

ВЫСОКОПРОИЗВОДИТЕЛЬНЫЙ ВИРТУАЛЬНЫЙ СКРИНИНГ В ENTERPRISE DESKTOP GRID НА БАЗЕ BOINC¹

Е.Е. Ивашко, Н.Н. Никитина, С. Меллер

В работе представлен опыт разработки и использования распределенной вычислительной инфраструктуры Enterprise Desktop Grid на базе программной платформы BOINC для проведения виртуального скрининга с использованием свободно распространяемого программного обеспечения и открытых баз данных моделей химических соединений. Согласно концепции Enterprise Desktop Grid, в рамках виртуальной единой вычислительной сети объединяются неспециализированные вычислители, принадлежащие одной организации или группе пользователей, напрямую заинтересованных в решении прикладной задачи. В работе приводится описание и оценка производительности Enterprise Desktop Grid, созданной в рамках совместного научно-исследовательского проекта Института экспериментальной дерматологии при университете г. Любек (Германия) и Института прикладных математических исследований Карельского научного центра РАН. Приведены выводы о применимости подхода, а также краткий обзор полученных результатов виртуального скрининга.

Ключевые слова: Desktop Grid, Enterprise Desktop Grid, распределенные вычисления, виртуальный скрининг.

Введение

В биологии и медицине в настоящее время выделяется ряд актуальных научно-исследовательских задач, решение которых требует значительных вычислительных ресурсов. Многие из этих задач обусловлены острой необходимостью в разработке новых лекарственных средств (часть из которых может стать настоящим прорывом в лечении редких или новых заболеваний) и повышении эффективности и безопасности клинически испытанных препаратов. Этапы испытания лекарств в лаборатории и клинических условиях предваряет сложный процесс отбора и подготовки «кандидатов». В ходе этого процесса используются различные методы как на этапе выбора макромолекулы-«мишени», играющей ключевую роль в протекании заболевания, так и на этапе подбора химического соединения (лиганда), которое должно связываться с молекулой-мишенью, подавляя или усиливая ее биологические функции.

Одним из этапов разработки лекарства является виртуальный скрининг, подразумевающий обработку больших баз данных моделей химических соединений с целью подбора наиболее подходящего для мишени лиганда. Как правило, оценка степени взаимодействия определенного лиганда с молекулой-мишенью проводится на основе специальных математических моделей и не требует большого количества вычислительных ресурсов. Однако отбор пяти-восьми наиболее подходящих лигандов из нескольких десятков миллионов кандидатов требует привлечения методов и средств высокопроизводительных вычислений. При этом оценка каждого лиганда-кандидата проводится независимо от остальных, что позволяет разбить исходную задачу на миллионы не связанных между собой подзадач.

В работе представлен опыт проведения виртуального скрининга с помощью вычислительных средств Enterprise Desktop Grid на базе программной платформы BOINC. Произво-

¹Статья рекомендована к публикации программным комитетом суперкомпьютерной конференции «Научный сервис в сети Интернет — 2014».

дительность созданной вычислительной инфраструктуры проиллюстрирована результатами, полученными в ходе выполнения реального исследовательского проекта, проводимого совместно сотрудниками Института экспериментальной дерматологии (ранее — Клиники дерматологии, аллергологии и венерологии) при университете г. Любек (Германия) и Института прикладных математических исследований Карельского научного центра РАН.

Работа организована следующим образом. В разделе 1 описана прикладная задача виртуального скрининга лекарств. Раздел 2 посвящен одному из способов решения данной задачи — использованию распределенной вычислительной системы типа Enterprise Desktop Grid на базе BOINC. В заключении приводится краткий обзор результатов, полученных в работе.

1. Виртуальный скрининг

Методы, используемые для подбора лигандов, принято делить на две группы: синтез принципиально новых молекул с желаемыми свойствами и подбор лигандов на основе структурных свойств мишени. Во втором случае исследование требует перебора больших баз данных химических соединений, поэтому не может быть целиком основано на физических экспериментах. На данном этапе целесообразно провести отбор потенциальных кандидатов при помощи молекулярного докинга — компьютерного моделирования процесса связывания лиганда с мишенью в трехмерном пространстве на основе специальных математических и физических моделей. Процесс компьютерного отбора лигандов для заданной мишени называется виртуальным скринингом [11]. Его эффективность для получения биологически значимых подмножеств лигандов доказывают успешные результаты целого ряда проектов [3, 7, 10].

В качестве входных данных для виртуального скрининга используются пространственные модели молекул белка и лигандов. В процессе молекулярного докинга подбирается взаимное геометрическое расположение молекул, минимизирующее свободную энергию Гиббса для данной молекулярной системы. Затем вычисляется значение так называемой оценочной функции, выражающее силу связывания при выбранном расположении молекул. Процесс повторяется для заданного множества лигандов; для этапа физических экспериментов выбираются те лиганды, для которых была предсказана наилучшая сила связывания.

