

О.Н. ДЕМЧУК, П.А. КАРПОВ, Я.Б. БЛЮМ
Институт пищевой биотехнологии и геномики НАН Украины, Киев
E-mail: demom79@gmail.com

ДОКИНГ НИЗКОМОЛЕКУЛЯРНЫХ ЛИГАНДОВ НА МОЛЕКУЛЕ РАСТИТЕЛЬНОГО FtsZ-БЕЛКА: ПРИМЕНЕНИЕ ТЕХНОЛОГИИ CUDA ДЛЯ УСКОРЕНИЯ РАСЧЕТОВ



Приведен обзор и анализ возможностей применения технологии CUDA для ускорения расчетов в области структурной биологии и биоинформатики. На примере работы с программой Hex 6.1 выполнен сравнительный анализ прироста производительности и качества результатов жесткого докинга низкомолекулярных соединений различных классов на поверхности FtsZ белка из Arabidopsis thaliana. Идентифицировано несколько потенциальных сайтов связывания бензимидазолов с растительным FtsZ.

© О.Н. ДЕМЧУК, П.А. КАРПОВ, Я.Б. БЛЮМ, 2012

ISSN 0564–3783. Цитология и генетика. 2012. № 3

Введение. Моделирование межмолекулярных взаимодействий является одной из наиболее востребованных вычислительных задач структурной биоинформатики [1]. Традиционно для расчетов такого типа ресурсозатратных задач *in silico* применяются вычислительные платформы, основанные на использовании как одиночных CPU (Central Processing Unit), так и многопроцессорной вычислительной среды (кластерные решения, грид) [2, 3]. Сравнительно недавно, благодаря появлению технологий расчетов с использованием графических процессоров (GPU – Graphics Processing Unit) [4] и, в частности, технологии CUDA (Compute Unified Device Architecture), GPU вычисления стали применяться для решения широкого спектра научных задач [3], в том числе для выполнения биоинформационных и молекулярно-биологических расчетов [5]. Это положительно сказалось на скорости реализации отдельных вычислительных операций и привело к значительному сокращению расчетного времени, необходимого для выполнения ряда алгоритмов и биоинформационных исследований в целом.

Так, применение GPU для выравнивания последовательностей ДНК и белков сокращает время расчетов с 10 [6, 7] до 100 раз [8] по сравнению с расчетами, выполненными при использовании центрального процессора (CPU). Показано, что применение GPU позволяет сократить время, необходимое для выполнения расчетов молекулярной динамики в 10–100 раз [9, 10], квантовой химии до 130 раз [11], а вейвлет-анализа масс-спектрографических данных – более чем в 200 раз [12]. Применение GPU для расчета доступной поверхности белка в 100–300 раз сокращает время вычислений в зависимости от размеров моделируемой системы [13].

Неудивительно, что количество прикладных программ, поддерживающих возможность молекулярных расчетов с применением GPU, стремительно возрастает [14]. Безусловным лидером в области развития технологии GPU вычислений в настоящее время является фирма nVidia и разработанная этой компанией технология CUDA [15]. Практически полный список приложений, поддерживающих технологию CUDA, представлен на странице «CUDA Zone» официального web-сайта компании nVidia (<http://www.nvidia.ru/>). В настоящее время

мя список насчитывает более 20 программ, предназначенных для решения задач биоинформатики и структурной биологии.

Так, для одного из наиболее востребованных в биоинформатике инструментов — BLAST (The Basic Local Alignment Search Tool, NCBI-BLAST www.ncbi.nlm.nih.gov) представлены решения с поддержкой технологии CUDA: GPU-BLAST (<http://eudoxus.chem.cmu.edu/gpublast/>), работающий в четыре раза быстрее при идентичности результатов поиска традиционному BLASTn [16], и CUDA-BLASTP [17], предназначенный для сканирования больших массивов аминокислотных последовательностей. Кроме того, в состав упомянутых решений входит подпрограмма для конвертации последовательностей в формат, который оптимизирован для использования в приложениях BLAST, адаптированных для CUDA (http://www.nvidia.com/object/blastp_on_tesla.html) [17].

Другая программа, CUDASW++ (<http://cudasw.sourceforge.net/>), предназначена для поиска в базах данных аминокислотных последовательностей с применением алгоритма Смита-Ватермана [18, 19]. Установлено, что при использовании графических ускорителей NVIDIA GeForce (GTX 280 и GTX 295) время поиска с помощью программы CUDASW++ сокращается в 10–50 раз по сравнению с NCBI-BLAST [18].

Для такой задачи, как поиск консервативных мотивов последовательностей, на основе пакета MEME (V. 3.5.4) [20] создана версия для CUDA — CUDA-MEME (mCUDA-MEME) (http://www.nvidia.com/object/meme_on_tesla.html) [21, 17].

Для оптимизации расчетов с использованием скрытых моделей Маркова (HMM) распараллеливание на GPU реализовано в программе GPU-HMMER (<http://www.mpihmm.org/>) (http://www.nvidia.com/object/hmmer_on_tesla.html), скорость работы которой в 60–100 раз выше, чем у HMMER v.3.0 [22, 23].

