

МОЛЕКУЛЯРНАЯ ДИНАМИКА В СИЛОВОМ ПОЛЕ FF14SB В ВОДЕ TIP4P-EW, И В СИЛОВОМ ПОЛЕ FF15IPQ В ВОДЕ SPC/Е_B: СРАВНИТЕЛЬНЫЙ АНАЛИЗ НА GPU И CPU*

© 2019 Д.А. Суплатов, Я.А. Шарапова, Н.Н. Попова, К.Е. Копылов,
Вл.В. Воеводин, В.К. Швядас

*Московский государственный университет имени М.В. Ломоносова
(119991 Москва, ул. Ленинские Горы, д. 1, стр. 40)*

*E-mail: d.a.suplatov@belozersky.msu.ru, sharapova@belozersky.msu.ru, popova@cs.msu.ru,
kopylov@mail.chem.msu.ru, voevodin@parallel.ru, vytas@belozersky.msu.ru*

Поступила в редакцию: 03.08.2018

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-центральной моделью воды TIP4P-Ew, а также нового многообещающего поля FF15IPQ с 3-центральной моделью воды SPC/Е_B. Были использованы классические процессоры Intel Xeon E5-2697 v3, а также GPU ускорители Tesla K40 (архитектура Kepler) и P100 (Pascal). Уменьшение количества атомов в ячейке на 25–31 % в результате использования 3-центральной 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 как альтернативы классическим CPU позволяет существенно увеличить длину траектории в повседневной практике.

Ключевые слова: классическая молекулярная динамика, AMBER, силовые поля FF14SB и FF15IPQ, модели воды TIP4P-Ew, TIP3P и SPC/Е_B, GPU-ускорители Kepler и Pascal.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Суплатов Д.А., Шарапова Я.А., Попова Н.Н., Копылов К.Е., Воеводин Вл.В., Швядас В.К. Молекулярная динамика в силовом поле FF14SB в воде TIP4P-Ew, и в силовом поле FF15IPQ в воде SPC/Е_B: сравнительный анализ на GPU и CPU // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2019. Т. 8, № 1. С. 71–88. DOI: 10.14529/cmse190105.

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

*Статья рекомендована к публикации программным комитетом Международной конференции «Суперкомпьютерные дни в России — 2018».

Карплусу, Майклу Левитту и Арье Уоршелу 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 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-центральной модели TIP4P-Ew вместо 3-центральной TIP3P); уменьшению шага интегрирования уравнения движения; усложнению электростатических расчетов (увеличению радиуса учета нековалентных взаимодействий и отключению псевдоструктур, которые представляют типовые химические блоки в упрощенном виде, например, редуцируют сопряженные многоатомные системы до нескольких псевдоатомов); 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 биомакромолекулярной системой 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 динамике 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.

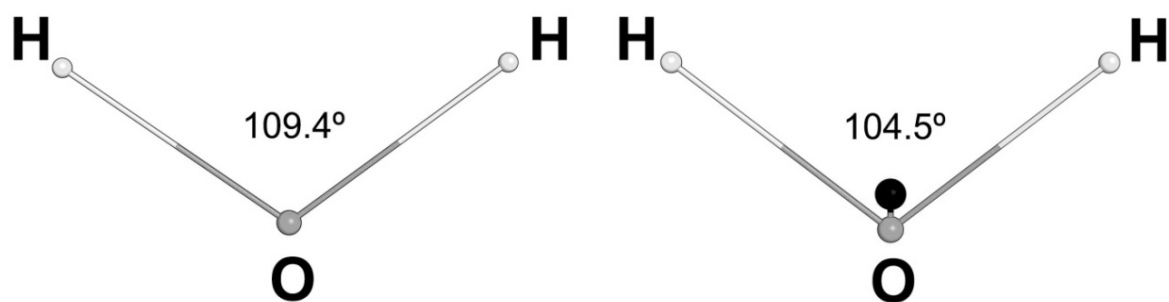
Силовые поля — это функции расчета энергии движения и взаимодействия атомов, а также наборы параметров для этих функций, характеризующих, в том числе, массу и заряд атомов, длины, углы и двугранные углы ковалентных взаимодействий атомов друг с другом, и т.д. Силовые поля AMBER предназначены для оригинальных 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, чтобы реконструировать, например, интерфейс взаимодействия между доменами. Расчет молекулярной динамики без растворителя можно исполнить, 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 которого, как правило, выступает вода. Растворитель играет важную роль в процессах, происходящих в живой клетке, поэтому модели воды для молекулярной динамики, как и сило-

вые поля, постоянно дорабатываются для повышения точности расчетов. Существует множество моделей воды, которые предназначены для использования в разных силовых полях, по-разному воспроизводят экспериментально изученные физико-химические свойства воды, и отличаются геометрией, зарядами, массами атомов, а также их количеством. Наиболее распространенными сегодня являются 3-центровые модели воды (пример приведен на рис. 1а). 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-центровые модели воды, содержащие один дополнительный псевдоатом. Для расчетов в периодической системе с явно заданным растворителем была разработана модель воды TIP4P-Ew (рис. 1б). Было показано, что она лучше описывает вращательное движение белков по сравнению с обычно используемой 3-центральной 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 вычислительную сложность эксперимента. В этом контексте представляет интерес 3-центровая модель воды SPC/E_b (рис. 1а), которая совместима с ранее упомянутым перспективным силовым полем FF15IPQ, а по своим характеристикам не уступает либо превосходит доступные аналоги [15]. Существуют также 2-, 5- и 6-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-центральной моделью воды TIP4P-Ew, а также нового перспективного силового поля FF15IPQ с 3-центральной моделью воды SPC/E_b. Проведен сравнительный анализ результатов — как с точки зрения скорости и масштабируемости молекулярной динамики на GPU и CPU, так и их биологической интерпретации.



