

## Молекулярная динамика в силовом поле FF14SB в воде TIP4P-Ew, и в силовом поле FF15IPQ в воде SPC/E<sub>b</sub>: сравнительный анализ на GPU и CPU\*

Д.А. Суплатов<sup>1</sup>, Я.А. Шарапова<sup>2</sup>, Н.Н. Попова<sup>3</sup>, К.Е. Копылов<sup>2</sup>,  
Вл.В. Воеводин<sup>3,4</sup>, В.К. Швядас<sup>1,2</sup>

НИИ ФХБ имени А.Н.Белозерского<sup>1</sup>,  
Факультет биоинженерии и биоинформатики<sup>2</sup>  
Факультет вычислительной математики и кибернетики<sup>3</sup>, НИВЦ<sup>4</sup>  
МГУ имени М.В.Ломоносова

Проведен сравнительный анализ вычислительной эффективности и масштабируемости молекулярной динамики (МД), реализованной в пакете AMBER, на реальных биологических системах с применением классического силового поля FF14SB с 4-х центровой моделью воды TIP4P-Ew, а также нового многообещающего поля FF15IPQ с 3-х центровой моделью воды SPC/E<sub>b</sub>. Уменьшение количества атомов в ячейке на 25-31% в результате использования 3-х центровой модели растворителя ускоряет расчет МД до 63% и ухудшает масштабируемость до 11%. При этом полученные результаты могут качественно отличаться, что говорит о необходимости совместного использования разных силовых полей при изучении биологических систем. Использование GPU-ускорителей позволяет существенно увеличить длину траектории.

*Ключевые слова:* классическая молекулярная динамика, AMBER, силовые поля FF14SB и FF15IPQ, модели воды TIP4P-Ew, TIP3P и SPC/E<sub>b</sub>, GPU-ускорители.

### 1. Введение

Молекулярная динамика — собирательное название семейства компьютерных алгоритмов для изучения структуры и функции биологических макромолекул, которые образуют один из ключевых методов современной биологии [1, 2]. Несмотря на ряд ограничений этого подхода не вызывает сомнений его полезность при решении широкого спектра научных задач — изучения стабильности, конформационных перестроек в белках, взаимодействия белков с лигандами, процесса формирования нативной пространственной структуры (фолдинга) белка и др. Роль молекулярной динамики в науке была окончательно признана в 2013 году присуждением Нобелевской премии Мартину Карплюс, Майклу Левитт и Ариелю Варшелу за «разработку многомерных моделей сложных химических систем».

Применение метода молекулярной динамики требует мощного вычислительного ресурса. Бурное развитие компьютерных технологий и совершенствование самого метода в последние годы способствовало его распространению и повседневному использованию в научной практике. Нужно отметить, что речь идет не столько об увеличении мощности компьютеров в целом, сколько о появлении специальных вычислительных устройств, ориентированных на решение конкретных научных задач. В 2008 году частная американская компания D.E. Shaw Research объявила о создании специализированного суперкомпьютера Anton, спроектированного на аппаратном уровне для решения единственной вычислительной задачи — расчета длинных молекулярно-динамических траекторий [3]. В 2014 году появилась улучшенная версия этой машины — Anton 2 [4]. С другой стороны, революцией стала разработка GPU-ускорителей (от англ. graphics processing unit) как эффективной альтернативы классическим процессорам при решении широкого спектра научных задач. Разработка и имплементация компьютерных алгоритмов для исполнения на GPU ускорителях, прежде всего ре-имплементация ранее предложенных

\* Работа выполнена при финансовой поддержке Российского Фонда Фундаментальных Исследований (грант №17-07-00751) и Российского Научного Фонда (грант №15-14-00069-П)

классических CPU кодов полезных алгоритмов, постепенно становится преобладающей стратегией для ускорения сложных вычислений в компьютерной биологии. Эта тенденция, которая раньше продвигалась в основном разработчиками GPU-оборудования, в последние годы твердо закрепилась в фундаментальной биологической науке [5]. Преимущества GPU заключаются в сочетании аппаратной универсальности, высокой мощности вычислительного устройства и относительно невысокой стоимости. Современные видеокарты могут быть установлены даже в индивидуальные рабочие станции и предоставляют значительное ускорение расчетов в ежедневной работе. К настоящему моменту все основные методы молекулярного моделирования, в том числе наиболее распространенный подход классической молекулярной динамики в периодической системе с явно заданным растворителем, частично или полностью реализованы на GPU [5].

Потребность молекулярного моделирования в больших ресурсах имеет конкретные причины. При использовании молекулярной динамики важна не столько скорость получения результатов, сколько их «правильность», т.е. способность метода адекватно описывать реальные процессы, происходящие в живых системах на уровне отдельных атомов. Повышение точности расчетов может быть достигнуто благодаря установке температуры, оптимальной для работы белков; увеличению времени наблюдения за системой (длины траектории); использованию более сложной молекулярной модели воды, которая могла бы более правильно воспроизводить физико-химические характеристики растворителя (например, 4-х центровой модели TIP4P-Ew вместо 3-х центровой TIP3P); уменьшению шага интегрирования уравнения движения; усложнению электростатических расчетов (увеличению радиуса учета нековалентных взаимодействий и отключению псевдо-структур, которые представляют типовые химические блоки в упрощенном виде, например, редуцируют сопряженные многоатомные системы до нескольких псевдо-атомов); увеличению отступа между поверхностью белка и краем молекулярной ячейки для исключения артефактов в периодической системе и др. Однако применение этих настроек приводит к существенному увеличению вычислительной сложности задачи. Поэтому для ускорения молекулярной динамики приходилось жертвовать точностью вычислений. Так, использование экстремальных температур (100-1000°C) позволяло ускорить структурные изменения в белках и уменьшить время наблюдения, однако существенно повышало вероятность появления артефактов. Именно поэтому появление доступных GPU-ускорителей и совместимых приложений молекулярной динамики было встречено с таким энтузиазмом – они позволяют увеличить время наблюдения за биомолекулярной системой (в среднем с наносекунд до микросекунд) при адекватных параметрах алгоритма (в том числе, физиологических значений температуры).