Однако наиболее востребованной технология CUDA является в области структурной биоинформатики [25], где сосредоточено наибольшее количество программ (NAMD 2.7 beta, VMD 1.8.7, Amber Alpha patch, HOOMD 0.8.1, Autodock 0.9, CHARMM Beta,

GROMACS 4.0, LAMMPS Alpha и др.), предназначенных для решения структурных молекулярно-биологических задач, таких как молекулярное моделирование, молекулярный докинг, симуляция молекулярной динамики и т.д. [27–30]. В 2007 г. группа исследователей из университета Иллинойса в Урбана-Шампейн (UIUC — «Biophysics Groupe») выпустила популярный программный пакет для расчетов молекулярной динамики NAMD с поддержкой технологии CUDA (<http://www.ks.uiuc.edu/Research/gpu/>) [31]. При осуществлении расчетов с использованием пакета NAMD и ресурса NCSA четыре графических ускорителя Tesla (<http://www.nvidia.com/>) превзошли по вычислительной мощности кластер из 16 четырехъядерных CPU (<http://www.ncsa.illinois.edu/UserInfo/Resources/Hardware/Intel64TeslaCluster/>).

В настоящее время реализована интеграция NAMD с популярной программой VMD («Biophysics Groupe» — www.ks.uiuc.edu/Research/vmd/), что при использовании CUDA позволяет осуществлять анализ, 3D-визуализацию и анимацию больших молекулярных систем. VMD (Visual Molecular Dynamics) также имеет поддержку CUDA с распараллеливанием на GPU от nVidia. «Biophysics Groupe» удалось интегрировать технологию CUDA в приложения для симуляции молекулярной динамики [32], отображения молекулярных орбиталей [26, 33], расчетов позиционирования ионов [9], симуляцию кулоновских взаимодействий [5] и обработку данных микроскопии [34]. Это позволило увеличить скорость работы соответствующих приложений от 20 до 100 раз (<http://www.ks.uiuc.edu/Research/gpu/>).

Недавно была разработана CUDA-версия пакета силовых полей AMBER (<http://ambermd.org/gpus/>). Применение GPU Tesla при моделировании молекулярной динамики в поле Amber для систем с использованием явного и неявного растворителя расчетное время сокращалось более чем в 10 раз по сравнению с сервером на основе четырехъядерного CPU [35].

Пакет силовых полей Gromacs (<http://www.gromacs.org/>) разработан для моделирования динамики и минимизации энергии белков, нуклеиновых кислот и липидов, а также пред-

назначен для моделирования динамики молекулярных систем, характеризующихся большим количеством сложных взаимодействий и связей. Начиная с версии 4.5, в Gromacs включена поддержка CUDA [36]. Несмотря на некоторые ограничения (размер анализируемой системы не более 200 000 атомов, зависимость от типа/модели графической карты, совместимость RAM) (http://www.gromacs.org/Downloads/Installation_Instructions/GPUs) по сравнению с расчетами на CPU, решение Gromacs для GPU позволяет значительно сократить время расчетов [36, 37]. Следует отметить, что при использовании CUDA могут наблюдаться незначительные отличия в результатах расчетов, выполненных с помощью разных графических ускорителей, что связано с изменениями, внесенными в работу ряда алгоритмов с целью адаптации работы Gromacs на GPU [36].

Таким образом, в настоящее время технологии вычислений с применением графических процессоров выходят на новый уровень интеграции с уже существующими технологиями суперкомпьютинга (кластерные технологии, грид). Сейчас в мире существуют более 700 кластеров, работающих на основании GPU (http://www.nvidia.ru/object/what_is_cuda_new_ru.html). Удачным примером такой вычислительной инфраструктуры, использующей преимущества графических адаптеров с поддержкой технологии CUDA в сочетании с технологией грид, является проект GPUGRID.net (<http://www.gpugrid.net/>) [38], представленный распределенными вычислительными ресурсами GPU серверов, которые оснащены графическими процессорами от nVidia (в том числе вычислительных ресурсов пользователей проекта). Проект предназначен для выполнения ресурсозатратных расчетов моделирования полноатомных макромолекулярных систем, решения вопросов молекулярной биологии, биоинформатики, медицины, материаловедения [38], и предоставляет исследователям новый класс относительно недорогих, но не менее эффективных суперкомпьютеров.

Важной областью структурной биологии и биоинформатики является моделирование межмолекулярных взаимодействий, или молекулярный докинг (стыковка). Будучи неотъем-

лемым этапом рационального дизайна новых лекарственных препаратов, биологически активных веществ, реконструкции структуры макромолекулярных комплексов и механизмов их формирования [30], молекулярный докинг является набором сложных расчетов (алгоритмов) взаимодействий 3D-структур белок-белковых или лиганд-белковых комплексов, начиная с моделирования отдельных несвязанных структур соответствующих белков и лигандов [39]. Это требует применения значительных вычислительных ресурсов, таких как суперкомпьютеры и распараллеленные вычислительные системы [40]. В настоящее время имеются несколько примеров успешного применения GPU для сокращения времени выполнения докинга [41–43]. Так, есть данные об успешном применении CUDA для оптимизации работы программы AutoDock [43; <http://www.vpac.org/files/OptimizingAutodockwithCUDA.pdf>] (www.gpuautodock.sf.net), работающей на основе Ламарковского генетического алгоритма (LGA) [45].

Другим примером является пакет Hex (<http://hex.loria.fr/>), предназначенный для выполнения и визуализации результатов гибкого молекулярного докинга белков и ДНК, а также жесткого докинга низкомолекулярных лигандов и белковых мишеней [42, 43] на основе применения полярного алгоритма быстрых преобразований Фурье (FFT) [43]. Начиная с версии 6.0, в пакет Hex включена поддержка технологии CUDA [42, 43].

