

TAMD
TRAJECTORY ANALYZER OF MOLECULAR DYNAMICS

Manual

TAMD
АНАЛИЗАТОР ТРАЕКТОРИЙ МОЛЕКУЛЯРНОЙ ДИНАМИКИ

Полное руководство

Содержание

1	Основные возможности	5
1.1	Отличия от VMD	5
2	Общий порядок работы с ПО	5
2.1	Открытие файлов	7
2.2	Перемещение вдоль траектории. Молекулярное кино	7
2.3	Расчет общих характеристик	8
2.4	Средство работы с графиками	9
3	Дополнительные режимы работы	9
3.1	Выделение групп атомов	9
3.2	Визуализация расширенных траекторий	10
3.4	Расширенные динамические контактные карты	11
3.5	Квазигармонический анализ	11
3.7	Визуализация траекторий с периодическими граничными условиями	13
3.8	Визуализация полостей	15
4.	Алгоритмы и их реализация	17
4.1	Чтение файлов траекторий молекулярной динамики	17
4.1.1	Структура траекторного файла	17
4.1.2	Буфер случайного доступа к траектории	18
4.1.3	Чтение траекторий из баз данных	19
4.2	Работа с PDB-файлом	20
4.3	Характеристики вдоль траектории	20
4.3.1	Расстояние между атомами (функция Distance)	21
4.3.2	Угол между атомами (функция Angle от 3 атомов)	21
4.3.3	Радиус инерции системы (функция R)	21
4.3.4	Угол между атомами (функция Angle от 4 атомов)	21
4.3.5	Угол между двумя подсистемами	22
4.3.6	Расстояние между диапазонами атомов (функция Distance от 4 атомов)	22
4.3.7	Флуктуации (колебания) аминокислотных остатков (функция Residue's fluctuations)	22
4.3.8	Угол между векторами (Atom1-Atom2, Atom3-Atom4) (функция Angle от 4 атомов)	22
4.3.9	Угол между проекциями аминокислот	22
4.3.10	Угол между четырьмя остатками (Acid2-Acid1, Acid3-A4)	22
4.3.11	Флуктуации C-alpha-атомов (C-alpha fluctuations)	22
4.3.12	Флуктуации C-alpha-атомов в координатах центра тяжести по всей траектории (C-alpha fluctuations by R (all trajectory))	23
4.3.13	Флуктуации C-alpha-атомов в координатах центра тяжести между кадрами a1 и a2 (C-alpha fluctuations by R (frames a1 and a2))	23
4.3.14	Флуктуации всех атомов в координатах центра тяжести (All atoms fluctuations by R (all trajectory))	23
4.3.15	Флуктуации всех атомов в координатах центра тяжести (All atoms fluctuations by R (frames a1 and a2))	23
4.3.16	Флуктуации заданных аминокислотных остатков в координатах центра тяжести (All atoms fluctuations by R (all trajectory SELECTED atoms))	23
4.3.17	Флуктуации заданных аминокислотных остатков в координатах центра тяжести (All atoms fluctuations by R (frames a1 and a2 SELECTED atoms))	23
4.3.18	Флуктуации C-alpha-атомов в координатах центра тяжести между кадрами a1 и a2 (C-alpha fluctuations by R (frames from a1 to a2))	24

