

Отчет о проведенных расчетах фолдинга белка виллина методом молекулярной динамики с использованием ресурсов информационно-вычислительного центра НГУ

Урсул А. С.

1. **Тема работы:** «Моделирование фолдинга белка виллина методом молекулярной динамики»
2. **Состав коллектива:**
 - Урсул Алексей Сергеевич, студент МедФ НГУ 4 курса – программист, исполнитель.
 - Бакулина Анастасия Юрьевна, к.б.н., зав.лабораторией структурной биоинформатики и молекулярного моделирования кафедры молекулярной биологии ФЕН НГУ – научный руководитель.
3. **Научное содержание работы:**
 - 3.1. **Постановка задачи:** На основе рассчитанных траекторий сворачивания белка виллина методами молекулярной динамики с образованием стабильных третичных структур, выполненных на суперкомпьютере НГУ, планируется провести анализ полученных состояний белка, и на основе сравнения по гомологии с третичными структурами известных белков из RCSB PDB (Protein Data Base) проверить метод поиска нативной структуры белка по его околонативному состоянию с использованием обширной базы гомологов. В рамках работы планируется выявить, имеется ли корреляция (и если да, то в какой степени) между первичной структурой белка (его аминокислотной последовательностью) и его полученной околонативной формой – промежуточной третичной формой, конфигурацию которой можно рассчитать на современных суперкомпьютерах.
 - 3.2. **Современное состояние проблемы:**

В данный момент одним из методов определения третичной структуры белка с известной аминокислотной последовательностью является моделирование по гомологии. Его суть заключается в сопоставлении аминокислотных последовательностей изучаемого белка и базы данных уже известных на данный момент белков из RCSB Protein Data Bank по существующим алгоритмам (COMA, COMPASS, PRC, CS-BLAST, HHsearch, HHblits, Mgenthreader, Sp3, Blastp, Pdbblast, Ffas-03, Pcons 5, Phyre, Fugue 2.0). Моделирование по гомологии становится возможным благодаря тому, что количество третичных белковых структур относительно небольшое и в данный момент уже практически не пополняется, при том, что количество первичных структур бесконечно большое.

Однако данный метод имеет свои недостатки - гомологичные структуры находятся не всегда, а уровень их соответствия бывает крайне низким.

Мы решили исследовать другой подход. Существуют программные пакеты для моделирования молекулярной динамики (Amber, Gromacs), которые позволяют смоделировать процесс сворачивания (фолдинга) белка в заданном растворе. Показано [1], что многие белки достаточно быстро сворачиваются до так называемой околонативной формы - третичной структуры, близкой к нативной, но не являющейся ею (для моделирования истинно нативной формы нужно значительно больше вычислительных мощностей). Мы предполагаем, что данную околонативную третичную структуру возможно использовать для достаточно достоверного предсказания нативной структуры. Это бы значительно облегчило распознавание третичных структур неизвестных ранее белков.

1. Shaw D. E. et al. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation (2008). Communications of the Acm 51, No. 7, 91.

3.3. Подробное описание работы, включая используемые алгоритмы:

Моделирование проводилось в средах специализированных пакетов Gromacs v4.6.7 с полем сил Amber03 (оптимальное поле выбиралось на основе результатов других исследователей [2]), и Amber с полями сил Amber03 и Amber14 в явном и неявном растворителе при температуре 300 К.

Процесс укладки белка в неявном растворителе запускался из развёрнутого состояния (предварительно денатурировали белок при 600 К).

Для моделирования в явном растворителе были использованы как развёрнутые структуры белка, так и структуры, полученные в процессе моделирования в явном растворителе.

Количество молекул воды в моделях варьирует от 2241 до 6469 в зависимости от размера бокса, используется модель воды TIP3P, для нейтрализации раствора в бокс добавлено 2 атома хлора.

1. Kubelka J, Eaton WA, Hofrichter J., "Experimental tests of villin subdomain folding simulations", J Mol Biol. 2003 Jun 13;329(4):625-30.
2. Irene Maffucci, Alessandro Contini, "An Updated Test of AMBER Force Fields and Implicit Solvent Models in Predicting the Secondary Structure of Helical, β -Hairpin, and Intrinsically Disordered Peptides", J. Chem. Theory Comput., 2016, 12 (2), pp 714–727
3. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE (2012) Systematic Validation of Protein Force Fields against Experimental Data. PLoS ONE 7(2): e32131. doi:10.1371/journal.pone.0032131

3.4. Полученные результаты

Методом молекулярной динамики было проведено моделирование укладки виллина с известной нативной пространственной структурой (pdb id: 1vii). С использованием Gromacs в неявном растворителе было получено 5 траекторий средней продолжительностью 6124 нс (макс. продолжительность

– 16211 нс) с минимальным среднеквадратичным отклонением (RMSD) от модельной структуры (pdb id: 1vii) в 3.831 Å.

С использованием Amber в неявном растворителе было получено 7 траекторий средней продолжительностью 5861 нс (макс. продолжительность – 10795 нс) с минимальным среднеквадратичным отклонением от модельной структуры в 2.863 Å. В результате моделирования в явном растворителе было получено 9 траекторий средней длительностью 2207 нс (макс. длительность – 2960 нс) с минимальным среднеквадратичным отклонением от модельной структуры в 1.778 Å.

На данный момент все расчёты на кластере завершены.

