ОТЗЫВ официального оппонента

на диссертацию на соискание ученой степени

кандидата физико-математических наук Армеева Григория Алексеевича

на тему: «ОПРЕДЕЛЕНИЕ СТРУКТУРЫ НУКЛЕОСОМ И ИХ КОМПЛЕКСОВ С БЕЛКАМИ ХРОМАТИНА МЕТОДАМИ МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ»

по специальности 03.01.09 – «математическая биология, биоинформатика»

В настоящее время методы вычислительного эксперимента широко применяются практически во всех областях науки и техники, являясь равноправным партнером традиционных экспериментальных подходов. В частности, в молекулярной биофизике и структурной биологии огромную роль играют методы компьютерного моделирования структуры и динамики сложных надмолекулярных систем. Получаемая в расчетах информация необходима для детального понимания на атомном уровне физических механизмов работы биологических макромолекул. Решение данной задачи создает основу для дальнейшего рационального конструирования новых молекул с заданными физико-химическими свойствами и определенным спектром биологической активности. Это позволит, например, значительно ускорить процесс создания новых лекарств, их средств доставки и т.д. Серьезной преградой на пути развития данного направления является большой размер изучаемых систем (часто – до 10^6 атомов), что требует применения высокопроизводительных вычислений и/или согласованного применения моделей различного уровня детализации. Достигнутый к настоящему времени прогресс на этом пути в основном связан с моделированием белков, водно-липидных систем, имитирующих биологические мембраны, а также белок-мембранных комплексов. Работ же, посвященных многомасштабному компьютерному моделированию нуклеиновых кислот, и, особенно, их ассоциатов с белками гораздо меньше, несмотря на то, что детальная информация об их структуре, динамике и механизмах действия является крайне востребованной как с

фундаментальной, так и с прикладной точек зрения. Диссертационная работа Армеева Г.А. посвящена созданию и применению новых эффективных вычислительных подходов к атомистическому молекулярному моделированию нуклеосом и их комплексов с белками хроматина. Поскольку нуклеосомы представляют из себя главный структурообразующий элемент хроматина, а экспериментальных данных о них явно недостаточно, привлечение структурно-динамической информации с использованием моделирования очень востребовано и своевременно. Это, несомненно, свидетельствует об актуальности темы диссертации.

Диссертация выполнена по «классической» схеме, изложена на 99 страницах и включает: Список сокращений, Введение, Обзор литературы (Глава 1), Материалы и методы (Глава 2), Результаты и обсуждение (Главы 3, 4), Раздел «Разработанные программы» (Глава 5), Заключение, Список литературы» (118 источников). Диссертация содержит 28 рисунков.

Во Введении обоснована актуальность темы диссертации, сформулированы цель и задачи исследования, описаны научная новизна, теоретическая и практическая значимость работы, а также степень достоверности результатов. Кроме того, представлены сведения об апробации диссертационной работы, о публикациях соискателя и о его личном вкладе в настоящее исследование. Глава 1 представляет собой литературный обзор, который затрагивает следующие основные тематики: 1) Структура и динамика нуклеосом (Раздел 1.1); 2) Изучение динамики нуклеосом методами МД (Раздел 1.2); 3) Методы молекулярного моделирования, основанные на интеграции разнородных экспериментальных данных (Раздел 1.3); 4) Ферстеровский резонансный перенос энергии (Раздел 1.4); 5) Определение контактов ДНК-белок с применением перекисного расщепления ДНК. Из обзора литературы следует, что, несмотря на большой интерес к структурной организации и динамическому поведению нуклеосом, число работ, посвященных разработке и применению методов компьютерного моделирования для их изучения, все еще

ограничено. Это означает, что рассматриваемая проблема остается очень актуальной. Одной из серьезных задач является выбор адекватных математических и физических моделей изучаемых систем. В первую очередь – для обеспечения адекватного исследования конфигурационного фазового пространства этих мезоскопических систем, особенно с учетом эффектов среды. Указанные вопросы автор рассматривает в Разделах 1.1 – 1.3. Поскольку решение поставленных в работе задач невозможно без привлечения данных экспериментов, большое внимание в обзоре литературы также уделено совместному использованию численного моделирования и методов флуоресцентной спектроскопии (включая FRET, Раздел 1.4) и перекисного расщепления ДНК (Раздел 1.5).