а) 3-центровая модель воды SPC/E_b

б) 4-центровая модель воды TIP4P-Ew.

Псевдоатом показан черным цветом

Рис. 1. Расположение атомов в разных моделях воды. Приведено значение угла Н-О-Н

1. Методы

1.1. Биологические системы и размеры ячеек

Метод классической молекулярной динамики применяли 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 \AA). В случае с NanA, исходя из длины междоменного линкера, использовали ячейку с минимальным отступом между белком и краем ячейки в 30 \AA (рис. 3а). В остальных случаях (MAPK, NanC, LDH) белки представляли собой устойчивые глобулярные структуры, что позволяло использовать стандартный отступ — не менее 12 \AA (рис. 3б). Для каждого случая были подготовлены две модели белка в двух силовых полях — FF14SB и FF15IPQ. Ячейки заполняли водой двух типов — для поля FF14SB использовали модель воды TIP4P-Ew, для поля FF15IPQ использовали модель воды SPC/Е_b. Размеры соответствующих биологических систем приведены в табл. 1. В колонке «изменение» указан коэффициент уменьшения количества атомов в системе с 3-центрковой моделью воды SPC/Е_b по сравнению с 4-центрковой моделью воды TIP4P-Ew. «Количество атомов воды в ячейке» указано в процентах от общего количества атомов. Для контрольных экспериментов также были подготовлены ячейки с MAPK в силовом поле FF14SB, заполненные водой TIP3P и SPC/Е_b, а также ячейки с LDH в силовом поле FF14SB, заполненные водой TIP3P и SPC/Е_b. Можно добавить, что силовое поле FF15IPQ совместимо только с моделью воды SPC/Е_b.

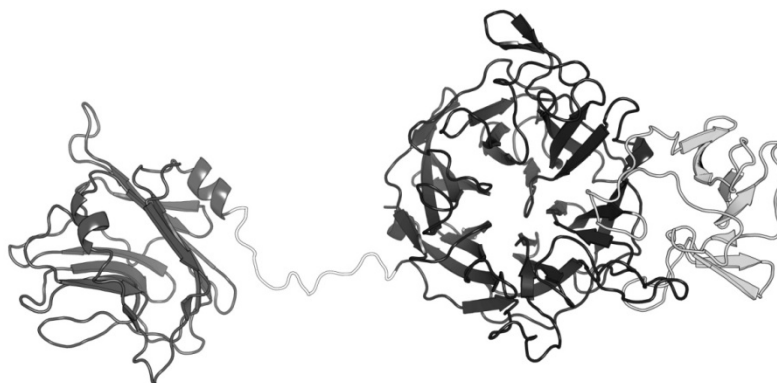
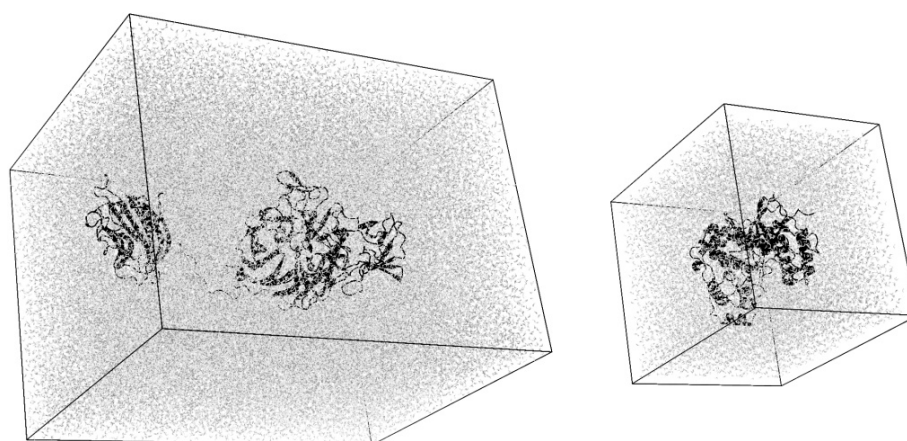


Рис. 2. Модель структуры нейраминидазы NanA из *Streptococcus pneumoniae* [18]. Отдельные домены и междоменный линкер окрашены в разные тона серого



а) NanA из *Streptococcus pneumoniae*

б) p38α MAP киназа человека

Рис. 3. Ячейки биологических макромолекулярных систем в воде

Таблица 1

Размеры исследованных биологических систем

Название системы	Количество атомов в ячейке			Количество атомов воды в ячейке	
	FF14SB TIP4P-Ew	FF15IPQ SPC/E _b	Изменение	FF14SB TIP4P-Ew	FF15IPQ SPC/E _b
МАРК	65941	51884	×1,27	91,4%	89,0%
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%

1.2. Вычислительный эксперимент

Метод классической молекулярной динамики в периодической системе с явно заданным растворителем применяли в соответствии с ранее описанным протоколом [16–18] с использованием реализаций алгоритма в пакете AMBER для GPU и CPU. Для изучения скорости и масштабируемости на разном числе видеокарт одного и нескольких узлов задачи запускали на 60 минут в режиме 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).

