

ОТЗЫВ официального оппонента

на диссертацию на соискание ученой степени

доктора физико-математических наук Шайтана Алексея Константиновича

на тему: «ИНТЕГРАТИВНОЕ МОДЕЛИРОВАНИЕ СТРУКТУРЫ И ДИНАМИКИ

БИОМАКРОМОЛЕКУЛЯРНЫХ КОМПЛЕКСОВ»

по специальности 03.01.09 – «математическая биология, биоинформатика»

Современное состояние исследований в области биомедицины, или в «науках о живом», требует понимания на молекулярном уровне процессов и явлений, происходящих в клетке. Более того, речь идет как о нормальном состоянии изучаемых систем, так и о развитии патологий различной природы, но обусловленных неправильной работой конкретных молекулярных/надмолекулярных объектов. Поэтому детальный анализ структуры, динамики и функционирования указанных систем – потенциальных будущих фармакологических мишеней – необходим для разработки соответствующих лекарств и средств терапии. Отметим, что в последние годы был достигнут значительный прогресс в развитии экспериментальных подходов, начиная со ставших уже привычными методов структурной биологии больших макромолекулярных комплексов (рентгеноструктурный анализ, крио-электронная микроскопия) и кончая физическими методами сверхвысокого разрешения, в т.ч. позволяющими наблюдать за одиночными молекулами в клетке. Безусловно, эта информация дает качественно новый уровень понимания устройства и работы клеточных молекулярных машин – сложнейших комплексов, включающих белки, нуклеиновые кислоты, липидные мембраны и т.д. Однако в большинстве случаев получаемая картина статична и, более того, представляет из себя среднее по ансамблю состояний, тогда как для понимания клеточных процессов часто требуется установить роль лишь небольшого числа наиболее важных молекулярных «актеров» - именно локальные эффекты могут критическим образом повлиять на интегральное поведение макроскопической системы, например, сдвигая в них химическое равновесие. Иными словами, для воссоздания полной картины сложнейших клеточных процессов необходимо дополнить экспериментальные модели результатами атомистического компьютерного моделирования.

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 10^7 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 отдельных белков и/или их олигомеров сравнительно невысокого порядка (n) (как правило, $n = 2 - 6$), 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. Кроме того, методами *in silico* мало изучены олигомеры макромолекул высоких порядков ($n > 10$). Ярким примером служат амилоидоподобные фибриллы, формирование которых обуславливает развитие т.н. «конформационных» заболеваний, включая болезнь Альцгеймера, Паркинсона и др.

Диссертационная работа Шайтана А.К. посвящена созданию и применению новых эффективных вычислительных подходов различного уровня сложности к изучению структуры, динамики и механизмов работы сложнейших надмолекулярных систем. Среди них - нуклеосомы и их комплексы с белками хроматина; агрегаты, образованные амилоидоподобными пептидами. Основная идея работы – на примере этих важнейших с точки зрения биомедицины объектов продемонстрировать возможности современных технологий компьютерного молекулярного моделирования – т.н. интегративного моделирования. Его особенность заключается в проведении вычислений с привлечением комплекса наиболее мощных компьютерных методов, физических моделей и алгоритмов, адекватных рассматриваемым объектам, а также их неразрывная связь и корректировка с учетом всего многообразия доступных экспериментальных данных. Как уже отмечалось выше, такой подход в настоящее время является наиболее адекватным и востребованным. Это, несомненно, свидетельствует об актуальности темы диссертации и ее научной значимости.

Диссертация изложена на 488 страницах и включает: Введение (в т.ч. список публикаций автора по теме диссертации); пять глав, каждая из которых содержит по

соответствующей теме: литературный обзор, изложение результатов работы, их обсуждение, а также описание материалов и методов; Заключение; Список сокращений и условных обозначений; Список литературы (677 источников). Диссертация содержит 106 рисунков и 9 таблиц.

Во Введении обоснована актуальность темы диссертации, сформулированы цель и задачи исследования, описаны научная новизна, теоретическая и практическая значимость работы, представлены объекты, методология и методы исследования, а также степень достоверности результатов. Сформулированы положения, выносимые на защиту. Кроме того, представлены сведения об апробации диссертационной работы, о публикациях соискателя и о его личном вкладе в настоящее исследование. В Главе 1 автор обсуждает современное состояние методов молекулярного моделирования применительно к решению биомедицинских задач, формулирует концепцию интегративного моделирования (ИМ), излагает кратко основные принципы метода молекулярной динамики (МД) и используемые в нем различные уровни приближений. Вторая глава посвящена описанию результатов проведенных автором вычислительных экспериментов, направленных на изучение структуры и динамики нуклеосом методами ИМ. Продемонстрирована высокая эффективность предложенных в диссертационной работе подходов *in silico* к анализу факторов, определяющих работу нуклеосом в клетке. Важно, что результаты моделирования были уточнены с привлечением большого набора экспериментальных данных, что и соответствует духу ИМ-технологии.