Изложенный в Главе 1 материал непосредственно связан с тематикой работы и четко изложен. Это, несомненно, является хорошим подспорьем для студентов и аспирантов, начинающих работать в данной области, а также для специалистов, желающих ознакомиться с современным состоянием дел в ней. Отмечу, что для раскрытия темы автор привлекает свежий библиографический материал – в Списке литературы много источников последних лет. Изложенная в обзоре литературы информация подводит читателя к логически обоснованным выводам относительно целей диссертационной работы и формулировке конкретных задач, которые необходимо решить.

Глава 2 («Методы молекулярного моделирования и анализа экспериментальных данных») посвящена описанию используемых автором вычислительных протоколов, математических моделей и компьютерных программ. Кроме того, технические детали двух из разработанных Армеевым Г.А. программ – FRETy и HYDROID – описаны в заключительной Главе 5. В большинстве случаев материал изложен четко и дает наглядное представление об использованных подходах. В то же время, у меня есть замечания, касающиеся подробности изложения ряда технических деталей вычислений (см. далее), – слишком краткое подчас их описание может затруднить заинтересованным исследователям самостоятельное воспроизведение изложенных в работе

результатов.

Представленные в диссертационной работе научные результаты (Главы 3 и 4) разбиты на восемь разделов – они относятся к изучению динамики нуклеосом методом полноатомной МД (Глава 3) и определению структуры нуклеосом и их комплексов с белками хроматина (Глава 4). Считаю, что основными результатами третьей главы являются следующие: 1) демонстрация стабильности полноатомной модели нуклеосомы в ходе МД в воде на протяжении 1 мкс; 2) Соответствие рассчитанных B-факторов различных элементов комплекса в МД-модели и в кристаллической структуре; 3) Неравномерное распределение молекул воды в нуклеосоме, слабая степень гидратации интерфейса взаимодействия гистонов H3-H4; 4) Гипотеза о наличии в нуклеосоме крупномасштабного движения «расщепления»; 5) Результаты моделирования начального этапа спонтанного «откручивания» ДНК от нуклеосомы, демонстрация формирования асимметричной укладки ДНК в областях входа и выхода из нуклеосомы. Важно, что, несмотря на сложность и громоздкость объекта исследования – полноатомной модели нуклеосомы в явно заданном водном растворе – автору удалось в целом воспроизвести в некоторых случаях имеющиеся (немногочисленные!) экспериментальные наблюдения. Более того, наличие таких моделей позволит в будущем облегчить интерпретацию экспериментальных данных низкого разрешения и предложить возможные механизмы работы нуклеосом.

Учитывая имеющиеся в настоящее время сложности получения структурной информации о нуклеосомах и их комплексах с белками хроматина в прямых биофизических экспериментах и, как следствие, крайнюю ограниченность таких данных, большое значение приобретает возможность применения альтернативных подходов к решению задач диссертационной работы. Среди них – т.н. интегративные технологии, в рамках которых имеющаяся (часто – косвенная!) экспериментальная информация об объекте служит в качестве ограничений в вычислительных экспериментах, а критерием