1.3. Оборудование

Минимизацию энергии, нагрев и уравнивание систем выполняли на локальном GPU-ускорителе GeForce GTX 980 Ti, а собственно оценку скорости молекулярной динамики на стадии «свободной динамики» (т.е. без применения позиционных ограничений) выполняли на двух разделах — «Compute» и «Pascal» — уникальной научной установки суперкомпьютере «Ломоносов-2». Один узел раздела «Compute» оснащен одним процессором Intel Xeon E5-2697 v3 (14 физических ядер), 64 Гб оперативной памяти и одной видеокартой Tesla K40 (архитектура Kepler). Один узел раздела «Pascal» оснащен одним процессором Intel Xeon Gold 6126 (12 физических ядер), 96 Гб оперативной памяти и двумя видеокартами Tesla P100 (архитектура Pascal). Основная сеть и сеть I/O — Infiniband FDR.

1.4. Программное обеспечение

Для расчета классической молекулярной динамики использовали программы пакета AMBER версий 14 и 16. Программы версии 14 использовали для вычислительного эксперимента на CPU и GPU раздела «Compute» суперкомпьютера «Ломоносов-2» (видеокарты Tesla K40 архитектуры Kepler), а программы версии 16 исполняли на GPU ускорителях раздела «Pascal» суперкомпьютера «Ломоносов-2» (видеокарты Tesla P100 архитектуры Pascal). Сборку программ осуществляли из исходного кода с использованием следующих компиляторов и библиотек: Intel icc/icpc/fort версии 15.0.3, MKL версии 11.1.3, CUDA версий 6.5 (для сборки AMBER версии 14) и 8.0 (для сборки AMBER версии 16), а также OpenMPI версий 1.10.7 и 2.1.1.

2. Результаты

2.1. Сравнение скорости и масштабируемости

Скорость вычислений в силовом поле FF15IPQ с 3-центральной моделью воды SPC/E_b во всех случаях была существенно выше, чем в силовом поле FF14SB с 4-центральной моделью воды TIP4P-Ew. При использовании GPU-ускоренной молекулярной динамики на видеокартах Tesla K40 (архитектура Kepler) уменьшение количества атомов в ячейке на 25–31 % в результате использования FF15IPQ+SPC/E_b вместо FF14SB+TIP4P-Ew ускоряет расчет в $\times 1,25$ – $\times 1,63$ раза в зависимости от размера биологической системы и количества использованных узлов (табл. 2, рис. 4). Наилучшее ускорение в $\times 1,44$ – $\times 1,63$ раза наблюдалось для системы NanA, самой большой как по общему количеству атомов, так и по процентному содержанию воды; наихудшее — в $\times 1,25$ – $\times 1,43$ раза — для самой маленькой системы MAPK. Для всех изученных систем наилучшее ускорение в результате использования FF15IPQ+SPC/E_b вместо FF14SB+TIP4P-Ew наблюдалось при запуске на двух узлах — в $\times 1,43$ – $\times 1,63$ раза, наихудшее — в $\times 1,25$ – $\times 1,44$ раза при запуске на восьми узлах. Можно также в целом отметить ухудшение масштабируемости вычислений при использовании FF15IPQ+SPC/E_b вместо FF14SB+TIP4P-Ew — до 11 % (табл. 3). Эффективность работы GPU-имплементации молекулярной динамики, собранной с использованием OpenMPI версий 1.10.7 и 2.1.1, различалась незначительно.

Результаты, полученные при использовании молекулярной динамики на GPU (Tesla K40) и CPU (Intel Xeon E5-2697 v3) узлах раздела «Compute», 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_b было эквивалентно значениям, полученным для GPU — до $\times 1,49$ раза. Однако скорость на GPU была существенно выше скорости на CPU — до 67 нс/день и 21 нс/день, соответственно. Максимальная масштабируемость в большинстве случаев была достигнута на 4 узлах (рис. 4 и 5, табл. 3). Однако для систем с процентным содержанием воды в ячейке более 91,3 % (MAPK, NanC, и NanA в FF14SB+TIP4P-Ew, а также NanA в FF15IPQ+SPC/E_b) производительность молекулярной динамики продолжала расти даже на большем числе узлов. Эффективность работы приложения, собранного разными версиями OpenMPI на 6–8 узлах при использовании только CPU принципиально различалась — масштабируемость сборки MPI компилятором версии 2.1.1 деградировала, в