3.5. Иллюстрации, визуализация результатов

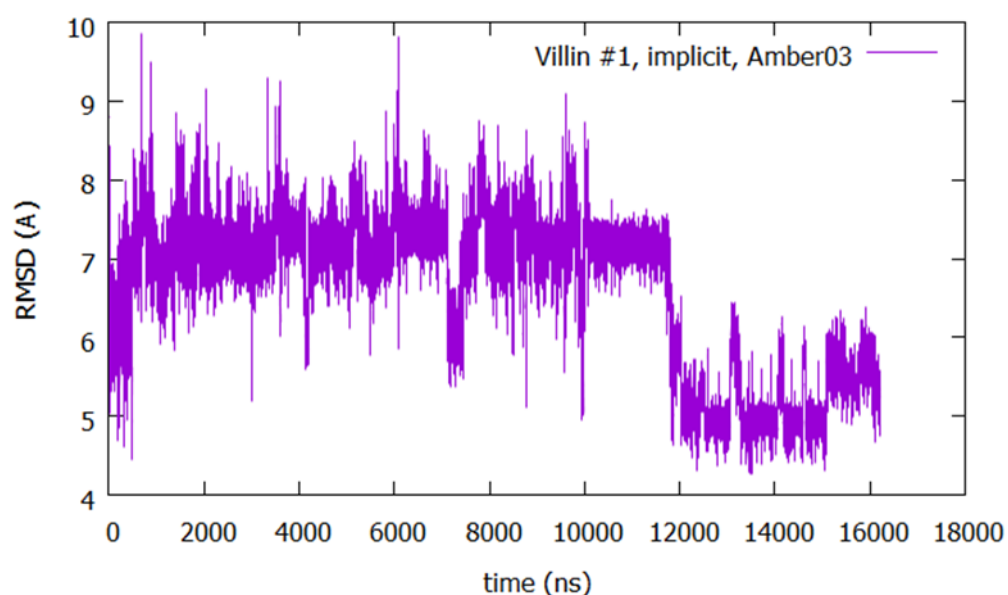


Рисунок 1. График зависимости среднеквадратичного отклонения (RMSD) от времени одной из траекторий

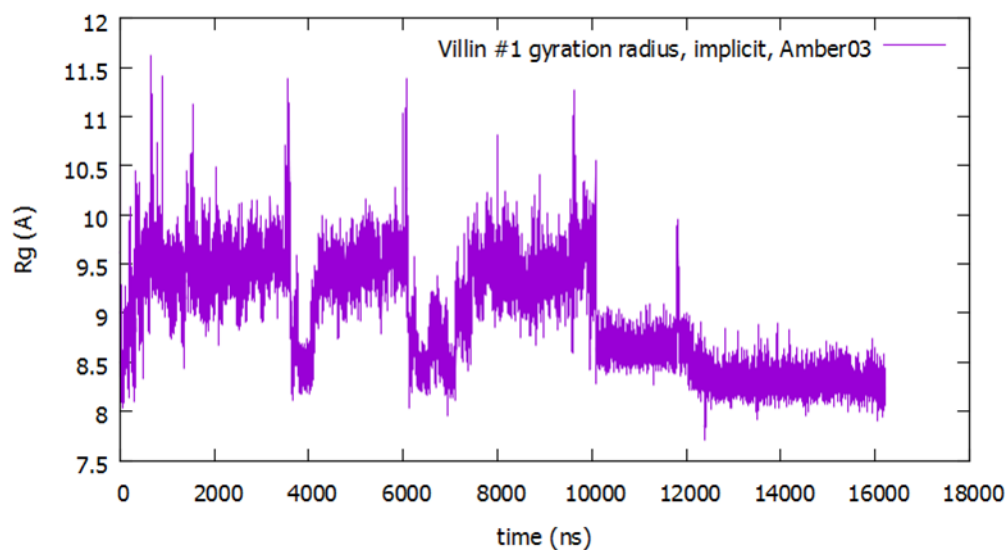


Рисунок 2. График зависимости радиуса инерции (R_g) от времени одной из траекторий

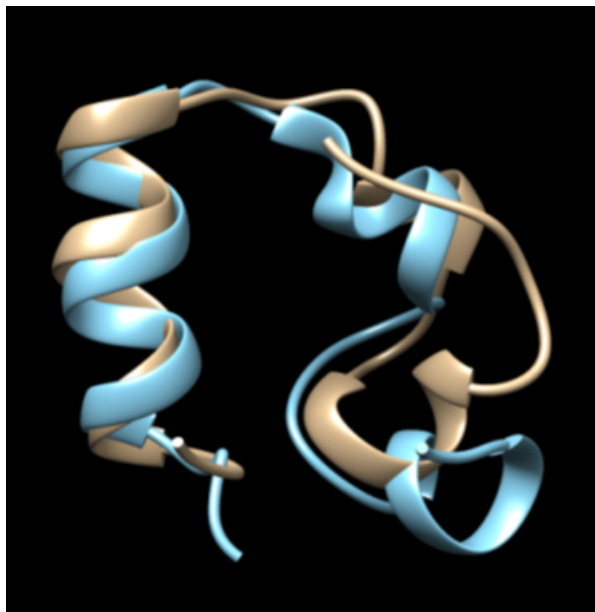


Рисунок 3. Визуальное сопоставление экспериментальной структуры (бежевая) и модельной (голубая)

4. Эффект от использования кластера в достижении целей работы

Вычислительные мощности ИВЦ НГУ позволили выполнить расчёты и наработать достаточно траекторий, необходимых для следующего этапа работы – анализа и сопоставления с существующими траекториями из PDB.

В качестве промежуточного результата работы на МНСК-2016 был представлен доклад «Моделирование фолдинга белка виллина методом молекулярной динамики» в секции «Биология» (подсекция «Биоинформатика») с занятием призового места (3-е место).