Существует целый ряд баз данных лигандов, в том числе, доступные в открытом доступе [8]. Примерами открытых баз данных моделей химических соединений, подготовленных для виртуального скрининга, являются ZINC [6] (более 35 млн. записей), ChemSpider [4] (более 30 млн. записей), CoCoCo [5] (порядка 7 млн. записей). В связи с большим объемом обрабатываемых баз данных и сложностью трехмерных моделей, проведение виртуального скрининга требует значительного количества вычислительных ресурсов. Помимо больших баз данных структурных моделей белков и лигандов, в открытом доступе также находится апробированное в ходе выполнения крупных проектов открытое программное обеспечение для молекулярного докинга.

Общую продолжительность виртуального скрининга можно проиллюстрировать на примере открытого ПО AutoDock Vina [9]. Данное ПО разработано для молекулярного докинга белковых молекул и лигандов, и по результатам специализированных тестов находится на одном уровне или превосходит по скорости и точности целый ряд эффективных программных решений, включая коммерческие.

Расчет энергии связывания 1 млн. лигандов с использованием AutoDock Vina с минимальной точностью на современном четырехъядерном персональном компьютере занимает порядка 1800 час. процессорного времени, и время расчетов растет с увеличением требуемой точности. Таким образом, проведение виртуального скрининга, например, по БД ZINC для единственной модели белка на современном настольном компьютере потребовало бы не менее 8,5 лет.

2. Enterprise Desktop Grid на базе BOINC

Проведение виртуального скрининга состоит из нескольких миллионов не связанных между собой подзадач. Эта особенность позволяет использовать не дорогостоящие мощности вычислительных кластеров, а более доступные и масштабируемые средства распределенных вычислений. В частности, ряд научно-исследовательских проектов (наиболее масштабный из которых — Docking@Home) использовали для виртуального скрининга системы Desktop Grid, основанные на принципах добровольных вычислений (volunteer computing).

Desktop Grid — это объединение в качестве единого логического «суперкомпьютера» большого количества неспециализированных вычислителей (офисных рабочих и персональных компьютеров, ноутбуков) относительно невысокой производительности. Основные достоинства технологии при соответствующей реализации — практически неограниченная масштабируемость и, следовательно, пиковая производительность, устойчивость к сбоям, минимальная стоимость создания и сопровождения. Однако такие системы подходят для решения только тех задач, которые могут быть разбиты на независимые и менее ресурсоемкие подзадачи.

Стандартом де-факто при организации добровольных вычислений является платформа BOINC (Berkeley Open Infrastructure for Network Computing) [2]. BOINC отличается простотой в установке, настройке и администрировании, обладает хорошими возможностями по масштабируемости, обеспечивает простое подключение вычислительных узлов и использование дополнительного ПО, дает возможности интеграции с другими грид-системами и др. Платформа имеет архитектуру «клиент–сервер», при этом клиентская часть может работать на компьютерах с различными аппаратными и программными характеристиками. Платформа BOINC была реализована для вычислений, необходимых для анализа космических радиосигналов, в рамках проекта SETI@Home на добровольно предоставляемых персональных компьютерах частных лиц и организаций, объединенных сетью Интернет. С момента создания на основе BOINC было реализовано множество проектов добровольных вычислений для исследований в различных отраслях наук, и на сегодняшний день суммарная пиковая мощность задействованных в вычислениях ресурсов составляет порядка 7,71 петафлопс.

Организация и сопровождение широкомасштабного проекта добровольных вычислений (наподобие Docking@Home) требуют большого внимания и сил для взаимодействия с сообществом и обеспечения функционирования проекта. Такой подход может оказаться нецелесообразным, если эксперименты проводятся нерегулярно, с промежутками на обработку результатов. Тогда более выгодной оказывается концепция Enterprise Desktop Grid, где в рамках виртуальной единой вычислительной сети собираются неспециализированные вычислители, принадлежащие одной организации или группе пользователей, напрямую заинтересованных в получении результатов экспериментов.

Enterprise Desktop Grid по сравнению с Desktop Grid, использующимися при проведении добровольных вычислений, имеют ряд преимуществ: вычислительные узлы являются надежными, что снижает уровень репликации и кворума; вычислительным процессом можно управлять централизованно; высокая скорость передачи данных по локальной сети и др. [1]. Возможности BOINC (репликация, кворумы, версионность приложений и разнообразие платформ) позволяют эффективно задействовать вычислительные ресурсы Enterprise Desktop Grid.