Таблица 2

Скорость расчета МД на GPU ускорителях Tesla K40 (в нс/день)

Название системы	N1			N2		
	FF14SB TIP4P-Ew	FF15IPQ SPC/E _b	Ускорение	FF14SB TIP4P-Ew	FF15IPQ SPC/E _b	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 _b	Ускорение	FF14SB TIP4P-Ew	FF15IPQ SPC/E _b	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 _b	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 3

Ускорение в параллельном режиме на GPU Tesla K40

Название системы		GPU			
		N2	N4	N6	N8
МАРК	FF14SB + TIP4P-Ew	1,11	1,60	1,64	1,69
	FF15IPQ + SPC/E _b	1,15	1,60	1,52	1,54
NanC	FF14SB + TIP4P-Ew	1,16	1,62	1,69	1,82
	FF15IPQ + SPC/E _b	1,20	1,66	1,62	1,64
LDH	FF14SB + TIP4P-Ew	1,29	1,78	1,77	1,76
	FF15IPQ + SPC/E _b	1,30	1,61	1,64	1,61
NanA	FF14SB + TIP4P-Ew	1,19	1,60	1,73	1,81
	FF15IPQ + SPC/E _b	1,26	1,62	1,72	1,69

то время как производительность сборки MPI компилятором версии 1.10.7 в некоторых случаях продолжала расти (рис. 5).

Отдельно было проведено изучение вычислительной эффективности и масштабируемости молекулярной динамики на видеокартах Tesla P100 (GPU-архитектура Pascal), которые появились на суперкомпьютере Ломоносов-2 после обновления в марте 2018 года (раздел «Pascal»). Вычислительная мощность одной Tesla P100 существенно выше, чем у Tesla K40, что принципиально сказывается на производительности и масштабируемости молекулярной динамики. Скорость расчета на двух видеокартах Tesla P100 од-

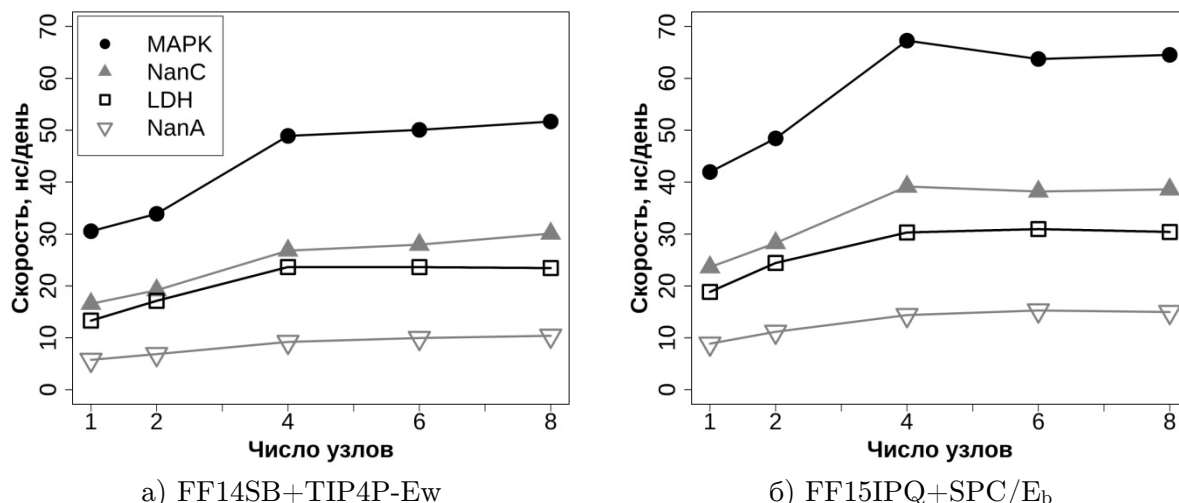


Рис. 4. Скорость расчета молекулярной динамики на GPU Tesla K40 (один узел — одна видеокарта)

