

Лекция 5: Методы численного моделирования биологических систем: Введение

1 Компьютерное моделирование как новый инструмент для научных исследований

Компьютерное моделирование берёт своё начало со времён развития первых компьютеров в 40-50х годах прошлого века, а первое применение компьютеров для Молекулярной Динамики датируется 1956 годом. В наше время, компьютерное моделирование стало стандартным инструментом, используемым исследователями для большого спектра научных задач. Нужда в применении вычислительного эксперимента возникает из-за большого числа научных задач, которые не поддаются аналитическому решению (с помощью "карандаша и бумаги").

Компьютерное моделирование тесно связано как с натурным экспериментом так и с теоритическими исследованиями: Эксперимент \leftrightarrow Моделирование \leftrightarrow Теория.

1. Модели, используемые в компьютерном моделировании базируются на теоритическом приближении. Соответственно, моделирование может быть использовано для проверки теории.
2. Компьютерное моделирование само по себе не предлагает физического понимания. Результаты моделирования необходимо обрабатывать при помощи теоритических знаний.

3. Компьютерное моделирование не может полностью заменить эксперимент, так как в моделировании результат зависит от изначально заданной модели. Поэтому, моделирование очень сильно связано с экспериментом и требует экспериментальных данных для проверки результата. Часто, эксперимент предоставляет начальные условия для компьютерного моделирования.

Моделирование может быть использовано для предсказания свойств материалов в экстремальных условиях (высокой температуре, давлении) или для изучения свойств сложных систем (например, сворачивание белков или их агрегация). Последнее особенно важно, так как благодаря этим возможностям моделирования, оно сейчас находится на острие научных исследований.

2 Молекулярное моделирование

Рентгеновская кристаллография может предоставить неподвижную картину взаимного расположения аминокислот в нативном состоянии белка. Однако в физиологической среде большинство белков могут менять свою структуру благодаря локальным движениям атомов белка под действием обычной внешней среды (нормальная кислотность, температура, давление, ионная сила) [1]. Более того, во многих случаях локальные структурные изменения (например, переход из одного подсостояния в другое) могут играть столь же важную роль в биологической функции белка, что и глобальные конформационные изменения (такие как диссоциация белок-белковых комплексов или денатурация белков). В некоторых случаях удаётся получить несколько промежуточных структур, кристаллизуя белок при различных условиях среды или используя методы Ядерного Магнитного Резонанса (ЯМР). Но даже в таких случаях получить информацию о механизме того или иного конформационного изменения, его кинетических и термодинамических свойствах практически невозможно. Вычислительные методы, такие как Молекулярная Динамика (МД) и Динамика Ланжевена (ДЛ), способны заполнить этот пробел, предоставив динамическую картину микромолекулярных движений, дополнив экспериментальные результаты. Эти продвину-

тые вычислительные методы могут предоставить нам максимально детальные знания о динамике биологических молекул, движении их структурных и функциональных фрагментов, поведении элементов вторичной структуры [1-5].

Как правило, при моделировании биологической задачи N тел, приходится искать компромисс между уровнем детализации и скоростью расчёта. Выбор правильного уровня детализации позволяет исследователям решать конкретные задачи, возникающие для той или иной биологической системы. Для некоторых из них необходимо использовать более детальное полноатомное приближение (МД), иногда же более точные результаты можно получить при помощи упрощённой модели (ДЛ). В то время, как МД более точный подход к моделированию, ДЛ позволяет моделировать на биологически-важных временных интервалах (мс-с). Численные процедуры, используемые в обоих методах, очень похожи. Взаимодействия между частицами (например, силы Ван-дер-Ваальса, электростатические взаимодействия, ковалентные связи) описываются при помощи потенциальной функции (силового поля). Эта функция зависит от координат всех частиц (атомов и групп атомов), их типов, зарядов и описываемых степеней свободы (вибраций, поворотов, торсионных вращений и т.д.). Используя силовое поле, рассчитываются атомарные и молекулярные силы, действующие на все частицы в системе. Затем, частицы перемещаются согласно законам движения (Ньютона или Ланжевена).