Для проведения виртуального скрининга в рамках совместного научно-исследовательского проекта Клиники дерматологии университета г. Любек (Германия) и Института прикладных математических исследований Карельского научного центра РАН была построена Enterprise Desktop Grid на базе BOINC. В представленной работе проводилось исследование свойств одного из белков в рамках научной работы Института экспериментальной дерматологии университета г. Любек. Для виртуального скрининга была задана пространственная модель белка и несколько целевых подмножеств лигандов из базы данных ZINC.

К концу июля 2014 г. в состав Enterprise Desktop Grid вошли 52 вычислительных узла — настольные персональные компьютеры (ПК) сотрудников Института экспериментальной дерматологии, ресурсы вычислительного кластера (незадействованные другими задачами), серверы и ПК сотрудников Института прикладных математических исследований Карельского научного центра РАН. Пиковая вычислительная мощность системы (по оценкам BOINC) составила 4801 гигафлопс. Такое количество вычислительных узлов обеспечило комфортный анализ поступающих результатов. На первом этапе экспериментов был проведен виртуальный скрининг по 8 279 803 лигандам с использованием ПО AutoDock Vina, было отобрано более 17 тыс. лигандов для дальнейших экспериментов, 41 лекарственное средство рекомендовано для тестирования в лабораторных условиях. Общая продолжительность виртуального скрининга составила 27 недель, а суммарное процессорное время расчетов — 31,7 лет. В настоящее время на Enterprise Desktop Grid продолжают расчеты нескольких десятков тысяч лигандов из целевых множеств.

На текущий момент полученные результаты позволяют говорить об успешности проекта. Отобранные лиганды-кандидаты пройдут экспериментальную проверку в лаборатории университета г. Любек.

Заключение

В работе представлены результаты разработки и апробации системы Enterprise Desktop Grid для проведения виртуального скрининга одного из белков. Для организации вычислений использована программная платформа BOINC, в качестве вычислительных узлов задействованы компьютеры Института экспериментальной дерматологии университета г. Любек (Германия) и Института прикладных математических исследований Карельского научного центра РАН. Подход показал применимость, позволив в краткие сроки организовать вычислительную сеть, и обеспечив исследователей необходимыми результатами моделирования.

На первом этапе виртуального скрининга получен ряд биологически значимых результатов и оценена вычислительная эффективность системы. В дальнейшем планируется продолжение экспериментов на основе данных, полученных из лаборатории универси-

тета г. Любек, а также разработка и реализация алгоритмов повышения эффективности Enterprise Desktop Grid в задачах виртуального скрининга.

Работа выполнена при финансовой поддержке РФФИ (проект 13-07-00008), Программы стратегического развития Петрозаводского государственного университета и Германской службы академических обменов DAAD.

Литература

1. Ивашко, Е.Е. Desktop Grid корпоративного уровня / Е.Е. Ивашко — Программные системы: теория и приложения. — 2014. — No. 1(19). — С. 183–190.
2. Anderson, D.P. BOINC: A system for public-resource computing and storage / D.P. Anderson // R. Buyya (Editor), Fifth IEEE/ACM International Workshop on Grid Computing. — 2004. — P. 4–10.
3. Breakthrough in the fight against childhood cancer. URL: http://www.worldcommunitygrid.org/about_us/viewNewsArticle.do?articleId=342 (дата обращения: 14.11.2014).
4. ChemSpider: Search and share chemistry. URL: <http://www.chemspider.com> (дата обращения: 14.11.2014).
5. Del Rio, A. CoCoCo: a free suite of multiconformational chemical databases for highthroughput virtual screening purposes / A. Del Rio, A.J. Moura Barbosa, F. Caporuscio, G.F. Mangiatordi — Molecular BioSystems. — 2010. — No. 6. — P. 2122–2128.
6. Irwin, J.J. ZINC: A Free Tool to Discover Chemistry for Biology / J.J. Irwin, T. Sterling, M.M. Mysinger, et al. — Journal of Chemical Information and Modeling. — 2012. — No. 52(7). — P. 1757–1768.
7. Matter, H. Applications and Success Stories in Virtual Screening / H. Matter, C. Sotriffer — Virtual Screening: Principles, Challenges, and Practical Guidelines (ed. C. Sotriffer). Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2011. — P. 319–358.
8. Moura Barbosa, A.J. Freely accessible databases of commercial compounds for highthroughput virtual screenings / A.J. Moura Barbosa, A. Del Rio — Current Topics in Medicinal Chemistry. — 2012. — No. 12. — P. 866–877.
9. Trott, O. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading / O. Trott, A.J. Olson — Journal of Computational Chemistry. — 2010. — No. 31. — P. 455–461.
10. Villoutreix, B.O. Structure-based virtual ligand screening: recent success stories / B.O. Villoutreix, R. Eudes, M.A. Miteva — Combinatorial chemistry & high throughput screening. — 2009. — No. 12(10). — P. 1000–1016.
11. Walters, W.P. Virtual screening an overview / W.P. Walters, M.T. Stahl, M.A. Murcko — Drug Discovery Today. — 1998. — Vol. 3, No. 4. — P. 160–178.