4.3.19 Расстояния между парами C-Alpha-атомов для кадра a1 (C-alpha pairs distances for frame a1)	24
4.3.20 Карта Рамачандрана (Ramachandran map (phi, psi))	24
4.4 Средство работы с графиками TAMD-Grapher	24
4.4.1 Вывод графиков	25
4.4.2 Изменение масштаба графиков	25
4.4.3 Изменение цвета графиков	25
4.4.4 Изменение представления графиков	25
4.4.5 Сохранение графика в виде рисунка	26
4.4.6 Экспорт графика в Excel	26
4.4.7 Плотность распределения	26
4.4.8 Быстрое преобразование Фурье	27
4.4.9 Выход на плато	27
4.5 Контактные карты	27
4.5.1 Булева контактная карта	28
4.5.2 Построение монохромной контактной карты	28
4.5.3 Расширение возможностей контактных карт при анализе структурированных систем при помощи построения полноцветной контактной карты	29
4.5.5 Программная реализация	30
4.5.6 Пример использования	30
4.5.7 Методика проведения исследования. Интерпретация результатов	31
4.5.8 Динамика	32
4.5.9 Ускорение вывода контактных карт	32
4.5.10 Ленты контактов	32
4.5.10 О перспективах программы TAMD	35
4.6 Коэффициент вариации молекулярной динамики. Распознавание процессов на исследуемой системе без предварительных сведений	35
4.6.1 Использование коэффициента формы при анализе траекторий молекулярной системы	35
4.6.2 Алгоритм	36
4.6.3 Исследование protein L при помощи коэффициента вариации молекулярной динамики	37
4.6.4 Физический смысл коэффициента вариации. Обоснование	38
4.6.5 Реализация и возможности расчета	39
4.7 Матрица среднеквадратичных отклонений атомов	40
4.7.1 Построение матрицы среднеквадратичных отклонений атомов	41
4.7.2 Режимы построения матрицы среднеквадратичных отклонений	41
4.7.2 Вывод матрицы среднеквадратичных отклонений в режиме реального времени	42
4.8 Тензор инерции и его применение в Анализаторе траекторий	42
4.8.1 Тензор инерции в функциях Анализатор траекторий	43
4.8.2 Визуализация системы в координатах тензора инерции	43
4.9 Работа с подсистемами	43
4.9.1 Заполнение списка подсистем	44
4.9.2 Использование подсистем	45
4.9.3 Хранение подсистем в СУБД	45
4.10 Опции программы	47
4.10.1 Коллекция типов атомов	48
4.10.2 Опции отображения трехмерной структуры	49
4.10.3 Опции отображения разреза	50
4.10.4 Опции отображения полостей	51
4.10.5 Прочие опции отображения и опции программы	51

4.11 Анализ расширенных компонент	52
4.11.1 Визуализация расширенных компонент траектории на молекулярном кино	52
4.11.2 Анализ расширенных компонент вдоль всей траектории	53
4.12 Восстановление траекторий	54
4.12.1 Объяснение работы с повреждённой траекторией.....	55
4.12.2 Признаки поверженной траекторий.....	55
4.12.3 Алгоритм восстановления траекторий	55
4.12.4 Порядок восстановления файлов траекторий	55
4.13 Статистическая обработка энергетических траекторий и прочих файлов с данными	57
4.13.1 Порядок проведения анализа.....	58
4.14 Запись видео.....	60
4.14.1 Запись видео в интерактивном режиме.....	61
4.14.2 Наложение текста при записи видео	62
4.14.3 Запись видео в автоматическом режиме	62
4.15 Построение твердотельных моделей	63
4.15.1 Описание применяемых конструкций Python for Blender	64
4.15.2 Порядок экспорта моделей в формат Python for Blender	65
4.15.2 Многоцветная печать трёхмерных моделей	66
4.15.3 Исполнение скриптов Python в программе Blender	67
4.15.4 Подготовка моделей в формате G-CODE.....	69
4.16 Конструирование системы для поиска наилучшего расположения двух подсистем	71
4.16.1 Порядок работы с TAMD.....	72
4.16.2 Особенности реализации	74
5. TAMD-Script. Сценарный анализ траекторий молекулярной динамики	75
5.1 Назначение	75
5.2 Поддержка сценариев анализа траекторий	75
5.3 БАЗОВЫЕ КОМАНДЫ TAMD-SCRIPT.....	76
6. Методика работы с программой	77
6.3 Возможности TAMD-Script	77
6.3.1 Обработка всех файлов в директории	77
6.3.2 Расчет отклонений C-alpha-атомов вдоль всей траектории	77
6.3.3 Расчет матрицы отклонений C-alpha-атомов.....	79
6.3.4 Визуализация матрицы отклонений C-alpha-атомов средствами внешней программы.....	80
6.3.5 Средство визуализации и интерактивного анализа матрицы отклонений C-alpha-атомов	81
6.3.5 Сценарий анализа лево- и правозакрученных альфаспиральных белков в опытах их разворачивания под действием постоянной силы, приложенной к концам.....	85
РЕЗУЛЬТАТЫ И ВЫВОДЫ	86
Приложение 1. Минимальные требования к системе	87
Приложение 2. Установка программы	87

1 Основные возможности

1. Показ молекулярного кино в трехмерном виде больших молекулярных объектов (100 000 атомов) с настройкой выводимой информации.
2. Возможность работы с подструктурами.
3. Расчет характеристик вдоль траектории, основанных на геометрии системы.
4. Средство работы с графиками функций, в т.ч. статистической их обработки.
5. Расчет коэффициента вариации молекулярной динамики.
6. Квазигармонический анализ. (в скорейшей разработке)
7. Интерактивное построение расширенных динамических контактных карт больших молекулярных объектов при поддержке технологии CUDA.
8. Работа с расширенными траекториями молекулярной динамики.
9. Средство исправления траекторий (поврежденных в результате системных ошибок).
10. Средство построения отчетов об ошибках.
11. Реализации ресурсоемких алгоритмов в 3-х видах: для одного процессора, для многопроцессорной системы, для массивно-параллельной ЭВМ (CUDA).

