



Е.Е. Ивашко<sup>1</sup>, Н.Н. Никитина<sup>1,2</sup>, S.Möller<sup>2</sup>

## ВИРТУАЛЬНЫЙ СКРИНИНГ В ENTERPRISE DESKTOP GRID НА БАЗЕ ПЛАТФОРМЫ VOINC\*

(<sup>1</sup>ФГБУН Институт прикладных математических исследований Карельского  
научного центра РАН, г. Петрозаводск, Россия; <sup>2</sup>University of Lübeck,  
Department of Dermatology, Lübeck, Germany)

В работе представлен опыт разработки и использования инфраструктуры Enterprise Desktop Grid на базе VOINC для проведения виртуального скрининга. Представлено описание и оценка производительности Enterprise Desktop Grid, созданной в рамках совместного научно-исследовательского проекта отделения дерматологии университета г. Любек (UKSH) и Института прикладных математических исследований Карельского научного центра РАН (ИПМИ). Приведен обзор результатов виртуального скрининга.

### **Введение**

В биологии и медицине выделяется ряд вычислительно-напряженных научно-исследовательских задач. Многие из этих задач обусловлены острой необходимостью в разработке новых лекарственных средств и повышении эффективности и безопасности клинически испытанных лекарств. 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 лиганда с мишенью проводится на основе специальных математических моделей и не требует большого количества вычислительных ресурсов. Однако отбор нескольких 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.

### **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, поэто-

---

\*Работа поддержана грантом РФФИ 13-07-00008 а



му 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. 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.

Существуют базы данных (БД) лигандов, например, ZINC (более 35 миллионов записей), ChemSpider (более 30 миллионов записей), CoCoCo (порядка 7 миллионов записей) [5]. В связи с большим объемом обрабатываемых баз данных и сложностью трехмерных моделей, проведение виртуального скрининга требует значительного количества вычислительных ресурсов. Помимо больших БД 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 [6]. Расчет энергии связывания 1 миллиона лигандов 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 по БД ZINC для единственной модели белка на настольном компьютере потребовало бы не менее 8.5 лет.

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 – Docking@Home) использовали для виртуального скрининга системы Desktop Grid, основанные на принципах добровольных вычислений (volunteer computing).

Desktop Grid – это объединение в качестве единого логического «супер-компьютера» 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) [7]. 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.

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 EDG, где в рамках виртуальной единой вычислительной сети собираются неспециализированные вычислители, принадлежащие одной организации или группе пользователей, напрямую заинтересованных в получении результатов экспериментов.

EDG по сравнению с Desktop Grid, используемыми 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. [8]. Возможности BOINC (репликация, кворумы, версионность приложений и разнообразие платформ) позволяют эффективно задействовать вычислительные ресурсы EDG.

Для проведения виртуального скрининга в рамках совместного научно-исследовательского проекта между UKSH и ИПМИ была построена EDG на базе BOINC. В представленной работе проводилось исследование свойств одного из белков в рамках научной работы UKSH. Для виртуального скрининга была задана пространственная модель белка и несколько целевых подмножеств лигандов из базы данных ZINC.

В состав EDG вошли 42 вычислительных узла – настольные ПК сотрудников клиники, ресурсы вычислительного кластера (незадействованные другими задачами), серверы и настольные ПК сотрудников ИПМИ. Такое количество вычислительных узлов обеспечило комфортный анализ поступающих результатов. На первом этапе экспериментов был проведен виртуальный скрининг по 4 293 345 лигандам с использованием ПО AutoDock Vina, было отобрано 17 360 лигандов для дальнейших экспериментов, 41 лекарственное средство рекомендовано для тестирования в лабораторных условиях. Пропускная способность Desktop Grid составила 2762 лиганда в час. Общая продолжительность виртуального скрининга составила 15 недель, суммарное процессорное время расчетов за вычетом избыточных вычислений – 8.5 лет.



### **Заключение**

В работе представлены результаты разработки и апробации системы EDG для проведения виртуального скрининга. Для организации вычислений использована программная платформа BOINC, в качестве вычислительных узлов задействованы компьютеры UKSH и ИПМИ. 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 университета г. Любек, а также разработка и реализация алгоритмов повышения эффективности EDG в задачах виртуального скрининга.

### **Литература**

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](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. 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.
7. 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.
8. Ивашко, Е. Е. Desktop Grid корпоративного уровня [Текст] / Е. Е. Ивашко // *Программные системы: теория и приложения*. – 2014. – №1(19). – С. 183-190.