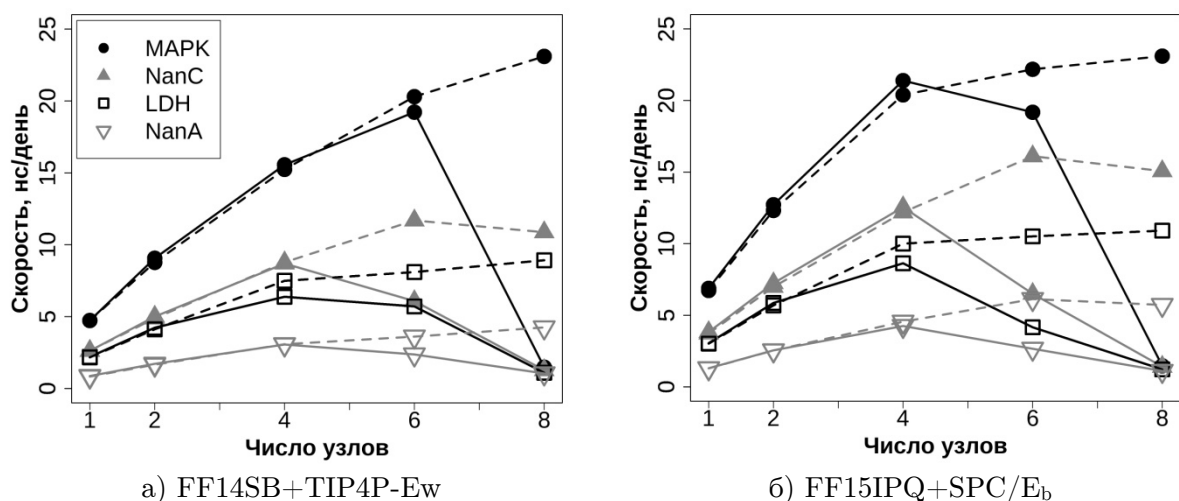


Рис. 5. Скорость расчета молекулярной динамики на CPU. Сплошной линией показаны результаты работы приложения, собранного с использованием OpenMPI версии 2.1.1, пунктирной линией — OpenMPI версий 1.10.7

ного узла раздела «Pascal» (содержит две видеокарты на каждом узле) была существенно выше лучшей скорости расчетов, полученной на 4–8 узлах с Tesla K40 раздела «Compute» (содержит одну видеокарту на каждом узле) — в $\times 2,32$ – $\times 2,77$ и $\times 2,57$ – $\times 3,02$ для набора параметров FF14SB+TIP4P-Ew и FF15IPQ+SPC/Eb, соответственно (рис. 6). Ускорение на двух Tesla P100 одного узла по сравнению с одной Tesla P100 составило $\times 1,21$ – $\times 1,37$, однако производительность на двух и более узлах во всех случаях была существенно ниже производительности на одной видеокарте, что объясняется различиями в механизме обмена данными между устройствами, установленными в одном и разных узлах. Две видеокарты одного узла способны обмениваться данными в режиме peer-to-peer, т.е. непосредственно через соединение PCI-E минуя центральный процессор, что снижает время ожидания и в итоге приводит к ускорению. Однако вычислительная мощность Tesla P100 настолько велика, что обмен данными становится скоростью-лимитирующей стадией расчета при совместном использовании видеокарт на разных узлах, что делает невозможным дальнейшее ускорение. Ускорение на Tesla P100 при ис-

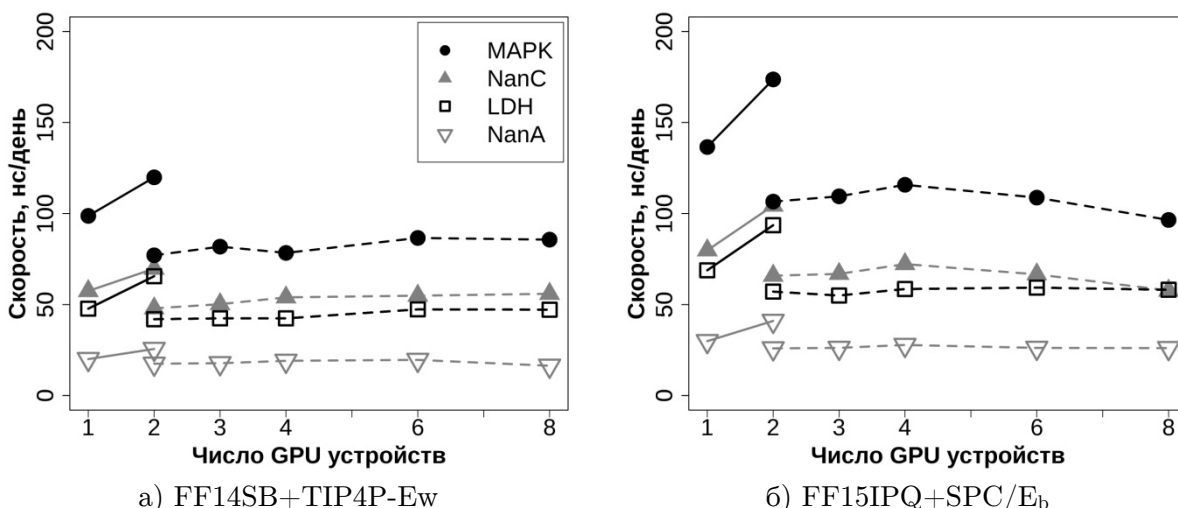


Рис. 6. Скорость расчета молекулярной динамики на GPU Tesla P100 (один узел — две видеокарты). Сплошной линией показаны результаты запуска приложения на одном или двух GPU ускорителях одного узла (в режиме peer-to-peer), пунктирной линией — на двух и более GPU ускорителях двух и более узлов

пользовании FF15IPQ+SPC/E_b вместо FF14SB+TIP4P-Ew были эквивалентны значениям, полученным на Tesla K40 — $\times 1,38$ – $\times 1,49$ и $\times 1,43$ – $\times 1,60$ на одной и двух видеокартах одного узла, соответственно. Эффективность работы на Tesla P100 приложения, собранного с использованием OpenMPI версий 1.10.7 и 2.1.1, различалась незначительно.