Поэтому целью нашего исследования стал анализ реального прироста скорости вычислений и производительности докинга за счет GPU-оптимизации работы сферического полярного алгоритма преобразований Фурье, являющегося основным ресурсозатратным этапом расчетов молекулярной стыковки в программе Hex [43], при использовании CUDA версии программного пакета Hex 6.1 на примере докинга трехмерных моделей низкомолекулярных соединений классов бензимидазолов, динитроанилинов и фенилкарбаматов на модели цитозольной формы растительного белка FtsZ2-2 (UniProt: Q9LXJ0, At3g52750) из *Arabidopsis thaliana* [46].

Материалы и методы. Молекулярный докинг осуществляли с использованием ранее

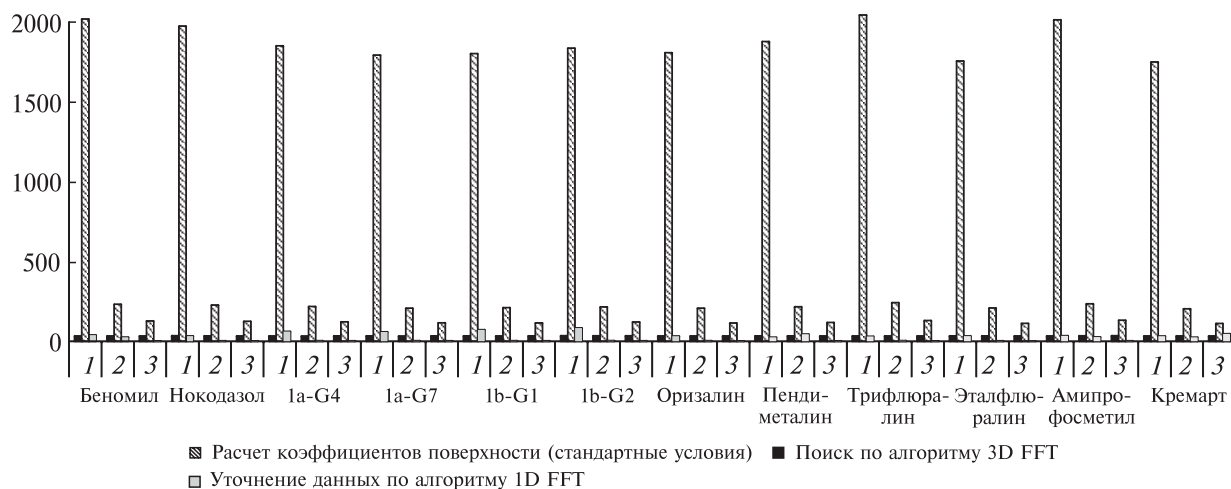
реконструированной нами модели цитозольной формы растительного белка FtsZ2-2 из *Arabidopsis thaliana* [47] и трехмерных моделей низкомолекулярных соединений класса бензимидазолов – беномила (метил-N-[1-(бутилкарбамоил)бензимидазол-2-ил]карбамат) (CID: 28780), нокодазола (метил-N-[6-(тиофен-2-карбонил)-1H-бензимидазол-2-ил]карбамат) (CID: 4122) и четырех 2,5,6-трехзамещенных бензимидазолов, которые оказались лидерами по антибактериальной активности в исследованиях группы И. Оджимы – 1a-G4, 1a-G7, 1b-G1 и 1b-G2 [48], а также представителей динитроанилинов: оризалина (4-(дипропиламино)-3,5-динитробензенсульфонамид) (CID: 29393), пендиметалина (3,4-диметил-2,6-динитро-N-пентан-3-иланилин) (CID: 38479), трифлюралина (2,6-динитро-N,N-дипропил-4-(трифторметил)анилин) (CID: 5569) и эталфлюралина (N-этил-N-(2-метилпроп-2-энил)-2,6-динитро-4-(трифторметил)анилин) (CID: 41381) и соединений класса фосфоаминов – амипрофосметила (N-[метокси-(4-метил-2-нитрофеноксифосфинотиол)-пропан-2-амин] (CID: 100524) и кремарта (N-[этокси-(5-метил-2-нитрофеноксифосфинотиол)бутан-2-амин] (CID: 37419). Исходные модели пространственных структур гербицидов взяты из репозитория базы данных «PubChemCompound» (NCBI – <http://www.ncbi.nlm.nih.gov/pccompound/>) [49]. Файлы координат трехзамещенных бензимидазолов, как и файлы топологии всех используемых лигандов, получены с помощью сервера PRODRG2 (<http://davapc1.bioch.dundee.ac.uk/prodrg/>) [50].

Вычислительный сервер собрали на основе материнской платы ASUS P5K-SE/EPU (www.asus.com), построенной на чипсете Intel P35 (www.intel.com) платформы LGA775 (FC-LGA6), с инсталляцией 4 Гб оперативной памяти Kingston DDR2-SDRAM 400 МГц. В качестве CPU использовали центральный процессор Intel Core 2 Quad Q6600 (2400 МГц), имеющий четыре логических ядра с объемами кэшей L1 кода, L1 данных и L2, равными 32 Кб, 32 Кб и 2×4096 Мб соответственно, и поддерживающий наборы инструкций x86, x86-64, MMX, SSE, SSE2, SSE3, SSSE3 (<http://ark.intel.com/>).