Считаю, что основными результатами этой главы являются следующие: 1) построены реалистичные (т.е. согласующиеся с данными экспериментов) атомистические модели нуклеосом, продемонстрирована их стабильность в ходе МД в воде на протяжении свыше 10 мкс; 2) Сделан вывод о ключевой роли электростатических взаимодействий, определяющих динамику ДНК в нуклеосомах; 3) Выявлены молекулярные особенности взаимодействия ДНК-гистоны, установлена важная роль флуктуаций в поведении системы; 4) Сформулирована гипотеза о наличии в нуклеосоме крупномасштабных структурных перестроек в микросекундном диапазоне, что, вероятно, указывает на возможность «скольжения нуклеосом вдоль ДНК по винтовому механизму». Отмечу, что наличие таких моделей и полученные результаты их динамики позволят в будущем облегчить интерпретацию экспериментальных данных низкого разрешения и предложить возможные механизмы работы нуклеосом.

В Главе 3 описаны результаты применения созданного комбинированного – экспериментального и теоретического – подходов к построению модели наиболее энергетически выгодного расположения нуклеосомы на большом участке ДНК. Поскольку одними лишь методами моделирования (в частности, МД) данная задача не решается, автором

был предложен ИМ-подход, сочетающий применение вычислительных методов с экспериментальными. В качестве последних были успешно применены данные по измерению гидроксильного расщепления ДНК (футпринтингу). Для решения этой сложной проблемы под руководством соискателя был разработан целый ряд оригинальных программных средств, прежде всего, пакет HYDROID. Важно, что эта программа описана в статье, опубликованной в высокорейтинговом журнале Nature Protocols, что свидетельствует о мировом признании разработки. В ходе проведенных исследований были получены следующие фундаментальные результаты: (1) Показана высокая эффективность совместного применения футпринтинга ДНК и молекулярного моделирования к построению атомистических моделей ДНК-белковых комплексов большого размера, При этом точность позиционирования ДНК в нуклеосоме достигает одной пары нуклеотидов. (2) Выявлены основные физические взаимодействия, обуславливающие формирование и дальнейшее функционирование нуклеосом.

Развивая тему на еще более сложные объекты, в Главе 4 автор рассматривает ИМ-подходы к моделированию комплексов нуклеосом с белками хроматина. В этом случае, помимо экспериментальных данных футпринтинга ДНК, используются результаты электронной микроскопии и FRET-микроскопии. Показано, что с помощью предложенного комплексного подхода удастся изучать пространственную супрануклеосомную структуру хроматина. На сегодняшний день таких работ в мире крайне мало. Таким образом, автор является одним из первопроходцев в этой предметной области. Несомненно, еще предстоит ответить на многие вопросы, в том числе связанные и с техникой расчетов, используемыми приближениям, оценкой воспроизводимости результатов. Но это – начало пути, которое позволяет подключать к работе все новые коллективы и ресурсы, уточнять модели первого уровня и т.д.

В пятой главе продемонстрированы возможности ИМ-подхода к моделированию структуры и динамики другого класса мезоскопических биомолекулярных комплексов – агрегатов амилоидоподобных фибрилл. В результате автором были предложены атомистические модели указанных сложнейших объектов, образующихся при самосборке олигопептидов. Поскольку при их моделировании необходимо использовать допущения и приближения различного уровня, важнейшая задача – проверка и корректировка результатов расчетов путем их согласования с экспериментальными данными (оптической спектроскопии, рентгеновской дифракции, атомно-силовой микроскопии и т.д.). Шайтан А.К. успешно справился с этой задачей – в работе предложены реалистичные модели фибрилл из диблок олигомеров тетрапептида-(Thr-Val)₃-пептидов и амилоидоподобных фибрилл EF-C на базе 12-членного пептида из белка gp41 HIV-1. Важно, что предложенные модели необходимы для дальнейшего уточнения и использования с целью установления детальных молекулярных

механизмов формирования фибрилл, что сопровождается как нормальные физиологические процессы в клетке, так и развитие опасных патологий.

