УСОВЕРШЕНСТВОВАНЫЙ АЛГОРИТМ ДЛЯ РАСПРЕДЕЛЁННЫХ ВЫЧИСЛЕНИЙ ПРИ МОДЕЛИРОВАНИИ ФОЛДИНГА БЕЛКОВЫХ МОЛЕКУЛ

Московский А.А. Вановский В.В. Грановский А.А. Немухин А.В.
Московский Государственный Университет, Химический факультет, г. Москва.

Московский А.А.
Вановский В.В.
Грановский А.А.
Немухин А.В.
Московский Государственный Университет им. М.В. Ломоносова

В последние годы для моделирования фолдинга белковых молекул применяют распределённые вычислительные комплексы, состоящие из нескольких тысяч компьютеров, подключённых к глобальной сети Интернет – проекты Folding@Home[1], MD@Home[2]. При этом удаётся адекватно описать динамику сравнительно небольших пептидных цепей с использованием атомно-силовых полей для вычисления потенциальной энергии системы.

Проблема заключается в том, что для описания динамики формирования белковых глобул требуется моделировать траекторию движения молекулы длиной около миллисекунды. С использованием метода молекулярной динамики, на современных персональных компьютерах такое моделирование заняло бы несколько сотен лет. Использование суперкомпьютеров с массовым параллелизмом (до сотен или даже тысяч процессоров) позволяет достичь ускорения в несколько сотен раз, в то время как для практического применения расчёт необходимо ускорить как минимум ещё на порядок.

Для решения этой задачи применяется исследование нескольких десятков тысяч независимых молекулярно-динамических траекторий длиной порядка нескольких десятков наносекунд, что позволяет организовать параллельное моделирование этих траекторий на распределённых вычислительных системах, включающих до 40000 компьютеров [1]. Длина единичной траектории ограничивает область применения этого подхода – если для свертывания белковой цепи в глобулу требуется большее время, то правильная конформация белковой молекулы не будет найдена.

В данной работе развивается двухуровневый подход к распараллеливанию вычислений в методе молекулярной динамики. Для увеличения длины отдельной молекулярно-динамической траектории предлагается проводить расчёт энергии, сил и следующего шага по времени также при помощи параллельных вычислений. Это значительно усложняет как написание параллельной программы, так и процесс управления ресурсами в распределённой вычислительной системе. Использование общепринятых на сегодняшний день средств создания параллельных программ (MPI) становится практически невозможным. Значительно более эффективным оказывается использование средств, специально созданных для использования в гетерогенных вычислительных средах, таких как Т-система.

Создаваемый алгоритм управления ресурсами должен принимать в расчёт «удалённость» компьютеров друг от друга. В процессе вычислений будут создаваться задания двух типов

а) расчёт участка молекулярно-динамической траектории

б) расчёт сил и вклада в энергию и силы, действующие на атомы, для фрагмента белковой молекулы.

Задания второго типа, порождаемые в процессе выполнения заданий первого типа, должны передаваться лишь машинам, способным достаточно эффективно взаимодействовать друг с другом по сети. Предполагается проверить эффективность создаваемого алгоритма и его работоспособность в реальной сети Интернет при помощи полигона для тестирования распределённых приложений PlanetLab (www.plant-lab.org).

Проект осуществляется при поддержке компании Интел.

Литература

1. *B. Zagrovic, C. D. Snow, M. R. Shirts and V. S. Pande*. Simulation of Folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing.// Journal of Molecular Biology, 323, 927-937 (2002)
2. http://www.md-at-home.ru/description.htm