Перед тем, как исследовать свой объект с использованием молекулярной динамики необходимо определиться по следующим ключевым вопросам – (1) выбрать программу (метод), (2) силовое поле и (3) модель растворителя для проведения вычислительного эксперимента. Конечно, есть и другие параметры, значения которых необходимо подбирать под конкретную задачу (например, размер ячейки с растворителем, о чем будет более подробно сказано ниже), однако, указанные три пункта будут иметь наиболее существенное влияние на вычислительную сложность эксперимента и биологическую интерпретацию его результатов.

В настоящий момент существует множество пакетов программ для молекулярной динамики, наиболее известными из которых являются AMBER, CHARMM, GROMACS и NAMD. Каждый из этих продуктов обладает своими преимуществами и существенно отличается от других – по набору доступных методов, параметров (в том числе силовых полей), средств анализа результатов, доступности исходного кода, и др. При том, что выбор программного продукта для молекулярной динамики диктуется конкретной задачей пакет AMBER [6, 7], возможно, является оптимальным с точки зрения сочетания доступности и функциональности для решения широкого круга задач. Пакет программ AMBER богат на реализации различных биологических алгоритмов – классической и ускоренной молекулярной динамики, метадинамики, молекулярной динамики в вакууме, неявно заданном растворителе, методов QM/MM (сочетание классической динамики и методов квантовой химии) и т.д. Некоторые из них доступны в двух версиях – для исполнения на GPU или CPU. Важным преимуществом AMBER является наличие дополнительного пакета AMBERTOOLS с широким набором силовых полей, моделей растворителя, а также различных программ для подготовки биологической системы к молекулярной ди-

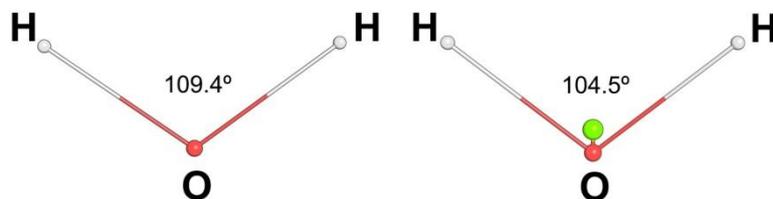
намике и анализа результатов. В частности, пользователям AMBER доступно программное обеспечение для «параметризации» новых молекул, которые отсутствуют в силовом поле по умолчанию (а это практически все низкомолекулярные соединения), например, субстратов ферментативных реакций или ингибиторов ключевых ферментов метаболизма патогенных бактерий, что значительно расширяет возможности метода молекулярной динамики для изучения живых систем.

Силовые поля – это функции расчета энергии движения и взаимодействия атомов, а также наборы параметров для этих функций, характеризующих, в том числе, массу и заряд атомов, длины, углы и двугранные углы ковалентных взаимодействий атомов друг с другом, и т.д. Силовые поля AMBER предназначены для оригинальных программ из пакета AMBER и, пожалуй, являются наиболее известными и широко используемыми в повседневной практике. Нужно отметить, что силовые поля AMBER в ограниченном виде поддерживаются программными продуктами сторонних производителей – например, пакет GROMACS версии 2018 года поддерживает устаревшие силовые поля AMBER, разработанные до 2003 года (FF94, FF96, FF99, FF99SB, FF99SB-ILDN, FF03 и FFGS) [8]. Силовые поля являются сложными аппроксимациями, предназначенными для описания динамических процессов в биологических системах, которые постоянно развиваются и обновляются для повышения точности расчетов. На момент написания статьи силовое поле FF14SB [9] является основным рекомендуемым полем для изучения динамики белков в пакете программ AMBER [10]. Для изучения биологических макромолекул небелковой природы (нуклеиновых кислот, липидов и т.д.) специально разработаны отдельные силовые поля, которые можно использовать в комбинации с полями для белков. Поле FF14SB является улучшенной (оптимизированной) версией силового поля FF99SB, полученного «классическим» способом. Процесс создания силовых полей заключается в выборе таких значений всех параметров, характеризующих сложную биологическую систему (зарядов, масс атомов, длин, углов и жесткости всех возможных связей и т.д.), которые позволяют получать результаты, в наибольшей степени согласующиеся с экспериментальными данными. Проблема в том, что таких параметров очень много. В этой связи, как правило, использовалась итеративная оптимизация только определенного подмножества параметров, например, ре-оптимизации параметров, характеризующих торсионные углы при заранее заданных и фиксированных зарядах атомов. Результат такого «классического» подхода, очевидно, зависит от точности неоптимизированных (фиксированных) параметров. Для преодоления этих ограничений были предложены новые подходы, позволяющие в полуавтоматическом режиме оптимизировать большое количество параметров силового поля. Одним из таких подходов является метод Implicitly Polarized Charge (IPolQ), с помощью которого было разработано поле FF14IPQ и его более новая версия FF15IPQ [11]. Разработчиками было отмечено, что FF15IPQ позволяет более точно моделировать ионные взаимодействия между противоположно заряженными аминокислотными остатками, вторичные структуры и глобулярные укладки белков, а также в целом предназначено для изучения биологических систем при расчете длинных траекторий. Поскольку одним из ключевых преимуществ современных GPU-ускорителей является возможность существенно увеличить время наблюдения за системой в повседневной практике, заявленные возможности силового поля FF15IPQ для изучения белков представляются интересными и актуальными.

