glu2), напрямую участвующих в катализе, может оказать существенное влияние на надёжность и достоверность моделей, однако до сих пор не проводилось попыток оценить  $pK_a$  этих групп экспериментально. Поэтому проводили МД всех четырёх возможных зарядовых состояний. Кроме того, если 8-oxoGua находится в анти-конформации (либо в пограничной области «высокой син») в большинстве известных моделей (рисунок 15), то относительно dAdo напротив повреждения таких данных нет. В связи с этим

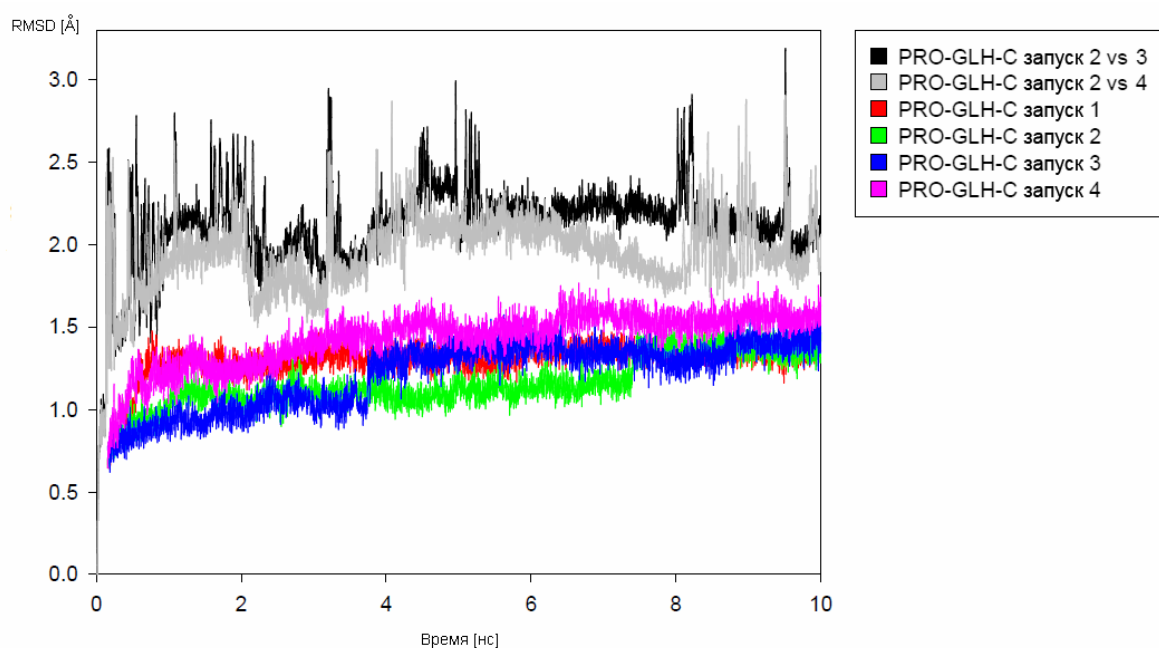


**Рисунок 15.** Схематичное положение повреждённого основания относительно плоскости рибозы в структурах, содержащих свободную ДНК, содержащую 8-оксогуанин (PDB ID: 183D), а также в комплексах Fpg с ДНК, содержащей поврежденные основания пуринового ряда (PDB ID: 1R2Y, 4CIS, 3C58 и 1XC8).

пару 8-oxoGua:Ade моделировали в обеих возможных конформациях аденина. Таким образом, исследовали 12 различных модельных комплексов.

Среднеквадратичное отклонение остова, рассчитанное вдоль траектории каждые 2 пс, быстро увеличивалось в течение первых 500 пс динамики и стабилизировалось до значения  $\sim 1,6 \text{ \AA}$ . Среднеквадратичное отклонение по остаткам показало, что остов ДНК имеет повышенную мобильность в сравнении с белком: средние значения для оснований ДНК выше на  $0,54 - 0,98 \text{ \AA}$  в зависимости от модели. В целом комплексы оказались стабильны вдоль всей траектории со среднеквадратичным радиусом  $\sim 20 \text{ \AA}$  для всех моделей. Угол изгиба ДНК был также стабилен (медиана  $55^\circ - 63^\circ$ ).

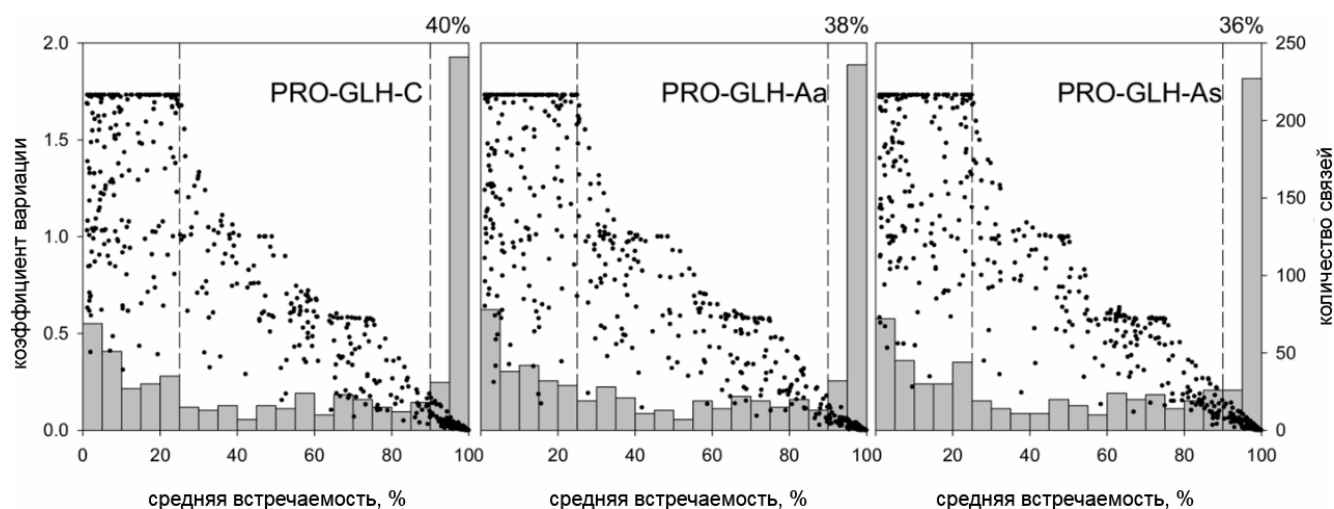
Для проверки согласованности результатов между независимыми компьютерными экспериментами выбрали три модели (PRO-GLH-C, PRO-GLH-Aa и PRO-GLH-As). Проводили три дополнительных моделирования методом МД для каждой из перечисленных моделей с различным протоколом разогрева стартовой структуры (см. раздел 3.2.2) для обеспечения независимых начальных условий для последующей глобальной оптимизации. Таким образом, получили 9 траекторий длительностью 10 нс каждая. Полученные 4 траектории для каждой из перечисленных моделей сравнивали между собой. Среднеквадратичные отклонения внутри одного и того же запуска (например, для модели PRO-GLH-C; рисунок 16) были сопоставимыми (1,0 – 1,5 Å в последние 8 нс динамики), а между запусками ожидаемо выше (1,9 – 2,2 Å), но не показывали существенного расхождения в мобильности.



**Рисунок 16.** Оценка воспроизводимости моделирования на основе сравнения среднеквадратичного отклонения (RMSD) остовов моделей PRO-GLH-C вдоль траекторий, полученных в результате повторных запусков моделирования. График попарных сравнений RMSD между снимками второй и третьей траекторий показан чёрным цветом; между снимками второй и четвёртой — серым; графики RMSD траекторий относительно первого снимка показаны красным (первая траектория, исходная), зелёным (вторая траектория), синим (третья траектория) и пурпурным (четвёртая траектория).

Паттерн водородных связей, включая специфичные для моделей, был также согласованным во всех четырёх запусках МД (рисунок 17); пороговое отсечение 90% по средней встречаемости показало, что стабильными оказываются все уотсон-криковские контакты и 79% связей в основной цепи, наблюдаемых в исходной структуре 1XC8.

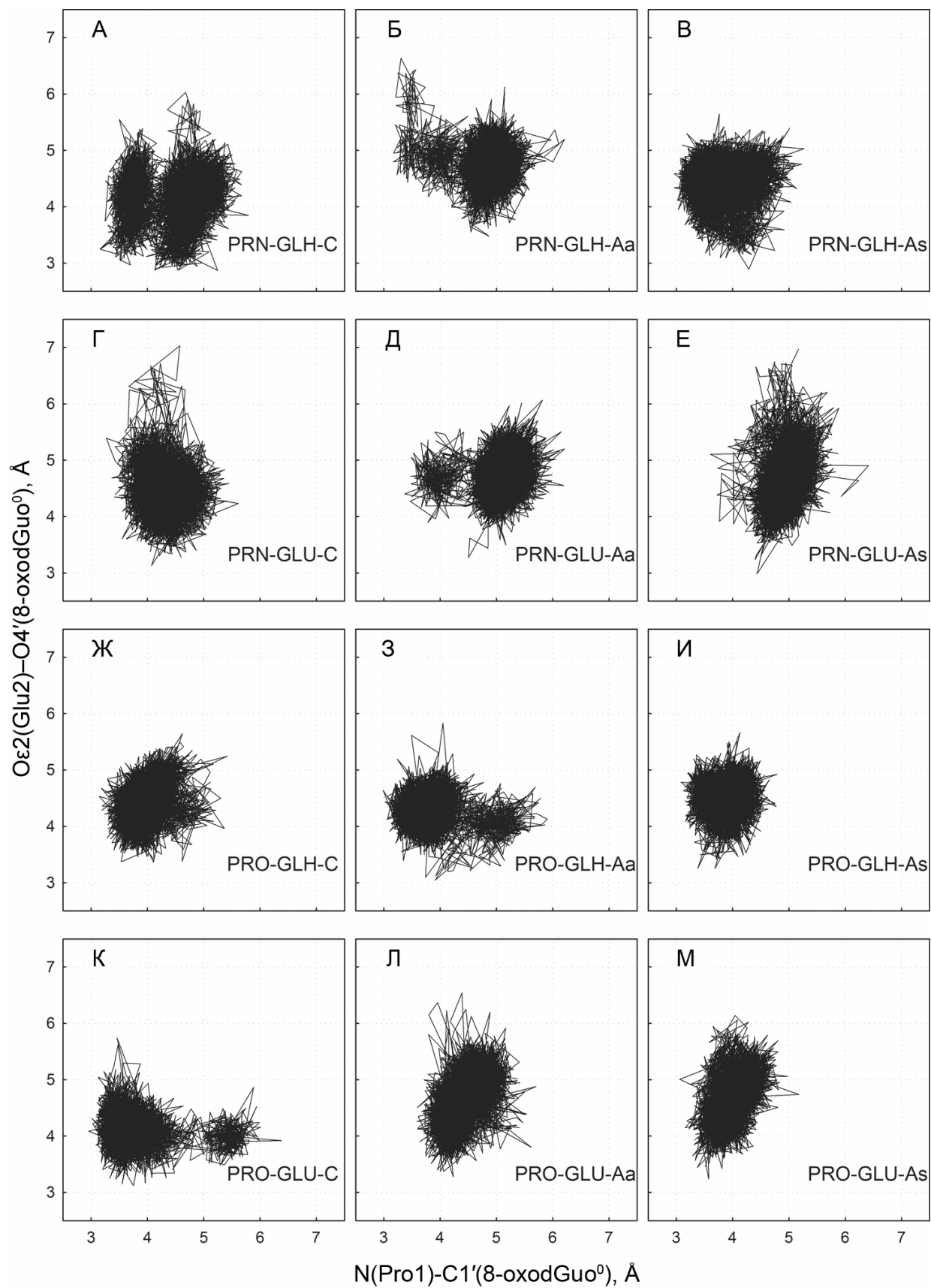
Анализ каталитической пары аминокислотных остатков Pro1 и Glu2, а также выбор модели зарядового состояния для последующего анализа проводили следующим образом. Исследовали две ключевые дистанции: N(Pro1)–C1'(8-oxodGuo<sup>0</sup>) и Oε2(Glu2)–O4'(8-oxodGuo<sup>0</sup>) (таблица 4, рисунок 18). Во всех С-моделях расстояние Oε2(Glu2) – O4'(8-oxodGuo<sup>0</sup>)



**Рисунок 17.** Воспроизводимость паттерна водородных связей в повторных моделированиях PRO-GLH. Точками обозначено соответствие коэффициентов вариации встречаемости конкретных водородных связей, вычисленных по четырём независимым экспериментам, средней встречаемости связи. Гистограммы показывают распределение средней встречаемости (в одинаковом масштабе). Проценты в верхней части графиков соответствуют доле водородных связей со средней встречаемостью выше 90%.

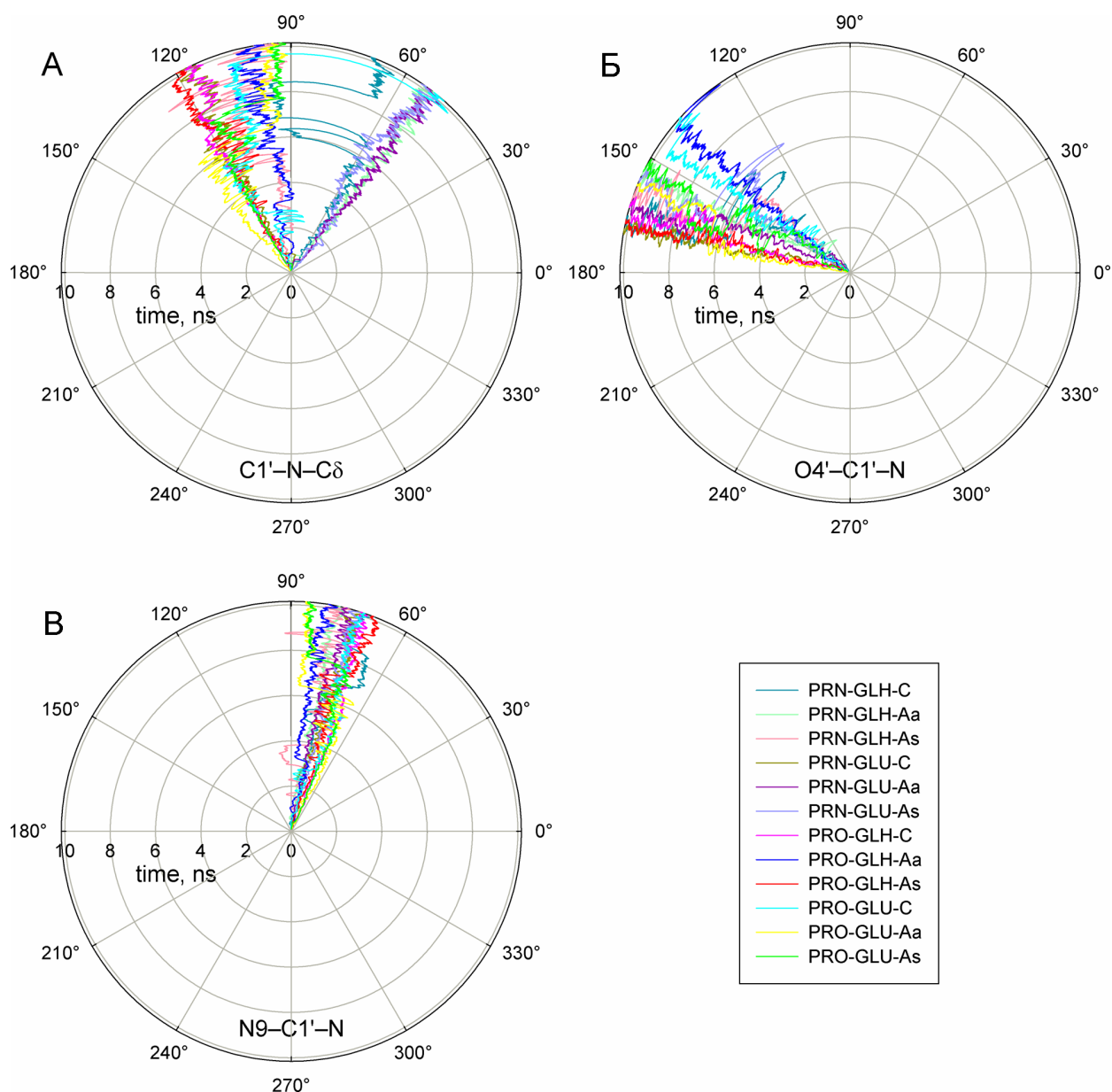
**Таблица 4.** Ключевые расстояния и углы, связанные с каталитически важными аминокислотными остатками Pro1 и Glu2, а также атомами нуклеозида. Приведены медианы значений вдоль траекторий; в скобках даны 90-процентные диапазоны. В квадратных скобках приведены значения, усреднённые по трём повторным расчётам динамики модели PRO-GLH.

Модель		Расстояние N(Pro1) – C1'(охоG <sup>0</sup> ), Å	Расстояние Oε2(Glu2) – O4'[охоG <sup>0</sup> ], Å	Угол C1'(охоG <sup>0</sup> ) – N(Pro1) – Cδ(Pro1), градусы	Угол O4'(охоG <sup>0</sup> ) – C1'(охоG <sup>0</sup> ) – N(Pro1), градусы	Число снимков с оптимальной геометрией
С	PRN-GLH	3,73 (3,41–4,04) 4,74 (4,29–5,20)	4,20 (3,41–4,79)	61 (55–95)	150 (125–169)	409
	PRN-GLU	4,31 (3,83–4,88)	4,46 (3,87–5,12)	116 (80–122)	168 (145–171)	293
	PRO-GLH	3,97 (3,55–4,48)	4,42 (3,90–4,96)	119 (113–122)	165 (156–168)	1077 [788]
	PRO-GLU	3,59 (3,27–4,17) 5,43 (5,01–5,84)	4,13 (3,68–4,60)	107 (78–122)	145 (135–151)	141
А (анти)	PRN-GLH	4,86 (4,06–5,30)	4,75 (4,24–5,27)	52 (49–77)	153 (136–157)	0
	PRN-GLU	5,12 (4,35–5,57)	4,76 (4,22–5,29)	52 (49–75)	158 (132–165)	0
	PRO-GLH	3,79 (3,34–4,78)	4,27 (3,83–4,74)	100 (89–107)	141 (136–152)	73 [277]
	PRO-GLU	4,39 (3,96–4,97)	4,62 (3,92–5,33)	124 (94–133)	168 (156–172)	78
А (син)	PRN-GLH	3,82 (3,26–4,54)	4,44 (3,71–4,94)	101 (89–113)	155 (136–167)	225
	PRN-GLU	4,94 (4,52–5,38)	4,65 (4,01–5,48)	55 (48–62)	149 (135–160)	0
	PRO-GLH	3,92 (3,43–4,31)	4,49 (4,02–4,95)	119 (107–122)	167 (157–170)	961 [137]
	PRO-GLU	3,99 (3,57–4,43)	4,65 (4,02–5,34)	118 (93–121)	154 (144–164)	184



**Рисунок 18.** Популяции ключевых дистанций от каталитически важных аминокислотных остатков Pro1 и Glu2 до атомов нуклеозида. Пояснения в тексте.

демонстрировало распределение, близкое к нормальному, с медианами 4,20 Å, 4,46 Å, 4,42 Å и 4,13 Å соответственно для моделей PRN-GLH-C, PRN-GLU-C, PRO-GLH-C и PRO-GLU-C. Расстояние N(Pro1)–C1'(8-oxodGuo<sup>0</sup>) оказалось менее стабильным: близкое к нормальному распределение в моделях PRN-GLU-C и PRO-GLH-C (медианы 4,31 Å и 3,97 Å; рисунок 18, Г и Ж) превращалось в два отчётливых пика на гистограмме в моделях PRN-GLH-C and PRO-GLU-C (рисунок 18, А и К). В первом случае 24% времени медиана дистанции составляла 3,73 Å и 76% времени 4,74 Å. Во втором большую часть времени (95% траектории) модель находилась в конформации с коротким расстоянием (медиана 3,59 Å), но с кратковременными (5% траектории) увеличениями до 5,44 Å. В моделях, где напротив 8-охоGua находилось основание



**Рисунок 19.** Популяция ключевых углов, описывающих ориентацию иминогруппы Pro1 относительно разрываемых связей нуклеотида. Значения сглажены методом центрального скользящего среднего. Пояснения в тексте.

Ade, наилучшая стабильность обеспечивалась в зарядовых состояниях PRO-GLH (по расстоянию Oε2(Glu2)–O4'(8-oxodGuo<sup>0</sup>) медианы составляли 4,27 Å и 4,49 Å для *анти*-Ade и *син*-Ade соответственно; по расстоянию N(Pro1)–C1'(8-oxodGuo<sup>0</sup>) — 3,77 Å и 3,92 Å). В остальных зарядовых состояниях наблюдалось существенное отклонение по меньшей мере в одной из дистанций. Таким образом, иминогруппа Pro1 и карбоксильная группа боковой цепи Glu2 занимали наилучшее положение относительно позиций, оптимальных для образования важных для катализа связей, C1' и O4' остатка 8-oxodGuo<sup>0</sup>, в модели PRO-GLH вне зависимости от основания напротив повреждённого.

Угол атаки Pro1 на атом C1' определяет, какая из связей будет разорвана первой: C1'–O4' или C1'–N9. Оба варианта допустимы по механизму S<sub>N</sub>2, однако предпочтительной считается атака в направлении C1'–O4'. Действительно, анализ углов атаки вдоль траекторий (рисунок 19) показал, что медиана угла N9–C1'–N(Pro1) составляет 69°–82°, что находится далеко за пределами идеальной геометрии (180°). В то же время угол O4'–C1'–N(Pro1) демонстрировал значения, приемлемые для катализа (медианы в диапазоне 141°–168°), хотя в некоторых моделях были значения, неоптимальные для перекрывания орбиталей — меньше 155° (все модели PRN-GLH, PRN-GLU-As, PRO-GLH-Aa, PRO-GLU-C). Угол ориентации иминогруппы C1'–N(Pro1)–Cδ(Pro1), в идеале составляющий 107°, находился в допустимом диапазоне во всех PRO моделях, а также в PRN-GLH-As и PRN-GLU-C (медианы в диапазоне 98°–121°), а в остальных моделях конформация оказывалась неблагоприятной (53°–65°). Основание напротив повреждённого и тут не оказывало значительного влияния на угол. При сравнении моделей PRO-GLH-C и PRO-GLU-C обращает на себя внимание то, что в первом случае угол O4'–C1'–N(Pro1) гораздо ближе к идеальному (среднее значение 164° ± 8° против 144° ± 8°), а угол C1'–N(Pro1)–Cδ(Pro1) лишь чуть хуже (среднее значение 118° ± 8° против 106° ± 17°). Однако в модели PRO-GLU-C к концу моделирования этот угол сдвигался в диапазон 50°–85°, практически неприемлемый для атаки (рисунок 19, А).