адекватности метода является согласие результатов моделирования с наблюдаемыми опытным путем. Ряд таких подходов был создан и применен автором – результаты изложены в Главе 4. Важно, что, прежде, чем перейти к их описанию, Армеев Г.А. обсуждает проблему применимости и ограничений созданных методов (Раздел 4.1). Далее приведен пример использования разработанного подхода к оценке зависимости энергии деформации ДНК от нуклеотидной последовательности. Показано, что наиболее значимыми позициями в последовательности линкерных участков ДНК являются сайты -7, -4 и -3 (относительно входа в последовательность, позиционирующую нуклеосомы). При этом в т.н. НД-моделях связывания линкерной ДНК с нуклеосомой («на-диаде») в указанных позициях преобладают A/T, C/T, G/C, а в альтернативных (т.н. ВД, «вне-диады») моделях – пары C/G, C/T и G/T. На основании полученных данных были также предложены «оптимальные» последовательности линкерной ДНК. В Разделах 4.3 – 4.5 описаны результаты применения созданного комбинированного – экспериментального и теоретического – подходов к построению модели ДНК в комплексе нуклеосомы с линкерным гистоном (4.3), нуклеосомы со связанным гистоновым вариантом CenH3 (4.4) и для определения вероятных изменений в нуклеосоме под действием шаперона FACT. В качестве экспериментальной «поддержки» были успешно применены данные по измерению эффективности spFRET (4.3 и 4.5) и гидроксильного расщепления ДНК (4.4). Отмечу, что все перечисленные результаты опубликованы автором. Кроме того, при их получении были использованы разработанные Армеевым Г.А. уникальные эффективные программные средства - FRETy и HYDROID. В частности, программа HYDROID описана в статье, опубликованной в высокорейтинговом журнале Nature Protocols, что свидетельствует о мировом признании разработки.

Следуя сформулированным автором цели и задачам диссертационной работы, в каждом из представленных «блоков» результатов были продемонстрированы индивидуальные подходы к выполнению расчетов и анализу полученных данных. Степень разнообразия и уровень сложности

рассматриваемых объектов впечатляют. Речь идет как о моделировании сложнейших надмолекулярных систем (нуклеосом в среде), так и о комбинированном применении вычислительных методов с опорой на имеющиеся экспериментальные факты, взятые в качестве ограничений. Необходимо также сказать, что автор не только использовал готовые программные средства, но и создал новые инструменты анализа данных о структурной организации нуклеосом и их комплексов с белками. Отмечу, что разработанное программное обеспечение находится в открытом доступе - именно так, по моему мнению, и следует поступать авторам диссертационных работ, выполняемых по специальности «математическая биология, биоинформатика».

Считаю, что цель и задачи, сформулированные в работе, достигнуты в ходе её выполнения. Выводы и заключения, представленные автором, логически обоснованы и убедительно показывают новизну и применимость разработанного метода.

По материалам диссертации считаю необходимым сформулировать ряд вопросов и замечаний.

1. Основной недостаток работы, на мой взгляд, заключается в чересчур кратком, подчас – неинформативном описании вычислительных протоколов ряда методов моделирования и полученных результатов. В частности, автор не поясняет, как именно создавали модель нуклеосомы с «открученной» ДНК (Раздел 3.2.2, стр. 60) и насколько такое произвольное изменение исходной кристаллографической модели влияет на результат. Как показывают представленные здесь же данные, итоговые выводы могут кардинально измениться при выборе другого варианта «вмешательства» в модель нуклеосомы. Параметры силового поля, примененного для расчета энергии деформации ДНК (Раздел 2.2.1), взяты из литературы, но нет сведений о том, проверял ли автор настоящей работы их применимость к нуклеосомному комплексу ДНК-белок. В Разделе 3.2.1 приведенных данных МД явно недостаточно, чтобы делать вывод о наличии движения «расщепления

нуклеосомы» - длина траектории для этого слишком мала и, кроме того, нет сведений о воспроизводимости результатов в независимых стартах МД. В Разделе 3.2.3 не хватает важных технических деталей построения полноатомной модели нуклеосомы с удлиненными линкерными участками ДНК. Учитывая большой размер системы и относительно небольшие (~ 1 мкс) времена МД в явно заданной среде, результаты расчета, очевидно, сильно зависят от выбора стартовой конфигурации системы. Это же относится и к ряду других вычислительных МД-экспериментов! К сожалению, вопросы сходимости и устойчивости решений автор не обсуждает.

2. Выводы, сделанные в Разделе 3.2.3 о поведении модели с искусственно добавленными линкерными участками ДНК, достаточно тривиальны – такие участки ДНК оказались более подвижными, чем связанные с гистонами. Но что нового это дает? Насколько подробно при этом было исследовано конфигурационное фазовое пространство (см. замечание выше)?