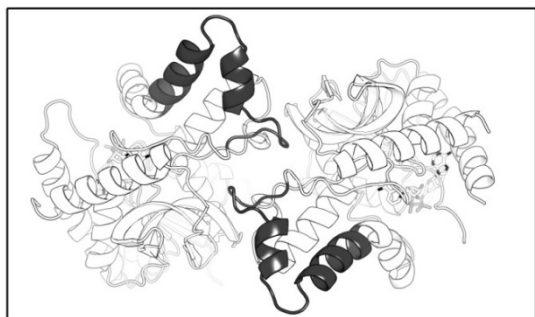
Для изучения влияния конкретной модели растворителя на вычислительную эффективность молекулярной динамики была дополнительно проведена серия контрольных экспериментов с системой MAPK, параметризованной в силовом поле FF14SB с разными 3-центровыми моделями растворителя. Сравнение показало, что скорость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

2.2. Биологическая интерпретация результатов

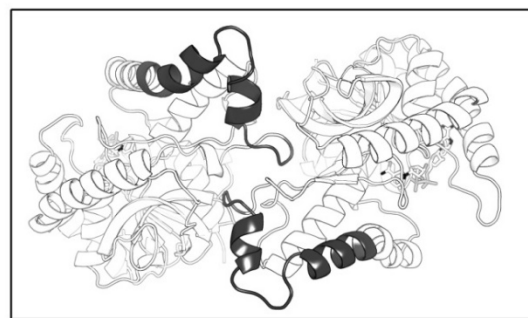
Результаты молекулярной динамики для MAPK, NanC и NanA в FF15IPQ+SPC/E_b и FF14SB+TIP4P-Ew были эквивалентны, однако при использовании первого набора параметров систематически наблюдалось более значительное среднеквадратичное отклонение (RMSD) структуры белка от первоначального положения в процессе длительной симуляции. Более подробно интерпретация результатов моделирования этих белков с биологической точки зрения изложена в отдельных 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 (ФВР). 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 ФВР критически необходимо для сборки четвертичной структуры фермента. Однако существуют и неаллостерические прокариотические лак-

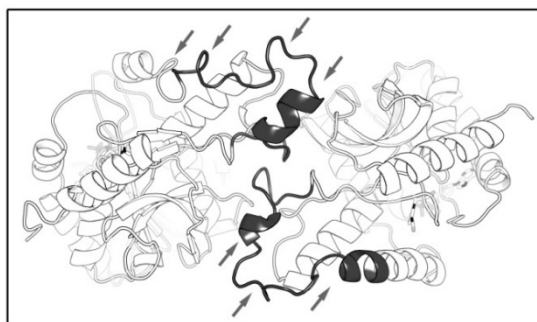
татдегидрогеназы, для которых роль FBP не описана, например, фермент из *Lactobacillus pentosus*, в котором стабилизация интерфейса между субъединицами достигается за счет сети солевых мостиков. Аллостерия в лактатдегидрогеназе из *Staphylococcus aureus* (объекте этого исследования) пока не описана и споры по этому поводу продолжают [19]. Молекулярное моделирование этого фермента в FF15IPQ+SPC/E_w выявило конформационную перестройку участка структуры, содержащего остатки 193-218, в течение 100 нс симуляции при температуре 300 К (рис. 7). Использование набора параметров FF14SB+TIP4P-Ew, а также FF14SB+TIP3P и FF14SB+SPC/E_w, 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 еще не описана [20].



а) Кристаллографическая структура LDH из банка данных PDB (код 3D4P)



б) Структура LDH через 100 нс молекулярной динамики в силовом поле FF14SB в воде TIP4P-Ew



в) Структура LDH через 100 нс молекулярной динамики при температуре 300 К в силовом поле FF15IPQ в воде SPC/E_w

Рис. 7. Существенные различия результатов молекулярной динамики структуры лактатдегидрогеназы из *Staphylococcus aureus* (LDH) в двух силовых полях. Участок структуры, содержащий остатки 193-218, показан черным цветом

Заключение

Использованные в работе силовые поля FF14SB/FF15IPQ и модели воды TIP4P-Ew/TIP3P/«SPC/E_b» 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 — расчете нековалентных взаимодействий каждого атома с другими атомами, расположенными внутри заданного радиуса. 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_b вместо FF14SB+TIP4P-Ew связаны, прежде всего, с разным количеством атомов в моделях растворителя. Использование 3-центральной модели растворителя вместо 4-центральной 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_b 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.

Расчет на видеокартах позволяет существенно увеличить длину траектории в повседневной практике — в среднем, с десятков наносекунд до нескольких микросекунд — при сохранении адекватных настроек метода. 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 (архитектура Kepler) и Tesla P100 (архитектура Pascal), открывает принципиально новые горизонты для научной работы в области структурной биологии и биоинформатики.