Расчёты с использованием эмпирического алгоритма PROPKA [178] позволили оценить pK<sub>a</sub> обеих каталитически важных групп Fpg на основе структур различных видов (таблица 5). Выяснилось, что pK<sub>a</sub> Pro1 заметно снижена (на 0,35–2,99) в сравнении со стандартным значением для N-конца полипептида, а pK<sub>a</sub> Glu2 значительно выше (на 1,54–3,64) в сравнении со стандартным значением для боковой цепи Glu. Похожие значения pK<sub>a</sub> наблюдаются для эндонуклеазы V бактериофага T4 (DenV), также использующей в катализе N-концевую аминогруппу и карбоксильную группу Glu [147]. Структуры свободного Fpg и фермента, связанного с неповреждённой ДНК, продемонстрировали более низкие значения pK<sub>a</sub> Glu2, что

может свидетельствовать о специфической активации этой группы при выворачивании повреждённого основания.

**Таблица 5.** Данные  $pK_a$  а.о. Pro1 и Glu2 некоторых структур Fpg.

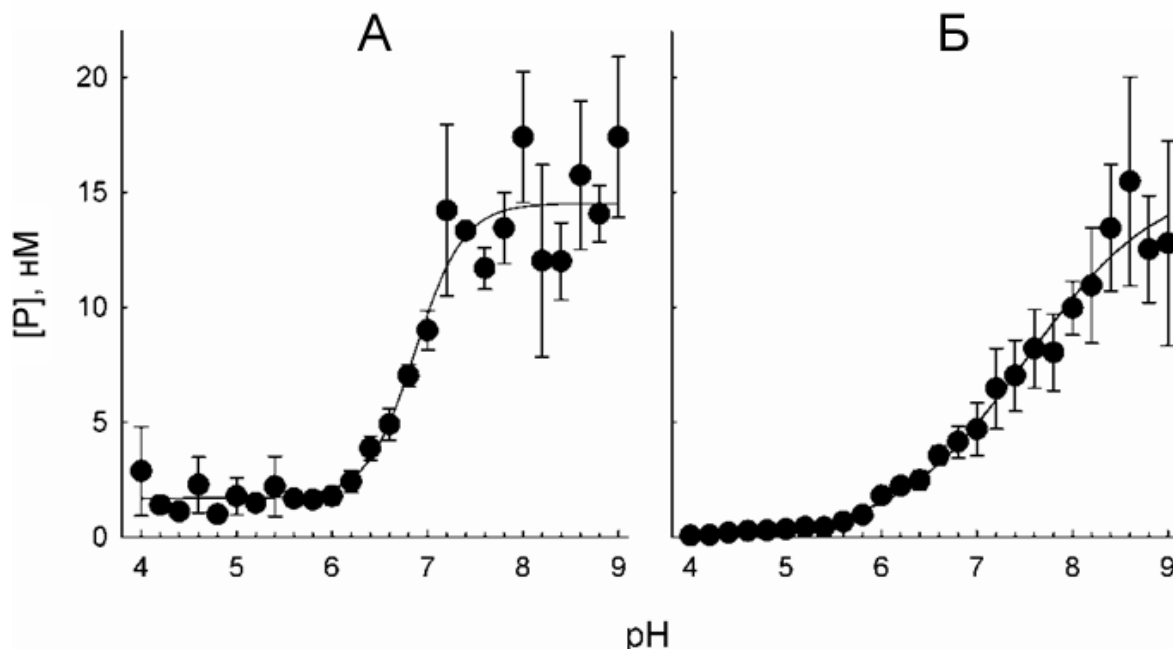
<i>PDB ID</i>	$pK_a$ Pro1	$pK_a$ Glu2	Фермент	Тип структуры
(стандартное)	8,00	4,50	—	—
(модель)	6,00	6,72	<i>Lla</i> -Fpg	комплекс с дцДНК (8-охоGua:Cyt)
1XC8	5,01	7,62	<i>Lla</i> -Fpg	комплекс с дцДНК (карба-Fapy-Gua:Cyt)
1EE8	7,15	6,40	<i>Tth</i> -Fpg	свободный фермент
1PM5	6,36	8,14	<i>Lla</i> -Fpg	комплекс с дцДНК (Fur:Cyt)
1PJ1	7,65	8,05	<i>Lla</i> -Fpg	комплекс с дцДНК (АП-п:Cyt)
1L1T	7,39	7,14	<i>Bst</i> -Fpg	комплекс с дцДНК (АП-в:Cyt)
1L2C	7,35	7,43	<i>Bst</i> -Fpg	комплекс с дцДНК (АП-в:Thy)
1L2D	6,87	7,27	<i>Bst</i> -Fpg	комплекс с дцДНК (АП-в:Gua)
2F5O	6,95	6,22	<i>Bst</i> -Fpg	сшивка с неповрежд. дцДНК
2F5P	6,38	6,41	<i>Bst</i> -Fpg	сшивка с неповрежд. дцДНК
2F5N	7,05	6,04	<i>Bst</i> -Fpg	сшивка с неповрежд. дцДНК

Для экспериментальной оценки функциональной значимости зарядовых состояний каталитической пары проводили анализ рН-профиля активности фермента *Eco*-Fpg, подразумевая в целом консервативный механизм катализа у него и *Lla*-Fpg. Наличие двух функционально важных ионизируемых остатков в активном центре фермента, один из которых должен быть протонирован, в то время как другой — депротонирован для проявления активности, должен придавать графику рН-зависимости колоколообразную форму. Такая форма была ранее показана для алкиладенин-ДНК-гликозилазы человека, монофункциональной и структурно отличающейся от Fpg (она использует для катализа пару аминокислотных остатков His–Glu) [179].

Однако Fpg продемонстрировал единственный переход в диапазоне рН от 4 до 9 (рисунок 20). Это наблюдалось как в условиях одного оборота фермента (рисунок 20, А; скоростьлимитирующей стадией является катализ), так и в стационарных условиях (рисунок 20, Б; скоростьлимитирующими стадиями являются как катализ, так и высвобождение продукта реакции). Рассчитанные для описанных условий значения  $pK_a$  составили 6,9 и 7,5 соответственно. Увеличение  $pK_a$  в условиях стационарной кинетики происходит, вероятно, из-за влияния рН на высвобождение продукта. Принимая во внимание абсолютную необходимость депротонирования Pro1 для протекания нуклеофильной реакции, можно сделать вывод, что



единственный перегиб кривой увеличения активности с ростом pH свидетельствует о том, что переход между зарядовыми состояниями Glu2 не лимитирует скорость реакции. В частности, можно предположить, что такой переход может достаточно легко осуществляться при помощи молекул воды, попадающих в активный центр (см. раздел 4.2.4).



**Рисунок 20.** Графики pH-зависимости активности *Eco-Frg* (A) в условиях одного оборота фермента; 500 нМ Frg, 100 нМ субстрат, 0° С; (Б) в стационарных условиях; 2 нМ Frg, 100 нМ субстрат, 37° С.

С учётом всех описанных факторов, наиболее структурно обоснованным представляется набор моделей PRO-GLH, т.е. моделей с заряженной N-концевой иминогруппой Pro1 (в дальнейшем активируемой внешним акцептором) и нейтральной боковой карбоксильной группой Glu2, выступающей в роли донора протона. После депротонирования остатка Pro1 модели PRN-GLU и PRN-GLH также проводят значительное время в каталитически компетентной конформации.

#### 4.2.3. Специфичность Frg к основанию напротив повреждённого

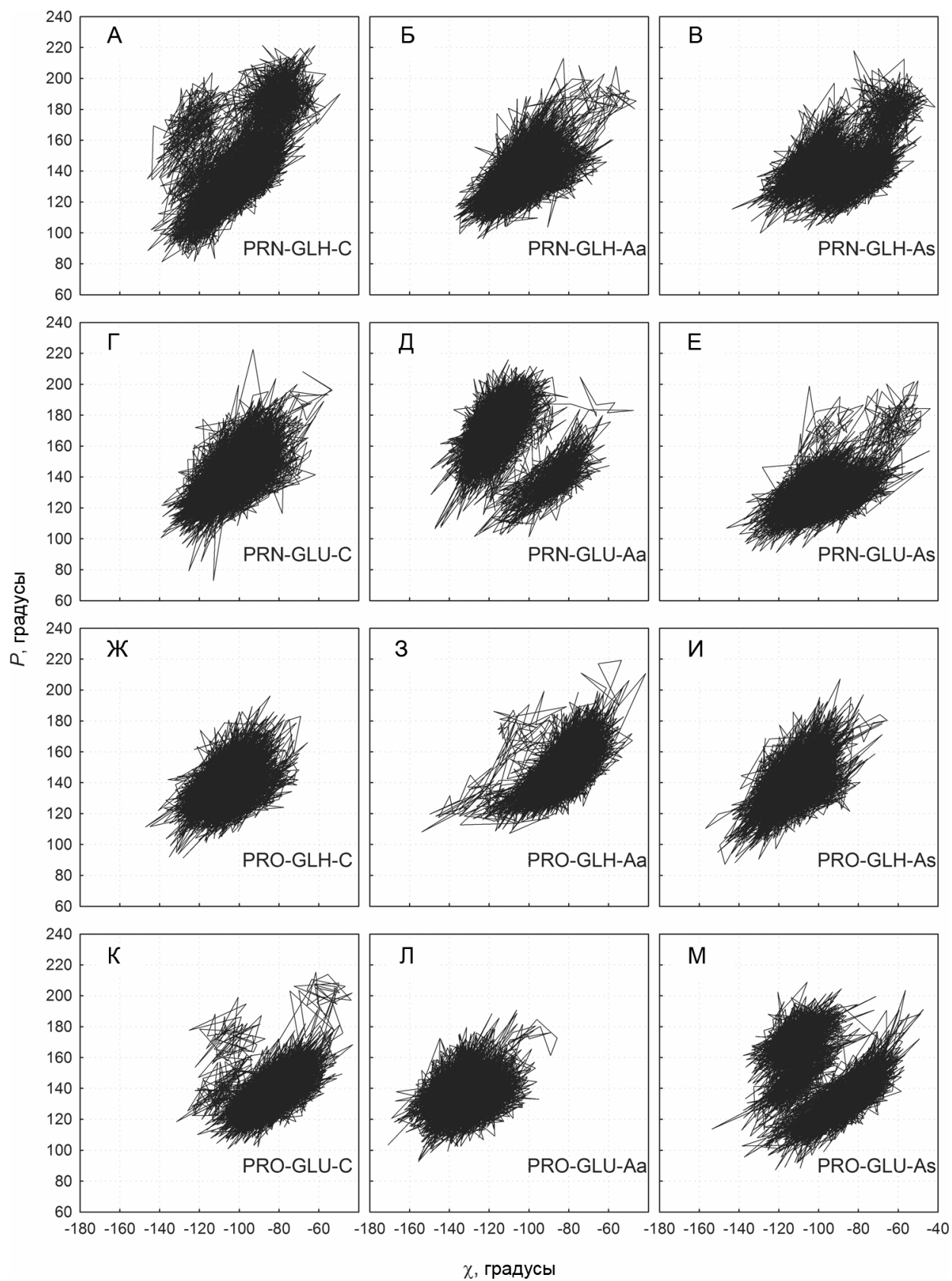
Функционирование GO-системы сильно зависит от специфичности Frg к основанию напротив повреждённого: так как 8-охоGua эффективно образует пары как с Cyt, так и с Ade, фермент должен вырезать 8-охоGua только из пары с Cyt, иначе произойдёт мутация. Поэтому весьма важно рассмотреть взаимодействия в предкаталитическом комплексе Frg, содержащем 8-охоGuo как напротив Cyt, так и напротив Ade. Общепринятый в настоящее время механизм реакции вовлекает нуклеофильную атаку C1' повреждённого дезоксинуклеозида неподеленной электронной парой иминогруппы N-концевого остатка Pro1. Это приводит к разрыву гликозидной связи, раскрытию рибозного кольца и образованию ковалентного интермедиата

между ДНК и Fpg с последующими двумя последовательными шагами элиминирования 3'- и 5'-концевых фосфатных групп, и далее гидролизом основания Шиффа. Однако остаётся немало вопросов относительно первых этапов реакции. Например, неясен механизм узнавания охoGua в активном центре фермента, а также механизм переноса протона в процессе реакции. До сих пор не был исследован экспериментально или методом моделирования комплекс Fpg с ДНК, содержащей пару охoGua:Ade. Исследование методом молекулярной динамики комплексов Fpg с повреждённой ДНК, содержащей различные основания напротив охoGua (Ade или Cyt) позволяет узнать больше о структурных детерминантах специфичности к основанию напротив повреждённого, а также на механизм узнавания и выщепления охoGua.

Для выявления факторов, отвечающих за более низкое сродство Fpg к основанию Ade напротив повреждённого (в 20 – 30 раз ниже, чем к Cyt [38]), проводили сравнение моделей C, Aa и As. Выяснилось, что замена не приводила к значительным изменениям энергии связывания и конформационной динамики. Угол выворачивания 8-охoGua был примерно одинаковый во всех моделях (медиана в диапазоне  $74^\circ - 87^\circ$ ), и повреждённое основание во всех случаях размещалось в активном центре фермента. Конформация 8-охoGua переходила из *высокой анти* в *анти* в процессе минимизации энергии (угол  $\chi$  изменился от  $-87^\circ$  до  $-115^\circ$ ) и оставалась такой в большинстве моделей. Однако некоторые модели продемонстрировали уход в *син* (но по-прежнему близко к *высокой анти*), а именно PRO-GLH-Aa и PRO-GLU-C ( $\chi = -83^\circ \pm 12^\circ$ ), а модель PRO-GLU-Aa в ещё сильнее выраженную *анти*-конформацию ( $\chi = -131^\circ \pm 11^\circ$ ). Петля, гомологичная известной O<sup>6</sup>-петле вокруг 8-охoGua в модели 1R2Y Bst-Fpg (Thr220–Tyr224), присутствует и в Lla-Fpg (Ser218–Tyr222), формируя ту же сеть контактов, несмотря на то, что повреждённое основание находится в *анти*-конформации вместо *син*. Это вполне согласуется с известными литературными данными [58, 134].

Сравнение угла псевдовращения с углом  $\chi$  повреждённого основания 8-охoGua показало наличие высокой корреляции флуктуаций внутри модели и практически нулевую корреляцию между моделями. В 8 из 12 моделей распределение обоих углов было однородным (рисунок 21), в моделях PRN-GLU-Aa и PRO-GLU-As наблюдали два кластера (рисунок 21, Д и М), в PRN-GLH-C распределение угла  $\chi$  было однородным, а псевдовращение распадалось на две зоны (рисунок 21, А), а в PRN-GLH-As — наоборот, две чёткие зоны видели у распределения угла  $\chi$  (рисунок 21, В). Это свидетельствует о высокой энергетической связанности конформаций сахара и основания, определяющих ориентацию 8-охoGua, а влияние аминокислотного окружения, по-видимому, минимально.

В обычных комплексах с субстратами, где Cyt находится напротив повреждения, это основание формировало две водородные связи с остатком Arg109 после того, как он внедрялся

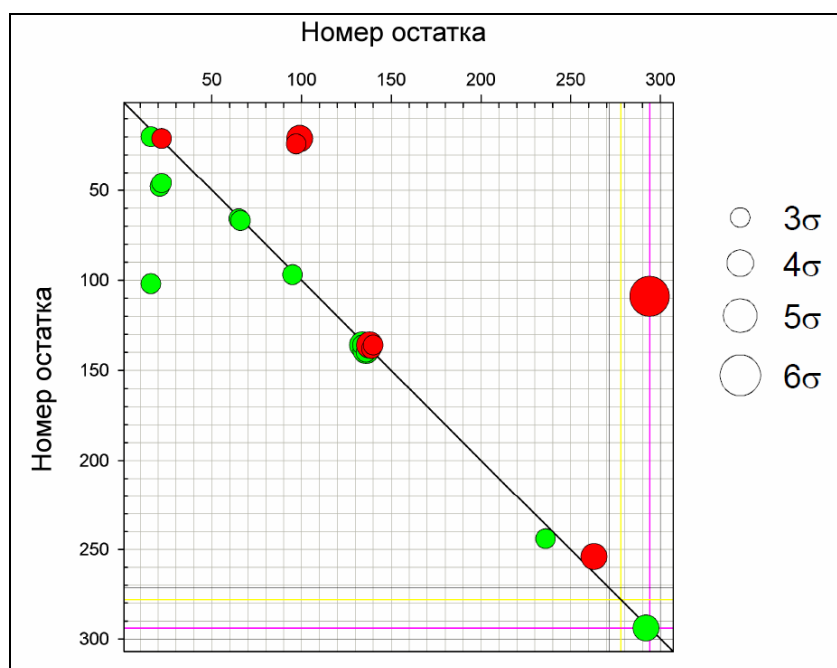


**Рисунок 21.** Популяции углов псевдовращения ( $P$ ) и гликозидного угла ( $\chi$ ) повреждённого основания 8-oxodGuo. Пояснения в тексте.

в ДНК:  $N\epsilon(\text{Arg109})\text{--}O^2(\text{Cyt}^{(0)})$  и  $N\eta 2(\text{Arg109})\text{--}N3(\text{Cyt}^{(0)})$ . Эти связи, как считается, отвечают за специфическое узнавание Cyt. При замене последнего на *син*-Gua остаток Arg109 чуть менее эффективно связывается с N7 и O<sup>6</sup> соответственно, а Thy сохраняет лишь одну связь  $N\epsilon(\text{Arg109})\text{--}O^2(\text{Thy}^{(0)})$ , в то время как вторая не образуется из-за замены акцептора на донора. В исследуемых моделях  $N\epsilon(\text{Arg109})$  and  $N\eta 2(\text{Arg109})$  формировали стабильные связи ( $\sim 3$  Å) с вышеупомянутыми партнёрами вне зависимости от зарядовых состояний Pro1 и Glu2. Однако при замене Cyt на Ade паттерн связей был иным. В моделях *анти*-Ade ε-аминогруппа Arg109 изменяла ориентацию через  $\sim 3$  нс динамики и формировала стабильную водородную связь с  $N7(\text{Ade}^{(+1)})$ , находящемся с 3'-конца от основания напротив повреждённого. Это происходило во всех зарядовых состояниях, кроме PRN-GLU. При этом остаток  $\text{Ade}^{(+1)}$  принимал *син*-конформацию.  $\text{Ade}^{(0)}$  при этом наблюдался в двух ориентациях: в моделях PRN-GLU-Aa и PRO-GLU-Aa оставался в *анти*-конформации внутри цепи, стабилизированной водородной связью  $N^6(\text{Ade}^{(0)})\text{--}O2P(\text{Ade}^{(+1)})$ , а в моделях PRN-GLH-Aa и PRO-GLH-Aa выталкивался в большую бороздку и поворачивался, приближаясь к *син*-конформации, при этом вступая в стекинг-взаимодействие с гуанидиновой группой Arg109. В моделях *син*-Ade основание демонстрировало большую стабильность. Во всех случаях формировалась водородная связь  $N\eta 2(\text{Arg109})\text{--}N7(\text{Ade}^{(0)})$ , хотя она частично нарушалась в модели PRN-GLU-As, когда в качестве акцептора иногда выступал атом N3. В целом же стабильность обеспечивалась сетью дополнительных водородных связей, вызванных изгибом ДНК, например,  $N^6(\text{Ade}^{(0)})\text{--}O^4(\text{Thy}^{+1})$  и  $N^6(\text{Ade}^{(0)})\text{--}O2P(\text{Ade}^{(+2)})$ . Учитывая склонность *анти*- $\text{Ade}^{(0)}$  выдвигаться из цепи, вполне вероятно, что он принимает *син*-конформацию в комплексе фермента с неканонической парой 8-oxodGuo-dAdo, в то время как в свободной ДНК он находится в *анти*-конформации [180, 181].

Во всех известных структурах фермента Fpg, связанного с ДНК, определённый консервативный остаток Phe (в случае *Lla*-Fpg это Phe111) стабильно интеркалирует через малую бороздку между основанием напротив повреждённого и основанием с 3'-стороны от него (считая вдоль повреждённой цепи). Однако в исследуемых моделях Phe111 оказался довольно мобилен. В 8 из 12 моделей этот аминокислотный остаток возвращался назад в малую бороздку, что сопровождалось значительным поворотом  $\text{Ade}^{(+1)}$ , поддерживающим стекинг. В результате пара  $T^{-1}:A^{(+1)}$  сильно искажалась. Только в двух моделях (PRN-GLH-C и PRN-GLU-Aa) интеркаляция Phe111 была стабильной. Вероятно, наблюдаемый выход Phe111 из стекинга связан с ранними событиями скольжения Fpg по ДНК, когда фермент сканирует основания неповреждённой цепи, пытаясь внедрить остаток Phe111.

Для поиска внутри- и межмолекулярных взаимодействий, важных в узнавании пары 8-охоGua:Cyt, составили список водородных связей, присутствующих в более чем 1% траекторий со средней энергией более 1,2 ккал/моль. Таких связей было около 600 в каждой модели (минимум 599 в модели PRN-GLH-Aa; максимум 634 в модели PRO-GLH-Aa). При этом во всех моделях менее 50% таких связей присутствовали в 90 и более процентах траектории, и менее 25% присутствовали менее чем в 25 и менее процентах траектории. Первую выборку можно считать репрезентативной и функционально важной, в то время как вторая — вероятно, продукт случайных конформационных колебаний. Попарное сравнение обеих выборок по коэффициенту среднеквадратичной сопряжённости не выявило существенного вклада в общее распределение ни разных зарядовых состояний Pro1 и Glu2, ни основания напротив повреждённого. Затем проводился поиск связей, которые различались между моделями C, Aa и As на разницу более  $3\sigma$  по встречаемости (рисунок 22). Только несколько водородных связей демонстрировали разную стабильность при сравнении C и A (вне зависимости от конформации последнего). Неудивительно, что некоторые из них формировались Arg109 с Cyt, присутствовавшим только в моделях C.