3. На мой взгляд, нуждается в пояснениях и протокол расчета энергии деформации ($E_{def.}$) линкерной ДНК в двух моделях связывания – «на диаде» (НД) и «вне диады» (ВД). Во-первых, как именно оценивали $E_{def.}$? Лишь для нуклеотидов линкера и с использованием силового поля в обобщенных координатах или вклад соседних фрагментов белка и ДНК также учитывали? Но в каком силовом поле? Насколько адекватной является ВД-модель – по утверждению автора она была построена с помощью молекулярного докинга? Что значит «Z-оценка» положения последовательности? Опять возникает вопрос об устойчивости и единственности полученных решений!

4. Детали построения модели нуклеосомы с линкерным гистоном описаны поверхностно. Так, не указано, сколько и каких ограничений на расстояния было использовано при минимизации энергии системы. Как ограничения были получены из анализа экспериментальных данных? От ответа на эти вопросы зависит и качество итоговой модели. Кроме того, неясно, как автор справился с проблемой преодоления множественных локальных минимумов при решении задачи оптимизации для системы со столь значительным числом степеней

свободы? Одним из выводов данного раздела является то, что «модель схожа по конфигурации линкеров с моделью, построенной в работе [72]», однако никаких количественных критериев «схожести» не приводится.

5. Бóльшая часть Раздела 4.4 посвящена обсуждению проблемы определения позиционирования ДНК на нуклеосомах, а собственно результаты настоящего исследования представлены лишь в трех предложениях (стр. 70), без необходимых для понимания качества результата деталей. Как в других – отмеченных выше – случаях, подобная лаконичность зачастую не позволяет Читателю оценить надежность/корректность выводов, а заинтересованным коллегам – самостоятельно воспроизвести эти результаты.

6. На стр. 41 автор утверждает: «… при нормальных условиях для биологических систем зеселен относительно небольшой диапазон микросостояний. По этой причине зачастую достаточно изучить лишь состояния, близкие к кристаллическим». Этот вывод мне представляется спорным, особенно учитывая, что эффекты среды часто играют ключевую роль и, следовательно, состояния в кристалле могут отличаться от нативных.

7. На мой взгляд, выводы 1-3 (стр. 83) сформулированы расплывчато – их сложно понять без дополнительного ознакомления с материалами работы.

Технические замечания:

8. Поскольку структура и динамика ДНК и ее комплексов с белками существенно зависят от корректного учета электростатических взаимодействий в среде, вопрос о расположении ионов в стартовой структуре и об ионизационном состоянии функциональных групп в нуклеосоме является крайне важным. Считаю, что этому должно быть уделено особое внимание в Разделе «Методы», но этой информации в работе нет.

9. Значения квантового выхода для красителей Cy3 и Cy5 измеряли по разным реперным соединениям, причем находящимся в разных средах (стр. 51). Насколько это корректно, учитывая, что указанные данные потом применяются для расчета отношения величин квантовых выходов?

10. В формуле (11) не поясняется, чему соответствуют переменные ε_{ij} и μ_i.

11. Оценка эффективности FRET (стр. 47) произведена в расчете МД меток, прикрепленных с помощью подвижных линкеров к неподвижному остову фрагмента ДНК. Полученный при этом «ансамбль потенциальных положений Cy3 и Cy5 относительно меченых нуклеотидов» является сильно упрощенным, т.к. в водном растворе остов может обладать высокой конформационной подвижностью. Насколько учет этих эффектов повлияет на результаты интерпретации данных FRET?