Наконец, необходимо определиться со средой, в которой будет проведен эксперимент по моделированию. Она может отсутствовать, в таком случае расчет будет проводиться в вакууме. Такой подход используют при создании сложных моделей белков – «прижимая» две части структуры в вакууме можно индуцировать комплементарность между ними и таким образом реконструировать, например, интерфейс взаимодействия между доменами. Расчет молекулярной динамики без растворителя можно произвести, не имея значительных вычислительных ресурсов, и его результаты могут быть полезны в отдельных случаях, однако для описания работы белков в физиологических условиях этот вариант, очевидно, не подходит. Существует отдельный класс методов для молекулярного моделирования в неявно заданном растворителе, которые применяют для ускорения расчетов при изучении очень больших биомолекулярных систем, например, белков в составе липидной мембраны, однако эти подходы имеют ряд существенных ограничений [12]. Наиболее распространенным подходом к изучению динамики биологических макромолекул является использование явно заданного растворителя в роли которо-

го, как правило, выступает вода. Растворитель играет важную роль в процессах, происходящих в живой клетке, поэтому модели воды для молекулярной динамики, как и силовые поля, постоянно развиваются и обновляются для повышения точности расчетов. Существует множество моделей воды, которые предназначены для использования в разных силовых полях, по-разному воспроизводят физико-химические свойства воды, и отличаются геометрией, зарядами, массами атомов, а также их количеством. Наиболее распространенными сегодня являются 3-х центровые модели воды (Рис. 1, слева). Так, вместе с силовым полем FF14SB чаще всего используют модель воды TIP3P, состоящую, как и настоящая вода, из трех атомов – кислорода (O) и двух атомов водорода (H). В попытке улучшить точность моделирования были предложены 4-х центровые модели воды, содержащие один дополнительный псевдоатом. Для расчетов в периодической системе с явно заданным растворителем была разработана модель воды TIP4P-Ew (Рис. 1, справа). Было показано, что она лучше описывает вращательное движение белков по сравнению с обычно используемой 3-х центровой моделью TIP3P [13], а также более корректно воспроизводит физические свойства жидкого растворителя при физиологических значениях температуры и давления [14]. Добавление псевдоатомов в модель растворителя может помочь более точно оценить поведение белка в воде, однако существенно увеличивает количество и плотность атомов в ячейке, что сказывается на вычислительной сложности эксперимента. В этом контексте представляет интерес 3-х центровая модель воды SPC/E<sub>b</sub> (Рис. 1, слева), которая совместима с ранее упомянутым перспективным силовым полем FF15IPQ, а по своим характеристикам не уступает либо превосходит доступные аналоги [15]. Существуют также 2-х, 5-ти и 6-ти центровые модели воды, однако на практике используются значительно реже.

В этой работе метод классической молекулярной динамики в периодической системе с явно заданным растворителем, реализованный в пакете AMBER, был применен для анализа реальных биологических систем – с применением классического силового поля FF14SB с 4-х центровой моделью воды TIP4P-Ew, а также нового перспективного силового поля FF15IPQ с 3-х центровой моделью воды SPC/E<sub>b</sub>. Проведен сравнительный анализ результатов – как с точки зрения скорости и масштабируемости молекулярной динамики на GPU и CPU, так и их биологической интерпретации.



**Рис. 1.** Расположение атомов в 3-х центровой модели воды SPC/E<sub>b</sub> (слева) и 4-х центровой модели воды TIP4P-Ew (справа). Псевдоатом в модели TIP4P-Ew показан зеленым цветом. Для моделей приведено значение угла H-O-H.

## 2. Методы

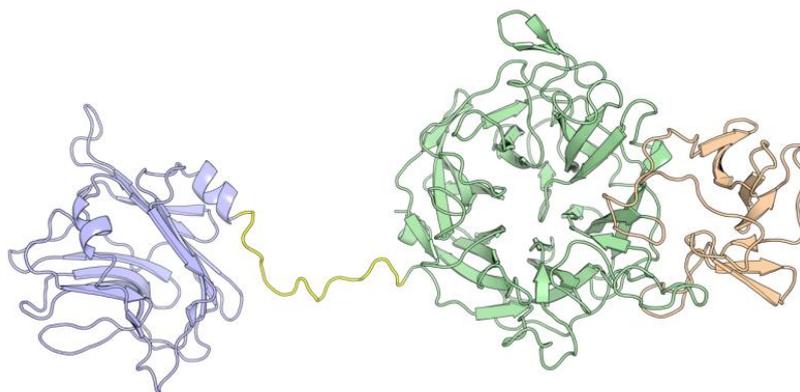
### 2.1 Биологические системы и размеры ячеек

Метод молекулярной динамики был применен к четырем белкам: MAPK – митоген-активируемая протеинкиназа человека (PDB код 1R3C); NanC – нейраминидаза C из *Streptococcus pneumoniae* (PDB код 4YW2); LDH – лактатдегидрогеназа из *Staphylococcus aureus* (PDB код 3D4P); NanA – нейраминидаза A из *Streptococcus pneumoniae* (PDB коды 2YA8 и 4ZXK). Перечисленные белки представляют интерес в качестве мишеней для разработки лекарств от различных заболеваний человека и существенно отличаются по размеру и геометрии. Так, NanA – модульный белок, состоящий из доменов, которые не образуют стабильной глобулярной структуры, а соединены между собой гибким междоменным линкером (Рис. 2). Для динамики этой макромолекулы в периодической системе необходимо поместить ее в «ячейку» таких размеров, чтобы белок в процессе движения и конформационных перестроек не взаимодействовал со своей копией в соседней ячейке при заданном радиусе учета нековалентных

