

Программная система высокопроизводительного виртуального скрининга на базе Enterprise Desktop Grid

Ивашко Е. Е.

ФГБУН Институт прикладных математических исследований
Карельского научного центра РАН

Аннотация. В работе представлен проект разработки программной системы высокопроизводительного виртуального скрининга, работающей в Enterprise Desktop Grid на базе BOINC.

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

Введение

С ростом скоростей доступа в Интернет и локальных сетей передачи данных, ростом производительности персональных компьютеров, растет и актуальность реализации вычислительноемких проектов на базе Desktop Grid. Desktop Grid — это объединение в качестве единого логического «суперкомпьютера» большого количества неспециализированных вычислителей (офисных рабочих и персональных компьютеров, ноутбуков и даже смартфонов) относительно невысокой производительности. Такой подход сужает класс доступных для эффективного решения задач. При этом основные преимущества технологии при соответствующей реализации — практически неограниченные масштабируемость и, следовательно, пиковая производительность грид-системы, устойчивость к сбоям и минимальная стоимость создания и сопровождения вычислительной сети.

Одно из перспективных направлений создания Desktop Grid — это реализация технологии распределенных вычислений в рамках локальных коммуникационных сетей предприятий и организаций, т. е. Desktop Grid корпоративного уровня или Enterprise Desktop Grid. Такой подход позволяет организации без существенных инвестиций в оборудование и инфраструктуру получить необходимые вычислительные мощности.

Desktop Grid корпоративного уровня по своим характеристикам занимают промежуточное место между сервисными Grid и классическими Desktop Grid, расширяя класс доступных приложений (благодаря повышению скорости передачи данных и возможности обмена большими объемами данных), снижая гетерогенность и вероятность отказа вычислительных узлов, а также повышая управляемость самой сети.

Виртуальный скрининг

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. 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 [1]. Его эффективность для получения биологически значимых подмножеств лигандов доказывают успешные результаты целого ряда проектов [2-4].

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 [5]. Примерами открытых баз данных моделей химических соединений, подготовленных для виртуального скрининга, являются ZINC [6] (более 35 миллионов записей), ChemSpider [7] (более 30 миллионов записей), CoCoCo [8] (порядка 7 миллионов 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 миллиона лигандов с использованием AutoDock Vina с минимальной точностью на современном 4х-ядерном персональном компьютере занимает порядка 1800 часов 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.

Программная система высокопроизводительного виртуального скрининга

В рамках проекта разрабатывается программная система высокопроизводительного виртуального скрининга, использующая для своей работы Enterprise Desktop Grid на базе

BOINC. В проекте проводится модификация стандартного BOINC-проекта для упрощения настройки, развертывания в локальной сети, запуска экспериментов и анализа результатов.

Промежуточные результаты проекта будут представлены на конференции НСКФ-2015.

Список литературы:

1. Walters, W. P. Virtual screening – an overview [Текст] / W. P. Walters, M. T. Stahl, M. A. Murcko // *Drug Discovery Today*. – 1998. – Vol.3, №4. – P. 160-178.
2. Villoutreix, B. O. Structure-based virtual ligand screening: recent success stories [Текст] / B. O. Villoutreix, R. Eudes, M. A. Miteva // *Combinatorial chemistry & high throughput screening*. – 2009. – №12(10). – P. 1000-1016.
3. Matter, H. Applications and Success Stories in Virtual Screening [Текст] / H. Matter, C. Sotriffer // *Virtual Screening: Principles, Challenges, and Practical Guidelines* (ed. C. Sotriffer). – Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2011.
4. Breakthrough in the fight against childhood cancer. [Электронный ресурс]. URL: http://www.worldcommunitygrid.org/about_us/viewNewsArticle.do?articleId=342 (дата обращения: 1.05.2014)
5. Moura Barbosa, A. J. Freely accessible databases of commercial compounds for high-throughput virtual screenings [Текст] / A. J. Moura Barbosa, A. Del Rio // *Current Topics in Medicinal Chemistry*. - 2012. - №12. - P. 866-877.
6. ZINC: A Free Tool to Discover Chemistry for Biology [Текст] / J. J. Irwin, T. Sterling, M. M. Mysinger и др. // *Journal of Chemical Information and Modeling*. - 2012. - №52(7). - P. 1757-1768.
7. ChemSpider: Search and share chemistry. [Электронный ресурс]. : <http://www.chemspider.com> (дата обращения: 1.05.2014)
8. CoCoCo: a free suite of multiconformational chemical databases for high-throughput virtual screening purposes [Текст] / A. Del Rio, A. J. Moura Barbosa, F. Caporuscio, G. F. Mangiatordi // *Molecular BioSystems*. - 2010. - №6. - P. 2122–2128.
9. Trott, O. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading [Текст] / O. Trott, A. J. Olson // *Journal of Computational Chemistry*. – 2010. – №31. – P. 455-461.
10. Anderson, D. P. BOINC: A system for public-resource computing and storage [Текст] / D. P. Anderson // R. Buyya (Editor), *Fifth IEEE/ACM International Workshop on Grid Computing*. – 2004. – P. 4-10.
11. Ивашко, Е. Е. Desktop Grid корпоративного уровня [Текст] / Е. Е. Ивашко // *Программные системы: теория и приложения*. – 2014. – №1(19). – С. 183-190.