Как отмечалось выше, вместо единого обзора литературы автор предпочел представить современное состояние проблемы в каждой из пяти глав, посвященных конкретной теме исследования. Учитывая широкий круг рассматриваемых в работе независимых задач, считаю это решение целесообразным, поскольку оно дает возможность сосредоточиться на анализе конкретной темы диссертационной работы. В своей совокупности этот материал является высокоинформативным и хорошо структурированным. Он четко изложен и прекрасно иллюстрирован. Это, несомненно, является хорошим подспорьем для научных сотрудников, аспирантов и студентов, причем как опытных, так и 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.

Строго говоря, идеология интегративного моделирования (ИМ) не является новой (хотя само ее название появилось недавно). В большинстве серьезных работ последних лет, посвященных вычислительному анализу конкретных биомолекулярных ансамблей, авторы вынуждены были применять в качестве ограничений тот или иной набор экспериментальных данных, часто взятых из разнородных источников. Учитывая огромное число степеней свободы этих систем, трудности корректного задания стартовых конфигураций в моделировании, эмпирический характер силовых полей (со всеми вытекающими последствиями), отсутствие высокоэффективных методов исследования фазового пространства и пр., применение в качестве «реперных точек» фактов, известных из экспериментов, является совершенно необходимым в современных проектах в области структурной молекулярной биофизики. Вместе с тем, для каждого класса биологических объектов методы ИМ имеют свои особенности и, более того, для ряда объектов они ранее практически не были предложены, либо имели ограниченные возможности. Заслугой автора диссертационной работы является расширение границ применимости ИМ-технологии к двум очень интересным и востребованным классам систем – нуклеосомам и амилоидным агрегатам. Шайтану А.К. не только удалось на нескольких конкретных примерах показать надежность ИМ-подхода к описанию указанных сложнейших систем и его преимущество перед стандартными методами моделирования, но и критически проанализировать достоинства и недостатки новой методологии, создать оригинальные средства разработки и применения ИМ и выбрать из уже известных оптимальные на сегодняшний день, а также обобщить

полученный опыт и дать детальные четкие рекомендации коллегам-экспериментаторам. Поэтому считаю, что основное требование, предъявляемое к докторским диссертациям, - создание нового направления в научных исследованиях по конкретной тематике – выполнено.

Важнейший показатель высокого научного уровня диссертационной работы – список публикаций автора по данной теме. Он впечатляет, поскольку статьи Шайтана А.К. (причем в большинстве из них он является первым/последним автором) вышли в высокорейтинговых профильных изданиях, включая Nature Protocols, Nature Structural and Molecular Biology, Nucleic Acids Research, Cell Reports и др. Таким образом, все основные результаты работы прошли тщательное рецензирование и, я уверен, обсуждение с редакторами и анонимными рецензентами. Поэтому у меня нет сомнений в значимости и корректности опубликованных результатов. Именно поэтому некоторые отмеченные ниже погрешности изложения материала, по-видимому, объясняются желанием автора резюмировать в диссертации большое число опубликованных данных, не вдаваясь во многие технические детали.

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. Материал, изложенный в Главе 2, представляет из себя набор модулей, каждый из которых соответствует опубликованным автором одной или нескольким работам по конкретной тематике. Считаю такой формат не слишком удачным с точки зрения целостного восприятия материала, поскольку неизбежно возникают повторы, например, при описании природы и устройства нуклеосом, их роли и пр., изложении методических подходов и пр.

2. Недостатком работы, посвященной моделированию нуклеосом (Глава 2), является излишне краткое, подчас неинформативное описание вычислительных процедур и ряда полученных результатов. Кроме того, мало внимания уделяется обсуждению погрешностей моделирования и возможной чувствительности результатов молекулярной динамики (МД) к выбору исходных конфигураций рассматриваемых сложнейших объектов. В частности, расчеты МД проведены для ряда систем, содержащих изменения в исходной кристаллографической модели нуклеосомы, но при этом автор не поясняет, как подобные модификации структуры влияют на результаты МД. Тем более, что представленные здесь же данные показывают, что итоговые выводы могут измениться при выборе другого варианта «вмешательства» в модель нуклеосомы. Наконец, при обсуждении некоторых результатов вместо их исчерпывающего описания даны ссылки на опубликованные статьи автора и на интернет-ресурсы, что затрудняет анализ диссертационной работы.