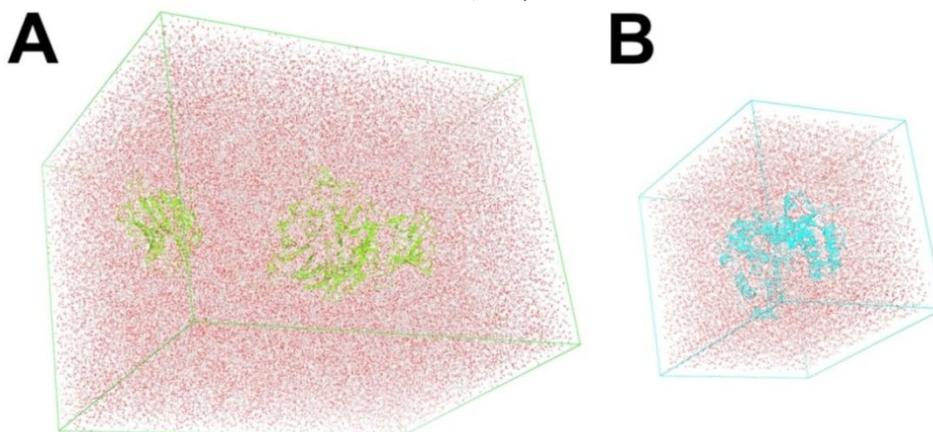
взаимодействий (в этой работе – 10 Å). В случае с NanA, исходя из длины междоменного линкера, использовалась ячейка с минимальным отступом между белком и краем ячейки в 30 Å (Рис. 3, А). В остальных случаях (MAPK, NanC, LDH) белки представляли собой устойчивые глобулярные структуры, что позволяло использовать стандартный отступ – не менее 12 Å (Рис. 3, В). Для каждого случая были подготовлены две модели белка в двух силовых полях – FF14SB и FF15IPQ. Ячейки заполняли водой двух типов – для поля FF14SB использовалась модель воды TIP4P-Ew, для поля FF15IPQ использовалась модель воды SPC/E<sub>b</sub> (Таблица 1). Для контрольных экспериментов также были подготовлены ячейки с MAPK в силовом поле FF14SB, заполненные водой TIP3P и SPC/E<sub>b</sub> (Таблица 2), а также ячейки с LDH в силовом поле FF14SB, заполненные водой TIP3P и SPC/E<sub>b</sub>. Можно добавить, что силовое поле FF15IPQ совместимо только с моделью воды SPC/E<sub>b</sub>.

## 2.2 Вычислительный эксперимент

Метод классической молекулярной динамики в периодической системе с явно заданным растворителем применяли в соответствии с протоколом из L. Pierce и др. [16] с использованием реализаций алгоритма в пакете AMBER для GPU и CPU. Для изучения скорости и масштабируемости на разном числе узлов (от 1 до 8) задачи запускали на 60 минут на GPU или CPU и оценивали среднюю скорость расчета траектории в нс/день. Для получения биологически значимых результатов системы изучали в течение длительного периода времени (от 100 до 1000 нс) с использованием GPU-реализации метода на 4-х узлах суперкомпьютера.



**Рис. 2.** Модель структуры нейраминидазы NanA из *Streptococcus pneumoniae*, построенная с использованием молекулярного моделирования на основании подструктур отдельных доменов из базы данных PDB (коды 2YA8 и 4ZXK). Каталитический, инсерционный и лектиновый домены окрашены в зеленый, пшеничный и синий цвета, соответственно.



**Рис. 3.** Ячейки биологических макромолекулярных систем в воде: (А) NanA из *Streptococcus pneumoniae*; (В) p38α MAP киназа человека.

**Таблица 1.** Количество атомов в ячейке биологических систем NanC, LDH и NanA. В колонке «изменение» указан коэффициент уменьшения количества атомов в системе с 3-х центровой моделью воды SPC/E<sub>b</sub> по сравнению с 4-х центровой моделью воды TIP4P-Ew. Количество атомов воды в ячейке указано в процентах от общего количества атомов.

Название системы	Количество атомов в ячейке			Количество атомов воды в ячейке	
	FF14SB TIP4P-Ew	FF15IPQ SPC/E <sub>b</sub>	Изменение	FF14SB TIP4P-Ew	FF15IPQ SPC/E <sub>b</sub>
NanC	119709	93438	×1.28	91.3%	88.9%
LDH	140069	111932	×1.25	86.3%	82.8%
NanA	347145	265232	×1.31	96.9%	95.9%

**Таблица 2.** Количество атомов в ячейке биологической системы MAPK. В скобках указан коэффициент уменьшения количества атомов в системе с 3-х центровыми моделями воды SPC/E<sub>b</sub> и TIP3P по сравнению с 4-х центровой моделью воды TIP4P-Ew. Количество атомов воды в ячейке указано в процентах от общего количества атомов.

Набор параметров	Количество атомов в ячейке	Количество атомов воды в ячейке
FF14SB + TIP4P-Ew	65941	91.4%
FF14SB + TIP3P	51890 (×1.27)	89.0%
FF14SB + SPC/E <sub>b</sub>	51882 (×1.27)	89.0%
FF15IPQ + SPC/E <sub>b</sub>	51884 (×1.27)	89.0%

## 2.3 Оборудование

Минимизацию энергии, нагрев и уравнивание систем выполняли на локальном GPU-ускорителе GeForce GTX 980 Ti, а финальную стадию «свободной динамики» (т.е. без применения позиционных ограничений) выполняли на уникальной научной установке суперкомпьютере «Ломоносов-2». Один узел суперкомпьютера оснащен одним процессором Intel Xeon E5-2697 v3 (14 физических ядер), 64 Гб оперативной памяти и одной видеокартой Tesla K40. Основная сеть и сеть I/O – Infiniband FDR.

## 3. Результаты

### 3.1 Сравнение скорости и масштабируемости

Скорость вычислений в силовом поле FF15IPQ с 3-х центровой моделью воды SPC/E<sub>b</sub> во всех случаях была существенно выше, чем в силовом поле FF14SB с 4-х центровой моделью воды TIP4P-Ew. При использовании GPU-ускоренной молекулярной динамики уменьшение количества атомов в ячейке на 25-31% в результате использования FF15IPQ+SPC/E<sub>b</sub> вместо FF14SB+TIP4P-Ew ускоряет расчет в ×1.25 - ×1.63 раза в зависимости от размера биологической системы и количества использованных узлов (Таблица 3). Наилучшее ускорение в ×1.44 - ×1.63 раза наблюдалось для системы NanA, самой большой как по общему количеству атомов, так и по процентному содержанию воды; наихудшее – в ×1.25 - ×1.43 раза – для самой маленькой системы MAPK. Для всех изученных систем наилучшее ускорение в результате использования FF15IPQ+SPC/E<sub>b</sub> вместо FF14SB+TIP4P-Ew наблюдалось при запуске на двух узлах – в ×1.43 - ×1.63 раза, наихудшее – в ×1.25 - ×1.44 раза при запуске на восьми узлах.