При выполнении молекулярных расчетов использовали видеоадаптеры ASUS EN9800GT (www.asus.com) и ZOTAC GeForce GTX 470 (www.zotac.com), работающие в программной среде 32-битной версии пакета для разработчиков «cudatoolkit 3.2.7 win32» и «cudatoolkit 3.2.16 win32» соответственно. Видеоадаптер ASUS EN9800GT построен на чипсете nVIDIA GeForce® 9800 GT и оснащен 512 Мб оперативной памяти GDDR3 SDRAM с разрядностью 256-бит. Частота памяти составляет 1800 МГц с полосой пропускания 57,6 Гб/с. Эта видеокарта имеет 112 Stream-процессоров, характеризуется частотой ядра 1200 МГц и частотой шейдерного блока 1500 МГц со скоростью наложения текстур 33,6 млрд текстур за 1 с. Видеоадаптер ZOTAC GeForce GTX 470 построен на чипсете nVIDIA GeForce® GTX 470 и оснащен 1280 Мб оперативной памяти GDDR5 SDRAM с разрядностью 320-бит, которая работает на частоте 3348 МГц с пропускной способностью 133,9 Гб/с. Видеокарта имеет 448 Stream-процессоров, работающих на частоте 1215 МГц со скоростью наложения текстур 34 млрд текстур за 1 с.

Докинг низкомолекулярных соединений рассчитывали путем поочередного использования вычислительных мощностей центрального процессора (CPU режим) и видеоадаптеров (GPU режим), установленных на указанную платформу.

Жесткий докинг [51] выполняли с использованием программного пакета Hex 6.1 (<http://hex.loria.fr/>) [43, 52] при поддержке технологии CUDA (<http://www.nvidia.com/>), который позволяет проводить вычисления как в CPU, так и в GPU режимах [43]. При выполнении расчетов во всех вариантах использовали стандартный набор параметров контроля (пункт «Docking» из меню «Controls»): Correlation Type – Shape + Electrostatic; FFT Mode – 3D; Post Processing – None; Grid Dimensions – 0.5 Å; Distance Range – 60 Å; Scan Step – 0.5 Å; Solutions – 500. Параметр FFT Mode-3D свидетельствует об использовании в процессе докинга алгоритма быстрого преобразования Фурье (FFT – fast Fourier transform) в режиме трехмерной корреляции по форме и электростатическому



Сравнение длительности выполнения отдельных операций с помощью программного пакета Hex 6.1 во время расчетов докинга низкомолекулярных лигандов на молекуле FtsZ белка с использованием вычислительной мощности ядра центрального процессора (CPU) и ядер графических ускорителей (GPU); по вертикали – время расчета, с; по горизонтали – гербициды; 1 – GPU; 2 – GPU 9800GT; 3 – GPU 470GTX

взаимодействию. Установки параметра Solutions позволяют сохранять информацию о первых 500 ориентациях лиганда [42, 53]. Остальные параметры пакета Hex использовали в режиме по умолчанию (http://hex.loria.fr/manual/hex_manual.html).

Результаты исследований и их обсуждение. Влияние аппаратной платформы на скорость молекулярных вычислений. Анализ лог-файлов процесса докирования в программе Hex 6.1 позволяет выделить два основных этапа, которые в свою очередь можно разделить еще на несколько подэтапов. В течение первого этапа происходит контурирование поверхностей молекул белка-мишени (FtsZ2-2 из *Arabidopsis thaliana*) и лиганда, отбор поверхностей молекул, их внутренних объемов, а также четырехкратный расчет коэффициентов поверхностей при стандартном значении ($N = 25$) уровня пространственной (3D) детализации.

Второй этап заключается в поиске возможных пространственных ориентаций лиганда относительно поверхности белка с использованием 3D алгоритма быстрых преобразований Фурье (3D FFT) [43] и оценке каждой из сгенерированных конформаций лиганда в составе комплекса. Последующий процесс включает расчет коэффициента вектора

ротации, запуск процесса верификации и оценку выборки конформаций, полученных в результате работы алгоритма одномерных (1-D) быстрых преобразований Фурье (1D FFT) [43].

На гистограмме показано время выполнения (этапы) докинга при использовании разных аппаратных платформ как с использованием ядер центрального процессора (CPU), так и графических ускорителей (GPU) разных поколений. Основное время выполнения первого этапа приходится на расчет коэффициентов поверхностей, поскольку предшествующие подготовительные стадии длятся очень короткий промежуток времени – порядка нескольких секунд – и не вносят существенного вклада в общее время расчета докинга. Установлено, что выполнение отдельных стадий первого этапа характеризуется одинаковой длительностью как при расчетах на CPU, так и GPU. Учитывая незначительную общую продолжительность первого этапа по сравнению со вторым (в среднем 35 с против 30–35 мин в случае с CPU и 2–4 мин в случае с GPU), можно сделать вывод об отсутствии существенных преимуществ использования графических процессоров при проведении расчетов на первом этапе докинга.