Так как молекулярное моделирование основывается на эмпирической функции взаимодействия между частицами, результаты моделирования должны быть сравнены с тем или иным экспериментом. Это делается для того, чтобы убедиться в точности параметров и функций силового поля. Но, в силу того, что временные шкалы эксперимента (мс-мин) и доступная при моделировании в полноатомном разрешении (нс-мкс) не совпадают, прямое сравнение провести практически невозможно. С ростом вычислительных возможностей компьютерной техники, становится возможным увеличить время моделирования, что иногда позволяет найти новые недостатки современных силовых полей [6, 7]. Последние постоянно претерпевают изменения: появляются новые наборы параметров, изменяется функциональная часть, вводится зависимость между параметрами разных частей потен-

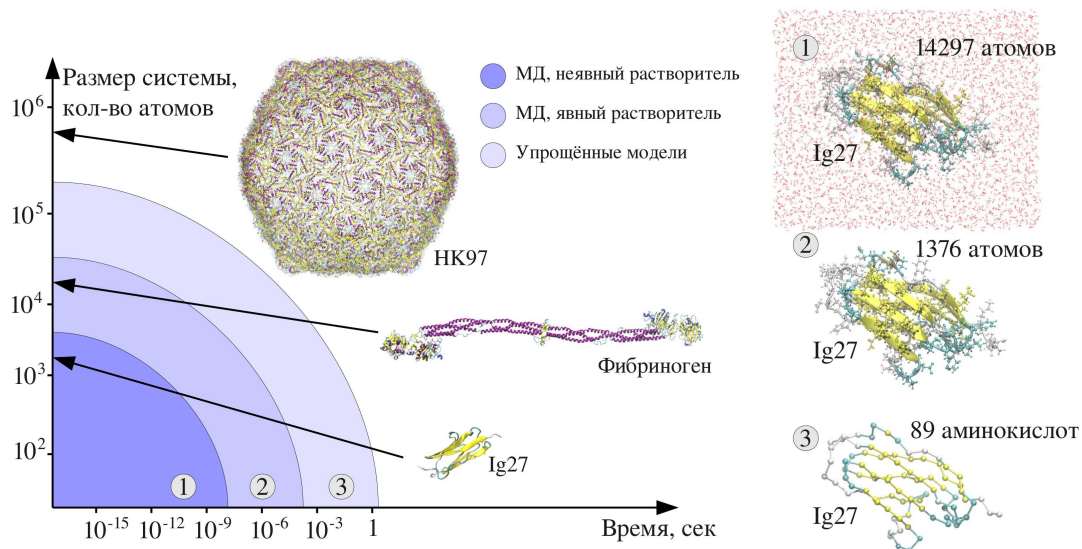


Рис. 1: Примерные временные интервалы достигаемые на одном ЦП для различных размеров систем и разных способов моделирования. На графике изображены различные системы: домен титина *Ig27*, мономер фибриногена и капсид *НК97*, для которых стрелкой показаны примерные размеры в атомах. Справа показаны размеры систем, получаемых при моделировании в явном растворителе (сверху), в неявном растворителе (посередине) и при использовании упрощённой C_α -модели. Видно, системы размера фибриногена уже невозможно моделировать на экспериментальном временном интервале с использованием ЦП.

циальной функции [8]. Кроме улучшения аппаратной части, можно применять и другие способы увеличения производительности вычислений. Среди таких способов использование неявного растворителя и упрощение моделей биомолекул. При использовании упрощённых моделей число рассматриваемых степеней свободы уменьшается, так как остаются только те, которые важны для решаемой задачи. Обычно в упрощённых моделях одна аминокислота белка или одна нуклеиновая кислота ДНК/РНК рассматривается как отдельная частица или пара частиц [9–11]. В силу того, что при этом структурные единицы системы увеличиваются в размерах, шаг по времени также может быть увеличен. Из-за того, что при применении данного подхода вычислительная точность ниже, чем в полноатомном разрешении, упрощённые модели часто разрабатываются для моделирования конкретных