Результаты, полученные при использовании GPU и CPU реализаций алгоритма, существенно различались по двум показателям – абсолютным значениям скорости, а также масштабируемости вычислений на большом числе узлов. При использовании только классических процессоров ускорение расчетов в результате использования FF15IPQ+SPC/E<sub>b</sub> было эквивалентно значениям, полученным для GPU – до ×1.49 раза (Таблица 4). Однако скорость на GPU была существенно выше скорости на CPU – до 67 нс/день и 21 нс/день, соответственно. Максималь-

ная масштабируемость в большинстве случаев была достигнута на 4 узлах (Таблица 5, Рис. 3). Однако для систем с процентным содержанием воды в ячейке более 91.3% (МАРК, NanC, и NanA в FF14SB+TIP4P-Ew, а также NanA в FF15IPQ+SPC/E<sub>b</sub>) производительность молекулярной динамики на GPU продолжала расти даже на большем числе узлов, в то время как результаты на CPU были нестабильны для всех систем уже при использовании шести узлов. Можно также отметить ухудшение масштабируемости вычислений при использовании FF15IPQ+SPC/E<sub>b</sub> вместо FF14SB+TIP4P-Ew – до 11% на GPU (Таблица 5).

Для изучения влияния конкретной модели растворителя на вычислительную эффективность молекулярной динамики была дополнительно проведена серия контрольных экспериментов с системой МАРК, параметризованной в силовом поле FF14SB с разными 3-х центровыми моделями растворителя (Таблица 6). Сравнение показало, что скорость и масштабируемость не зависели от конкретного силового поля или модели воды, а определялись количеством центров в растворителе.

### 3.2 Биологическая интерпретация результатов

Результаты молекулярной динамики для МАРК, NanC и NanA в FF15IPQ+SPC/E<sub>b</sub> и FF14SB+TIP4P-Ew были эквивалентны, однако при использовании первого набора параметров систематически наблюдалось более значительное среднеквадратичное отклонение (RMSD) структуры белка от первоначального положения в процессе длительной симуляции (например, Рис. 4). Более подробно интерпретация результатов моделирования этих белков с биологической точки зрения изложена в соответствующих публикациях [17, 18].

Существенные различия были выявлены при моделировании системы LDH – лактатдегидрогеназы из *Staphylococcus aureus*. В структурах многих бактериальных лактатдегидрогеназ (например, из *Bifidobacterium longum*, *Bacillus subtilis*, *Geobacillus stearothermophilus*) на границе между субъединицами гомотетрамера есть аллостерический сайт, в котором связывается аллостерический активатор фруктозо-1,6-дифосфат (FBP). Считается, что эти ферменты существуют в растворе в виде ансамбля конформеров с разной функциональностью и соотношение между этими структурными формами зависит от концентрации аллостерического эффектора. Для многих лактатдегидрогеназ присутствие FBP критически необходимо для сборки четвертичной структуры фермента. Однако, существуют и неаллостерические прокариотические лактатдегидрогеназы, для которых роль FBP не описана, например, фермент из *Lactobacillus pentosus*, в котором стабилизация интерфейса между субъединицами достигается за счёт сети солевых мостиков. Аллостерия в лактатдегидрогеназе из *Staphylococcus aureus* (объекте этого исследования) пока не описана и споры по этому поводу продолжаются [19]. Молекулярное моделирование этого фермента в FF15IPQ+SPC/E<sub>b</sub> выявило конформационную перестройку участка структуры, содержащего остатки 193-218, в течение 100 нс симуляции при температуре 300 К (Рис. 5). Использование набора параметров FF14SB+TIP4P-Ew, а также FF14SB+TIP3P и FF14SB+SPC/E<sub>b</sub>, не выявило никаких структурных изменений в этой области белка в идентичных условиях, что в целом согласуется с представлением о том, что классическим силовым полям AMBER, в том числе полю FF14SB, свойственна переоценка некоторых типов взаимодействий, в том числе возникающих в элементах вторичной структуры, а также солевых мостиках, приводящая к артефактам стабилизации структуры [2, 11]. Участок 193-218 в структуре лактатдегидрогеназы из *Staphylococcus aureus* расположен на границе между субъединицами гомотетрамера и входит в состав аллостерического центра связывания FBP в родственных лактатдегидрогеназах. Роль этого участка и его конформационных перестроек в структуре и функции фермента требует дальнейшего изучения, однако полученный результат моделирования представляет интерес в свете все более активного в последние годы изучения аллостерии и выявления новых путей регуляции в ферментах, для которых аллостерия ранее описана не была [20].

**Таблица 3.** Сравнение скорости расчета МД на GPU с использованием 3-х центральной модели воды SPC/E<sub>b</sub> в силовом поле FF15IPQ и 4-х центральной модели воды TIP4P-Ew в силовом поле FF14SB. Значения скоростей приведены в нс/день.

Название системы	N1			N2		
	FF14SB TIP4P-Ew	FF15IPQ SPC/E <sub>b</sub>	Ускорение	FF14SB TIP4P-Ew	FF15IPQ SPC/E <sub>b</sub>	Ускорение
МАРК	30.54	41.95	×1.37	33.89	48.43	×1.43
NanC	16.56	23.61	×1.43	19.17	28.26	×1.47
LDH	13.31	18.85	×1.42	17.13	24.42	×1.43
NanA	5.75	8.87	×1.54	6.86	11.16	×1.63

Название системы	N4			N6		
	FF14SB TIP4P-Ew	FF15IPQ SPC/E <sub>b</sub>	Ускорение	FF14SB TIP4P-Ew	FF15IPQ SPC/E <sub>b</sub>	Ускорение
МАРК	48.91	67.26	×1.38	50.07	63.72	×1.28
NanC	26.80	39.15	×1.46	27.95	38.19	×1.37
LDH	23.64	30.30	×1.28	23.62	30.95	×1.31
NanA	9.21	14.38	×1.56	9.96	15.27	×1.53

Название системы	N8		
	FF14SB TIP4P-Ew	FF15IPQ SPC/E <sub>b</sub>	Ускорение
МАРК	51.67	64.52	×1.25
NanC	30.09	38.61	×1.28
LDH	23.44	30.40	×1.30
NanA	10.39	14.97	×1.44

**Таблица 4.** Сравнение скорости расчета МД на CPU с использованием 3-х центральной модели воды SPC/E<sub>b</sub> в силовом поле FF15IPQ и 4-х центральной модели воды TIP4P-Ew в силовом поле FF14SB. Значения скоростей приведены в нс/день.