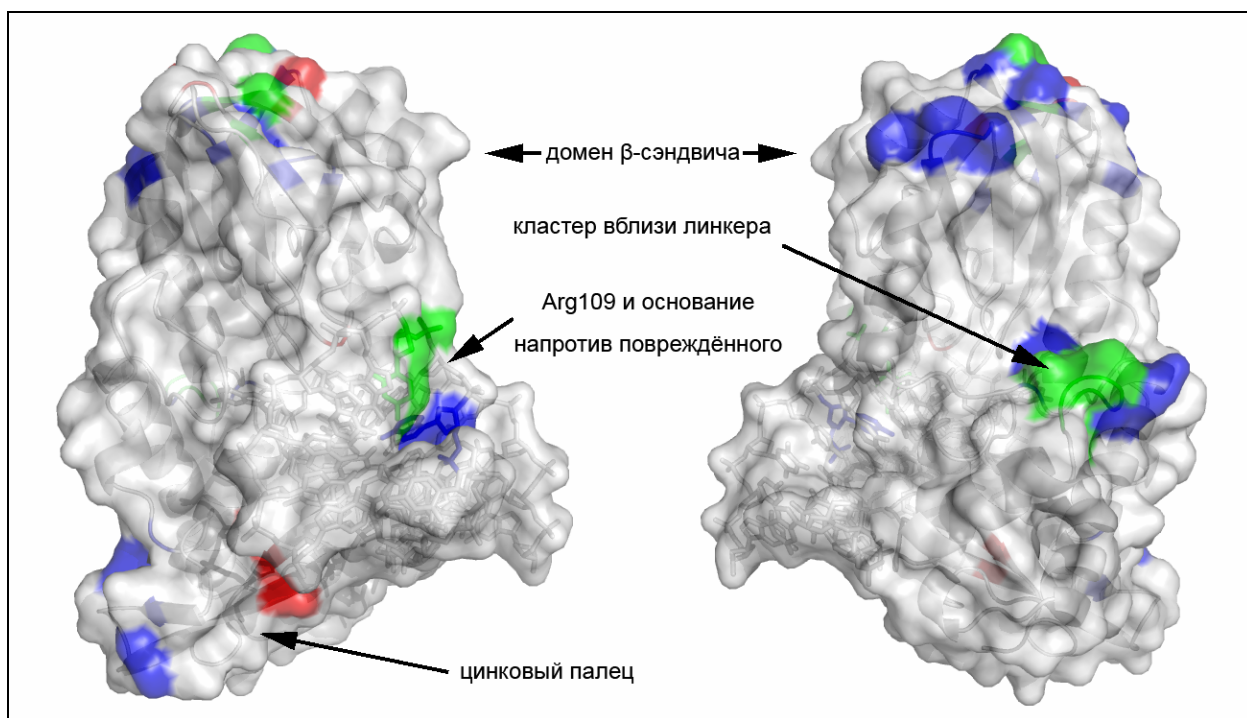


**Рисунок 22.** Карта сравнения специфических водородных связей, образуемых аминокислотными остатками (с разницей по встречаемости больше либо равной  $3\sigma$ ) в моделях, содержащих Cyt (показаны красным цветом) и Ade (показаны зелёным цветом) напротив 8-охоGua. Остатки 1–271 принадлежат ферменту, 272–285 повреждённой цепи ДНК, 286–299 комплементарной цепи ДНК, 300 — атом цинка, 301–307 — молекулы воды, содержащиеся в кристаллической структуре. Жёлтой линией отмечено положение повреждённого основания (8-охоGua), пурпурной — положение основания напротив повреждённого (Cyt/Ade).

С другой стороны, связи между  $dAdo^{(0)}$  и фосфатом  $dAdo^{(+2)}$ , упоминавшиеся ранее, присутствовали только в моделях A. При этом ни 8-оходGuo, ни  $O^6$ -петля, ни каталитическая

пара Pro1–Glu2 не образовывали специфических связей, а сравнение связей в зависимости от зарядовых состояний выявило лишь отдельные изолированные мостики вдали от активного центра.

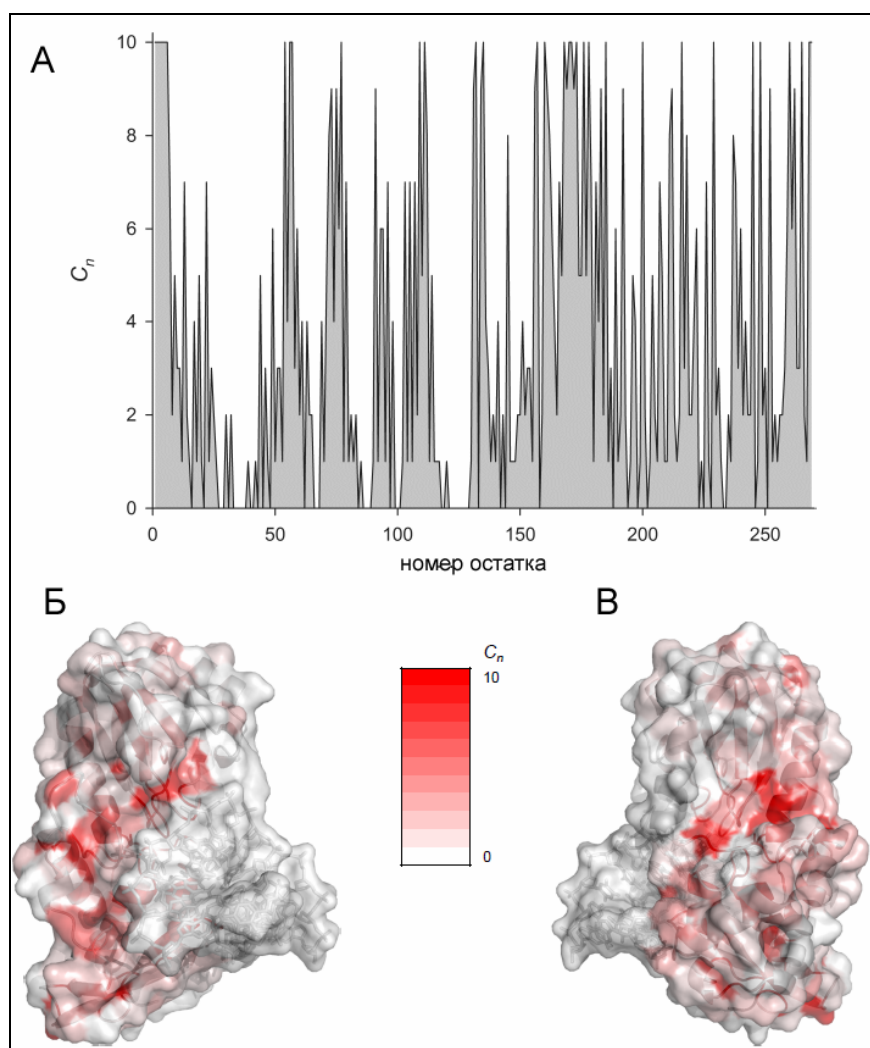
Наиболее заметным фактором специфичности Frg к основанию напротив повреждённого в исследуемых моделях оказался кластер в начале С-концевого домена, находящийся сразу за междоменным линкером: а. о. Glu134–Phe140 (рисунок 23). Многие из этих остатков вовлечены во множественные водородные связи, формирующие специфическую сеть. Например, Thr136 образует два мостика с Glu134, один с Asp139 и один с Phe140 — N(Thr136)–O(Glu134), O $\gamma$ (Thr136)–O(Glu134), N(Asp139)–O $\gamma$ (Thr136), N(Phe140)–O $\gamma$ (Thr136) — лишь в моделях А. В моделях С набор связей иной: O $\gamma$ (Thr136)–O $\epsilon$ 1/O $\epsilon$ 2(Glu138), N(Asp139)–O $\epsilon$ 1/O $\epsilon$ 2(Glu138), N(Phe140)–O(Thr136).



**Рисунок 23.** Поверхность модели PRO-GLH-C (снимок после 8 нс моделирования) с обозначением участков, где существуют С/А-специфичные водородные связи (две проекции, различающиеся поворотом на 180° вокруг вертикальной оси). Остатки, формирующие только С-специфичные связи, отмечены красным; остатки, формирующие только А-специфичные связи, отмечены синим; остатки, формирующие разные связи в С и А моделях, отмечены зелёным. См. также пояснения в тексте.

Карбоксильная группа Glu138 формировала другую С-специфичную водородную связь с Asp139: N(Asp139)–O $\epsilon$ 1/O $\epsilon$ 2(Glu138). В А-моделях при этом наблюдался мостик N(Tyr137)–O $\epsilon$ 1/O $\epsilon$ 2(Glu138). В итоге петля Glu134–Phe140 принимала разные конформации в моделях с Cyt и Ade. Важно тут то, что этот участок Frg высококонсервативен (рисунок 24), что подчёркивает функциональную важность региона, несмотря на его удалённость от активного центра. Степень консервативности  $C_n$  вычисляли на основе таксономически сбалансированной

выборки из 124 последовательностей бактериальных Frg (по две последовательности на таксономическое семейство) с помощью алгоритма AMAS [182].



**Рисунок 24.** Консервативность фермента Frg. (А) График зависимости степени консервативности ( $C_n$ ) от номера остатка. (Б) Вид модели PRO-GLH-C (8 нс), окрашенный в зависимости от степени консервативности. (В) Та же модель, развёрнутая на 180 градусов вокруг вертикальной оси.

Ещё один функционально значимый регион, где формировались различные специфические водородные связи — это  $\beta$ -шпилька цинкового пальца (рисунок 23). Этот структурный мотив Frg вовлечён в сканирование большой бороздки и узнавание повреждённого основания [36]. В основании цинкового пальца цепи соединялись водородной связью  $N\epsilon_2(\text{Gln}236)\text{--}O(\text{Lys}244)$  во всех А-моделях. При этом в С-моделях был более стабилен мостик  $N\zeta(\text{Lys}254)\text{--}N\delta_1(\text{His}263)$  между  $\beta$ -цепями в центре мотива. Несколько С/А-специфичных водородных связей были рассеяны в домене  $\beta$ -сэндвича вокруг С-конца длинной  $\alpha$ -цепи  $\alpha A$ , которая несёт каталитические остатки Pro1 и Glu2 на N-конце (рисунок 23). Функциональная значимость этого региона неясна, т.к. многие аминокислотные остатки,

участвующие здесь в формировании специфической сети водородных связей, находятся в петлях и не консервативны.

Интерес представляли также различия в сети водородных связей между различными конформациями Ade. Большинство различий обнаруживали между белком и ДНК, а также внутри ДНК, при этом пересечение с C/A-специфическими взаимодействиями практически отсутствовало. Наиболее заметные контакты — Tyr29/Arg31 и His91/Lys110 (координируют фосфаты, фланкирующие Ade напротив повреждённого нуклеотида), а также Lys155. Последний остаток важен для активности Fpg и контактирует с ДНК в нескольких нуклеотидах от мисмэтча, вероятно, помогая позиционировать цинковый палец. C/A-специфичный регион Glu134–Phe140, прилеющий к линкеру, не показал значительных различий между моделями Aa и As; однако сам линкер (остаток Lys129) вовлечён в специфические взаимодействия (с Glu115 в моделях Aa и с Tyr125 в моделях As).

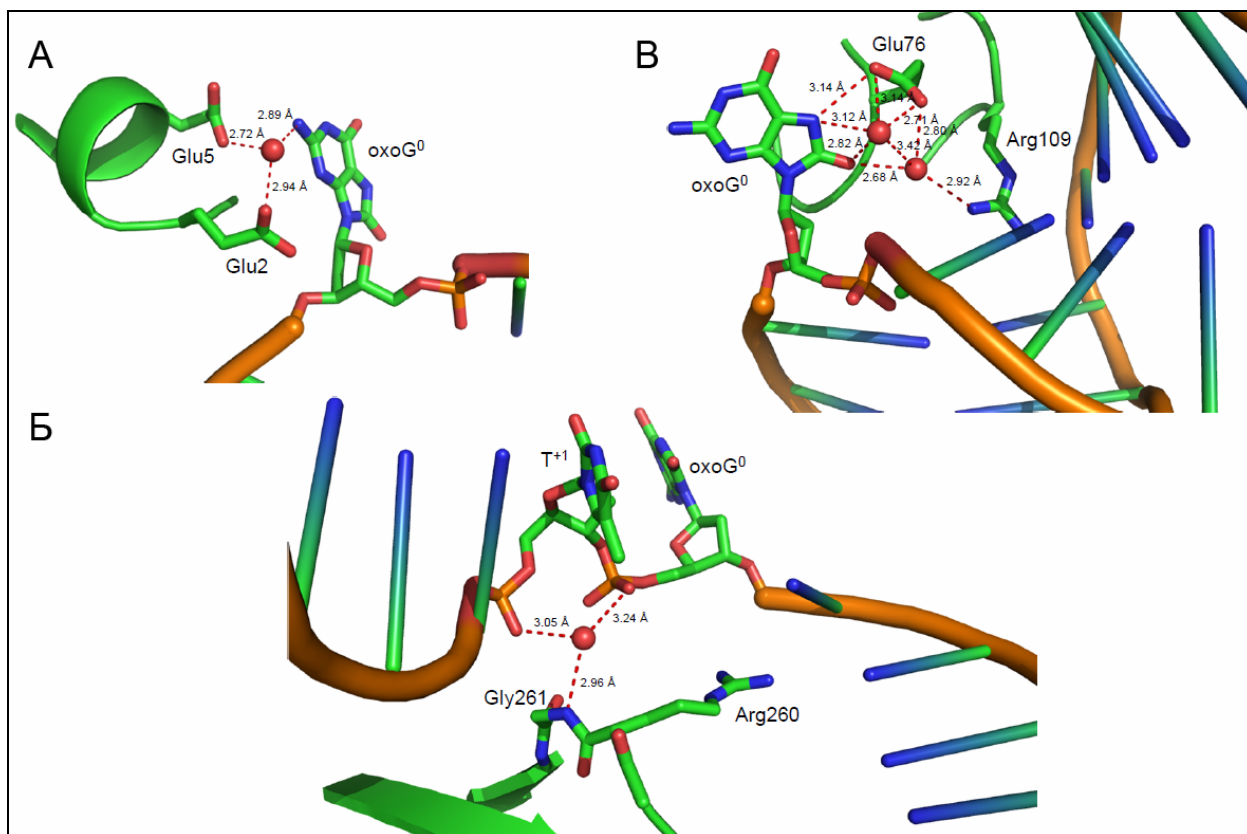
#### 4.2.4. Роль молекул воды в механизме действия Fpg

Оригинальная структура 1XC8 содержит 397 молекул воды, но лишь 22 находятся на ДНК-белковом интерфейсе и только 7 оказываются внутри (т.е. менее чем на 10% экспонированы наружу). Считается, что эти несколько молекул воды формируют сеть водородных связей в активном центре фермента и могут служить для переноса протона во время согласованного расщепления трёх связей, катализируемого Fpg [45]. Поэтому эти 7 молекул воды моделировали явно, а образуемые ими водородные связи тщательно анализировали. Интерес представляли «водные мостики», т.е. ситуации, где молекула воды формировала две или больше водородных связей в одном снимоте траектории. Если такие мостики оказывались стабильными, то могли быть вовлечены в поддержание стабильности всей структуры или обеспечение катализа. Значительной разницы между моделями в количестве образуемых водных мостиков не обнаружили: от 73 до 100 пар существовало по меньшей мере в одном снимоте, а также от 43 до 87 триплетов. Многие мостики строились не через одну и ту же молекулу воды, а через разные — две или более — занимающие одну и ту же позицию по очереди. Этот динамический обмен указывает на существование предпочтительных центров связывания растворителя в противовес ситуации, когда молекулы оказываются «зажаты» в структуре из-за высоких энергетических барьеров ухода. В 8 из 12 моделей акцепторы Oε1/Oε2(Glu76) и O<sup>8</sup>(8-oxodGuo<sup>0</sup>) стабильно формировали мостики с двумя молекулами воды. В некоторых моделях существовали опосредованные связи через воду между кислородами фосфатов 8-oxodGMP<sup>0</sup> и dTMP<sup>+1</sup>, а также между O2P(8-oxodGMP<sup>0</sup>) и Nη1(Arg109), хотя их



встречаемость оказалась ниже. Множественных мостиков, соединяющих триплеты, обнаружено не было.

Ранжирование водных мостиков по встречаемости (т.е. числу снапшотов, где пара или триплет образуют хотя бы одну водородную связь через воду) показало наличие нескольких триплетных кластеров, наблюдаемых в большинстве снапшотов (рисунок 25), при этом встречаемость остальных триплетов резко падала, в то время как встречаемость пар падала плавнее.



**Рисунок 25.** Основные кластеры формирования водных мостиков в модели PRO-GLH-C (снапшот после 8 нс моделирования). Цвет атомов соответствует химическому элементу (С – зелёный, N – синий, O – красный, P – оранжевый). Молекулы воды показаны в виде красных сфер. Остальные части комплекса показаны схематично или опущены для удобства восприятия. Пояснения в тексте.

Для определения предпочтительных сайтов связывания выставляли пороговое значение встречаемости 2000 для пар и 1500 для триплетов. Таким образом отбирали мостики, определяющие 75% всей сети. Наиболее часто встречающимся триплетом оказался  $O\epsilon_1/O\epsilon_2(\text{Glu}2)-O\epsilon_1(\text{Glu}5)-N^2(\text{охоGua}^0)$  (рисунок 25, А), присутствовавший в 11 из 12 моделей, а в одной (PRO-GLH-Aa) оказался немногим ниже порогового отсечения (1273) из-за предпочтительного связывания Glu2 и Glu5 через воду с  $N3(\text{охоGua}^0)$ . Ещё один кластер связывания воды обнаружили вблизи  $N\epsilon(\text{Arg}260)$  и  $N(\text{Gly}261)$  белка и фосфатами  $8\text{-охоGMP}^0$  и  $d\text{TMP}^{(+1)}$  в ДНК (рисунок 25, Б). Четыре триплета этого кластера ( $N\epsilon[\text{Arg}260]-N[\text{Gly}261]-$

O2P[dTMP<sup>+1</sup>], Nε[Arg260]–N[Gly261]–O2P[8-oxodGMP<sup>0</sup>], Nε[Arg260]–O2P[dTMP<sup>+1</sup>]–O2P[8-oxodGMP<sup>0</sup>] и N[Gly261]–O2P[dTMP<sup>+1</sup>]–O2P[8-oxodGMP<sup>0</sup>]) встречались в 8 или 9 моделях из 12, а в 1–2 моделях присутствовали альтернативные триплеты, вовлекающие O(Gly261). В каждом снимоте в этом сайте присутствовала обычно одна молекула воды, взаимодействующая с разными донорами и акцепторами описанных триплетов. Кроме того, Oε1/Oε2(Glu76) формировал триплетные мостики с Nη1(Arg109) и O2(dTMP<sup>+1</sup>) либо O<sup>8</sup>(8-oxodGMP<sup>0</sup>) в 9 из 12 моделей с участием двух молекул воды (рисунок 25, В). Остальные триплеты (включая те, чья встречаемость превышала пороговую) присутствовали в 1–3 моделях и, по-видимому, не имеют биологической значимости.

Обнаруженные стабильные триплеты позволяют оценить роль воды в механизме катализа, осуществляемого Fpg. Молекула воды между Oε1/Oε2(Glu2), Oε1(Glu5) и N<sup>2</sup>(8-oxoGua<sup>0</sup>) находится в позиции, близкой к идеальной для протонирования N3(8-oxoGua<sup>0</sup>) либо депротонирования N(Pro1) (в моделях PRO). Известно, что все хорошие субстраты Fpg содержат акцептор водородной связи в позиции, соответствующей N3 пуринов [45]. Хотя рK<sub>a</sub> атома N3 остатка 8-oxoGua очень низкая, это единственный сайт, способный протонироваться в водном растворителе. Согласованное депротонирование Pro1 и протонирование 8-oxoGua (напрямую или опосредованно через молекулу воды) может объяснить одновременную активацию нуклеофила и уходящей группы. Кроме того, эта же молекула воды может участвовать в процессе гидролиза основания Шиффа между N(Pro1) и C1'[8-oxoGua].

Молекула воды, нековалентно соединяющая фосфаты dTMP<sup>+1</sup> и 8-oxodGMP<sup>0</sup> с аминокислотными остатками белка, может играть роль в изломе двуцепочечной ДНК. Расстояние между атомами фосфора этих нуклеотидов во всех моделях заметно меньше, чем в канонической В-форме ДНК. Такое сжатие вблизи dTMP<sup>+1</sup>, наряду с вклиниванием Phe111 и интеркаляцией Met75 и Arg109, облегчает процесс изгиба ДНК примерно на 60° и способствует выворачиванию повреждённого основания.

Что касается стабильных водных мостиков, связывающих O<sup>8</sup> остатка 8-oxoGua — здесь трудно говорить о важном биохимическом значении, т.к. остаток Glu76, участвующего в их формировании, не является полностью консервативным. Многие формы Fpg содержат на его месте серин или треонин. Впрочем, водородные связи, опосредованные растворителем, ранее не исследовались подробно, поэтому такой мостик вполне может быть одним из многих способов узнавания ферментом экзоциклического кислорода повреждённого основания. Пластичность активного центра и в данном случае может быть решающим фактором, важным для широкой субстратной специфичности Fpg к окислительным повреждениям и, таким образом, предотвращения геномной нестабильности, вызванной окислительным стрессом.