свойств биомолекул. Например, модель SOP была разработана для описания механических свойств [11–14], существуют также модели для описания белок-белковых взаимодействий [15], фолдинга белков [9, 16] и так далее. Использование неявного растворителя также позволяет уменьшить количество степеней свободы. В данных схемах механическое воздействие воды описывается при помощи случайной силы, действующей на атомы системы, а энергетическое воздействие (гидрофобный эффект, заземление электростатических взаимодействий) описывается при помощи дополнительной потенциальной функции. При этом межатомные взаимодействия описываются также, как и когда вода представлена явно (при моделировании в явном растворителе). Применение неявного растворителя не только помогает ускорить вычислительный процесс, он также более привлекателен для некоторых задач, так как можно явно оценить энергетическую составляющую действия растворителя на биомолекулу, а также быстрее менять её макросостояния.

Список литературы

- [1] M. Karplus and J. A. McCammon, “Molecular dynamics simulations of biomolecules,” *Nature Struct. Biol.*, vol. 9, no. 9, pp. 646–652, 2002.
- [2] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, “Scalable molecular dynamics with NAMD,” *J. Comput. Chem.*, vol. 26, pp. 1781–1802, 2005.
- [3] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, “GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation,” *J. Chem. Theory Comput.*, vol. 4, pp. 435–447, 2008.
- [4] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, “CHARMM: The biomolecular simulation program,” *J. Comput. Chem.*, vol. 30, no. 10, pp. 1545–1614, 2009.
- [5] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, “The Amber biomolecular simulation programs,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1668–1688, 2005.
- [6] A. D. MacKerell, “Empirical force fields for biological macromolecules: Overview and issues,” *J. Comput. Chem.*, vol. 25, no. 13, pp. 1584–1604, 2004.
- [7] R. B. Best, N.-V. Buchete, and G. Hummer, “Are current molecular dynamics force fields too helical?,” *Biophys J.*, vol. 95, no. 1, pp. L07–L09, 2008.
- [8] T. A. Halgren and W. Damm, “Polarizable force fields,” *Curr. Opin. Struct. Biol.*, vol. 11, no. 2, pp. 236 – 242, 2001.
- [9] V. Tozzini, “Coarse-grained models for proteins,” *Curr. Opin. Struct. Biol.*, vol. 15, no. 2, pp. 144 – 150, 2005.

- [10] C. Clementi, “Coarse-grained models of protein folding: toy models or predictive tools?,” *Curr. Opin. Struct. Biol.*, vol. 18, no. 1, pp. 10–15, 2008.
- [11] C. Hyeon, R. I. Dima, and D. Thirumalai, “Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins,” *Structure*, vol. 14, pp. 1633–1645, 2006.
- [12] C. Hyeon and J. N. Onuchic, “Internal strain regulates the nucleotide binding site of the kinesin leading head,” *Proc. Natl. Acad. Sci. USA*, vol. 104, pp. 2175–2180, 2007.
- [13] R. I. D. M. Mickler, H. Dietz, C. Hyeon, D. Thirumalai, and M. Rief, “Revealing the bifurcation in the unfolding pathways of GFP using single molecule experiments and simulations,” *Proc. Natl. Acad. Sci. USA*, vol. 104, pp. 20268–20273, 2007.
- [14] R. I. Dima and H. Joshi, “Probing the origin of tubulin rigidity with molecular simulations,” *Proc. Natl. Acad. Sci. USA*, vol. 105, no. 41, pp. 15743–15748, 2008.
- [15] Y. C. Kim and G. Hummer, “Coarse-grained models for simulations of multiprotein complexes: Application to ubiquitin binding,” *J. Mol. Biol.*, vol. 375, no. 5, pp. 1416 – 1433, 2008.
- [16] S. Wallin, K. B. Zeldovich, and E. I. Shakhnovich, “The folding mechanics of a knotted protein,” *J. Mol. Biol.*, vol. 368, no. 3, pp. 884 – 893, 2007.