Название системы	N1			N2		
	FF14SB TIP4P-Ew	FF15IPQ SPC/E <sub>b</sub>	Ускорение	FF14SB TIP4P-Ew	FF15IPQ SPC/E <sub>b</sub>	Ускорение
МАРК	4.73	6.88	×1.46	9.06	12.72	×1.40
NanC	2.62	3.83	×1.46	5.00	7.24	×1.45
LDH	2.20	3.05	×1.39	4.23	5.86	×1.39
NanA	0.87	1.30	×1.49	1.72	2.55	×1.48

Название системы	N4			N6		
	FF14SB TIP4P-Ew	FF15IPQ SPC/E <sub>b</sub>	Ускорение	FF14SB TIP4P-Ew	FF15IPQ SPC/E <sub>b</sub>	Ускорение
МАРК	15.56	21.39	×1.38	19.22	19.18	×1.00
NanC	8.69	12.55	×1.44	6.08	6.50	×1.07
LDH	6.38	8.63	×1.35	5.71	4.16	×0.73
NanA	3.06	4.25	×1.39	2.37	2.66	×1.12

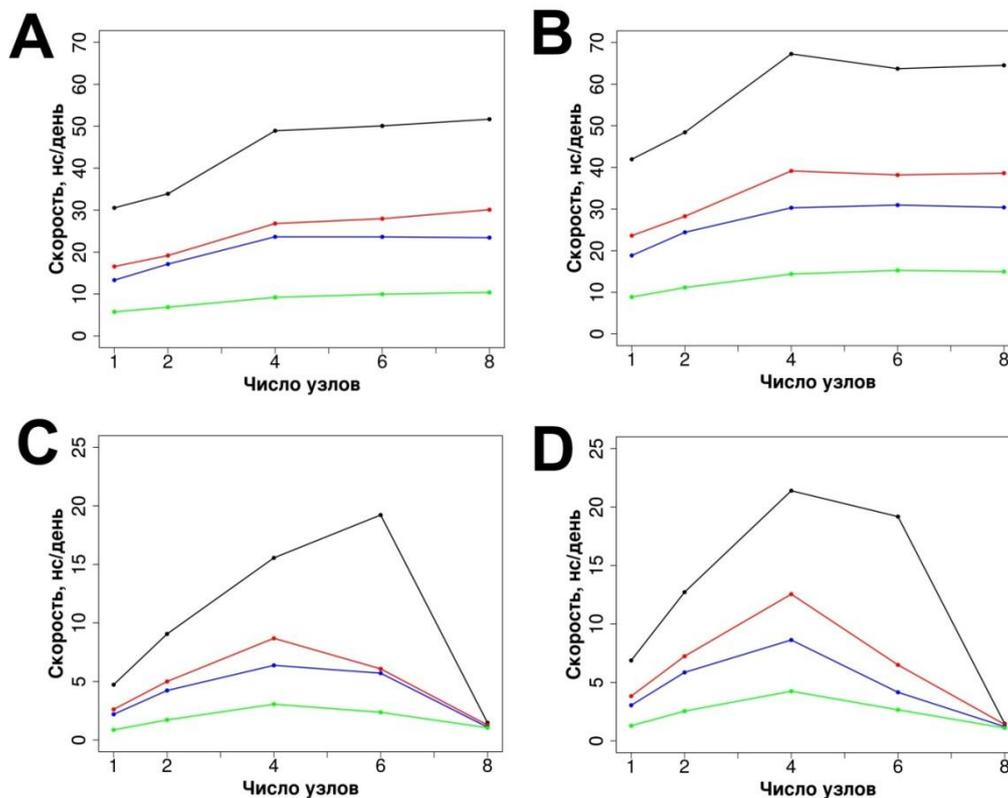
Название системы	N8		
	FF14SB TIP4P-Ew	FF15IPQ SPC/E <sub>b</sub>	Ускорение
МАРК	1.48	1.47	×0.99
NanC	1.28	1.41	×1.10
LDH	1.10	1.22	×1.11
NanA	1.05	1.11	×1.06

**Таблица 5.** Ускорение в параллельном режиме на GPU и CPU по сравнению с одним узлом.

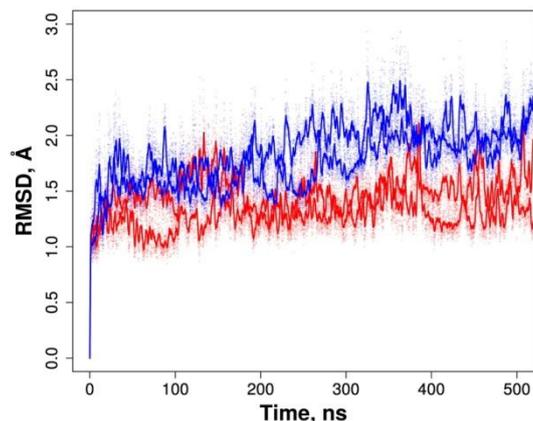
Название системы		GPU				CPU			
		N2	N4	N6	N8	N2	N4	N6	N8
МАРК	FF14SB + TIP4P-Ew	1.11	1.60	1.64	1.69	1.92	3.29	4.06	0.31
	FF15IPQ + SPC/E <sub>b</sub>	1.15	1.60	1.52	1.54	1.85	3.11	2.79	0.21
NanC	FF14SB + TIP4P-Ew	1.16	1.62	1.69	1.82	1.91	3.32	2.32	0.49
	FF15IPQ + SPC/E <sub>b</sub>	1.20	1.66	1.62	1.64	1.89	3.28	1.70	0.37
LDH	FF14SB + TIP4P-Ew	1.29	1.78	1.77	1.76	1.92	2.90	2.60	0.50
	FF15IPQ + SPC/E <sub>b</sub>	1.30	1.61	1.64	1.61	1.92	2.83	1.36	0.40
NanA	FF14SB + TIP4P-Ew	1.19	1.60	1.73	1.81	1.98	3.52	2.72	1.21
	FF15IPQ + SPC/E <sub>b</sub>	1.26	1.62	1.72	1.69	1.96	3.27	2.05	0.85

**Таблица 6.** Сравнение скорости расчета МД в системе МАРК с использованием 3-х центровых моделей воды TIP3P и SPC/E<sub>b</sub> и 4-х центральной модели воды TIP4P-Ew. Скорость указана в нс/день.