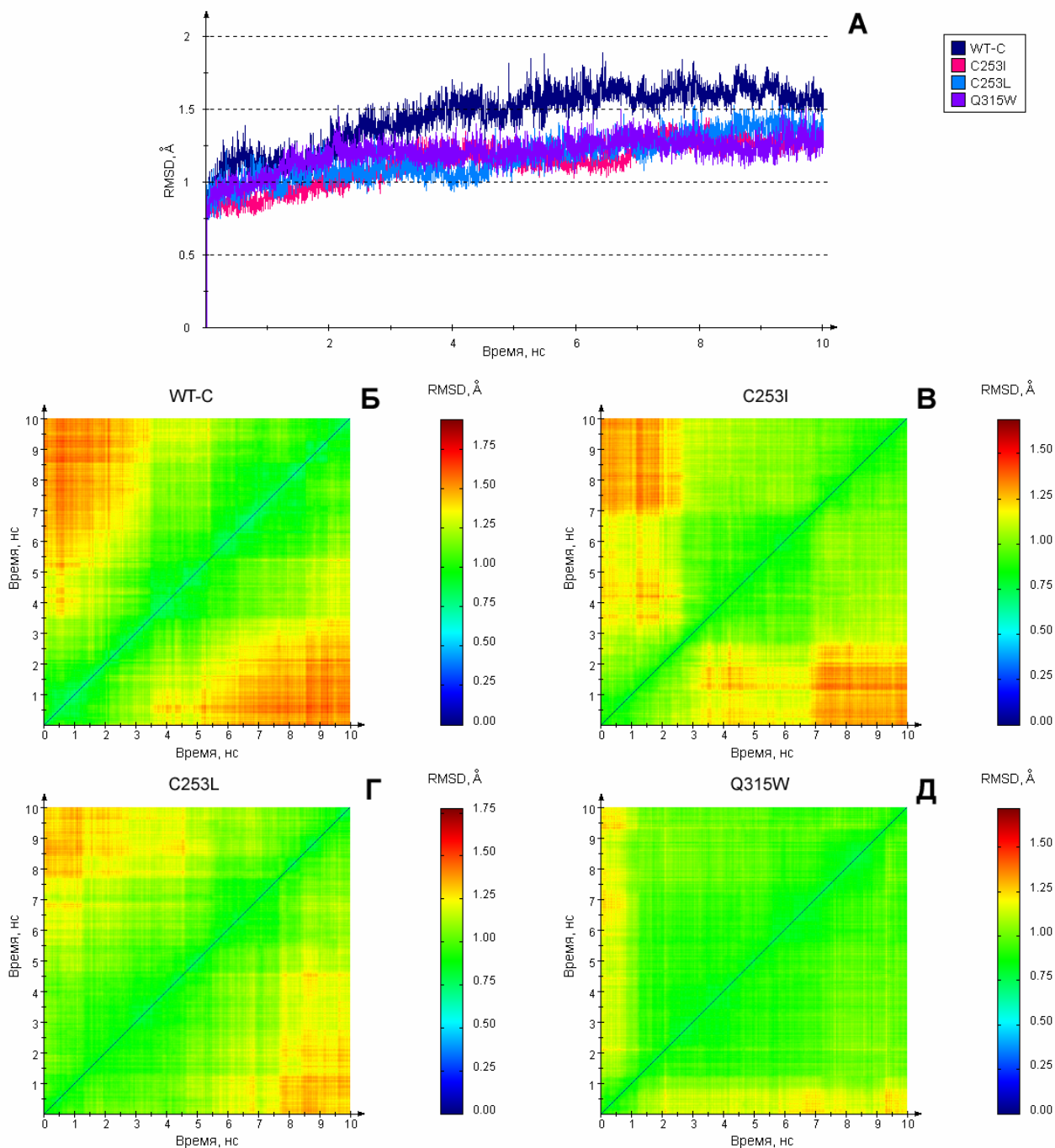
### 4.3. Исследование структуры и активности OGG1

В работе также проведено исследование структуры и динамики функционального аналога прокариотического фермента Fpg — 8-оксогуанин-ДНК-*N*-гликозилазы человека. Исследованы структурные причины значительного снижения активности мутантных форм белка, содержащих объёмные аминокислотные заместители в активном центре, в отношении субстратов, содержащих 8-оксогуанин и АП-сайт, в рамках концепции пластичности, ранее показанной для Fpg. Также, как и в случае Fpg, выявлены детерминанты специфичности к основанию напротив повреждённого. Полученные структурные данные и предсказания нашли экспериментальное подтверждение в кинетических экспериментах.

#### 4.3.1. Структурные основы снижения активности мутантных форм OGG1 на субстратах, содержащих 8-оксогуанин

Существующие структурные данные дают противоречивые ответы на вопрос о влиянии объёмных аминокислотных заместителей в активном центре фермента [183, 184]. Поэтому были сконструированы мутантные формы с заменами Cys253 на лейцин либо изолейцин, а также Gln315 на триптофан. Эти аминокислотные остатки занимают дополнительный объём в активном центре, поэтому представлялось интересным проанализировать конформационное пространство, доступное ферменту с вывернутым повреждённым основанием, готовым к выщеплению.

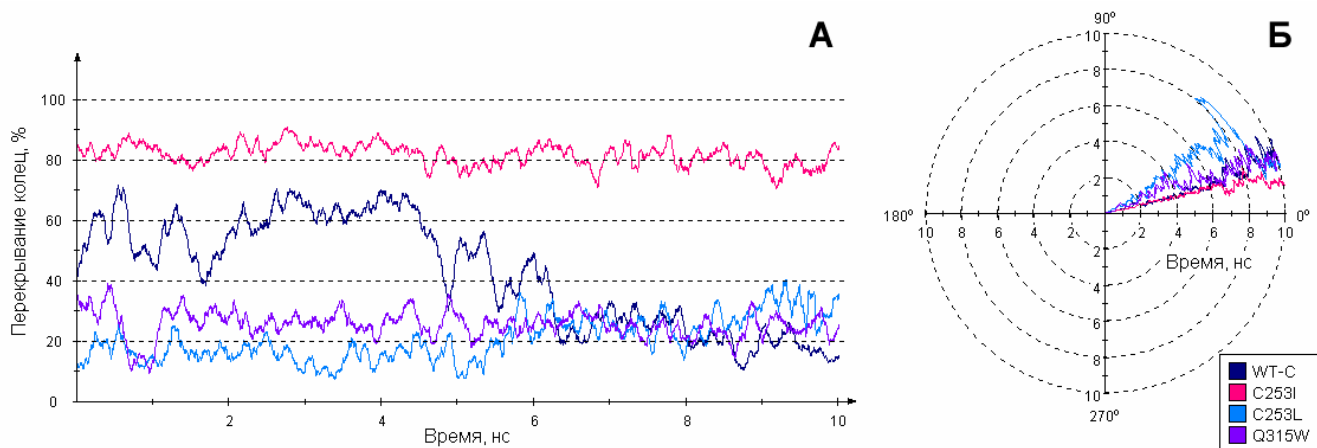
Известно, что на динамику активного центра влияют заряды ионизируемых групп внутри него; поэтому несколько ключевых аминокислотных остатков OGG1 моделировали в особых зарядовых состояниях. Так, аминокислотная группа Lys249 должна быть депротонированной, т.к. для атаки важна неподелённая электронная пара атома азота. Остаток Asp268 соседствует с атомом O4', несущим частичный отрицательный заряд и получающим протон в процессе раскрытия рибозного цикла [139], поэтому скорее всего находится в протонированной форме. Таким образом, Lys249 и Asp268 моделировали в нейтральной форме во всех структурах. Остаток Cys253 в моделях дикого типа (WT-C, таблица 3) и Q315W, согласно квантовомеханическим данным [141], моделировали в виде тиолат-аниона. Замена заряженного Lys249 на нейтральный остаток нарушала формирование диполя Lys249–Cys253, который, как предполагается, присутствует в активном центре белка дикого типа [141], но эти остатки по-прежнему удерживались рядом благодаря водородной связи (2,35 ккал/моль), присутствовавшей в 99% траектории.



**Рисунок 26.** Среднеквадратичное отклонение (RMSD) моделей OGG1 (субстрат 8-охоGua). (A) RMSD тяжёлых атомов, измеренное вдоль траектории. График, соответствующий модели WT-C, показан синим цветом; C253I — пурпурным; C253L — голубым; Q315W — сиреневым. (Б)–(Д) Двумерные графики RMSD тяжёлых атомов, построенные методом попарного сопоставления снапшотов траектории, для моделей WT-C, C253I, C253L и Q315W соответственно.

Одно- и двумерные графики RMSD показали, что все модели уравнивались в течение 5–7 нс и были достаточно стабильны после 7,5 нс (рисунок 26). Среднеквадратичное отклонение позиций тяжёлых атомов не превышало 1,9 Å. Введение объёмного заместителя в центр узнавания повреждения, как и ожидалось, смещало остаток 8-охоGua. В модели WT-C Phe319 формировал одну из стенок центра, вступая в стекинг-взаимодействие с вывернутым из

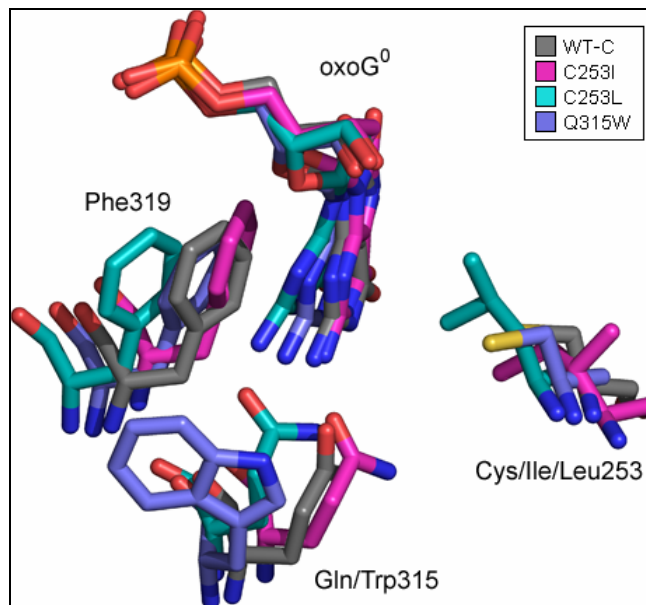
цепи остатком 8-охоGua. В модели C253I 80% площади ароматического кольца Phe319 перекрывалось с конденсированной системой колец 8-охоGua на протяжении всей траектории, в то время как в модели C253L перекрывание не превышало 20%, а в модели Q315W — 25% (рисунок 27, А). В модели дикого типа наблюдалось среднее перекрывание, начинавшееся с 60% перекрывания и постепенно снижавшееся до 20% при 4-6 нс. Угол наклона кольца Phe319 относительно 8-охоGua в модели C253L на  $\sim 15^\circ$  превышал значения в остальных моделях, за исключением угла в модели Q315W, который также демонстрировал существенные отклонения, местами достигающие тех же значений, хотя в целом превышал значение угла в модели дикого типа на  $5^\circ$  (рисунок 27, Б).



**Рисунок 27.** Структурные параметры центра узнавания повреждения моделей OGG1 (субстрат 8-охоGua), измеренные вдоль траекторий. Графики, соответствующие модели WT-C, показаны синим цветом; C253I — пурпурным; C253L — голубым; Q315W — сиреневым. (А) Процент перекрывания ароматических колец Phe319 и 8-охоGua. (Б) Угол между плоскостями ароматических колец Phe319 и 8-охоGua. Значения сглажены методом центрального скользящего среднего.

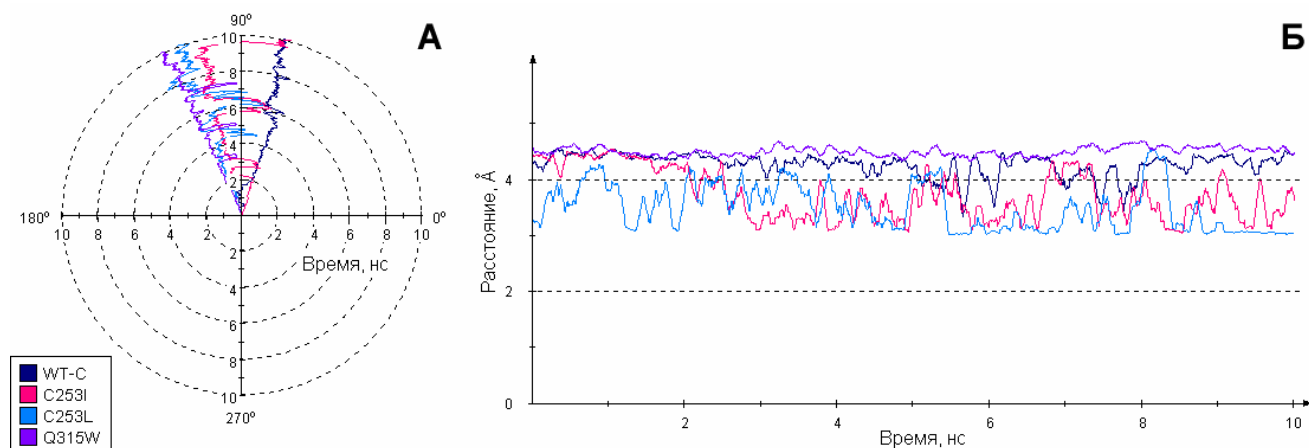
Стерические затруднения, возникшие при введении в активный центр остатков лейцина и изолейцина, разрешались разными путями. В модели C253I боковая цепь изолейцина была свёрнута таким образом, что оказывалась в целом параллельной плоскости 8-охоGua, что могло дополнительно ограничивать подвижность основания за счёт ван-дер-ваальсовых взаимодействий. В модели C253L боковая цепь была вытянута и сдвигала основание глубже в сторону стенки кармана, сформированной Phe319 (рисунок 28). В модели Q315W основание сдвигалось из-за некоторого вытягивания вперёд остатка Cys253 по сравнению с моделью дикого типа. Несмотря на освобождение дополнительного пространства вблизи Phe319, стекинг-взаимодействие повреждённого основания с ним было частично нарушено. В то же время уотсон-криковские контакты с Gln315 оказались стабильными во всех моделях, содержащих этот остаток. Дистанции  $N1(8\text{-охоG})\text{--}O\epsilon 1(\text{Gln315})$  и  $N^2(8\text{-охоG})\text{--}O\epsilon 1(\text{Gln315})$  сохраняли значение  $3,0 \text{ \AA}$  (с некоторыми кратковременными отклонениями в моделях

мутантных форм). То же самое относилось к водородной связи N7(8-охоG)–O(Gly42): 3,0–3,1 Å в течение 99% времени во всех моделях.



**Рисунок 28.** Суперпозиция репрезентативного снапшота траектории (5,1 нс), показывающая положения некоторых аминокислотных остатков, ориентирующих и удерживающих 8-охоGua в активном центре фермента. Остаток Leu253 заметно выдвинут вперёд, что в результате смещает повреждённое основание и остаток Phe319, формирующий заднюю стенку активного центра. Атомы углерода изображены серым (модель WT-C), пурпурным (C253I), бирюзовым (C253L) или сиреневым (Q315W) цветом; атомы кислорода — красным цветом; атомы азота — синим цветом; атомы серы — жёлтым цветом; атомы фосфора — оранжевым цветом; атомы водорода не показаны.

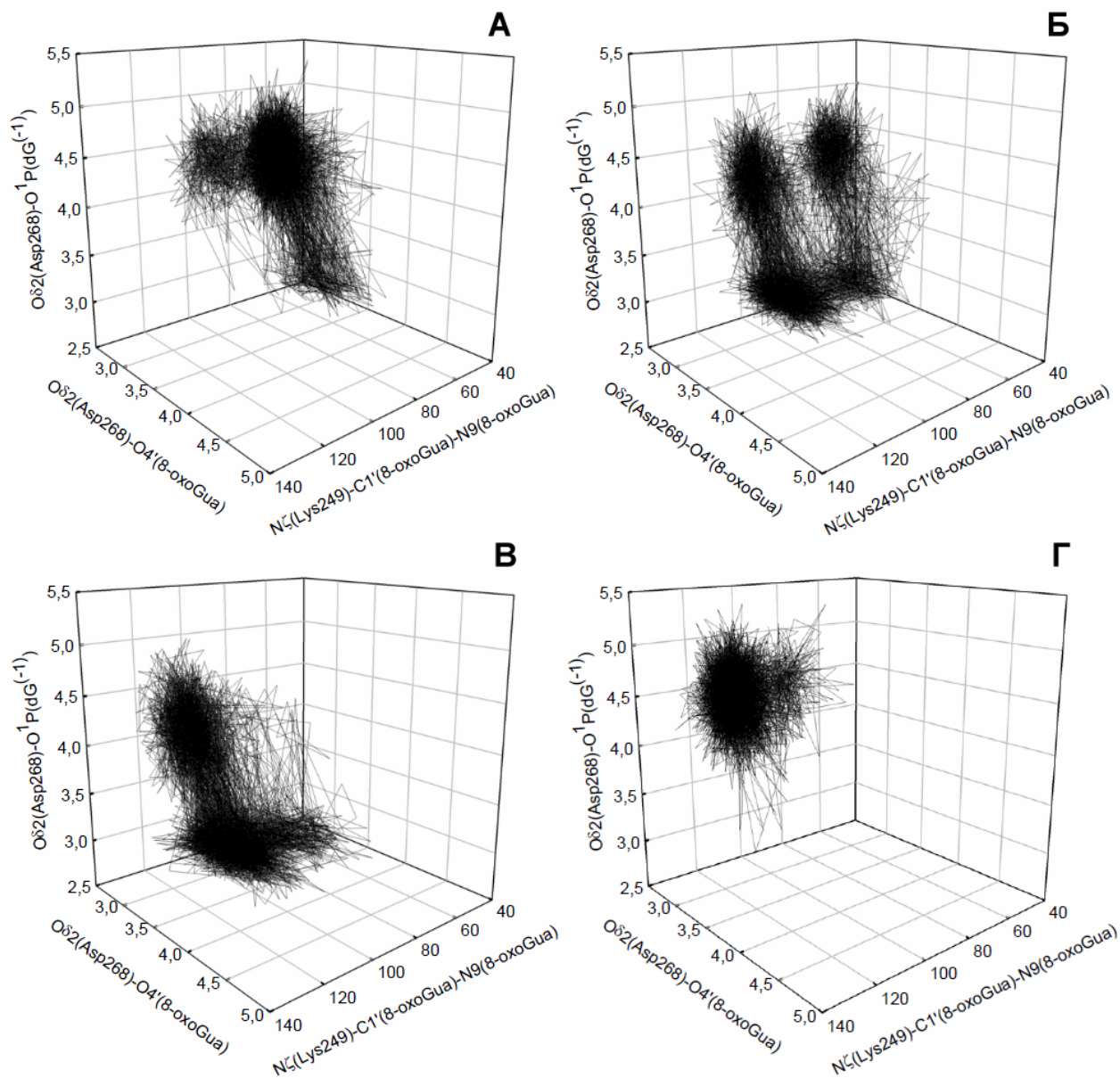
Внимательное исследование структурных параметров ДНК показало, что мутации практически не влияли на конформацию как основания 8-охоGua, так и соседних с ним. Торсионные углы валентных связей и фазовый угол дезоксирибозы в этих остатках практически совпадали во всех моделях, с небольшими отличиями угла  $\beta$  в C253L (поворот примерно на  $20^\circ$  во второй половине траектории) и Q315W (стабильно больше соответствующего угла модели дикого типа на  $10^\circ$ ), а также угла  $\chi$  в C253I (стабильно  $-90^\circ$ , в то время как в модели Q315W стабильно  $-75^\circ$ , а в моделях дикого типа и C253L изменялся до  $-70^\circ$  во второй половине траектории). Таким образом, мутации, которые, как ожидалось, заслоняли центр узнавания и связывания повреждённого основания и препятствовали размещению в нём 8-охоGua, не препятствовали движению нуклеотида, но вместо этого искажали структуру самого 8-охоGua-связывающего центра белка. Расстояние N $\zeta$ (Lys249)–C1'(8-охоGua) держалось на уровне 4 Å во всех моделях, будучи в среднем на 0,2 – 0,4 Å меньше в модели WT-C. Однако угол N $\zeta$ (Lys249)–C1'(8-охоGua)–N9(8-охоGua) стабильно отличался между моделями дикого типа ( $75^\circ$ ) и моделями мутантных форм ( $\sim 110^\circ$ ; рисунок 29, А). Этот угол важен для правильного перекрытия орбиталей во время нуклеофильного замещения 8-охоGua



**Рисунок 29.** Параметры каталитически важных дистанций и углов моделей OGG1, измеренные вдоль траекторий (субстрат 8-охоGua). Графики, соответствующие модели WT-C, показаны синим цветом; C253I — пурпурным; C253L — голубым; Q315W — сиреневым. (А) Угол  $N\zeta(\text{Lys249})\text{--}C1'(\text{8-охоGua})\text{--}N9(\text{8-охоGua})$ . (Б) Альтернативная водородная связь между Asp268 и  $O^1P\ dG^{(-1)}$ . Значения сглажены методом центрального скользящего среднего.

аминогруппой Lys249; такие реакции допускают отклонение не более чем  $10\text{--}15^\circ$  от оптимального положения [185]. Однако ориентация в направлении  $N\zeta(\text{Lys249})\text{--}C1'(\text{8-охоGua})\text{--}O4'(\text{8-охоGua})$  была близка во всех моделях, что согласуется с идеей о том, что именно N-гликозидная связь, а не связь  $C1'\text{--}O4'$ , разрывается первой. Боковая группа второй каталитически важной аминокислоты, Asp268, в модели дикого типа оставалась вблизи  $O4'$  в течение 67% времени и формировала альтернативную водородную связь с  $O^1P$  соседнего остатка dG на протяжении 7% траектории. Однако в моделях мутантных форм (кроме Q315W) Asp268 в основном контактировал с  $O^1P\ dG^{(-1)}$ , а не с  $O4'$  (45% и 39% времени соответственно для C253I, 60% и 39% времени для C253L, (рисунок 29, Б). Любопытно, что модель Q315W демонтирует стабильную водородную связь  $O\delta2(\text{Asp268})\text{--}O4'(\text{8-охоGua})$  на протяжении всей траектории. Трёхмерный график популяции каталитически значимых параметров — угол  $N\zeta(\text{Lys249})\text{--}C1'(\text{8-охоGua})\text{--}N9(\text{8-охоGua})$ , расстояния  $O\delta2(\text{Asp268})\text{--}O4'(\text{8-охоGua})$  и  $O\delta2(\text{Asp268})\text{--}O^1P(dG^{(-1)})$  — демонстрировал группировку возможных конформаций в четыре основных кластера (рисунок 30). Только один из них (верхний правый) состоит из пригодных для катализа конформаций, т.е. содержит благоприятный угол атаки Lys249 и стабилизированный Asp168. Этот кластер был наиболее заселён в модели дикого типа (рисунок 30, А), менее заселён в модели C253I (рисунок 30, Б) и почти пуст в модели C253L (рисунок 30, В). В модели Q315W популяция оказалась довольно компактной, однако была сильно смещена в верхний левый угол, что соответствует повышенным значениям угла  $N\zeta(\text{Lys249})\text{--}C1'(\text{8-охоGua})\text{--}N9(\text{8-охоGua})$  (рисунок 29, А; рисунок 30, Г). Таким образом, сочетание неблагоприятной ориентации Lys249 и сниженного стабилизационного контакта Asp168 с  $O4'$  может отвечать за падение каталитической активности мутантных форм OGG1.