Как было отмечено, второй этап (рисунок) представлен двумя подэтапами: 1) по-

Результаты докинга низкомолекулярных лигандов
на молекуле растительного FtsZ-белка

Лиганд	Устройство расчета	Общее время докинга	Количество наилучших ориентаций	E_{\min}	E_{\max}	Количество найденных кластеров
				кДж/моль		
Бензимидазолы						
Беномил	CPU	34 мин, 35 с	1782	-167,6	-119,1	103
	GPU 9800GT	4 мин, 30 с	1681	-167,6	-119,7	102
	GPU 470GTX	2 мин, 21 с	1685	-167,6	-119,7	102
Нокодазол	CPU	33 мин, 43 с	970	-188,1	-140,0	55
	GPU 9800GT	4 мин, 02 с	940	-188,1	-140,1	56
	GPU 470GTX	2 мин, 19 с	1002	-183,0	-140,0	56
1a-G4	CPU	32 мин, 10 с	3863	-220,1	-146,6	113
	GPU 9800GT	3 мин, 47 с	3501	-220,1	-115,48	113
	GPU 470GTX	2 мин, 17 с	3503	-220,1	-115,48	113
1a-G7	CPU	31 мин, 8 с	3110	-206,4	-154,5	116
	GPU 9800GT	3 мин, 33 с	2894	-206,4	-155,4	116
	GPU 470GTX	2 мин, 10 с	2899	-206,4	-155,4	116
1b-G1	CPU	31 мин, 31 с	2534	-213,3	-151,6	139
	GPU 9800GT	3 мин, 36 с	2419	-213,3	-152,1	139
	GPU 470GTX	2 мин, 11 с	2423	-213,3	-152,1	139
1b-G2	CPU	32 мин, 17 с	2605	-226,6	-160,5	126
	GPU 9800GT	3 мин, 56 с	2477	-226,6	-161,1	126
	GPU 470GTX	2 мин, 16 с	2479	-226,6	-161,1	126
Динитроанилины						
Оризалин	CPU	30 мин, 56 с	9718	-293,1	-186,4	78
	GPU 9800GT	3 мин, 42 с	4980	-293,1	-202,0	78
	GPU 470GTX	2 мин, 08 с	4979	-293,1	-202,0	78
Пендиметалин	CPU	31 мин, 58 с	3937	-161,6	-114,9	81
	GPU 9800GT	4 мин, 33 с	3130	-161,6	-116,5	81
	GPU 470GTX	2 мин, 11 с	3133	-161,6	-116,5	81
Трифлюралин	CPU	34 мин, 51 с	1478	-193,6	-147,2	44
	GPU 9800GT	4 мин, 18 с	1424	-193,6	-147,4	44
	GPU 470GTX	2 мин, 43 с	1429	-193,6	-147,4	44
Эталфлюралин	CPU	30 мин, 04 с	1490	-184,0	-144,1	59
	GPU 9800GT	3 мин, 42 с	1430	-184,0	-144,6	59
	GPU 470GTX	2 мин, 05 с	1433	-184,0	-144,5	59
Фосфоамины						
Амипрофосметил	CPU	34 мин, 25 с	6798	-209,8	-141,5	65
	GPU 9800GT	4 мин, 34 с	4673	-209,8	-147,2	65
	GPU 470GTX	2 мин, 26 с	4675	-209,8	-147,2	65
Кремарт	CPU	29 мин, 58 с	218	-147,8	-120,8	8
	GPU 9800GT	4 мин, 01 с	193	-147,8	-121,9	7
	GPU 470GTX	2 мин, 51 с	194	-147,8	-121,9	7

Примечание. E_{\min} – электростатическая энергия наилучшей ориентации лиганда на поверхности белка-мишени, E_{\max} – электростатическая энергия наихудшей из 500 ориентаций лиганда.

иск по 3-D алгоритму быстрых преобразований Фурье (3D FFT); 2) оценка ориентаций лиганда по алгоритму одномерных быстрых преобразований Фурье (1D FFT). Из-за незначительной продолжительности (в пределах до 2 с как для CPU, так и для обеих GPU) при построении гистограммы время расчета коэффициентов векторов ротации было исключено. На основании анализа результатов расчетов нами сделан вывод о том, что наиболее продолжительной и ресурсозатратной стадией расчета докинга является 3D FFT-поиск. При использовании CUDA именно на этой стадии докинга ряда низкомолекулярных соединений на молекуле растительного белка FtsZ2-2 из *A. thaliana* происходит значительный прирост производительности расчетов за счет существенного (в среднем в 8 раз) сокращения времени расчета. Так, при использовании центрального процессора общее время расчета докинга составляет от 30 до 35 мин, тогда как использование режима CUDA сокращает это время до 2–4,5 мин в зависимости от используемого видеоадаптера. Несмотря на то, что количество процессоров в чипсете 470GTX в 4 раза больше, чем в чипсете 9800GT, переход на более новый графический ускоритель обеспечивает прирост производительности докинга всего в 1,5–2 раза.

Оценка и анализ результатов докинга. По условиям докинга программа Hex сохраняла только первые 500 лучших ориентаций лиганда. Эти ориентации организуются в кластеры, каждый из которых представляет собой координаты лиганда на поверхности белка-мишени в одном из возможных положений. Как показано в таблице, при довольно большом количестве (500 положений) допустимых ориентаций лиганда они формируют всего 40–140 кластеров. В случае с кремартром найдено всего около 200 возможных ориентаций, для которых дифференцировалось 7–8 кластеров.

Нами обнаружены отличия в количестве найденных лучших ориентаций лигандов (наиболее существенные в случае оризалина, пендиметалина и амипрофосметила) в зависимости от используемых аппаратных устройств (таблица). Так, при анализе результатов расчетов с использованием центрального процессора (Intel Core 2 Quad Q6600) количество таких

ориентаций всегда было выше по сравнению с расчетами на графических ускорителях. Это можно объяснить некоторыми отличиями в программном коде алгоритмов, оптимизированных под технологию CUDA [42, 43, 53], однако следует отметить, что упомянутые отличия не влияют на конечный результат расчетов, так как для каждого из четырех трехзамещенных бензимидазолов и всех восьми гербицидов различные аппаратные устройства генерировали одинаковое количество кластеров с конечными лучшими позициями лигандов.