Ивашко Евгений Евгеньевич, к.ф.-м.н., научный сотрудник Лаборатории телекоммуникационных систем, Институт прикладных математических исследований Карельского научного центра РАН (Петрозаводск, Российская Федерация), ivashko@krc.karelia.ru.

Никитина Наталия Николаевна, стажер-исследователь Лаборатории информационных компьютерных технологий, Институт прикладных математических исследований Карельского научного центра РАН (Петрозаводск, Российская Федерация), nikitina@krc.karelia.ru.

Dr. Steffen Möller, PhD, research associate of the University of Lübeck, The Lübeck Institute of Experimental Dermatology (Lübeck, Germany), steffen.moeller@uksh.de.

Поступила в редакцию 19 ноября 2014 г.

Bulletin of the South Ural State University
Series "Computational Mathematics and Software Engineering"
2015, vol. 4, no. 1, pp. 57–63

DOI: 10.14529/cmse150105

HIGH-PERFORMANCE VIRTUAL SCREENING IN A BOINC-BASED ENTERPRISE DESKTOP GRID

E.E. Ivashko, Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences (Petrozavodsk, Russia)

ivashko@krc.karelia.ru,

N.N. Nikitina, Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences (Petrozavodsk, Russia)

nikitina@krc.karelia.ru,

S. Möller, University of Lübeck, The Lübeck Institute of Experimental Dermatology (Lübeck, Germany) steffen.moeller@uksh.de

In the paper we present the experience of developing and using an Enterprise Desktop Grid infrastructure based on BOINC software platform and designed for virtual screening using open-source software and open databases of chemical compounds models. According to the idea of Enterprise Desktop Grid, a virtual single computational network integrates non-specialized computers within an organization or group of users that are directly interested in solving an applied problem. We describe and evaluate the performance of the infrastructure created withing a joint research project between the Lübeck Institute of Experimental Dermatology (former Clinic of dermatology, allergology and venerology) of the University of Lübeck and the Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences. We summarize the project and briefly overview the results of the virtual screening.

Keywords: Desktop Grid, Enterprise Desktop Grid, distributed computing, virtual screening.

References

1. Ivashko E.E. Enterprise Desktop Grids // Programmnye Sistemy: Teoriya i Prilozheniya [Program Systems: Theory and Applications]. 2014. No. 1(19). P. 83–190.
2. Anderson D.P. BOINC: A system for public-resource computing and storage // R. Buyya (Editor), Fifth IEEE/ACM International Workshop on Grid Computing. 2004. P. 4–10.
3. Breakthrough in the fight against childhood cancer. URL: http://www.worldcommunitygrid.org/about_us/viewNewsArticle.do?articleId=342 (accessed: 14.11.2014).
4. ChemSpider: Search and share chemistry. URL: <http://www.chemspider.com> (accessed: 14.11.2014).

5. Del Rio A., Moura Barbosa A.J., Caporuscio F., Mangiatordi G.F. CoCoCo: a free suite of multiconformational chemical databases for high-throughput virtual screening purposes // *Molecular BioSystems*. 2010. No. 6. P. 2122–2128.
6. Irwin J.J., Sterling T., Mysinger M.M., et al. ZINC: A Free Tool to Discover Chemistry for Biology // *Journal of Chemical Information and Modeling*. 2012. No. 52(7). P. 1757–1768.
7. Matter H., Sotriffer C. Applications and Success Stories in Virtual Screening // *Virtual Screening: Principles, Challenges, and Practical Guidelines* (ed. C. Sotriffer). Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2011. P. 319–358.
8. Moura Barbosa A.J., Del Rio A. Freely accessible databases of commercial compounds for high-throughput virtual screenings // *Current Topics in Medicinal Chemistry*. 2012. No. 12. P. 866–877.
9. Trott O., Olson A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading // *Journal of Computational Chemistry*. 2010. No. 31. P. 455–461.
10. Villoutreix B.O., Eudes R., Miteva M.A. Structure-based virtual ligand screening: recent success stories // *Combinatorial Chemistry & High Throughput Screening*. 2009. No. 12(10). P. 1000–1016.
11. Walters W.P., Stahl M.T., Murcko M.A. Virtual screening an overview // *Drug Discovery Today*. 1998. Vol. 3, No. 4. P. 160–178.

Received November 19, 2014.