**Рисунок 30.** Трёхмерный график популяции каталитически значимых дистанций и углов моделей OGG1 (субстрат 8-охоGua). (А) Модель дикого типа; (Б) модель C253I; (В) модель C253L; (Г) модель Q315W. Пояснения в тексте.

Контакты с противоположной цепью ДНК, не содержащей повреждённого основания, также очень важны для формирования михаэлисовского комплекса с OGG1, что выражено в специфичности фермента к основанию  $dC^{(0)}$  напротив 8-охоGua [62, 186]. Для фиксации правильной конформации противоположной цепи наиболее важны два структурных фактора. Во-первых, это внедрение ароматического кольца боковой цепи Tyr203, находящегося между  $dC^{(0)}$  и основанием с 5'-конца ( $dC^{(+1)}$ ), способствующие изгибу ДНК. Во-вторых, это остатки Asn149, Arg154 и Arg204, формирующие контакты с  $dC^{(0)}$ . Анализ динамики наклона и перекрытия кольца Tyr203 с соседними цитозинами показал, что во всех четырёх структурах не происходит значительных перестроек в этом регионе. Наибольшее искажение ДНК среди



моделей мутантных форм наблюдали в районе нуклеозидов  $dC^{(0)}$  и особенно  $dT^{(-1)}$ . Однако оно оказывало влияние главным образом на остов, т.к. все шесть структурных параметров уотсон-криковской пары («shear», «stretch», «stagger», «buckle», «propeller» и «opening») в целом не отличались между моделями дикого типа и мутантных форм.

Гуанидиновая группа Arg154 в модели C253I отклонялась от основания  $dC^{(0)}$  практически в начале моделирования, а в модели Q315W вообще не образовывала водородных связей с  $dC^{(0)}$ , формируя лишь неспецифическую связь с фосфатом  $dG^{(-2)}$ . Любопытно, что в модели C253L остаток Asn149 находился далеко от цитозина в первые 80% времени, но затем принимал сходную с моделью WT-C конформацию.

В целом и экспериментальные данные, и результаты компьютерного моделирования хорошо согласовывались с концепцией пластичности активного центра ферментов и многомерным профилем свободной энергии ферментативного катализа. Пластичность — характерная черта множества ферментов с более или менее широким спектром субстратов, либо катализирующих разные реакции, эволюционировавших таким образом, чтобы адаптироваться к разнообразным конформациям и стабилизировать множественные (хотя и похожие) переходные состояния. В этом случае замена критически важного аминокислотного остатка часто не оказывается пагубной для, по крайней мере, некоторых функций фермента, как можно было бы ожидать из анализа энергетического профиля отсутствующих взаимодействий; активный центр оказывается достаточно гибким, чтобы частично компенсировать несоответствие. Например, показано, что гликозилаза MutY *E. coli* компенсирует отсутствие Lys20, ответственного за слабую  $\beta$ -элиминирующую активность, использованием другой аминокислоты активного центра, Lys142 [72]. В свою очередь аминокислотные замены в эндонуклеазе VIII *E. coli*, приводящие к нарушению ДНК-белковых взаимодействий и значительно снижающие гликозилазную активность, не оказывают влияния на АП-лиазную активность из-за реконструирования активного центра, приводящего к закреплению субстрата в каталитически выгодной конформации (Жарков Д.О., личное сообщение). Гликозилаза OGG1 активна на довольно широком спектре модифицированных оснований, включая 8-охоGua, 8-охоAde, Fapy-Gua и некоторые другие замещённые пурины, и катализирует как удаление оснований, так и  $\beta$ -элиминирование, причём последнее может происходить как в присутствии, так и в отсутствие выщепляемого основания в активном центре [62]. Активный центр достаточно пластичен для сдвигов Lys249 и Cys253, что выражается в сохранении значительной остаточной активности. Т.е. несмотря на заслонение центра связывания повреждённого основания и снижение активности фермента, последний не теряет способности к его поиску и

узнаванию. Это, вероятно, объясняется внутренней гибкостью белка и высоким энергетическим сродством к вывернутому остатку 8-охоGua после связывания фермента с субстратной ДНК.

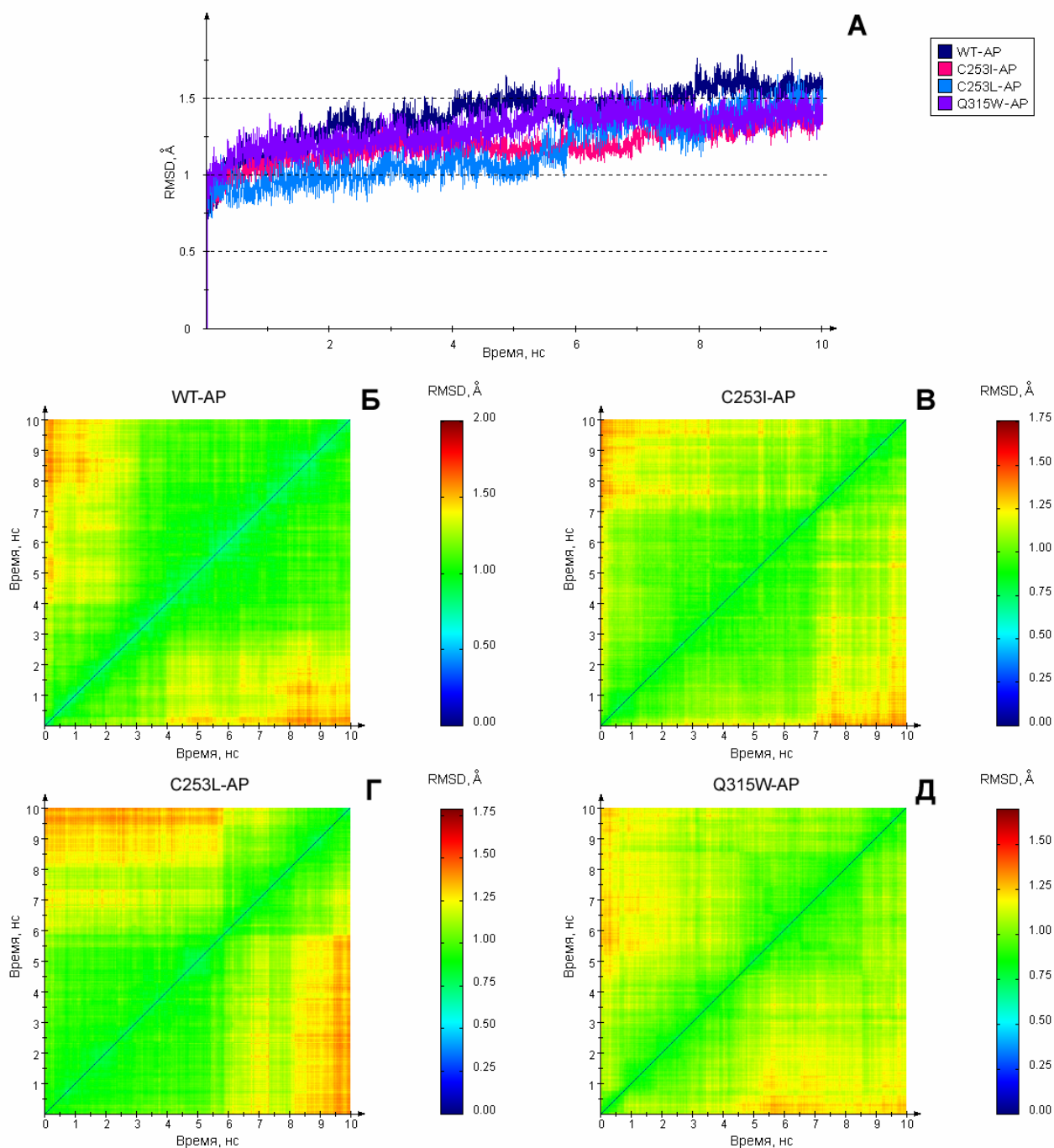
#### 4.3.2. Структурные основы снижения активности мутантных форм OGG1 на субстратах, содержащих АП-сайт

Замещение 8-оксогуанина на АП-сайт, т.е. удаление азотистого основания из активного центра фермента, должно приводить к появлению дополнительного свободного пространства для объемных заместителей, формирующих стенки активного центра, и его общей релаксации, из-за чего активность мутантных форм на таком субстрате будет снижаться в меньшей степени. Было проведено моделирование методом МД для такого субстрата для фермента дикого типа (модель WT-AP) и тех же самых мутантных форм (модели C253I-AP, C253L-AP и Q315W-AP). Приведённые в разделе 4.3.1 соображения легли в основу выбора зарядовых состояний и здесь: Lys249 и Asp268 моделировались в нейтральной форме (первый остаток в депротонированной, а второй в протонированной), а остаток Cys253 в моделях WT-AP и Q315W-AP моделировался в форме тиолат-аниона.

Одно- и двумерные графики RMSD показали, как и в случае моделей, содержащих 8-охоGua, что все модели уравнивались в течение 5–7 нс и были довольно стабильны после 7,5 нс (рисунок 31). Среднеквадратичное отклонение позиций тяжёлых атомов не превышало 1,8 Å.

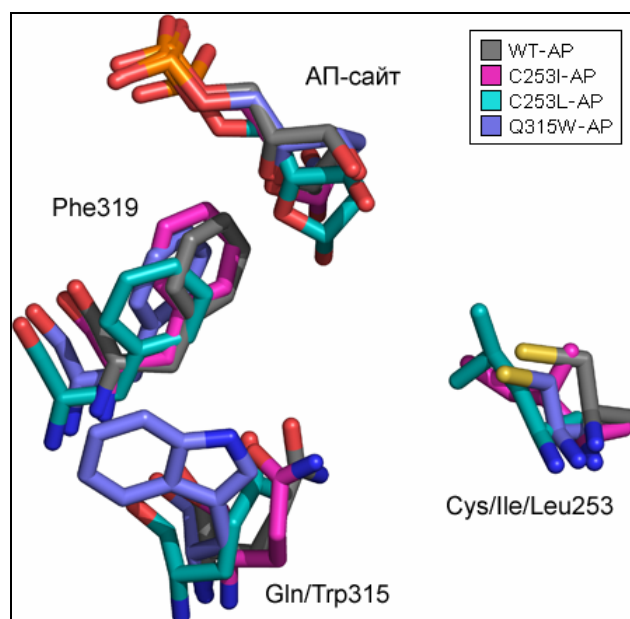
Как показано на примере повреждённого основания 8-охоGua, мутации в активном центре фермента не препятствуют движению нуклеотида, но искажают сам центр из-за возникающих стерических затруднений. Однако в случае АП-сайта большая их часть исчезает, т.к. азотистое основание отсутствует. Анализ таких моделей показал, что специфические водородные связи у Gly42 и Gln315 не формируются, равно как отсутствует какое-либо стекинг-взаимодействие у остатка Phe319, поэтому эти аминокислотные остатки проявляют значительно большую подвижность. Остаток Leu253 в модели C253L-AP точно также выдвигается в активный центр, однако, не находя преграды в виде 8-оксогуанина, не приводит к деформации задней стенки (рисунок 32).

Однако описанная избыточная пластичность активного центра приводит к серьёзным последствиям: субстрат, в данном случае АП-сайт, не может в нём правильно расположиться, а каталитически важные аминокислотные остатки не фиксируются вблизи атома C1'. Это было в большей степени характерно для мутантных форм фермента: в моделях C253I-AP и C253L-AP остаток Asp268 часто взаимодействовал с O<sup>1</sup>P dG<sup>(-1)</sup>, теряя контакт с O4' АП-сайта в последние



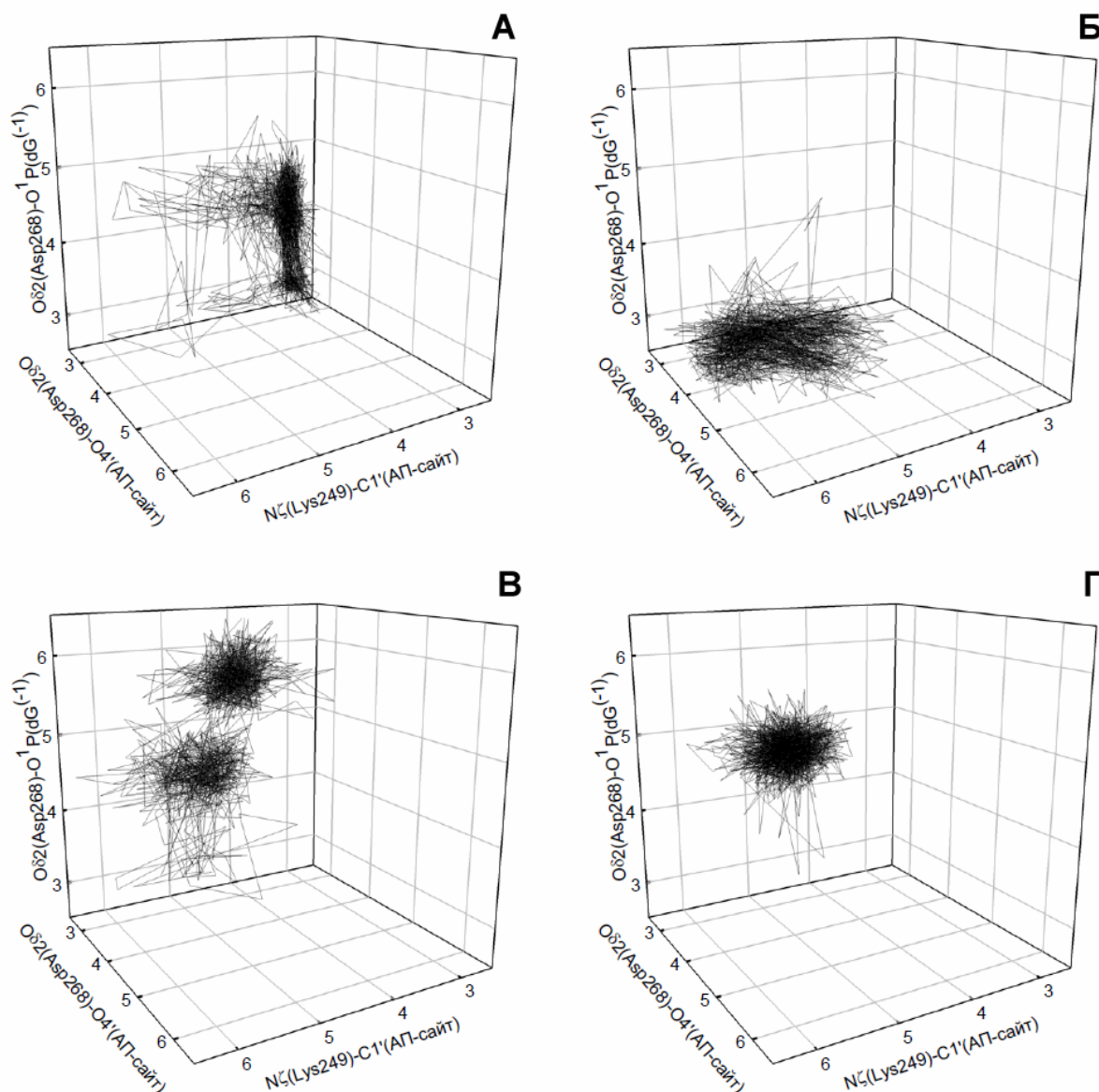
**Рисунок 31.** Среднеквадратичное отклонение (RMSD) моделей OGG1 (субстрат АП-сайт). (А) RMSD тяжёлых атомов, измеренное вдоль траектории. График, соответствующий модели WT-AP, показан синим цветом; C253I-AP — пурпурным; C253L-AP — голубым; Q315W-AP — сиреневым. (Б)–(Д) Двумерные графики RMSD тяжёлых атомов, построенные методом попарного сопоставления снапшотов траектории, для моделей WT-AP, C253I-AP, C253L-AP и Q315W-AP соответственно.

наносекунды динамики; Q315W-AP, впрочем, демонстрировал стабильную динамику, в целом соответствующую поведению модели дикого типа. Говорить об угле атаки аминокислоты Lys249 в случае АП-сайта не приходится, т.к. вследствие мутаротации полуацетальный гидроксил при



**Рисунок 32.** Суперпозиция репрезентативного снимка траектории (5,1 нс), показывающая положения некоторых аминокислотных остатков, обычно ориентирующих и удерживающих 8-охоGua в активном центре фермента; в данном случае в роли субстрата выступает АП-сайт. Остаток Leu253 также несколько выдвинут вперед, но в отсутствие азотистого основания не влияет на остаток Phe319, формирующий заднюю стенку активного центра. Атомы углерода изображены серым (модель WT-AP), пурпурным (C253I-AP), бирюзовым (C253L-AP) или сиреневым (Q315W-AP) цветом; атомы кислорода — красным цветом; атомы азота — синим цветом; атомы серы — желтым цветом; атомы фосфора — оранжевым цветом; атомы водорода не показаны.

атоме C1' не имеет единой пространственной ориентации. Однако каталитически значимым остаётся расстояние  $N\zeta(\text{Lys249})\text{--}C1'(\text{АП-сайт})$ , которое вследствие повышенной пластичности активного центра сильно менялось в процессе динамики во всех моделях, содержащих АП-сайт, что сильно отличает их от моделей, содержащих 8-охоGua, где это расстояние было стабильным. Таким образом, для анализа выбирали три каталитически значимых расстояния:  $N\zeta(\text{Lys249})\text{--}C1'(\text{АП-сайт})$ ,  $O\delta 2(\text{Asp268})\text{--}O4'(\text{АП-сайт})$  и  $O\delta 2(\text{Asp268})\text{--}O^1P(dG^{(-1)})$ , а для минимизации эффектов избыточной подвижности остатков в отсутствие азотистого основания выбирали наиболее стабильные последние 2,5 нс динамики. Трёхмерный график популяции этих расстояний показал, что благоприятная для катализа конформация складывается главным образом в модели дикого типа (рисунок 33, А), где дистанция между атомом C1' АП-сайта и аминогруппой Lys249 лежит в области 3,5 Å. В модели C253I-AP (рисунок 33, Б) при общем увеличении этой дистанции (~5,5 Å) возникала стабильная водородная связь между Asp268 и фосфатной группой  $dG^{(-1)}$ , поэтому предполагается, что активность этого мутанта в отношении субстратов, содержащих АП-сайт, будет сниженной в наибольшей степени. Популяция конформаций модели C253L-AP распадалась на два выраженных кластера, а в модели Q315W-AP, напротив, была довольно компактной (рисунок 33, В, Г); расстояние  $N\zeta(\text{Lys249})\text{--}C1'(\text{АП-сайт})$  в обоих случаях составляло в среднем 4,5 Å.



**Рисунок 33.** Трёхмерный график популяции каталитически значимых дистанций и углов моделей OGG1 (субстрат АП-сайт). (А) Модель дикого типа; (Б) модель C253I-AP; (В) модель C253L-AP; (Г) модель Q315W-AP. Пояснения в тексте.

Это позволяет предположить, что в отношении субстратов, содержащих АП-сайт, активность этих мутантов будет несколько выше, чем в отношении субстратов с 8-охоGua, причём большая устойчивость активного центра Q315W-AP, вероятно, позволяет говорить и о более высокой активности этой мутантной формы, в противовес модели Q315W, где, несмотря на аналогичную компактность, возникла крайне неблагоприятная для нуклеофильной атаки ориентация аминогруппы Lys249. Все эти предположения впоследствии нашли подтверждение в кинетических экспериментах.

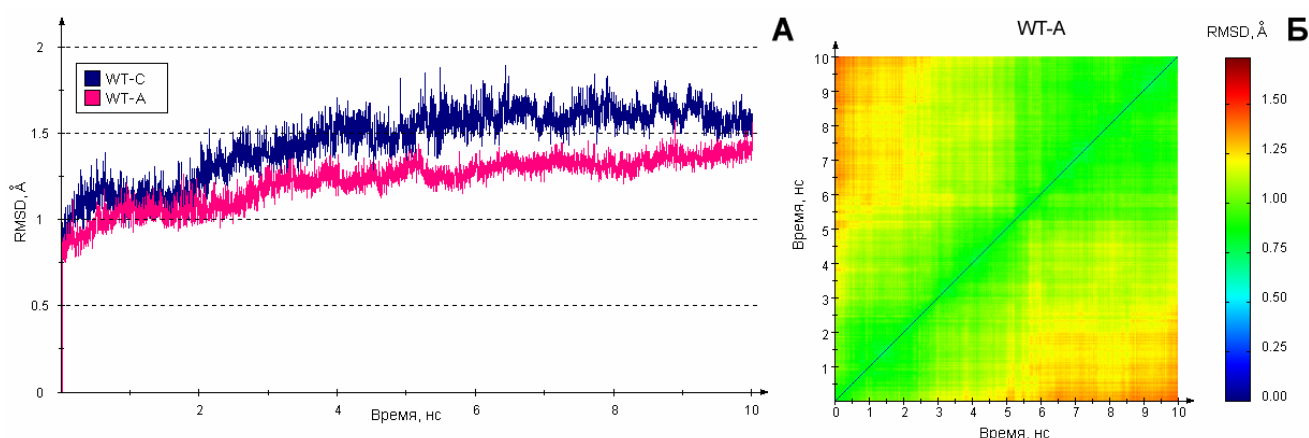
Контакты с противоположной цепью ДНК, не содержащей АП-сайта, в целом соответствовали тем, что наблюдались в моделях с 8-охоGua. Гуанидиновая группа Arg154 в

модели Q315W-AP точно так же не образовывала водородных связей с  $dC^{(0)}$ . Однако в модели C253I-AP, в отличие от C253I, группа находилась вблизи этого нуклеотида значительно больший процент времени (водородная связь  $N\eta1(\text{Arg154})-O^2(dC^{(0)})$  существовала 67% времени против 9%, а водородная связь  $N\eta2(\text{Arg154})-N3(dC^{(0)})$  — 62% времени против 5%). В остальном положения остатков Asn149, Arg154 и Arg204 не осложнялись какими-либо конформационными перестройками.