Кроме того, при использовании программного пакета Hex 6.1 для первых 500 ориентаций лиганда, как правило, наблюдали совпадение энергетических показателей лучшей и худшей конформаций, рассчитанных как с использованием центрального процессора, так и графических ускорителей. При этом обнаруженные частные отличия являются незначительными, и они не могут повлиять на конечные результаты расчета докинга низкомолекулярных соединений с молекулой целевого белка.

Анализ полученных трехмерных моделей комплексов FtsZ-лиганд, сгенерированных при помощи программного пакета Hex 6.1, позволил нам идентифицировать в случае бензимидазолов несколько потенциальных сайтов их связывания с поверхностью молекулы белка FtsZ2-2 из *A. thaliana*. Следует отметить, что большинство кластеров каждого из бензимидазольных соединений локализованы в районе спирали H7 между N- и C-концевыми доменами, который аналогичен известному сайту связывания таксола молекулой β -тубулина [54]. В дальнейшем все идентифицированные нами сайты подлежат оценке на сродство с веществами бензимидазольной природы.

Выводы. На примере докинга ряда соединений класса бензимидазолов, динитроанилинов и фенилкарбаматов на молекулу белка FtsZ была подтверждена аутентичность результатов расчетов, выполненных с помощью GPU, результатам, полученным с использованием центрального процессора. Показано, что использование распараллеливания на GPU и, в частности, технологии CUDA (nVidia – www.nvidia.com) позволяет значительно со-

кратить время, необходимое для выполнения ряда молекулярных расчетов при значительном сокращении себестоимости исследований в области структурной биологии и биоинформатики. Идентифицированы также потенциальные сайты связывания бензимидазолов с поверхностью молекулы белка FtsZ2-2 из *A. thaliana*.

О.М. Демчук,
П.А. Карпов, Я.Б. Блюм

DOCKING OF LOW-MOLECULAR LIGANDS ON THE PLANT FtsZ-PROTEIN WITH APPLICATION OF CUDA-ACCELERATED CALCULATIONS

This article provides review and analysis of opportunities for application of the CUDA technology for acceleration of computations in structural biology and bioinformatics. On the example of work with the Hex 6.1 program, comparative analysis of increase in the speed and quality of results of hard-docking of a number of low-molecular compounds on the surface of the FtsZ protein from *Arabidopsis thaliana* was performed. Several potential benzimidazole – plant FtsZ protein binding sites were identified.

О.М. Демчук,
П.А. Карпов, Я.Б. Блюм

ДОКІНГ НИЗЬКОМОЛЕКУЛЯРНИХ ЛІГАНДІВ НА ПОВЕРХНІ РОСЛИННОГО FtsZ БІЛКА: ЗАСТОСУВАННЯ ТЕХНОЛОГІЇ CUDA ДЛЯ ПРИСКОРЕННЯ ОБЧИСЛЕНЬ

Наводиться огляд та аналіз можливостей застосування технології CUDA для прискорення обчислень в галузі структурної біології та біоінформатики. На прикладі роботи з програмою Hex 6.1 здійснено порівняльний аналіз приросту продуктивності та якості результатів жорсткого докінгу низькомолекулярних сполук різних класів на поверхні FtsZ-білка із *Arabidopsis thaliana*. Ідентифіковано декілька потенційних сайтів зв'язування бензимидазолів із рослинним FtsZ.

СПИСОК ЛІТЕРАТУРИ

- Ritchie D.W., Venkatraman V. Ultra-fast FFT protein docking on graphics processors // Bioinformatics. – 2010. – 26, № 19. – P. 2398–2405.
- Chunlian L., Xicheng W., Wen L., Jincheng Z., Ensheng Y. A parallel computing of drug molecular docking design // Proc. Int. Conf. on Parallel Algorithms and Computing Environments (ICPACE). – Hong Kong, 2003. – P. 229–231.
- Olšak M., Filipovič J., Prokop M. FastGrid – the accelerated autogrid potential maps generation for mo-

lecular docking // MEMICS 2009 : Fifth Doctoral Workshop on Mathematical and Engineering Methods in Computer Science. – Brno, 2009. – P. 160–167.