Работа выполнена при поддержке грантов РФФИ №17-07-00751 и РНФ №15-14-00069-П с использованием оборудования Центра коллективного пользования сверхвысокопроизводительными 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., Швьядас 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. Molecular Dynamics Simulations and Computer-Aided Drug Discovery «Computer-Aided Drug Discovery» (Wei Zhang) Springer, New York. 2016. P. 1–31. DOI: 10.1007/7653_2015_41.
3. Shaw D.E. et al. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation // Communications of the ACM. July 2008. Vol. 51, No. 7. P. 91–97. DOI: 10.1145/1250662.1250664.
4. Shaw D.E. et al. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer // International Conference for High Performance Computing, Networking, Storage and Analysis (SC '14): Proceedings of the International Conference (New Orleans, LA, USA, November 16–21, 2014). IEEE Press. 2014. pp. 41–53. DOI: 10.1109/SC.2014.9.
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. DOI: 10.1093/bib/bbw058.
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. DOI: 10.1021/ct400314y.
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. DOI: 10.1021/ct200909j.
8. Документация к программному обеспечению GROMACS. URL: <http://manual.gromacs.org/documentation/2018/user-guide/force-fields.html> (дата обращения: 25.07.2018).
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. DOI: 10.1021/acs.jctc.5b00255.
10. Документация к программному обеспечению AMBER. URL: <http://ambermd.org/doc12/Amber17.pdf>, P. 33 (дата обращения: 25.07.2018).
11. Debic 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. DOI: 10.1021/acs.jctc.6b00567.
12. Onufriev A. Implicit Solvent Models in Molecular Dynamics Simulations: A Brief Overview // Annual Reports in Computational Chemistry. 2008. Vol. 4. P. 125–137. DOI: 10.1016/S1574-1400(08)00007-8.
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. DOI: 10.1021/jp0761564.

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. DOI: 10.1063/1.1683075.
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. DOI: 10.1021/jp301100g.
16. Pierce L. 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. DOI: 10.1021/ct300284c.
17. Suplatov D., Kopylov K., Sharipova Y., Švedas V. Human p38 α 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. Sharipova 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. Vol. 285, No. 13. P. 2428–2445. 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. DOI: 10.1371/journal.pone.0054293.
20. Суплатов Д.А., Швядас В.К. Изучение функциональных и аллостерических сайтов в суперсемействах белков // *Acta Naturae*. 2015. Т. 7, №. 4. С. 39–52.
21. Воеводин Вл.В., Жуматий С.А., Соболев С.И., Антонов А.С., Брызгалов П.А., Никитенко Д.А., Стефанов К., Воеводин В.В. Практика суперкомпьютера «Ломоносов» // *Открытые системы*. 2012. Т. 7, С. 36–39.

Суплатов Дмитрий Андреевич, к.х.н., с.н.с. НИИ физико-химической биологии имени А.Н. Белозерского, Московский государственный университет имени М.В. Ломоносова (Москва, Российская Федерация)

Шарипова Яна Александровна, аспирант факультета биоинженерии и биоинформатики, Московский государственный университет имени М.В. Ломоносова (Москва, Российская Федерация)

Попова Нина Николаевна, к.ф.-м.н., доцент факультета вычислительной математики и кибернетики, Московский государственный университет имени М.В. Ломоносова (Москва, Российская Федерация)

Копылов Кирилл Евгеньевич, аспирант факультета биоинженерии и биоинформатики, Московский государственный университет имени М.В. Ломоносова (Москва, Российская Федерация)

Воеводин Владимир Валентинович, чл.-корр. РАН, д.ф.-м.н., зав. кафедрой суперкомпьютеров и квантовой информатики факультета вычислительной математики и кибернетики, зам. директора Научно-исследовательского вычислительного центра, Московский государственный университет имени М.В. Ломоносова (Москва, Российская Федерация)

Швядас Витас Каятоно, д.х.н., профессор факультета биоинженерии и биоинформатики и НИИ физико-химической биологии имени А.Н. Белозерского, Московский государственный университет имени М.В. Ломоносова (Москва, Российская Федерация)

**MOLECULAR DYNAMICS
IN THE FORCE FIELD FF14SB IN WATER TIP4P-EW,
AND IN THE FORCE FIELD FF15IPQ IN WATER SPC/E_b:
A COMPARATIVE ANALYSIS ON GPU AND CPU**

© 2019 D.A. Suplatov, Ya.A. Sharapova, N.N. Popova, K.E. Kopylov,
Vl.V. Voevodin, V.K. Švedas

Lomonosov Moscow State University

(Leninskye gory 1-40, Moscow, 119991 Russia)

*E-mail: d.a.suplatov@belozersky.msu.ru, sharapova@belozersky.msu.ru, popova@cs.msu.ru,
kopylov@mail.chem.msu.ru, voevodin@parallel.ru, vytyas@belozersky.msu.ru*

Received: 03.08.2018

A comparative analysis of computational efficiency and scalability of molecular dynamics (MD) implemented in the AMBER package was carried out on real biological systems using the classical force field FF14SB with the 4-site water model TIP4P-Ew, as well as the new promising force field FF15IPQ with the 3-site water model SPC/E_b. The Intel Xeon E5-2697 v3 processors, as well as GPU accelerators Tesla K40 (Kepler architecture) and P100 (Pascal) were used. Reduction of the number of atoms in a cell by 25–31 % as a result of implementing a 3-site solvent model speeds up the MD calculations by up to 63 % and decreases scalability by about 11 %. The obtained results can be qualitatively different, what indicates the need for joint use of different force fields at studying biological systems. The use of GPU-accelerators as an alternative to classical CPUs provides an opportunity to significantly increase the length on MD trajectories in the daily laboratory practice.