Таким образом, расположение каталитически важной аминокислотной группы Lys249 в активном центре фермента OGG1 оказывается решающим и в случае субстратов, содержащих АП-сайт.

### 4.3.3. Специфичность OGG1 к основанию напротив повреждённого

Известно, что фермент OGG1 демонстрирует высокую специфичность к основанию напротив повреждённого, предпочитая вырезать 8-охоGua из пары с Cyt, в то время как выщепление повреждённого основания из пары с Ade проходит с низкой эффективностью [62]. С целью выявления структурных основ такой специфичности проводили сравнительное моделирование белка OGG1 в комплексе с ДНК, содержащей как пару 8-охоGua:Cyt (модель WT-C, см. раздел 4.3.1 и таблица 3), так и 8-охоGua:Ade (WT-A). Модели были подготовлены по описанным выше соображениям: остатки Lys249 и Asp268 моделировали нейтральными, а Cys253 — в виде тиолат-аниона. В отличие от моделей Fpg, где остаток аденина моделировался как в *син*-, так и в *анти*-конформации, в данном случае рассматривали только *анти*-Ade вариант. Это связано с наличием большого числа аминокислотных остатков вблизи основания напротив повреждённого, которые охватывают его со всех сторон, и поворот Ade в *син*-конформацию выглядит стерически затруднённым из-за жёсткого столкновения с белком.



**Рисунок 34.** Среднеквадратичное отклонение (RMSD) моделей OGG1 с субстратами, содержащими разные основания напротив 8-охоGua. (А) RMSD тяжёлых атомов, измеренное вдоль траектории. График, соответствующий модели WT-C, показаны синим цветом; WT-A — пурпурным. (Б) Двумерный график RMSD тяжёлых атомов, построенный методом попарного сопоставления снимков траектории, для модели WT-A.

Стабильность уравновешенных моделей также анализировали с помощью одно- и двумерных графиков RMSD (рисунок 26, Б; рисунок 34), причём модель WT-A демонстрировала меньшее среднеквадратичное отклонение позиций тяжёлых атомов, не превышающее 1,6 Å (значения сглажены методом центрального скользящего среднего). Стекинг-взаимодействие Phe319 с остатком 8-охоGua в целом было сходным, при этом в модели WT-A оно было более стабильным на уровне 50% с углом наклона  $\sim 15^\circ$ . Стабильны были также водородные связи 8-охоGua с Gln315 (3,0 Å) и Gly42 (также 3,0 Å).

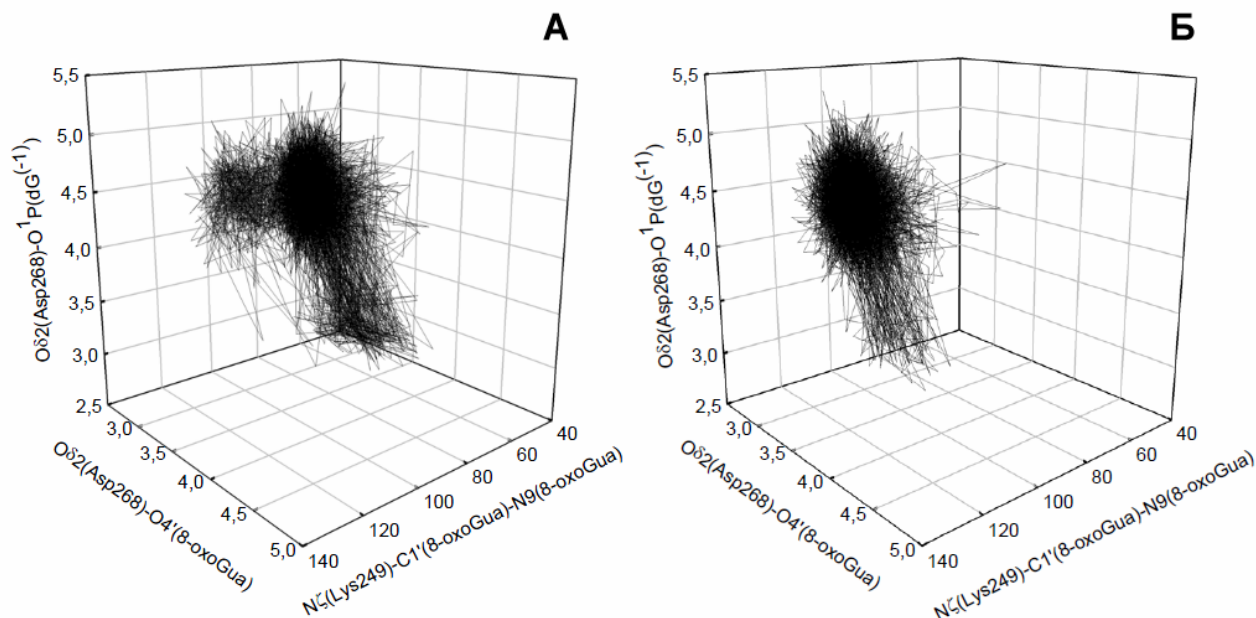
Конформация основания 8-охоGua, так и соседних с ним, также в целом совпадала в обеих моделях. Исключение составлял угол  $\chi$  в соседних нуклеотидах ( $dG^{-1}$  и  $dA^{+1}$ ), где отклонения достигали  $10\text{--}15^\circ$ . Конформация основания напротив повреждённого несколько различалась: угол  $\gamma$   $dA^{(0)}$  на  $9^\circ$  превышал соответствующий угол  $dC^{(0)}$ , угол  $\zeta$  — был меньше на  $11^\circ$  соответственно.

Контакты аминокислотных остатков Asn149, Arg154 и Arg204 с противоположной цепью ДНК осуществлялись в моделях WT-C и WT-A похожим образом; при этом в качестве доноров и акцепторов основания  $dC^{(0)}$  в последней модели выступали аналогично расположенные атомы  $dA^{(0)}$ . Стабильными были водородные связи  $O\delta 1(\text{Asn149})\text{--}N^4(dC^{(0)})/N^6(dA^{(0)})$ ,  $N\eta 1(\text{Arg204})\text{--}O^2(dC^{(0)})/N3(dA^{(0)})$  и  $N\eta 2(\text{Arg204})\text{--}N3(dC^{(0)})/N1(dA^{(0)})$ , они существовали на протяжении 96–99% траектории. Интересно, однако, что остаток Arg154, образующий связи с экзоциклическим кислородом  $dC^{(0)}$ , в модели WT-A отклонялся от остатка аденина и взаимодействовал с  $O^2(dT^{+1})$  и атомом кислорода рибозы  $dG^{+2}$ , лишь в 29% траектории образуя водородную связь с  $N1(dA^{(0)})$ .

Расстояние  $N\zeta(\text{Lys249})\text{--}C1'(8\text{-охоGua})$ , как и в случае мутантных форм (см. раздел 4.3.1), держалось на уровне 4 Å, а угол  $N\zeta(\text{Lys249})\text{--}C1'(8\text{-охоGua})\text{--}N9(8\text{-охоGua})$  также стабильно отличался ( $105^\circ$  в модели WT-A против  $75^\circ$  в модели WT-C). Исключительная важность этого угла для катализа была описана ранее. Точно так же ориентация в направлении  $N\zeta(\text{Lys249})\text{--}C1'(8\text{-охоGua})\text{--}O4'(8\text{-охоGua})$  в целом совпадала ( $137^\circ$ ). Положение боковой группы Asp268 было стабильным в обеих моделях: она находилась вблизи  $O4'$ , причём альтернативная водородная связь с  $O^1P dG^{-1}$  в модели WT-A формировалась лишь в 2% траектории (против 7% в модели WT-C). Аналогично популяциям каталитически значимых параметров мутантных форм, строили трёхмерный график по трём параметрам: угол  $N\zeta(\text{Lys249})\text{--}C1'(8\text{-охоGua})\text{--}N9(8\text{-охоGua})$ , расстояния  $O\delta 2(\text{Asp268})\text{--}O4'(8\text{-охоGua})$  и  $O\delta 2(\text{Asp268})\text{--}O^1P(dG^{(-1)})$ . В сравнении с популяцией модели WT-C (рисунок 35, А), модель WT-A демонстрировала уход в верхний левый угол (рисунок 35, Б), что соответствовало неблагоприятной ориентации Lys249 относительно повреждённого основания.



Это позволяет сделать вывод в том, что и здесь решающим каталитически значимым параметром, влияющим на специфичность OGG1 к основанию напротив повреждённого, является ориентация аминокетильной группы Lys249.

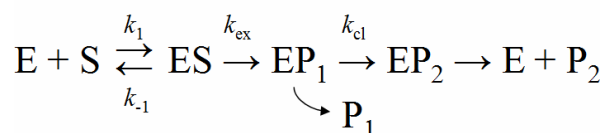


**Рисунок 35.** Трёхмерный график популяции каталитически значимых дистанций и углов моделей комплексов OGG1 дикого типа с субстратами, содержащими разные основания напротив 8-оксоGua. (А) Модель WT-C; (Б) модель WT-A. Пояснения в тексте.

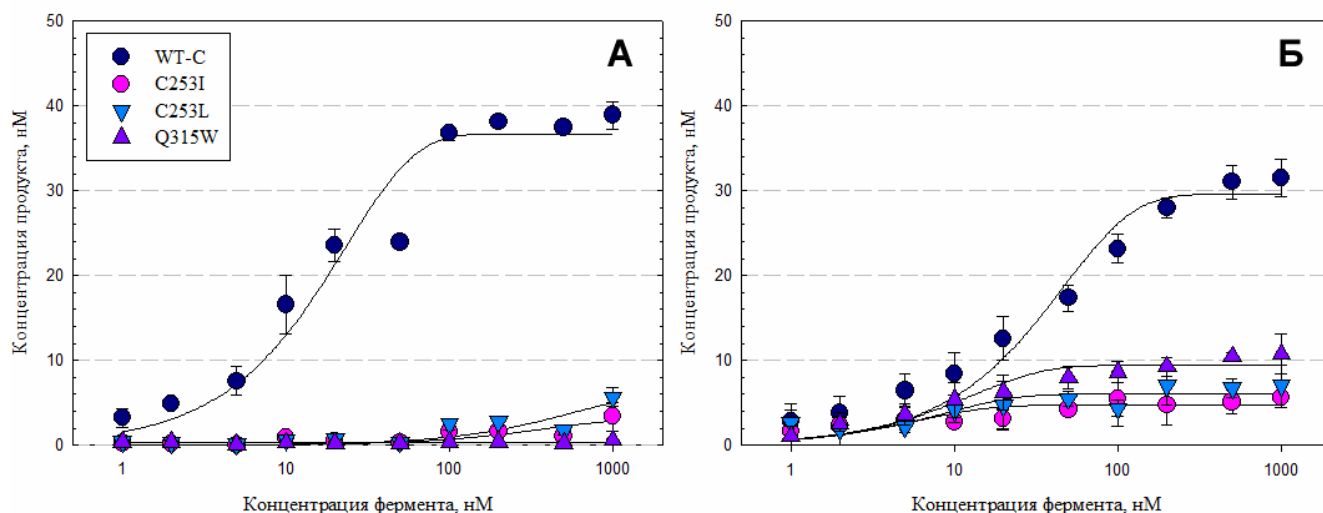
#### 4.3.4. Кинетические параметры реакции мутантных форм OGG1 с субстратами, содержащими 8-оксогуанин или АП-сайт

Для оценки влияния замен аминокислотных остатков в активном центре OGG1 на активность фермента методом сайт-направленного мутагенеза были получены и выделены в чистом виде варианты белка, несущие замены C253I, C253L и Q315W. Полученные мутантные формы белка, как и ожидалось, обладали заметно сниженной активностью по отношению к выщеплению 8-оксоGua из ДНК (рисунок 36, А).

Фермент OGG1 последовательно катализирует реакцию гидролиза *N*-гликозидной связи и β-элиминирования, при этом второй этап реакции протекает намного медленнее первого [62, 187]. В связи с этим в литературе кинетический механизм действия OGG1 обычно описывают двухступенчатой схемой, в которой скорость реакции лимитируется последней стадией:







**Рисунок 36.** Зависимость расщепления субстрата от концентрации фермента OGG1 (см. раздел 3.3.11). (А) Расщепление субстрата, содержащего 8-оксогуанин; (Б) расщепление субстрата, содержащего АП-сайт. Легенда для графиков А и Б общая.

Кажущиеся значения констант,  $k_{ex}$  и  $k_{cl}$ , при условии  $k_{cl} \ll k_{ex}$  можно определить в экспериментах двух типов — кинетики одного оборота и кинетики фазы всплеска соответственно. В первом случае фермент берется в значительном избытке по отношению к субстрату, и константа  $k_{ex}$  для фермента OGG1 отражает эффективность гидролиза *N*-гликозидной связи. Для выявления апурин-апиримидинового продукта разрыва фосфодиэфирного остова в этом случае индуцируют обработкой щелочью или сильноосновными аминами (пиперидин). Константа  $k_{cl}$  определяется при избытке субстрата по отношению к ферменту и представляет собой комбинированный параметр скорости  $\beta$ -элиминирования и распада фермент-субстратного комплекса.

**Таблица 6.** Значения кажущихся каталитических констант OGG1 для субстрата, содержащего 8-охоGua.

<i>Модель</i>	$k_{ex}$	<i>отн. WT-C</i>	$k_{cl}$	<i>отн. WT-C</i>
WT-C	$0,097 \pm 0,009$	1	$0,005 \pm 0,001$	1
C253I	$0,0090 \pm 0,0005$	$0,09 \pm 0,01$	$0,0011 \pm 0,0006$	$0,2 \pm 0,1$
C253L	$0,0130 \pm 0,0007$	$0,13 \pm 0,01$	$0,0011 \pm 0,0007$	$0,2 \pm 0,1$
Q315W	нет реакции	-	нет реакции	-

Для фермента OGG1 дикого типа измеренные значения  $k_{ex}$  и  $k_{cl}$  составляли  $0,097 \pm 0,009$  мин<sup>-1</sup> и  $0,005 \pm 0,001$  мин<sup>-1</sup> соответственно, что согласуется с природой скоростьлимитирующего шага реакции. Замены C253I и C253L примерно на порядок снижали значение  $k_{ex}$  (таблица 6), а замена Q315W приводила к полной инактивации фермента и отсутствию продукта реакции. Такие последствия мутации вполне ожидаемы для замен,

мешающих связыванию 8-охоGua в активном центре фермента, что не позволяет реагирующей группе Lys249 принять нужную для нуклеофильной атаки ориентацию. Влияние мутаций на  $k_{cl}$  было менее выраженным, однако эта константа также снижалась примерно в 5 раз (таблица 6). Очевидно, конформационные искажения, возникающие из-за стерических взаимодействий 8-охоGua с объемными группами в положении 253, влияют и на эффективность реакции  $\beta$ -элиминирования. Интересно, что в качестве возможного акцептора для уходящего в реакции  $\beta$ -элиминирования протона при C2' ряд авторов рассматривает само основание 8-охоGua, некоторое время после выщепления существующее в виде аниона [139]. В таком случае эффективность  $\beta$ -элиминирования может снижаться не только за счет конформационных искажений дезоксирибозы в активном центре, но и из-за снижения сродства 8-охоGua к OGG1 после выщепления.

При действии фермента OGG1 и его вариантов на субстраты, содержащие АП-сайт, активность мутантных форм была также снижена, но проявлялась более выражено, чем в случае субстрата с 8-охоGua (рисунок 36, Б). Реакция в этом случае сводится только к процессу  $\beta$ -элиминирования, причем, очевидно, акцептором протона не может выступать выщепленное основание. Кажущаяся константа  $k_{cl}$ , определенная методом кинетики одного оборота, в этом случае имеет физический смысл константы скорости реакции  $\beta$ -элиминирования, не осложненной высвобождением продукта. Замены C253I, C253L и Q315W оказывали небольшое влияние на значение  $k_{cl}$ , снижая его примерно вдвое (таблица 7), что согласуется с отсутствием необходимости связывания основания в активном центре фермента.

**Таблица 7.** Значения кажущихся каталитических констант OGG1 для субстрата, содержащего АП-сайт.

<i>Модель</i>	$k_{cl}$	<i>отн. WT-C</i>
WT-C	$0,012 \pm 0,001$	1
C253I	$0,0048 \pm 0,0005$	$0,40 \pm 0,05$
C253L	$0,0055 \pm 0,0009$	$0,46 \pm 0,08$
Q315W	$0,0071 \pm 0,0007$	$0,59 \pm 0,08$

Примечательно, что константы снижались в ряду WT-C  $\gg$  Q315W > C253L > C253I, что соответствует структурным данным о популяции каталитически благоприятных конформаций, описанной в разделе 4.3.2.

Для OGG1 дикого типа и вариантов C253I и C253L эти значения были в 2,5–5 раз выше, чем  $k_{cl}$  для 8-охоGua-субстратов, что свидетельствует о вкладе как  $\beta$ -элиминирования, так и распада фермент-субстратного комплекса в комбинированную константу  $k_{cl}$ . Примечательно, однако, что в условиях избытка АП-субстрата мутантные формы фермента после начального

всплеска практически не показывали увеличения количества продукта со временем. Возможно, мутации, блокирующие активный центр, также влияют на реакцию гидролиза основания Шиффа, которая должна произойти для распада фермент-субстратного комплекса. Полное отсутствие структурных данных об этой стадии реакции не позволяет рассмотреть возможные причины влияния на нее конформационных искажений в активном центре.

## 5. ЗАКЛЮЧЕНИЕ

В настоящей работе предпринята успешная попытка развить методы молекулярной динамики в области упрощения и автоматизации проведения исследования и комплексного анализа полученных результатов. Разработанные программные средства представляют собой законченные продукты, доступные для широкого использования, и уже находят применение в сфере структурной биологии. С привлечением созданных программ проведено актуальное исследование молекулярно-динамических аспектов активности и субстратной специфичности 8-оксогуанин-ДНК-гликозилаз бактерий и человека. Выявлены структурные детерминанты, позволяющие пролить свет на молекулярные основы узнавания повреждённого основания, а также основания напротив повреждённого, и таким образом на функционирование GO-системы в целом. Полученные данные компьютерного моделирования также получили в работе экспериментальное подтверждение.

Результаты проведённого исследования не только позволяют углубить фундаментальное понимание механизмов действия ДНК-гликозилаз ЭРО. Новые структурные данные, а также разработанные программы и алгоритмы могут в дальнейшем лечь в основу методики рационального дизайна ферментов, а именно разработки ДНК-гликозилаз с изменённой субстратной специфичностью для биоаналитических и генноинженерных приложений. Другой важнейший аспект применения знаний о структуре ферментов — это разработка ингибиторов ЭРО для борьбы с онкозаболеваниями, т.к. это одна из многообещающих стратегий борьбы с опухолевыми клетками, утратившими альтернативные пути репарации. Фермент OGG1, как и прочие компоненты GO-системы, всё чаще рассматриваются как терапевтические мишени, подавление которых селективно повышает чувствительность клеток опухоли к цитотоксичному действию химио- и радиотерапии.

Нужно также отметить, что методология, лежащая в основе работы, не ограничивается ДНК-гликозилазами. В частности, разработанный метод поиска сайтов связывания молекул воды, использованный при анализе специфичности Frg, может быть в дальнейшем обобщён для поиска сильно связанной сольватной оболочки, формирующейся вблизи биополимеров и их комплексов. Такие инструменты востребованы как при анализе интерфейсов взаимодействия биополимеров, так и в рамках разработки высококачественных молекулярно-континуальных моделей растворителя для нужд современной молекулярной динамики. Однако подобные обобщения представляют собой дальнейшие стадии исследования и выходят за рамки диссертационной работы.

## 6. ВЫВОДЫ

1. Разработаны новые программные средства: графический пользовательский интерфейс GUI-BioPASED к пакету BISON/BioPASED, упрощающий формирование заданий для молекулярного моделирования и автоматизирующий первичную верификацию моделей; программа MDTRA с графическим пользовательским интерфейсом, осуществляющая анализ молекулярно-динамических траекторий на принципах эргономичности (проектная организация и конвейер данных), расширяемости (внешнее конфигурирование и пользовательские подпрограммы), эффективной визуализации результатов (немедленный просмотр и экспорт в векторные форматы) и низких системных требований (поддержка многопоточности, GPGPU, оптимизация использования оперативной памяти). В рамках тестирования созданных программ охарактеризован паттерн водородных связей и выявлены районы повышенной конформационной мобильности в молекуле 8-оксогуанин-ДНК-гликозилазы Fpg из *Escherichia coli*.

2. Исследована динамика каталитически важных аминокислотных остатков Pro1 и Glu2 в активном центре 8-оксогуанин-ДНК-гликозилазы Fpg из *Lactococcus lactis* в разных состояниях ионизации. Наиболее оптимальная для катализа пространственная конфигурация реагирующих групп для природного субстрата — пары 8-оксогуанин–цитозин — наблюдалась при нахождении Pro1 в заряженной форме, а Glu2 в нейтральной; после необходимого депротонирования Pro1 допустимы оба состояния ионизации Glu2. Показано, что при связывании ферментом наименее оптимального субстрата — пары 8-оксогуанин–аденин — в наибольшей степени изменяется динамика удалённого от активного центра высококонсервативного кластера в начале С-концевого домена. Установлено, что дезоксиаденозин в парах 8-оксогуанин–аденин в комплексе с Fpg более стабилен в *син*-конформации, чем в *анти*-конформации, характерной для таких пар в свободной ДНК. Обнаружены сайты прочного связывания молекул воды и установлена их потенциальная роль в механизме излома субстратной ДНК, выворачивания повреждённого основания и катализа.