- Hwu W.W. GPU computing gems emerald edition (Applications of GPU Computing Series). – Waltham : Morgan Kaufmann, 2011. – 886 p.
- Owens J.D., Houston M., Luebke D., Green S., Stone J.E., Phillips J.C. GPU computing // Proc. IEEE. – 2008. – 96, № 5. – P. 879–899.
- Manavski S.A., Valle G. CUDA compatible GPU cards as efficient hardware accelerators for Smith-Waterman sequence alignment // BMC Bioinformatics. – 2008. – 9, Suppl. 2. – P. 10.
- Schatz M.C., Trapnell C., Delcher A.L., Varshney A. High-throughput sequence alignment using Graphics Processing Units // BMC Bioinformatics. – 2007. – 8. – P. 474.
- Suchard M.A., Rambaut A. Many-core algorithms for statistical phylogenetics // Bioinformatics. – 2009. – 25, № 11. – P. 1370–1376.
- Stone J.E., Phillips J.C., Freddolino P.L., Hardy D.J., Trabuco L.G., Schulten K. Accelerating molecular modeling applications with graphics processors // J. Comp. Chem. – 2007. – 28. – P. 2618–2640.
- Van Meel J.A., Arnold A., Frenkel D. et al. Harvesting graphics power for MD simulations // Mol. Simul. – 2008. – 34, № 3. – P. 259–266.
- Ufimtsev I.S., Martinez T.J. Graphical processing units for quantum chemistry // Comp. Sci. Eng. – 2008. – 10, № 6. – P. 26–34.
- Hussong R., Gregošandt A. Highly accelerated feature detection in proteomics data sets using modern graphics processing units // Bioinformatics. – 2009. – 25. – P. 1937–1943.
- Dynerman D., Butzlaff E., Mitchell J.C. CUSA and CUDE: GPU accelerated methods for estimating solvent accessible surface area and desolvation // J. Comput. Biol. – 2009. – 16, № 4. – P. 523–537.
- Farber R.M. Topical perspective on massive threading and parallelism // J. Mol. Graph. Model. – 2011. – 30. – P. 82–89.
- Sanders J., Kandrot E. CUDA by Example : An Introduction to General-Purpose GPU Programming. – Boston : Addison-Wesley, 2011. – 290 p.
- Vouzis P.D., Sahinidis N.V. GPU-BLAST: using graphics processors to accelerate protein sequence alignment // Bioinformatics. – 2011. – 27, № 2. – P. 182–188.
- Liu W., Schmidt B., Muller-Wittig W. CUDA-BLASTP: Accelerating BLASTP on CUDA-enabled graphics hardware // Comput. Biol. Bioinform. – 2011. – 8, № 6. – P. 1678–1684.
- Liu Y., Maskell D., Schmidt B. CUDASW++:

- optimizing Smith-Waterman sequence database searches for CUDA-enabled graphics processing units // BMC Res. Notes. – 2009. – **2**. – P. 73.
19. *Smith T., Waterman M.* Identification of common molecular subsequences // J. Mol. Biol. – 1981. – **147**. – P. 195–197.
 20. *Bailey T.L., Williams N., Mischel C., Li W.W.* MEME: discovering and analyzing DNA and protein sequence motifs // Nucl. Acid Res. – 2006. – **34** (Web Server issue). – W. 369–373.
 21. *Liu Y., Schmidt B., Liu W., Maskell D.L.* CUDA-MEME: accelerating motif discovery in biological sequences using CUDA-enabled graphics processing units // Pattern Recognition Lett. – 2010. – **31**, № 14. – P. 2170–2177.
 22. *Walters J.P., Balu V., Kompalli S., Chaudhary V.* Evaluating the use of GPUs in liver image segmentation and HMMER database searches // IEEE Int. Symp. Parallel & Distributed : Proc. – Rome, 2009 – P. 1–12.
 23. *Eddy S.R.* Profile hidden Markov models // Bioinformatics. – 1998. – **14**. – P. 755–763.
 24. *Eddy S.R.* HMMER User's Guide. – Ashburn : Howard Hughes Med. Inst., 2010. – 93 p.
 25. *Dunning T.H.Jr., Harrison R.J., Feller D., Xanthreas S.S.* Promise and challenge of high-performance computing, with examples from molecular modelling // Philos. Transact. A Math. Phys. Eng. Sci. – 2002. – **360**, № 1795. – P. 1079–1105.
 26. *Stone J.E., Hardy D.J., Saam J., Vandivort K.L., Schulten K.* GPU-accelerated computation and interactive display of molecular orbitals // GPU Computing Gems / Ed. W.-M. Hwu. – Waltham : Morgan Kaufmann Publ., 2011. – Ch. 1. – P. 5–18.
 27. *Stone J.E., Hardy D.J., Isralewitz B., Schulten K.* GPU algorithms for molecular modeling // Scientific Computing with Multicore and Accelerators / Eds J. Dongarra, D.A. Bader, J. Kurzak. – London, Chapman & Hall / CRC Press. – 2011. – Ch. 16. – P. 351–371.
 28. *Hardy D.J., Stone J.E., Vandivort K.L., Gohara D., Rodrigues C., Schulten K.* Fast molecular electrostatics algorithms on GPUs // GPU Computing Gems / Ed. Wen-Mei Hwu. – Waltham : Morgan Kaufmann Publ., 2011. – Ch 4. – P. 43–58.
 29. *Myung H.J., Sakamaki R., Oh K.J., Narumi T., Yasuoka K., Lee S.* Accelerating molecular dynamics simulation using graphics processing unit // Accelerating Molecular Dynamics Simulation // Bull. Korean Chem. Soc. – 2010. – **31**, № 12. – P. 3639–3643.
 30. *Gu J., Bourne P.E.* Structural Bioinformatics. – New York : Wiley-Blackwell, 2009. – 1035 p.
 31. *Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kalü L., Schulten K.* Scalable molecular dynamics with NAMD // J. Comput. Chem. – 2005. – **26**. – P. 1781–1802.
 32. *Levine B.G., Stone J.E., Kohlmeyer A.* Fast analysis of molecular dynamics trajectories with graphics processing units-radial distribution function histogramming // J. Comput. Phys. – 2011. – **230**. – P. 3556–3569.
 33. *Stone J.E., Saam J., Hardy D.J., Vandivort K.L., Hwu W.-m.W., Schulten K.* High performance computation and interactive display of molecular orbitals on GPUs and multi-core CPUs // Proc. 2nd Workshop on General-Purpose Processing on Graphics Processing Units, ACM Int. Conf. – New York, 2009. – P. 9–18.
 34. *Arkhipov A., Hüve J., Kahms M., Peters R., Schulten K.* Continuous fluorescence microphotolysis and correlation spectroscopy using 4Pi microscopy // Biophys. J. – 2007. – **93**. – P. 4006–4017.
 35. *Yang H., Zhou Q., Li B., Wang Y., Luan Z., Qian D., Li H.* GPU acceleration of Dock6's Amber scoring computation // Adv. Exp. Med. Biol. – 2010. – **680**. – P. 497–511.
 36. *Apol E., Apostolov R., Berendsen H.J.C., van Buuren A., Bjelkmar P., van Drunen R., Feenstra A., Groenhof G., Kasson P., Larsson P., Meulenhoff P., Murtola T., Pall S., Pronk S., Schulz R., Shirts M., Sijbers A., Tieleman P., Hess B., van der Spoel D., Lindahl E.* GROMACS User Manual. Version 4.5.4. www.gromacs.org. Groningen, 2010. – 348 p.
 37. *Friedrichs M.S., Eastman P., Vaidyanathan V., Houston M., LeGrand S., Beberg A.L., Ensign D.L., Bruns C.M., Pande V.S.* Accelerating molecular dynamic simulation on graphics processing units // J. Comp. Chem. – 2009. – **30**, № 6. – P. 864–872.
 38. *Buch I., Harvey M.J., Giorgino T., Anderson D.P., De Fabritiis G.* High-throughput all-atom molecular dynamics simulations using distributed computing // J. Chem. Inf. Model. – 2010. – **50**, № 3. – P. 397–403.
 39. *Halperin I., Ma B., Wolfson H., Nussinov R.* Principles of docking: an overview of search algorithms and a guide to scoring functions // Proteins. – 2002. – **47**. – P. 409–443.
 40. *Roh Y., Lee J., Park S., Kim J.-I.* A molecular docking system using CUDA. // ICHIT '09 Proc. 2009 Int. Conf. on Hybrid Information Technology. – ACM New York, 2009. – P. 28–33.
 41. *Ritchie D.W.* Recent progress and future directions in protein-protein docking. // Curr. Protein Pept. Sci. – 2008. – **9**. – P. 1–15.
 42. *Ritchie D.W.* Hex 6.3 User Manual. Vavdoeuvreles-Nancy: 2010. – 65 p.
 43. *Ritchie D.W., Venkatraman V.* Ultra-fast FFT protein docking on graphics processors // Bioinformatics. – 2010. – **26**, № 19. – P. 2398–2405.