1.1 Отличия от VMD

Программный комплекс TAMM с точки зрения программного продукта создан с нуля и никак не переплетается с кодовой базой VMD. Помимо того, что это две независимых программных продукта одного направления, ниже перечислены основные отличия:

- Поддержка расширенных траекторий молекулярной динамики с произвольным числом расширенных компонентов, включая модуль трехмерной визуализации.
- Интерактивный инструмент работы с расширенными динамическими контактными картами и построением количества контактов.
- Квазигармонический анализ.
- Быстрое получение простых характеристик вдоль траектории.
- Модуль работы с графиками, полученными в процессе анализа траекторий.

2 Общий порядок работы с ПО

Алгоритмически порядок работы с «Анализатором траектории» можно представить на следующей диаграмме.

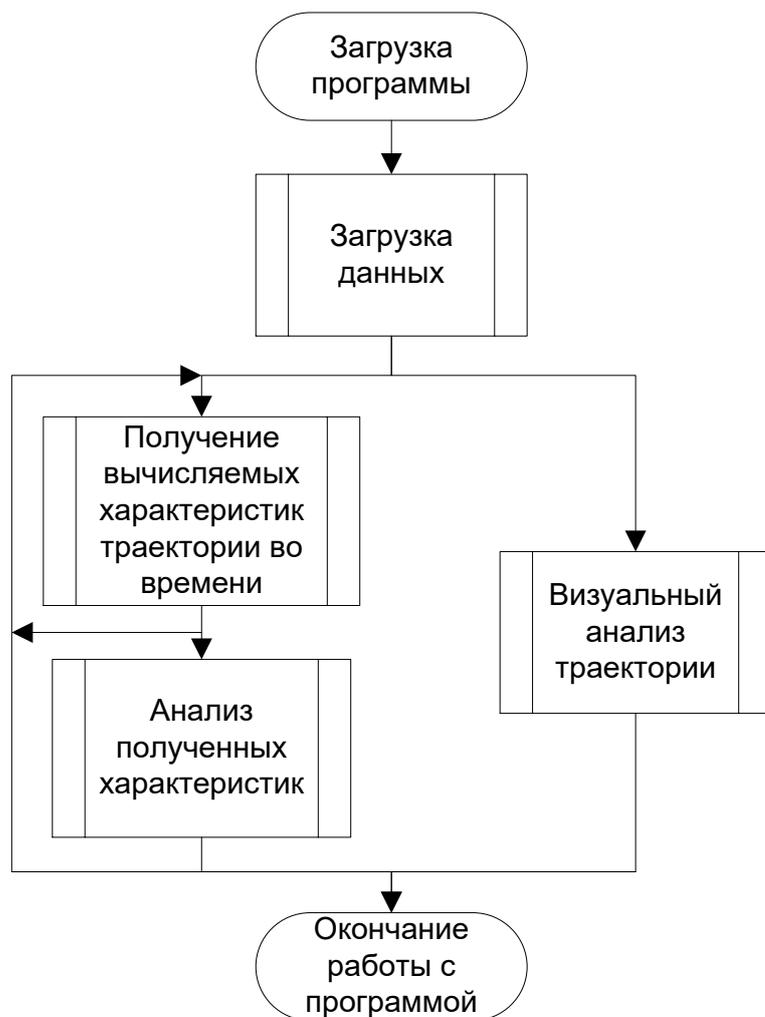


Рисунок 1. Схема работы с программой

В целом, работа с программой аналогична работе с большинством Windows-приложений. Главное окно программы содержит в себе:

- панель инструментов для исполнения некоторых действий (на кнопках имеются подсказки);
- ортогональную проекцию молекулы (или групп атомов);
- панель параметров;
- дополнительные окна;
- бегунок, позволяющий переходить на требуемый момент времени.