3. Обнаружено, что снижение активности мутантных форм 8-оксогуанин-ДНК-гликозилазы OGG1 человека с заменами в активном центре (C253I, C253L, Q315W), в отношении субстратов, содержащих 8-оксогуанин или апурин-апиримидиновый сайт, происходит за счет деформации активного центра, а не изменений в позиционировании субстратного нуклеотида. Упомянутые мутантные формы белка выделены в рекомбинантном виде, определены стационарные кинетические параметры катализируемых ими реакций; изменение активности мутантных форм соотнесено со степенью отклонения активного центра

от каталитически компетентной конформации. Показано, что ориентация аминокислотной группы Lys249 служит определяющим структурным фактором активности фермента OGG1 в отношении исследованных субстратов.

## 7. СПИСОК ЛИТЕРАТУРЫ

1. Huffman, J.L., Sundheim, O., Tainer, J.A. DNA base damage recognition and removal: new twists and grooves // *Mutat. Res.* – 2005. – V. 577. – N 1-2. – P. 55–76.
2. Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., Ellenberger, T. *DNA Repair and Mutagenesis.* – Washington, D.C.: ASM Press, 2006. – 1118 pp.
3. Зенков, Н.К., Ланкин, В.З., Меньщикова, Е.Б. Окислительный стресс: Биохимический и патофизиологический аспекты. Москва: МАИК Наука/Интерпериодика, 2001. – 343 с.
4. von Sonntag, C. *Free-Radical-Induced DNA Damage and Its Repair: A Chemical Perspective.* Berlin - Heidelberg: Springer, 2006. – 523 pp.
5. Burrows, C.J., Muller, J.G. Oxidative nucleobase modifications leading to strand scission // *Chem. Rev.* – 1998. – V. 98. – N 3. – P. 1109–1151.
6. Slupphaug, G., Kavli, B., Krokan, H.E. The interacting pathways for prevention and repair of oxidative DNA damage // *Mutat. Res.* – 2003. – V. 531. – N 1-2. – P. 231–251.
7. Regulus, P., Duroux, B., Bayle, P.A., Favier, A., Cadet, J., Ravanat, J.L. Oxidation of the sugar moiety of DNA by ionizing radiation or bleomycin could induce the formation of a cluster DNA lesion // *Proc. Natl. Acad. Sci. U.S.A.* – 2007. – V. 104. – N 35. – P. 14032–14037.
8. Dedon, P.C. The chemical toxicology of 2-deoxyribose oxidation in DNA // *Chem. Res. Toxicol.* – 2008. – V. 21. – N 1. – P. 206–219.
9. David, S.S., Williams, S.D. Chemistry of glycosylases and endonucleases involved in base-excision repair // *Chem. Rev.* – 1998. – V. 98. – N 3. – P. 1221–1261.
10. Purmal, A.A., Kow, Y.W., Wallace, S.S. Major oxidative products of cytosine, 5-hydroxycytosine and 5-hydroxyuracil, exhibit sequence context-dependent mispairing *in vitro* // *Nucleic Acids Res.* – 1994. – V. 22. – N 1. – P. 72–78.
11. Vaisman, A., Woodgate, R. Unique misinsertion specificity of poliota may decrease the mutagenic potential of deaminated cytosines // *EMBO J.* – 2001. – V. 20. – N 22. – P. 6520–6529.
12. Steenken, S. Purine bases, nucleosides and nucleotides: aqueous solution redox chemistry and transformation of their radical cations and  $e^-$  and OH adducts // *Chem. Rev.* – 1989. – V. 89. – N 3. – P. 503–520.
13. Kamiya H., Kasai H. Substitution and deletion mutations induced by 2-hydroxyadenine in *Escherichia coli*: effects of sequence contexts in leading and lagging strands // *Nucleic Acids Res.* – 1997. – V. 25. – N 2. – P. 304–311.

14. Culp, S.J., Cho, B.P., Kadlubar, F.F., Evans, F.E. Structural and conformational analyses of 8-hydroxy-2'-deoxyguanosine // *Chem. Res. Toxicol.* – 1989. – V. 2. – N 6. – P. 416–422.
15. Aida, M., Nishimura, S. An ab initio molecular orbital study on the characteristics of 8-hydroxyguanine // *Mutat. Res.* – 1987. – V. 192. – N 2. – P. 83–89.
16. ESCODD (European Standards Committee on Oxidative DNA Damage), Gedik, C.M., Collins, A. Establishing the background level of base oxidation in human lymphocyte DNA: Results of an interlaboratory validation study // *FASEB J.* – 2005. – V. 19. – N 1. – P. 82–84.
17. Halliwell, B., Gutteridge, J. *Free Radicals in Biology and Medicine.* – Oxford, U.K: Oxford University Press, 2007. – 888 pp.
18. Lu, T., Pan, Y., Kao, S.-Y., Li, C., Kohane, I., Chan, J., Yankner, B.A. Gene regulation and DNA damage in the ageing human brain // *Nature* – 2004. – V. 429. – N 6994. – P. 883–891.
19. Atamna, H., Cheung, I., Ames, B.N. A method for detecting abasic sites in living cells: Age-dependent changes in base excision repair // *Proc. Natl Acad. Sci. U.S.A.* – 2000. – V. 97. – N 2. – P. 686–691.
20. Василенко, Н.Л., Невинский, Г.А. Ферменты прямой, эксцизионной и коррекционной систем репарации высших и низших организмов и их биологическая роль // *Молекулярная биология.* – 2003. – Т. 37. – № 6. – С. 1–17.
21. Eisen, J.A., Hanawalt, P.C. A phylogenomic study of DNA repair genes, proteins, and processes // *Mutat. Res.* – 1999. – V. 435. – N 3. – P. 171–213.
22. Zharkov, D.O. Base excision DNA repair // *Cell. Mol. Life Sci.* – 2008. – V. 65. – N. 10. – P. 1544–1565.
23. Demple, B., Harrison, L. Repair of oxidative damage to DNA: enzymology and biology // *Annu. Rev. Biochem.* – 1994. – V. 63. P. 915–948.
24. Izumi, T., Wiederhold, L.R., Roy, G., Roy, R., Jaiswal, A., Bhakat, K.K., Mitra, S., Hazra, T.K. Mammalian DNA base excision repair proteins: their interactions and role in repair of oxidative DNA damage // *Toxicology.* – 2003. – V.193. – N 1–2. – P. 43–65.
25. Ischenko, A.A., Saparbaev, M.K. Alternative nucleotide incision repair pathway for oxidative DNA damage // *Nature.* – 2002. – V. 415. – N 6868. – P. 183–187.
26. Жарков, Д.О. Структура и конформационная динамика гликозилаз эксцизионной репарации оснований ДНК // *Молекулярная биология.* – 2007. – Т. 41. – N 5. – С. 772–786.
27. McCullough, A.K., Dodson, M.L., Lloyd, R.S. Initiation of base excision repair: Glycosylase mechanisms and structures // *Annu. Rev. Biochem.* – 1999. – V. 68. – P. 255–285.



28. Gilboa, R., Zharkov, D.O., Golan, G., et al. 2002. Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA // *J. Biol. Chem.* – 2002. – V. 277. – N 22. – P. 19811–19816.
29. Fromme, J.C., Verdine, G.L. 2002. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM // *Nat. Struct. Biol.* – 2002. – V. 9. – N 7. – P. 544–552.
30. Verdine, G.L., Norman, D.P.G. Covalent trapping of protein-DNA complexes // *Annu. Rev. Biochem.* – 2003. – V. 72. – P. 337–366.
31. Zharkov, D.O., Ishchenko, A.A., Douglas, K.T., Nevinsky, G.A. Recognition of damaged DNA by *Escherichia coli* Fpg protein: Insights from structural and kinetic data // *Mutat. Res.* – 2003. – V. 531. – N 1–2. – P. 141–156.
32. Zharkov, D.O., Grollman, A.P. Combining structural and bioinformatics methods for the analysis of functionally important residues in DNA glycosylases // *Free Radic. Biol. Med.* – 2002. – V. 32. – N 12. – P. 1254–1263.
33. Zharkov, D.O. Predicting functional residues in DNA glycosylases by analysis of structure and conservation // *Practical Bioinformatics* / Ed. J.N. Bujnicki. – Springer-Verlag, 2004. – P. 243–261.
34. Lingaraju, G.M., Sartori, A.A., Kostrewa, D., Prota, A.E., Jiricny, J., Winkler, F.K. A DNA glycosylase from *Pyrobaculum aerophilum* with an 8-oxoguanine binding mode and a noncanonical helix-hairpin-helix structure // *Structure.* – 2005. – V. 13. – N 1. – P. 87–98.
35. Eichman, B.F., O'Rourke, E.J., Radicella, J.P., Ellenberger, T. Crystal structures of 3-methyladenine DNA glycosylase MagIII and the recognition of alkylated bases // *EMBO J.* – 2003. – V. 22. – N 19. – P. 4898–4909.
36. Zharkov, D.O., Shoham, G., Grollman, A.P. Structural characterization of the Fpg family of DNA glycosylases // *DNA Repair.* – 2003. – V. 2. – N 8. – P. 839–862.
37. Lee, D.-H., Liu, Y., Lee, H.-W., Xia, B., Brice, A.R., Park, S.-H., Balduf, H., Dominy, B.N., Cao, W. A structural determinant in the uracil DNA glycosylase superfamily for the removal of uracil from adenine/uracil base pairs // *Nucleic Acids Res.* – 2015. – V. 43. – N 2. – P. 1081–1089.
38. Tchou, J., Bodepudi, V., Shibutani, S., Antoshechkin, I., Miller, J., Grollman, A.P., Johnson, F. Substrate specificity of Fpg protein: Recognition and cleavage of oxidatively damaged DNA // *J. Biol. Chem.* – 1994. – V. 269. – N 21. – P. 15318–15324.

39. Рыхлевская, А.И., Кузнецова, С.А. Моно- и бифункциональные ДНК-гликозилазы, участвующие в репарации окислительных повреждений в ДНК // Молекулярная биология. – 2000. – Т. 34. – N 6. – С. 1007–1024.
40. Guan, Y., Manuel, R.C., Arvai, A.S., et al. MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily // Nat. Struct. Biol. – 1998. – V. 5. – N 12. – P. 1058–1064.
41. Fromme, J.C., Banerjee A., Huang S.J., et al. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase // Nature. – 2004. – V. 427. – N 6975. – P. 652–656.
42. Luncsford, P.J., Chang, D.Y., Shi, G., Bernstein, J., Madabushi, A., Patterson, D.N., Lu, A.L., Toth, E.A. A structural hinge in eukaryotic MutY homologues mediates catalytic activity and Rad9-Rad1-Hus1 checkpoint complex interactions // J. Mol. Biol. – 2010. – V. 403. – N 3. – P. 351–370.
43. Zharkov, D.O., Grollman, A.P. MutY DNA glycosylase: Base release and intermediate complex formation // Biochemistry. – 1998. – V. 37. – N 36. – P. 12384–12394.
44. Dizdaroglu, M., Burgess, S.M., Jaruga, P., Hazra, T.K., Rodriguez, H., Lloyd, R.S. Substrate specificity and excision kinetics of *Escherichia coli* endonuclease VIII (Nei) for modified bases in DNA damaged by free radicals // Biochemistry – 2001. – V. 40. – N. 40. – P. 12150–12156.
45. Zharkov, D.O., Golan, G., Gilboa, R., et al. Structural analysis of an *Escherichia coli* endonuclease VIII covalent reaction intermediate // EMBO J. – 2002. – V. 21. – N 4. – P. 789–800.
46. Грин, И. П., Жарков, Д. О. Эукариотические гомологи эндонуклеазы VIII: новые элементы системы эксцизионной репарации оснований ДНК // Биохимия. – 2011. – Т. 76. – № 1. – С. 99–114.
47. Liu, M., Bandaru, V., Holmes, A., Averill, A. M., Cannan, W., Wallace, S. Expression and purification of active mouse and human NEIL3 proteins // Protein Expr. Purif. – 2012. – V. 84. – N 1. – P. 130–139.
48. Takao, M., Oohata, Y., Kitadokoro, K., Kobayashi, K., Iwai, S., Yasui, A., Yonei, S., Zhang, Q.-M. Human Nei-like protein NEIL3 has AP lyase activity specific for single-stranded DNA and confers oxidative stress resistance in *Escherichia coli* mutant // Genes Cells. – 2009. – V. 14. – N 2. – P. 261–270.
49. Liu, M., Bandaru, V., Bond, J. P., Jaruga, P., Zhao, X., Christov, P. P., Burrows, C. J., Rizzo, C. J., Dizdaroglu, M., Wallace, S. S. The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo // Proc. Natl Acad. Sci. U.S.A. – 2010. – V. 107. – N 11. – P. 4925–4930.

50. Tajiri, T., Maki, H., Sekiguchi, M. Functional cooperation of MutT, MutM and MutY proteins in preventing mutations caused by spontaneous oxidation of guanine nucleotide in *Escherichia coli* // *Mutat. Res.* – 1995. – V. 336. – N 3. – P. 257–267.
51. Auffret, van der Kemp P., Thomas, D., Barbey, R., de Oliveira, R., Boiteux, S. Cloning and expression in *Escherichia coli* of the *OGG1* gene of *Saccharomyces cerevisiae*, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-*N*-methylformamidopyrimidine // *Proc. Natl Acad. Sci. U.S.A.* – 1996. – V. 93. – N 11. – P. 5197–5202.
52. Rosenquist, T.A., Zharkov, D.O., Grollman, A.P. Cloning and characterization of a mammalian 8-oxoguanine DNA glycosylase // *Proc. Natl Acad. Sci. U.S.A.* – 1997. – V. 94. – N 14. – P. 7429–7434.
53. Zharkov, D.O., Grollman, A.P. The DNA trackwalkers: Principles of lesion search and recognition by DNA glycosylases // *Mutat. Res.* – 2005. – V. 577. – N 1–2. – P. 24–54.
54. Perlow-Poehnelt, R.A., Zharkov, D.O., Grollman, A.P., et al. Substrate discrimination by formamidopyrimidine-DNA glycosylase: Distinguishing interactions within the active site // *Biochemistry.* – 2004. – V. 43. – N 51. – P. 16092–16105.
55. Sugahara, M., Mikawa, T., Kumasaka, T., et al. Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, *Thermus thermophilus* HB8 // *EMBO J.* – 2000. – V. 19. – N 15. – P. 3857–3869.
56. Fromme, J.C., Verdine, G.L. DNA lesion recognition by the bacterial repair enzyme MutM // *J. Biol. Chem.* – 2003. – V. 278. – N 51. – P. 51543–51548.
57. Serre, L., Pereira, de Jésus K., Boiteux, S., et al. Crystal structure of the *Lactococcus lactis* formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA // *EMBO J.* – 2002. – V. 21. – N 12. – P. 2854–2865.
58. Coste, F., Ober, M., Carell, T., et al. Structural basis for the recognition of the FapydG lesion (2,6-diamino-4-hydroxy-5-formamidopyrimidine) by formamidopyrimidine-DNA glycosylase // *J. Biol. Chem.* – 2004. – V. 279. – N 42. – P. 44074–44083.
59. Zaika, E.I., Perlow, R.A., Matz, E., et al. Substrate discrimination by formamidopyrimidine-DNA glycosylase: A mutational analysis // *J. Biol. Chem.* – 2004. – V. 279. – N 6. – P. 4849–4861.
60. Bjørås, M., Seeberg, E., Luna, L., Pearl, L.H., Barrett, T.E. Reciprocal «flipping» underlies substrate recognition and catalytic activation by the human 8-oxo-guanine DNA glycosylase // *J. Mol. Biol.* – 2002. – V. 317. – N 2. – P. 171–177.

61. Norman, D.P.G., Chung, S.J., Verdine, G.L. Structural and biochemical exploration of a critical amino acid in human 8-oxoguanine glycosylase // *Biochemistry*. – 2003. – V. 42. – N 6. – P. 1564–1572.
62. Zharkov, D.O., Rosenquist, T.A., Gerchman, S.E., et al. Substrate specificity and reaction mechanism of murine 8-oxoguanine-DNA glycosylase // *J. Biol. Chem.* – 2000. – V. 275. – N 37. – P. 28607–28617.
63. Bruner, S.D., Norman, D.P.G., Verdine, G.L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA // *Nature*. – 2000. – V. 403. – N 6772. – P. 859–866.
64. McCann, J.A.B., Berti, P.J. Adenine release is fast in MutY-catalyzed hydrolysis of G:A and 8-oxo-G:A DNA mismatches // *J. Biol. Chem.* – 2003. – V. 278. – N 32. – P. 29587–29592.
65. Bernards, A.S., Miller, J.K., Bao, K.K., Wong, I. Flipping duplex DNA inside out: A double base-flipping reaction mechanism by *Escherichia coli* MutY adenine glycosylase // *J. Biol. Chem.* – 2002. – V. 277. – N 23. – P. 20960–20964.
66. Stivers, J.T., Jiang, Y.L. A mechanistic perspective on the chemistry of DNA repair glycosylases // *Chem. Rev.* – 2003. – V. 103. – N 7. – P. 2729–2760.
67. Berti, P.J., McCann, J.A.B. Toward a Detailed Understanding of Base Excision Repair Enzymes: Transition State and Mechanistic Analyses of N-Glycoside Hydrolysis and N-Glycoside Transfer // *Chem. Rev.* – 2006. – V. 106. – N 2. – P. 506–555.
68. Bailly, V., Verly, W.G., O'Connor, T., Laval, J. Mechanism of DNA strand nicking at apurinic/apyrimidinic sites by *Escherichia coli* [formamidopyrimidine]DNA glycosylase. // *Biochem. J.* – 1989. – V. 262. – N 2. – P. 581–589.
69. Dodson, M.L., Michaels M.L., Lloyd R.S. Unified catalytic mechanism for DNA glycosylases // *J. Biol. Chem.* – 1994. – V. 269. – N 52. – P. 32709–32712.
70. Brinkmeyer, M.K., Pope, M.A., David, S.S. Catalytic Contributions of Key Residues in the Adenine Glycosylase MutY Revealed by pH-dependent Kinetics and Cellular Repair Assays // *Chem. Biol.* – 2012. – V. 19. – N 2. – P. 276–286.
71. Sobol, R.W. For MutY, It's All about the OG // *Chem. Biol.* – 2012. – V. 19. – N 3. – P. 313–314.
72. Manuel, R.C., Hitomi, K., Arvai, A.S., House, P.G., Kurtz, A.J., Dodson, M.L., McCullough, A.K., Tainer, J.A., Lloyd, R.S. Reaction Intermediates in the Catalytic Mechanism of *Escherichia coli* MutY DNA Glycosylase // *J. Biol. Chem.* – 2004. – V. 279. – N 45. – P. 46930–46939.
73. Adcock, S.A., McCammon, J.A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins // *Chem. Rev.* – 2006. – V. 106. – N 2. – P. 1589–1615.

74. McCammon, J.A., Harvey, S.C. Dynamics of Proteins and Nucleic Acids. – Cambridge University Press: Cambridge New York, 1987. – 248 pp.
75. Brooks III, C.L., Karplus, M., Pettitt, B.M. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics. Wiley: New York, 1988. – 260 pp.
76. Becker, O.M. Computational Biochemistry and Biophysics. M. Dekker: New York, 1987. – 489 pp.
77. Roux, B., Schulten, K. Computational studies of membrane channels // *Structure*. – 2004. – V. 12. – N 8. – P. 1343–1351.
78. Becker, O.M., Karplus, M. Guide to biomolecular simulations. Springer: Dordrecht, Netherlands, 2006. – 220 pp.
79. Klepeis, J.L., Lindorff-Larsen, K., Dror, R.O., Shaw, D.E. Long-timescale molecular dynamics simulations of protein structure and function // *Current Opinion in Structural Biology*. – 2009. – V. 19. – N 2. – P. 120–127.
80. Lavery, R., Maddocks, J. H., Pasi, M., Zakrzewska, K. Analyzing ion distributions around DNA // *Nucleic Acids Res.* – 2014. – V. 42. – N 12. – P. 8138–8149.
81. Krieger, E., Vriend, G. New ways to boost molecular dynamics simulations // *J. Comput. Chem.* – 2015. – V. 36. – N 13. – P. 996–1007.
82. Kollman, P.A. Free energy calculations: Application to chemical and biochemical phenomena // *Chem. Rev.* – 1993. – V. 93. – N 7. – P. 2395–2417.
83. Henzler-Wildman, K.A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., Pozharski, E., Wilson, M.A., Petsko, G.A., Karplus, M. Intrinsic motions along an enzymatic reaction trajectory // *Nature*. – 2007. – V. 450. – N 7171. – P. 838–844.
84. Kubelka, J., Chiu, T.K., Davies, D.R., Eaton, W.A., Hofrichter J. Sub-microsecond protein folding // *J. Mol. Biol.* – 2006. – V. 359. – N 3. – P. 546–553.
85. Roux, B. Faraldo-Gómez. On the importance of a funneled energy landscape for the assembly and regulation of multidomain Src tyrosine kinases // *Proc. Natl. Acad. Sci. U S A.* – 2007. – V. 104. – N 34. – P. 13643–13648.
86. Vásquez, V., Sotomayor, M., Cordero-Morales, J., Schulten, K., Perozo, E. A structural mechanism for MscS gating in lipid bilayers // *Science*. – 2008. – V. 321. – N 5893. – P. 1210–1214.
87. McDowell, S.E., Nad’á Spackova N., Sponer, J., Walter N.G. Molecular dynamics of RNA: in silico single molecule approach // *Biopolymers*. – 2007. – V. 85. – N 2. – P. 169–184.
88. Wong, C.F., McCammon, A.J. Protein simulation and drug design // *Advances in protein chemistry*. – 2003. – V. 66. – P. 87–121.