44. Kannan S., Ganji R. Porting Autodock to CUDA // WCCI 2010 IEEE World Congress on Computational Intelligence : CCIB Abstr. – Barcelona, 2010. – P. 3815–3822.
45. Norgan A.P., Coffman P.K., Kocher J.-P.A., Katzmann D.J., Sosa C.P. Multilevel Parallelization of AutoDock 4.2 // J. Cheminform. – 2011. – 3, № 12. – P.1–9.
46. Schmitz A.J., Glynn J.M., Olson B.J.S.C., Stokes K.D., Osteryoung K.W. Arabidopsis FtsZ2-1 and FtsZ2-2 are functionally redundant, but FtsZ-based plastid division is not essential for chloroplast partitioning or plant growth and development // Mol. Plant. – 2009. – 2. – P. 1211–1222.
47. Демчук О.Н., Ныпорко А.Ю., Блюм Я.Б. Построение трехмерных моделей FtsZ-белков *Arabidopsis thaliana* на основе кристаллической структуры архебактериального комплекса FtsZ-GDP // Цитология и генетика – 2006. – 40, № 1. – С. 10–20.
48. Kumar K., Awasthi D., Lee S-Y., Zanardi I., Ruzsicska B., Knudson S., Tonge P.J., Slayden R.A., Ojima I. Novel trisubstituted benzimidazoles, targeting *mtb* FtsZ, as a new class of antitubercular agents // J. Med. Chem. – 2011. – 54. – P. 374–381.
49. Kaiser J. Science resources. Chemists want NIH to curtail database // Science. – 2005 – 308, № 5723. – P. 774.
50. Schuettelkopf A.W., van Aalten D.M.F. PRODRG – a tool for high-throughput crystallography of protein-ligand complexes // Acta Crystal. – 2004. – D60. – P. 1355–1363.
51. May F.E.B., Church S.T., Major S., Westley B.R. The closely related estrogen-regulated trefoil proteins TFF1 and TFF3 have markedly different hydrodynamic properties, overall charge, and distribution of surface charge // Biochemistry. – 2003. – 42. – P. 8250–8259.
52. Ritchie D.W. Evaluation of protein docking predictions using Hex 3.1 in CAPRI Rounds 1 and 2 // Proteins Struct. Funct. Genet. – 2003. – 52. – P. 98–106.
53. Macindoe G., Mavridis L., Venkatraman V., Devignes M.-D., Ritchie D.W. HexServer: an FFT-based protein docking server powered by graphics processors // Nucl. Acid Res. – 2010. – 38. – P. 445–449.
54. Löwe J., Li H., Downing K.H., Nogales E. Refined structure of alpha beta-tubulin at 3.5 Å resolution // J. Mol. Biol. – 2001. – 313, № 5. – P. 1045–1057.

Поступила 26.10.11