3. Как справедливо отмечает автор (и в этом – суть ИМ-подхода!), без привлечения в качестве ограничений экспериментальных данных компьютерное моделирование таких сложных систем как нуклеосомы не позволяет пока достичь надежных результатов. В связи с этим особое внимание следовало бы уделить именно вопросу уточнения конкретных результатов расчетов в зависимости от данных экспериментов. На мой взгляд, было бы полезно показать в сравнении, что дает стандартный расчет, а что – ИМ-подход.
4. Как и в Главе 2, материал Главы 3 изложен в большой степени независимыми блоками, соответствующими публикациям по конкретной тематике. В результате имеются повторы в тексте, что затрудняет чтение.
5. При описании работы оператора с программой HYDROID часто упоминаются моменты, критическим (!) образом способные повлиять на результат. Кроме того, встречаются операции, которые необходимо проводить в «ручном» режиме. В какой степени это влияет на результаты профилирования и насколько программное обеспечение может быть использовано сторонними специалистами? Это важно с точки зрения как точности расчета, так и масштабирования программы и ее внедрения в практику исследований коллег, работающих в предметной области.
6. При построении стартовых моделей амилоидоподобных фибрилл автор использует большое число произвольно (практически «на глаз») выбранных преобразований бета-структурных тяжей. Критериями в данном случае служит число водородных связей, отсутствие стерических наталкиваний в системе, «плотность упаковки» тиофеновых фрагментов и т.д. Насколько адекватны подобные конфигурации? Времена полноатомного моделирования МД составляют всего 10 нс, что, конечно, недостаточно для уравнивания системы. Были ли предприняты попытки моделирования МД с независимых стартов? В какой степени результаты чувствительны к выбору исходной конфигурации? Пробовали ли делать выбор оптимальных состояний, основываясь на оценках свободной энергии системы?
7. При анализе паттернов водородных связей в фибриллах (Глава 5) речь идет лишь о взаимодействиях белок-белок, а водородные связи с растворителем и в целом параметры гидратации фибрилл не обсуждаются. Вместе с тем, эти эффекты могут иметь важное значение для сборки и стабильности пептидных агрегатов.
8. При обсуждении результатов моделирования (в частности, из пептидов EF-C) в качестве экспериментальных данных сравнения приводится лишь фотография профиля фибриллы, полученная методом атомно-силовой микроскопии (Рис. 5.37в). При этом автор утверждает, что «рассчитанная 2D картина дифракции фибрилл» согласуется с экспериментальными данными. Однако деталей такого согласия не приводится, при этом паттерны на Рис. 5.37в и 5.38г довольно сильно отличаются друг от друга.

Технические замечания:

9. В Главе 3 дано, на мой взгляд, излишне подробное описание программы HYDROID – местами оно, по сути, представляет собой руководство пользователя. Целесообразно было бы эти сведения дать либо в виде приложения, либо просто сослаться на соответствующий интернет-ресурс (ссылки на него и так есть).

10. В Главе 5 встречаются фрагменты текста, по-видимому, переведенные с английского языка с помощью машинных средств. В частности, об этом свидетельствуют выражения: «производственный МД прогон» (стр. 343, возможно, «калька» с productive MD run?), «тетрамер был помещен в коробку для моделирования» (стр. 382) и др.

К недостаткам работы относятся и некоторые погрешности оформления. Так, автор использует ряд неудачных, жаргонных и некорректных выражений, например: «межмолекулярная укладка пептидов» (стр. 14), «рентгеновская структура», «способность нуклеосом претерпевать определенные типы конформационных переходов разумно используется ... белками» (стр. 90), «полностью завернутое состояние» (стр. 99), «сложное динамическое взаимодействие между посттрансляционными модификациями гистонов» (стр. 128), «кристаллографические ионы» (стр. 129), «продвинутые навыки в языке Python» (стр. 190), «взаимодействие между внутренней геометрией... и геометрией, наложенной на ДНК...» (стр. 215) и пр. На стр. 341 вместо термина «постоянная длина» автор, по-видимому, имеет в виду персистентную длину.

Вместе с тем, отмеченные недостатки не снижают моей высокой оценки работы Шайтана А.К. 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, ДНК-белковых комплексов и фибриллярных пептидных агрегатов, 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 химической биологии и фундаментальной медицины СО 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 цитологии и генетики СО РАН и др.

Выносимые на защиту результаты опубликованы в виде 35 статей в журналах, индексируемых в международных библиографических базах данных Web of Science / Scopus, 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 Шайтан Алексей Константинович заслуживает присуждения ученой степени доктора физико-математических наук по специальности 03.01.09 – «Математическая биология, биоинформатика».

Официальный оппонент:

доктор физико-математических наук, профессор,

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

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. Ефре

Ученый секретарь ИБХ Р

д.ф.-м.н.

В.А. Олейников

23.03.2021 г.