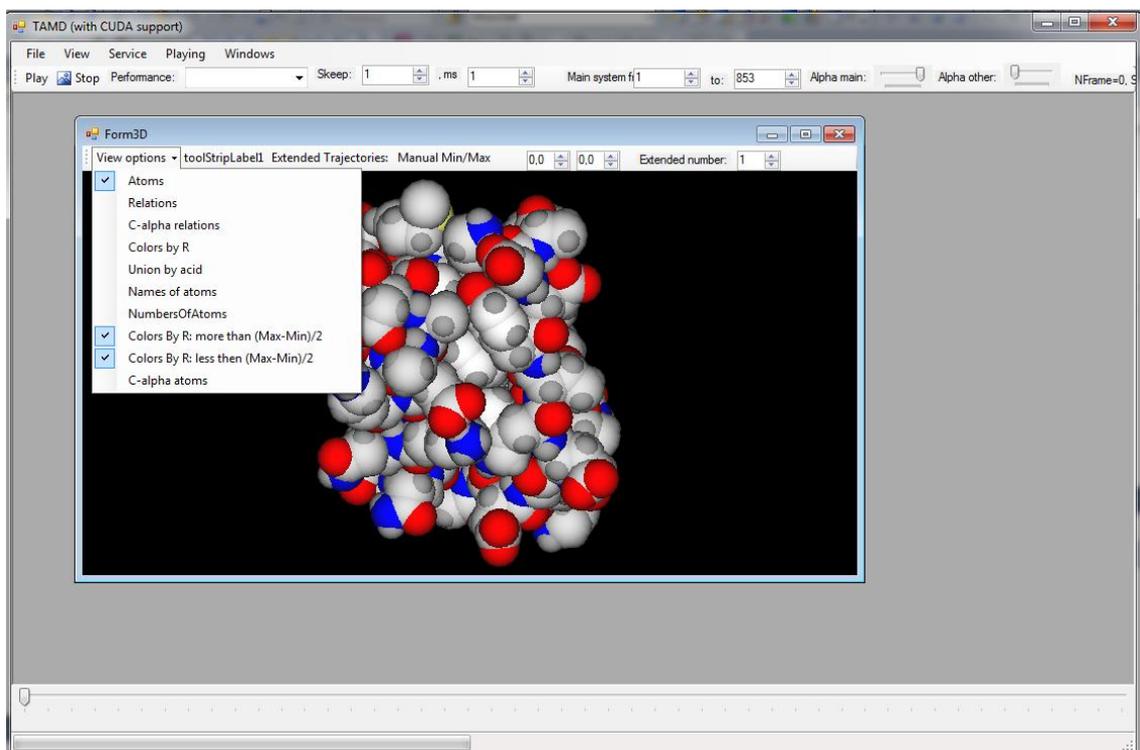


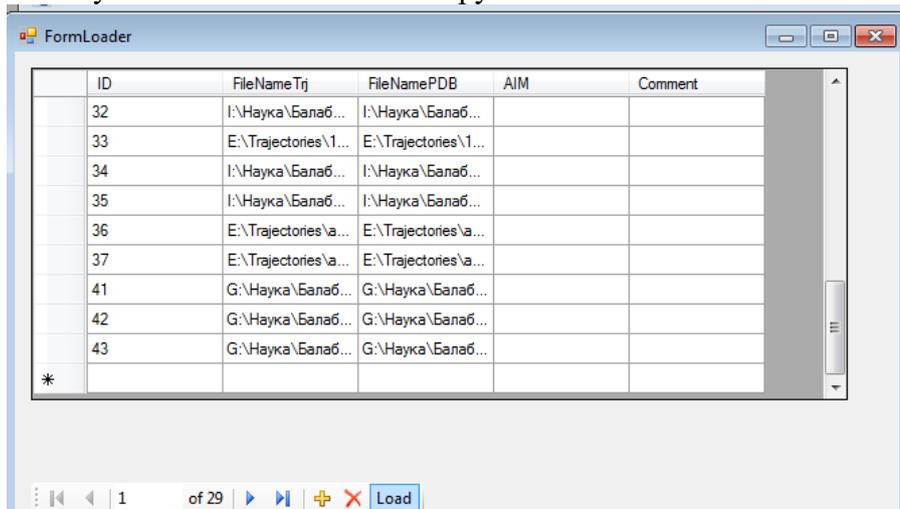
Рисунок 2. Работа с программой.

2.1 Открытие файлов

В программе предусмотрено 2 способа открытия файлов: непосредственный выбор из файловой системы и загрузка из истории. В любом случае загружаются два файла: файл траектории (*.trj) и файл pdb-структуры (*.pdb).