Keywords: classical molecular dynamics, AMBER, force fields FF14SB and FF15IPQ, water models TIP4P-Ew, TIP3P and SPC/E_b, GPU-accelerators Kepler and Pascal.

FOR CITATION

Suplatov D.A., Sharapova Y.A., Popova N.N., Kopylov K.E., Voevodin Vl.V., Švedas V.K. Molecular Dynamics in the Force Field FF14SB in Water TIP4P-Ew, and in the Force Field FF15IPQ in Water SPC/E_b: a Comparative Analysis on GPU and CPU. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2019. vol. 8, no. 1. pp. 71–88. (in Russian) DOI: 10.14529/cmse190105.

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

References

1. Suplatov D.A., Popova N.N., Kopylov K.E., Shegay M.V., Voevodin Vl.V., Švedas V.K. Hybrid Computing Clusters to Study Protein Structure, Function and Regulation. *Vestnik Yuzho-Uralskogo gosudarstvennogo universiteta. Seriya "Matematicheskoe modelirovanie i programmirovaniye"* [Bulletin of South Ural State University. Series: Mathematical Modeling, Programming & Computer Software]. 2017. vol. 6, no. 4. pp. 74–90. (in Russian) DOI: 10.14529/cmse170406.

- Godwin R.C., Melvin R., Salsbury F.R. *Molecular Dynamics Simulations and Computer-Aided Drug Discovery* “Computer-Aided Drug Discovery” (Wei Zhang) Springer, New York. 2016. pp. 1–31. DOI: 10.1007/7653_2015_41.
- Shaw D.E. et al. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *Communications of the ACM*. July 2008. vol. 51, no. 7. pp. 91–97. DOI: 10.1145/1250662.1250664.
- Shaw D.E. et al. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *International Conference for High Performance Computing, Networking, Storage and Analysis (SC '14): Proceedings of the International Conference (New Orleans, LA, USA, November 16-21, 2014)*. IEEE Press. 2014. pp. 41–53. DOI: 10.1109/SC.2014.9.
- Nobile M.S. et al. Graphics Processing Units in Bioinformatics, Computational Biology and Systems Biology. *Briefings in Bioinformatics*. 2016. vol. 18, no. 5. pp. 870–885. DOI: 10.1093/bib/bbw058.
- 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. pp. 3878–3888. DOI: 10.1021/ct400314y.
- 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. pp. 1542–1555. DOI: 10.1021/ct200909j.
- Dokumentacija k programnomu obespečeniju GROMACS* [GROMACS Software Manual]. Available at: <http://manual.gromacs.org/documentation/2018/user-guide/force-fields.html> (accessed: 25.07.2018).
- 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. pp. 3696–3713. DOI: 10.1021/acs.jctc.5b00255.
- Dokumentacija k programnomu obespečeniju AMBER* [AMBER Software Manual]. Available at: <http://ambermd.org/doc12/Amber17.pdf>, P. 33 (accessed: 25.07.2018).
- 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. pp. 3926–3947. DOI: 10.1021/acs.jctc.6b00567.
- Onufriev A. Implicit Solvent Models in Molecular Dynamics Simulations: A Brief Overview. *Annual Reports in Computational Chemistry*. 2008. vol. 4. pp. 125–137. DOI: 10.1016/S1574-1400(08)00007-8.
- Wong V., Case D.A. Evaluating Rotational Diffusion from Protein MD Simulations. *The Journal of Physical Chemistry B*. 2008. vol. 112, no. 19. pp. 6013–6024. DOI: 10.1021/jp0761564.
- 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. pp. 9665–9678. DOI: 10.1063/1.1683075.
- 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. pp. 6279–6287. DOI: 10.1021/jp301100g.

16. Pierce L. et al. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *Journal of Chemical Theory and Computation*. 2012. vol. 8, no. 9. pp. 2997–3002. DOI: 10.1021/ct300284c.
17. Suplatov D., Kopylov K., Sharapova Y., Švedas V. Human p38 α 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. vol. 285, no. 13. pp. 2428–2445. 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. pp. e54293. DOI: 10.1371/journal.pone.0054293.
20. Suplatov D., Švedas V. Study of Functional and Allosteric Sites in Protein Superfamilies. *Acta Naturae*. 2015. vol. 7, no. 4, pp. 34–45.
21. Voevodin V.I.V., Zhumatiy S.A., Sobolev S.I., Antonov A.S., Bryzgalov P.A., Nikitenko D.A., Stefanov K.S., Voevodin Vad.V. Practice of the “Lomonosov” Supercomputer. *Otkrytye Sistemy* [Open Systems]. 2012. vol. 7, pp. 36–39 (in Russian).