К недостаткам работы относятся и некоторые погрешности оформления. Так, автор использует ряд неудачных, жаргонных и некорректных выражений, например: «ДНК отрицательно сверхспирализована» (стр. 13), «времена прилета фотонов интегрируются по времени программным обеспечением микроскопа» (стр. 27), «отношение флуоресценции акцептора к полной флуоресценции» (стр. 28), «фотоотбеливание» (стр. 32) – по-видимому, речь идет о фотообесцвечивании (photobleaching ?), «пробег пика» (стр. 38), «генерализованные силовые поля» (стр. 40), «замедлять … ускорять скорость» (стр. 41), «вода и макромолекула термостатировались раздельно» (стр. 44), «квантовые выходы красок, прикрепленных к одноцепочечным олигонуклеотидам» (стр. 51), «наиболее амплитудный… вектор» (стр. 57) и пр.

Имеются погрешности в оформлении рисунков. Подписи к некоторым из них (например, 2, 6 (чему соответствует цветовая гамма?), 7 (что означает «изображение из открытых источников»?)) недостаточно информативны, в них не расшифровываются все детали, необходимые для понимания рисунка. Качество некоторых рисунков низкое, мелкие детали плохо видны – например, на рис. 7.

Вместе с тем, отмеченные недостатки не снижают моей высокой оценки работы Армеева Г.А. Высказанные замечания носят рекомендательный характер и служат для того, чтобы подчеркнуть сложность поставленной задачи, которая

была успешно решена автором. Работа выполнена на высоком методическом уровне, содержит новые интересные научные данные, хорошо оформлена и легко читается. Полученные автором результаты и разработанное программное обеспечение (находящееся в открытом доступе), наряду с богатым справочным материалом, несомненно, будут полезны не только исследователям, занимающимся биоинформатикой и структурной биологией нуклеиновых кислот и ДНК-белковых комплексов, но и специалистам в других областях – молекулярной биофизике, медицинской химии, биоинженерии. Подобные лаборатории и группы существуют в ВУЗах и научно-исследовательских организациях – на ряде естественнонаучных факультетов МГУ им. М.В. Ломоносова, СПбГУ и др., в Институте органической химии им. Н.Д. Зелинского РАН, в Институте биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова РАН, в Институте биомедицинской химии им. В.Н. Ореховича РАН, в Институте химической биологии и фундаментальной медицины СО РАН и др.

Выносимые на защиту результаты опубликованы в виде пяти статей в реферируемых международных журналах и пяти статей в профильных российских журналах (все издания индексируются в библиографических базах данных Web of Science / Scopus / RSCI), докладывались на российских и международных научных конференциях. Автореферат правильно отражает содержание диссертации. Диссертация отвечает требованиям, установленным Московским государственным университетом имени М.В. Ломоносова к работам подобного рода. Содержание диссертации соответствует паспорту специальности 03.01.09 – «математическая биология, биоинформатика» (по физико-математическим наукам), а также критериям, определенным пп. 2.1-2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова. Диссертация оформлена, согласно приложениям № 5, 6 Положения о диссертационном совете Московского государственного университета имени М.В. Ломоносова.

Таким образом, соискатель Армеев Григорий Алексеевич заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 03.01.09 – «математическая биология, биоинформатика».

Официальный оппонент:

доктор физико-математических наук,
Главный научный сотрудник
Лаборатории моделирования биомолекулярных систем
Федерального государственного бюджетного учреждения науки
Институт биоорганической химии
им. академиков М.М. Шемякина и Ю.А. Овчинникова
Российской академии наук (ИБХ РАН)

ЕФРЕМОВ Роман Гербертович

28 ноября 2018 г.

Контактные данные:
тел.: 7(903)7431656, e-mail: r-efremov@yandex.ru
Специальность, по которой официальным оппонентом
защищена диссертация:
03.01.02 – «биофизика»

Адрес места работы:
117997, г. Москва, ул. Миклухо-Маклая, д. 16/10,
Федеральное государственное бюджетное учреждение науки
Институт биоорганической химии
им. академиков М.М. Шемякина и Ю.А. Овчинникова
Российской академии наук (ИБХ РАН),
Лаборатория моделирования биомолекулярных систем

Тел.: 7(495)3305874; e-mail: efremov@nmr.ru

Подпись проф. Р.Г. Ефремова удостоверяю:

Ученый секретарь ИБХ РАН
д.ф.-м.н. В.А. Олейников

28.11.2018 г.