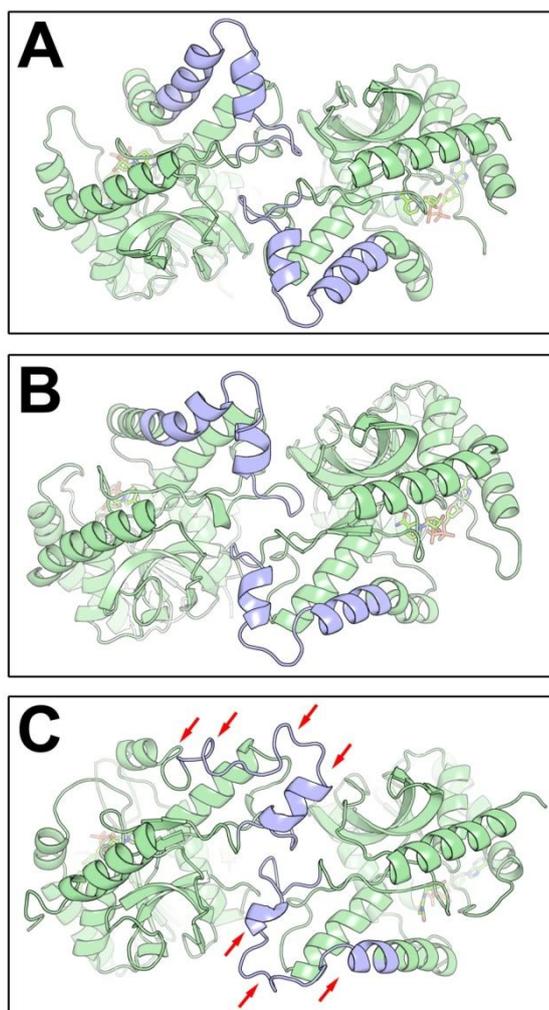
Число узлов	GPU				CPU			
	FF14SB			FF15IPQ SPC/E <sub>b</sub>	FF14SB			FF15IPQ SPC/E <sub>b</sub>
	TIP4P-Ew	TIP3P	SPC/E <sub>b</sub>		TIP4P-Ew	TIP3P	SPC/E <sub>b</sub>	
N1	30.54	42.45	42.23	41.95	4.73	6.95	6.86	6.88
N2	33.89	48.91	48.61	48.43	9.06	12.91	13.04	12.72
N4	48.91	66.92	67.03	67.26	15.56	21.42	21.65	21.39
N6	50.07	65.25	66.98	63.72	19.02	20.39	17.82	19.18
N8	51.90	63.31	64.44	64.52	1.48	1.44	1.46	1.47



**Рис. 3.** Скорость расчета молекулярной динамики в FF14SB+TIP4P-Ew на (A) GPU и (C) CPU; в FF15IPQ+SPC/E<sub>b</sub> на (B) GPU и (D) CPU для разных биологических систем: МАРК (черный), NanC (красный), LDH (синий), NanA (зеленый).



**Рис. 4.** Графики RMSD в течение 500 нс молекулярной динамики при температуре 300 К структуры нейраминидазы С из *Streptococcus pneumoniae* (NanC) – в силовом поле FF15IPQ в воде SPC/E<sub>b</sub> (синий), и в силовом поле FF14SB в воде TIP4P-Ew (красный). Для каждого случая приведены по две траектории. Точки соответствуют значениям среднеквадратичного отклонения, линии соответствуют усреднению этих значений в окрестности 2 нс.



**Рис. 5.** Существенные различия результатов молекулярного динамики структуры лактатдегидрогеназы из *Staphylococcus aureus* (LDH) в разных силовых полях. (А) Кристаллографическая структура LDH из банка данных PDB (код 3D4P); Структура LDH через 100 нс молекулярной динамики при температуре 300 К: (В) в силовом поле FF14SB в воде TIP4P-Ew; (С) в силовом поле FF15IPQ в воде SPC/E<sub>b</sub>. Участок структуры, содержащий остатки 193-218, показан синим цветом.

### 3. Выводы

Использованные в работе силовые поля FF14SB/FF15IPQ и модели воды TIP4P-Ew/TIP3P/«SPC/E<sub>b</sub>» представляют собой разные наборы параметров, однако уравнения, в которые значения этих параметров подставляются для вычисления энергии взаимодействия атомов и возникающего движения частиц, идентичны. Отличия в вычислительной эффективности молекулярной динамики с разными силовыми полями и моделями воды, в принципе, могут быть связаны с отличиями в геометрии, как самого растворителя, так и белка. Средние значения и жесткость углов/длин связей отличаются в разных моделях воды и силовых полях, что может сказаться на самой вычислительно сложной части алгоритма – расчёте нековалентных взаимодействий каждого атома с другими атомами, расположенными внутри заданного радиуса (в этом исследовании – 10 Å). Однако геометрия силовых полей и моделей воды различается не до такой степени, чтобы существенно повлиять на скорость. Поэтому полученные в работе результаты ускорения и масштабируемости при использовании FF15IPQ+SPC/E<sub>b</sub> вместо FF14SB+TIP4P-Ew связаны, прежде всего, с разным количеством атомов в моделях растворителя. Использование 3-х центровой модели растворителя вместо 4-х центровой модели (т.е. удаление одного псевдоатома из каждой молекулы воды в ячейке) приводит к уменьшению количества атомов в одинаковом объеме – изменяется как количество атомов в ячейке, так и плотность атомов (количество атомов в заданном радиусе), что существенно сказывается на расчете нековалентных взаимодействий и позволяет достичь ускорения до  $\times 1.63$  раз на GPU и до  $\times 1.49$  раз на CPU. Кроме того, в некоторых случаях использование нового перспективного силового поля FF15IPQ с единственной совместимой с ним моделью воды SPC/E<sub>b</sub> позволяет не только существенно ускорить расчет, но и получать качественно иные результаты, более интересные с точки зрения изучения конформационной подвижности в структурах белков. Однако не следует забывать, что поля серии IPQ (в том числе FF15IPQ) все еще находятся в процессе разработки и развития, в то время как FF14SB – основное рекомендованное поле для молекулярной динамики белков с помощью программ пакета AMBER. Таким образом, полученные результаты говорят о необходимости совместного использования разных силовых полей при изучении биологических систем для того, чтобы оценить зависимость полученных результатов от использованных параметров.