89. Schlick, Tamar. *Molecular modeling and Simulation*. Interdisciplinary Applied Mathematics. Springer: New York, 2002. – 768 pp.
90. Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations // *J. Comp. Chem.* – 1983. – V. 4. – N 2. – P. 787–217.
91. Reed, T.M., Gubbins, K.E. *Applied Statistical Mechanics*. McGraw: New York, 1973. – 496 pp.
92. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules // *J. Am. Chem. Soc.* – 1995. – V. 117. – N 19. – P. 5179–5197.
93. Bayly, C.I., Ciepak, P., Cornell, W.D., Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP mode // *J. Phys. Chem.* – 1993. – V. 97. – N 40. – P. 10269–10280.
94. Cornell, W.D., Ciepak, P., Bayly, C.I., Kollman, P.A. Application of the RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation // *J. Am. Chem. Soc.* – 1993. – V. 115. – N 21. – P. 9620–9631.
95. Wang, L., Wolf, R. M., Caldwell, J. W., Kollman, P. A., Case, D. A. Development and Testing of a General Amber Force Field // *J. Comput. Chem.* – 2004. – V. 25. – N 9. – P. 1157-1174.
96. Jorgensen, W. L., Maxwell, D. S., Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids // *J. Am. Chem. Soc.* – 1996. – V. 118. – N 45. – P. 11255-11236.
97. Adri van Duin, C.T., Dasgupta, S., Lorant, F., Goddard III, W.A. ReaxFF: A Reactive Force Field for Hydrocarbons // *J. Phys. Chem.* – 2001. – V. 105. – N 41. – P. 9396–9409.
98. Гилл, Ф., Мюррей, У., Райт, М. *Практическая оптимизация*. М.: Мир, 1985. – 509 с.
99. Рапапорт, Д. К. *Искусство молекулярной динамики / под научной редакцией Р. Г. Ефремова; перевод с англ. яз. А. Н. Дьяконовой*. М.-Ижевск: НИЦ «Регулярная и хаотическая динамика», Ижевский институт компьютерных исследований, 2012. – 632 с.
100. Beeman, D. Some multistep methods for use in molecular dynamics calculations // *J. Comput. Phys.* – 1976. – V. 20. – N 2. – P. 130–139.
101. Brünger, A.T., Brooks III, C.L., Karplus, M. Active Site Dynamics of Ribonuclease // *Proc. Natl. Acad. Sci. U S A.* – 1985. – V. 82. – N 24. – P. 8458–8462.
102. Ryckaert, J.-P., Ciccotti, G., Berendsen, H.J.C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes // *J. Comput. Phys.* – 1977. – V. 23. – N 3. – P. 327–341.

103. Yoneya, M., Berendsen, H.J.C, Hirasawa, K. A Noniterative Matrix Method for Constraint Molecular-Dynamics Simulations // *Molecular Simulations* - 1994. - V. 13. - N 6. - P. 395–405.
104. Forester, T.R., Smith, W. SHAKE, Rattle, and Roll: Efficient Constraint Algorithms for Linked Rigid Bodies // *J. Comput. Chem.* - 1998. - V. 19. - N 1. - P. 102–111.
105. Gonnet, P. P-SHAKE: A quadratically convergent SHAKE in  $O(n^2)$  // *J. Comput. Phys.* - 2007. - V. 220. - N 2. - P. 740–750.
106. Hess, B., Bekker, H., Berendsen, H.J.C, Fraaije, J.G.E.M. LINCS: A Linear Constraint Solver for Molecular Simulations // *J. Comput. Chem.* - 1997. - V. 18. - N 12. - P. 1463–1472.
107. Lazaridis, T., Karplus, M. Effective Energy Function for Proteins in Solution // *Proteins.* - 1999. - V. 35. - N 2. - P. 133–152.
108. Lazardis, T., Karplus, M. Thermodynamics of protein folding: a microscopic view // *Biophys. Chem.* - 2003. - V. 100. - N 1-3. - P. 367–395.
109. Decherchi, S., Masetti, M., Vyalov, I., Rocchia, W. Implicit solvent methods for free energy estimation // *Eur. J. Med. Chem.* - 2015. - V. 91. - P. 27–42.
110. Zhang, L.Y., Gallicchio, E., Friesner, R.A., Levy, R.M. Solvent Models for Protein-Ligand Binding: Comparison of Implicit Solvent Poisson and Surface Generalized Born Models with Explicit Solvent Simulations // *J. Comput. Chem.* - 2001. - V. 22. - N 6. - P. 591–607.
111. Vorobjev, Y.N. Advances in implicit models of water solvent to compute conformational free energy and molecular dynamics of proteins at constant pH // *Adv. Prot. Chem. Struct. Biol.* - 2011. - V. 85. - P. 281–322.
112. Reddy, Ch. K., Das, A., Jayaram, B. Do Water Molecules Mediate Protein-DNA Recognition? // *J. Mol. Biol.* - 2001. - V. 314. - N 3. - P. 619–632.
113. Schneider, B., Černý, J., Svozil, D., Čech, P., Gelly, J.-C., de Brevern, A. G. Bioinformatic analysis of the protein/DNA interface // *Nucleic Acids Res.* - 2014. - V. 42. - N 5. - P. 3381–3394.
114. Cipra, B.A. The Best of the 20th Century: Editors Name Top 10 Algorithms // *SIAM News* - 2000. - V. 33. - N 4. - P. 2.
115. De Groot, B., vanAalten, D., Scheek, R., Amadei, A., Vriend, G. Prediction of protein conformational freedom from distance constraints // *Proteins.* - 1997. - V. 29. - P. 240–251.
116. Car, R., Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory // *Phys. Rev. Lett.* - 1985. - V. 55. N 22. - P. 2471–2474.
117. McCammon, J.A., Lee, C.Y., Northrup, S.H. Side-Chain Rotational Isomerisation in Proteins: A Mechanism Involving Gating and Transient Packing Defects // *J. Am. Chem. Soc.* - 1983. - V. 105. - N 8. - P. 2232–2237.

118. Сандерс, Дж., Кэндрот, Э. Технология CUDA в примерах. М.: ДМК Пресс, 2011. – 232 с.
119. Боресков, А. В., Харламов, А. А. Основы работы с технологией CUDA. М.: ДМК Пресс, 2011. – 232 с.
120. Jo, S., Kim, T., Iyer, V.G., Im, W. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM // *J. Comput. Chem.* – 2008. – V. 29. – N 11. – P. 1859–1865.
121. Hess, B., Kutzner, C., Spoel, D., Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation // *J. Chem. Theory Comput.* – 2008. – V. 4. – N 3. – P. 435–447.
122. Christen, M., Hünenberger, P.H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D.P., Heinz, T.N., Kastenholz, M.A., Kräutler, V., Oostenbrink, C., Peter, C., Trzesniak, D., van Gunsteren, W.F. The GROMOS software for biomolecular simulation: GROMOS05 // *J. Comput. Chem.* – 2005. – V. 26. – N 16. – P. 1719–1751.
123. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., Schulten, K. Scalable molecular dynamics with NAMD // *J. Comput. Chem.* – 2005. – V. 26. – N 15. – P. 1781–1802.
124. Humphrey, W., Dalke, A., Schulten, K. VMD – Visual Molecular Dynamics // *J. Molec. Graphics* – 1996. – V. 14. – N 1. – P. 33–38.
125. Case, D.A., Cerutti, D.S., Cheatham, T.E., III, Darden, T.A., Duke, R.E., Giese, T.J., Gohlke, H., Goetz, A.W., Greene, D., Homeyer, N., Izadi, S., Kovalenko, A., Lee, T.S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Mermelstein, D., Merz, K.M., Monard, G., Nguyen, H., Omelyan, I., Onufriev, A., Pan, F., Qi, R., Roe, D.R., Roitberg, A., Sagui, C., Simmerling, C.L., Botello-Smith, W.M., Swails, J., Walker, R.C., Wang, J., Wolf, R.M., Wu, X., Xiao, L., York, D.M., Kollman, P.A. AMBER 2017. San Francisco: University of California, 2017. – 951 pp.
126. Brooks, B. R., Brooks III, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, A., Caflisch, S., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., Karplus, M. CHARMM: The Biomolecular simulation Program // *J. Comput. Chem.* – 2009. – V. 30. – N 10. – P. 1545–1615.
127. Luo, N., Mehler, E., Osman, R. Specificity and Catalysis of Uracil DNA Glycosylase. A Molecular Dynamics Study of Reactant and Product Complexes with DNA // *Biochemistry.* – 1999. – V. 38. – N 29. – P. 9209–9220.



128. Ma, A., Hu, J., Karplus, M., Dinner, A.R. Implications of Alternative Substrate Binding Modes for Catalysis by Uracil-DNA Glycosylase: An Apparent Discrepancy Resolved // *Biochemistry*. – 2006. – V. 45. – N 46. – P. 13687–13696.
129. Dinner, A., Blackburn, G.M., Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis // *Nature*. – 2006. – V. 413. – N 6857. – P. 752–755.
130. Olufsen, M., Brandsdal, B.O., Smalås, A.O. Comparative unfolding studies of psychrophilic and mesophilic uracil DNA glycosylase: MD simulations show reduced thermal stability of the cold-adapted enzyme // *J. Mol. Graph. Model.* – 2007. – V. 26. – N 1. – P. 124–134.
131. Olufsen, M., Smalås, A.O., Brandsdal, B.O. Electrostatic interactions play an essential role in DNA repair and cold-adaptation of Uracil DNA glycosylase // *J. Mol. Model.* – 2008. – V. 14. – N 3. – P. 201–213.
132. Olufsen, M., Papaleo, E., Smalås, A.O., Brandsdal, B.O. Ion pairs and their role in modulating stability of cold- and warm-active uracil DNA glycosylase // *Proteins*. – 2008. – V. 71. – N 3. – P. 1219–1230.
133. Mi, R., Dong, L., Kaulgud, T., Hackett, K.W., Dominy, B.N., Cao, W. Insights from Xanthine and Uracil-DNA-Glycosylase Activities of Bacterial and Human SMUG1: Switching SMUG1 to UDG // *J. Mol. Biol.* – 2009. – V. 385. – N 3. – P. 761–778.
134. Song, K., Hornak, V., de los Santos, C., Grollman, A.P., Simmerling, C. Computational Analysis of the Mode of Binding of 8-Oxoguanine to Formamidopyrimidine-DNA Glycosylase // *Biochemistry*. – 2006. – V. 45. – N 46. – P. 10886–10894.
135. Amara, P., Serre, L., Castaing, B., Thomas, A. Insights into the DNA repair process by the formamidopyrimidine-DNA glycosylase investigated by molecular dynamics // *Protein Science*. – 2004. – V. 13. – N 8. – P. 2009–2021.
136. Amara, P., Serre, L. Functional flexibility of *Bacillus stearothermophilus* formamidopyrimidine DNA-glycosylase // *DNA Repair*. – 2006. – V. 5. – N 8. – P. 947–958.
137. Qi, Y., Spong, M.C., Nam, K., Banerjee, A., Jiralerspong, S., Karplus, M., Verdine, G.L. Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme // *Nature*. – 2009. – V. 462. – N 7274. – P. 762–766.
138. Qi, Y., Spong, M.C., Nam, K., Karplus, M., Verdine, G.L. Entrapment and Structure of an Extrahelical Guanine Attempting to Enter the Active Site of a Bacterial DNA Glycosylase, MutM // *J. Biol. Chem.* – 2010. – V. 285. – N 2. – P. 1468–1478.
139. Fromme, J.C., Bruner, S.D., Yang, W., Karplus, M., Verdine, G.L. Product-assisted catalysis in base-excision DNA repair // *Nat. Struct. Biol.* – 2003. – V. 10. – N 3. – P. 204–211.

140. Pinak, M. 8-Oxoguanine Lesioned B-DNA Molecule Complexed with Repair Enzyme hOGG1: A Molecular Dynamics Study // *J. Comput. Chem.* – 2003. – V. 24. – N 7. – P. 898–907.
141. Banerjee, A., Yang, W., Karplus, M., Verdine, G.L. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA // *Nature.* – 2005. – V. 434. – N 7033. – P. 612–618.
142. Schyman, P., Danielsson, J., Pinak, M., Laaksonen, A. Theoretical Study of the Human DNA Repair Protein HOGG1 Activity // *J. Phys. Chem. A.* – 2005. – V. 109. – N 8. – P. 1713–1719.
143. Calvarsi, M., Bottoni, A., Garavelli, M. Computational Clues for a New Mechanism in the Glycosylase Activity of the Human DNA Repair Protein hOGG1. A Generalized Paradigm for Purine-Repairing Systems? // *J. Phys. Chem. B.* – 2007. – V. 111. – N 23. – P. 6557–6570.
144. Anderson, P.C., Daggett, V. The R46Q, R131Q and R154H Polymorphs of Human DNA Glycosylase/ $\beta$ -Lyase hOgg1 Severely Distort the Active Site and DNA Recognition Site but do not Cause Unfolding // *J. Am. Chem. Soc.* – 2009. – V. 131. – N 27. – P. 9506–9515.
145. Brunk, E., Arey, J.S., Rothlisberger, U. Role of Environment for Catalysis of the DNA Repair Enzyme MutY // *J. Am. Chem. Soc.* – 2012. – V. 134. – N 20. – P. 8608–8616.
146. Jia, L., Shafirovich, V., Geacintov, N.E., Broyde, S. Lesion Specificity in the Base Excision Repair Enzyme hNeil1: Modeling and Dynamics Studies // *Biochemistry.* – 2007. – V. 46. – N 18. – P. 5305–5314.
147. Fuxreiter, M., Warshel, A., Osman, R. Role of Active Site Residues in the Glycosylase Step of T4 Endonuclease V. Computer Simulation Studies on Ionization States // *Biochemistry.* – 1999. – V. 38. – N 30. – P. 9577–9589.
148. Guliaev, A., Singer, B., Hang, B. Chloroethylnitrosourea-derived ethano cytosine and adenine adducts are substrates for *Escherichia coli* glycosylases excising analogous etheno adducts // *DNA Repair.* – 2004. – V. 3. – N 30. – P. 1311–1321.
149. Guliaev, A., Hang, B., Singer, B. Structural insights by molecular dynamics simulations into differential repair efficiency for ethano-A versus etheno-A adducts by the human alkylpurine-DNA N-glycosylase // *Nucleic Acid Res.* – 2002. – V. 30. – N 17. – P. 3778–3787.
150. Wang, P., Guliaev, A., Hang, B. Metal inhibition of human N-methylpurine-DNA glycosylase activity in base excision repair // *Toxicol. Lett.* – 2006. – V. 166. – N 3. – P. 237–247.
151. Vorobjev, Y.N., Hermans, J. SIMS, Computation of Smooth Invariant Molecular Surface // *Biophys. J.* – 1997. – V. 73. – N 2. – P. 722–732.
152. Vorobjev, Y.N., Scheraga, H.A. A fast multigrid boundary element method for Macromolecular electrostatics computations in solvent // *J. Comp. Chem.* – 1997. – V. 18. – N 4. – P. 569–583.

153. Vorobjev, Y.N., Almagro, J.C., Hermans, J. Discrimination between native and intentionally misfolded conformation of proteins: ES/IS, a new method for calculating conformational free energy // *Proteins*. – 1998. – V. 32. – N 4. – P. 399–413.
154. Vorobjev, Y.N., Vila, J.A., Scheraga, H.A. FAMBE-pH: A Fast and Accurate Method to Compute the Total Solvation Free Energies of Proteins // *J. Phys. Chem.* – 2008. – V. 112. – N 35. – P. 11122–11136.
155. Mallik, B., Masunov, A., Lazaridis, T. Distance and Exposure Dependent Effective Dielectric Function // *J. Comp. Chem.* – 2002. – V. 23. – N 11. – P. 1090–1099.
156. Kortemme, T., Morozov, A.V., Baker, D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes // *J. Mol. Biol.* – 2003. – V. 326. – N 4. – P. 1239–1259.
157. Manning, G. S. Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions I. Colligative Properties // *J. Chem. Phys.* – 1969. – V. 51. – N 3. – P. 924–933.
158. Ravishanker, G., Auffinger, P., Langley, D. R., Jayaram, B., Young, M. A., Beveridge, D. L. Treatment of Counterions in Computer Simulations of DNA. In: *Reviews in Computational Chemistry, Volume 11*. – New York: Wiley-VCH, 1997. – 373 pp.
159. Park, B. H., Levitt, M. Energy functions that discriminate X-ray and near native folds from well-constructed decoys // *J. Mol. Biol.* – 1996. – V. 258. – N. 2. – P. 367–392.
160. Kabsch, W. A solution of the best rotation to relate two sets of vectors // *Acta Crystallogr. A*. – 1976. – V. 32. – N 5. – P. 922–923.
161. Duwat, P., de Oliveira, R., Ehrlich, S. D., Boiteux, S. Repair of oxidative DNA damage in Gram-positive bacteria: The *Lactococcus lactis* Fpg protein // *Microbiology*. – 1995. – V. 141. – N. 2. – P. 411–417.
162. Schomaker, V., Waser, J., Marsh, R. E., Bergman, G. To Fit a Plane or a Line to a Set of Points by Least Squares // *Acta Cryst.* – 1959. – V. 12. – N 8. – P. 600–604.
163. Blow, D. M. To Fit a Plane to a Set of Points by Least Squares // *Acta Cryst.* – 1960. – V. 13. – N 2. – P. 168.
164. Sayle, R.A., Milner-White, E.J. RASMOL: biomolecular graphics for all // *Trends Biochem. Sci.* – 1995. – V. 20. – N 9. – P. 374.
165. Sambrook, J., Russell, D.W. *Molecular Cloning: A Laboratory Manual*. 3rd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2001.
166. Gill, S. C., von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data // *Anal. Biochem.* – 1989. – V. 182. – N 2. – P. 319–326.

167. Glykos, N. M. Carma: a molecular dynamics analysis program // J. Comput. Chem. – 2006. – V. 27. – N 14. – P. 1765–1768.
168. Seeber, M., Cecchini, M., Rao, F., Settanni, G., Caflisch, A. Wordom: a program for efficient analysis of molecular dynamics simulations // Bioinformatics – 2007. – V. 23. – N 19. – P. 2625–2627.
169. Verstraelen, T., Van Houteghem, M., Van Speybroeck, V., Waroquier, M. MD-tracks: a productive solution for the advanced analysis of molecular dynamics and Monte Carlo simulations // J. Chem. Inf. Model. – 2008. – V. 48. – N 12. – P. 2414–2424.
170. Mezei, M. Simulaid: a simulation facilitator and analysis program // J. Comput. Chem. – 2010. – V. 31. – N 14. – P. 2658–2668.
171. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., Beckstein, O. MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations // J. Comput. Chem. – 2011. – V. 32. – N 10. – P. 2319–2327.
172. Pearson, K. On lines and planes of closest fit to systems of points in space // Philos. Mag. – 1901. – V. 2. – N 6. – P. 559–572.
173. Shrake, A., Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin // J. Mol. Biol. – 1973. – V. 79. – N 2. – P. 351–371.
174. Dickerson, R. E. Definitions and nomenclature of nucleic acid structure components // Nucleic Acids Res. – 1989. – V. 17. – N 5. – P. 1797–1803.
175. Levine, J. R. Flex & Bison. – Sebastopol, CA: O'Reilly Media, 2009. – 304 pp.
176. Ierusalimsky, R., de Figueiredo, L. H., Celes, W. The implementation of Lua 5.0 // J. Univers. Comput. Sci. – 2005. – V. 11. – N 7. – P. 1159–1176.
177. Kuznetsov, N. A., Koval, V. V., Zharkov, D. O., Vorobjev, Y. N., Nevinsky, G. A., Douglas, K. T., Fedorova, O. S. Pre-steady-state Kinetic Study of Substrate Specificity of *Escherichia coli* Formamidopyrimidine-DNA Glycosylase // Biochemistry. – 2007. – V. 46. – N 2. – P. 424–435.
178. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M., Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of  $pK_a$  values // J. Chem.Theory Comput. – 2011. – V. 7. – N 7. – P. 2284–2295.
179. O'Brien, P. J., Ellenberger, T. Human alkyladenine DNA glycosylase uses acid-base catalysis for selective excision of damaged purines // Biochemistry. – 2003. – V. 42. – N 42. – P. 12418–12429.
180. Kouchakdjian, M., Bodepudi, V., Shibusani, S., Eisenberg, M., Johnson, F., Grollman, A. P., Patel, D. J. NMR structural studies of the ionizing radiation adduct 7-hydro-8-oxodeoxyguanosine (8-oxo-7H-dG) opposite deoxyadenosine in a DNA duplex. 8-Oxo-7H-

- dG(*syn*)•dA(*anti*) alignment at lesion site // *Biochemistry*. – 1991. – V. 30. – N 5. – P. 1403–1412.
181. McAuley-Hecht, K. E., Leonard, G. A., Gibson, N. J., Thomson, J. B., Watson, W. P., Hunter, W. N., Brown, T. Crystal structure of a DNA duplex containing 8-hydroxydeoxyguanine-adenine base pairs // *Biochemistry*. – 1994. – V. 33. – N 34. – P. 10266–10270.
182. Livingstone, C.D., Barton, G.J. Protein sequence alignments: A strategy for the hierarchical analysis of residue conservation // *Comput. Appl. Biosci.* – 1993. – V. 9. – N 6. – P. 745–756.
183. Radom, C. T., Banerjee, A., Verdine, G. L. Structural Characterization of Human 8-Oxoguanine DNA Glycosylase Variants Bearing Active Site Mutations // *J. Biol. Chem.* – 2007. – V. 282. – N 12. – P. 9182–9194.
184. Dalhus, B., Forsbring, M., Helle, I. H., Vik, E. S., Forstrøm, R. J., Backe, P. H., Alseth, I., Bjørås, M. Separation-of-Function Mutants Unravel the Dual-Reaction Mode of Human 8-Oxoguanine DNA Glycosylase // *Structure*. – 2011. – V. 19. – N 1. – P. 117–127.
185. Fersht, A. *Enzyme Structure and Mechanism*, 2nd Ed. – New York: W. H. Freeman & Co, 1985. – 475 pp.
186. Bjørås, M., Luna, L., Johnsen, B., Hoff, E., Haug, T., Rogness, T., Seeberg, E. Opposite base-dependent reactions of a human base excision repair enzyme on DNA containing 7,8-dihydro-8-oxoguanine and abasic sites // *EMBO J.* – 1997. – V. 16. – N 20. – P. 6314–6322.
187. Kuznetsov, N. A., Koval, V. V., Zharkov, D. O., Nevinsky, G. A., Douglas, K. T., Fedorova, O. S. Kinetics of substrate recognition and cleavage by human 8-oxoguanine-DNA glycosylase // *Nucleic Acids Res.* – 2005. – V. 33. – N 12. – P. 3919–3931.