В заключении можно подчеркнуть, что использование современных GPU-ускорителей для молекулярной динамики позволяет существенно увеличить время наблюдения за биологической макромолекулярной системой (т.е. длину траекторий) в повседневной практике – в среднем с десятков наносекунд до нескольких микросекунд – при адекватных параметрах алгоритма. Это означает, что для изучения даже больших и сложно организованных белков, таких как нейраминидаза А из *Streptococcus pneumoniae* (до 350 000 атомов в ячейке с водой) больше не нужно жертвовать качеством в угоду скорости. Уникальная научная установка суперкомпьютер «Ломоносов-2», оснащенная мощными GPU-вычислителями Tesla K40, открывает принципиально новые горизонты для научной работы в области структурной биологии с высоким потенциалом практического использования результатов в биотехнологии и медицине.

### Благодарность

Работа выполнена с использованием оборудования Центра коллективного пользования сверх-высокопроизводительными вычислительными ресурсами МГУ имени М.В. Ломоносова [21].

### Литература

1. Суплатов Д.А., Попова Н.Н., Копылов К.Е., Шегай М.В., Воеводин Вл.В., Швядас В.К. Гибридные вычислительные кластеры для изучения структуры, функции и регуляции белков // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2017. Т. 6, № 4. С. 74–90. DOI: 10.14529/cmse170406.
2. Godwin, R. C., Melvin, R., & Salsbury, F. R. Computer-Aided Drug Discovery (No. 34). New York, NY: Springer New York, 2016. 305 p. DOI:10.1007/978-1-4939-3521-5.

3. Shaw D. E. et al. Anton, a special-purpose machine for molecular dynamics simulation // Communications of the ACM. 2008. Vol. 51, No. 7. P. 91–97.
4. Shaw D. E. et al. Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer // Proceedings of the international conference for high performance computing, networking, storage and analysis. IEEE Press. 2014. P. 41–53.
5. Nobile M. S. et al. Graphics processing units in bioinformatics, computational biology and systems biology // Briefings in bioinformatics. 2016. Vol, 18. No. 5. P. 870–885.
6. Salomon-Ferrer R. et al. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald // Journal of chemical theory and computation. 2013. Vol. 9, No. 9. P. 3878–3888.
7. Götz A. W. et al. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born // Journal of chemical theory and computation. 2012. Vol, 8. No. 5. P. 1542–1555.
8. URL: <http://manual.gromacs.org/documentation/2018/user-guide/force-fields.html> (дата обращения: 30.05.2018).
9. Maier J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB // Journal of chemical theory and computation. 2015. Vol, 11. No. 8. P. 3696–3713.
10. URL: <http://ambermd.org/doc12/Amber17.pdf>, P. 33 (дата обращения: 30.05.2018).
11. Debiec K. T. et al. Further along the road less traveled: AMBER ff15ipq, an original protein force field built on a self-consistent physical model // Journal of chemical theory and computation. 2016. Vol. 12, No. 8. P. 3926–3947.
12. Onufriev A. Implicit solvent models in molecular dynamics simulations: A brief overview // Annual Reports in Computational Chemistry. Elsevier, 2008. Vol. 4, P. 125–137.
13. Wong V., Case D. A. Evaluating rotational diffusion from protein MD simulations // The Journal of Physical Chemistry B. 2008. Vol. 112, No. 19. P. 6013–6024.
14. Horn H. W. et al. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew // The Journal of chemical physics. 2004. Vol. 120, No. 20. P. 9665–9678.
15. Takemura K., Kitao A. Water model tuning for improved reproduction of rotational diffusion and NMR spectral density // The Journal of Physical Chemistry B. 2012. Vol. 116, No. 22. P. 6279–6287.
16. Pierce L. C. T. et al. Routine access to millisecond time scale events with accelerated molecular dynamics // Journal of chemical theory and computation. 2012. Vol. 8, No. 9. P. 2997–3002.
17. Suplatov D., Kopylov K., Sharapova Y., Švedas V. Human p38 $\alpha$  Mitogen-Activated Protein Kinase in the Asp168-Phe169-Gly170-in (DFG-in) state can bind allosteric inhibitor Doramapimod // Journal of Biomolecular Structure and Dynamics. 2018. DOI:10.1080/07391102.2018.1475260
18. Sharapova Y., Suplatov D., Švedas V. Neuraminidase A from Streptococcus pneumoniae has a modular organization of Catalytic and Lectin Domains separated by a flexible linker // The FEBS journal. 2018. DOI:10.1111/febs.14486.
19. Crooke A. K. et al. CspA-independent glucose regulation of lactate dehydrogenase 1 in Staphylococcus aureus // PLoS One. 2013. Vol. 8, No. 1. P. e54293.
20. Суплатов Д. А., Шведас В. К. Изучение функциональных и аллостерических сайтов в суперсемействах белков // Acta Naturae. 2015. Т. 7, №. 4. С. 39–52.
21. Воеводин В. В., Жуматий С. А., Соболев С. И., Антонов А. С., Брызгалов П. А., Никитенко Д. А., Стефанов К., Воеводин, В. Практика суперкомпьютера «Ломоносов» // Открытые системы, 2012, 7, 36-39.