Для непосредственного открытия файла необходимо воспользоваться меню File -> Open. Сначала выбирается файл траектории. Далее pdb-файл. Если файл траектории уже был загружен, программа предложит выбрать запомненный pdb-файл.

Для повторного открытия файлов предусмотрен режим загрузки из истории, меню File -> Loader. Необходимо выбрать интересующую пару файлов и нажать на кнопку Load, расположенную на нижней панели инструментов.

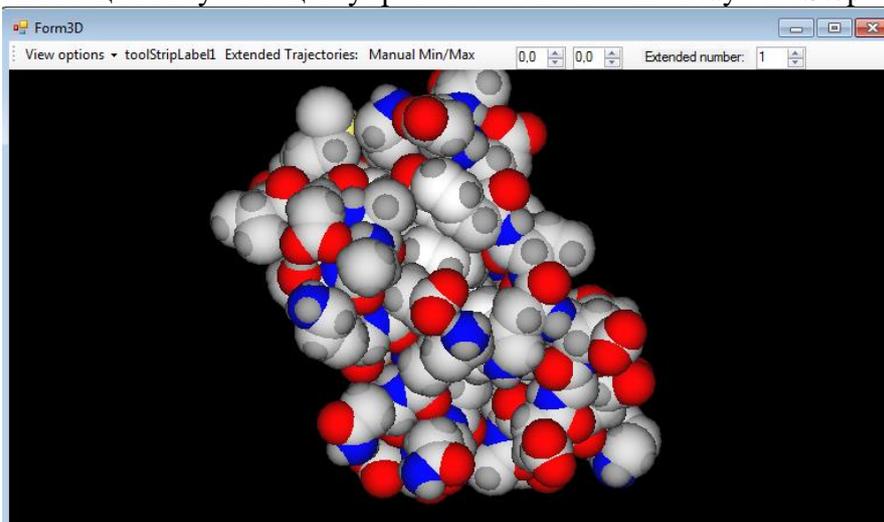


2.2 Перемещение вдоль траектории. Молекулярное кино

После того, как траектория открыта, с ней можно работать. Наиболее распространённый режим работы – просмотр молекулярного кино.

Окно перспективной/ортогональной проекции отображается выбором 3D Vision меню View.

Анимация визуализации управляется кнопками «Play» и «Stop» из главного окна.



Вращение изображения осуществляется при помощи перемещения мышью и одновременного нажатия на левую клавишу.

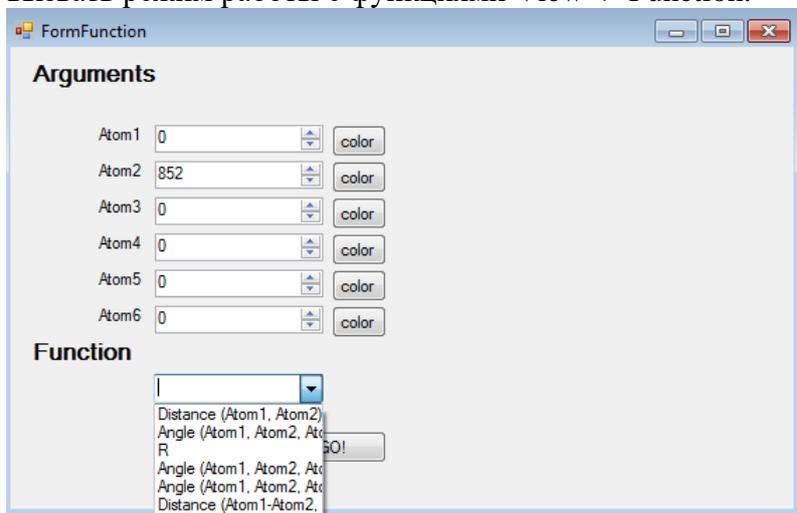
Перемещение изображения осуществляется при помощи перемещения мышью и одновременного нажатия на правую клавишу.

Масштабирование изображения осуществляется при помощи колеса мыши.

2.3 Расчет общих характеристик

Анализатор траекторий позволяет находить зависимость требуемой параметризованной величины вдоль траектории.

Для расчета характеристики требуется сначала загрузить траекторию. Далее следует вызвать режим работы с функциями View -> Function.



Для расчета необходимо выбрать тип функции из списка Function. Далее задать ее аргументы из полей Arguments. Лишние аргументы не влияют на расчет. Запуск расчета – кнопка Go.

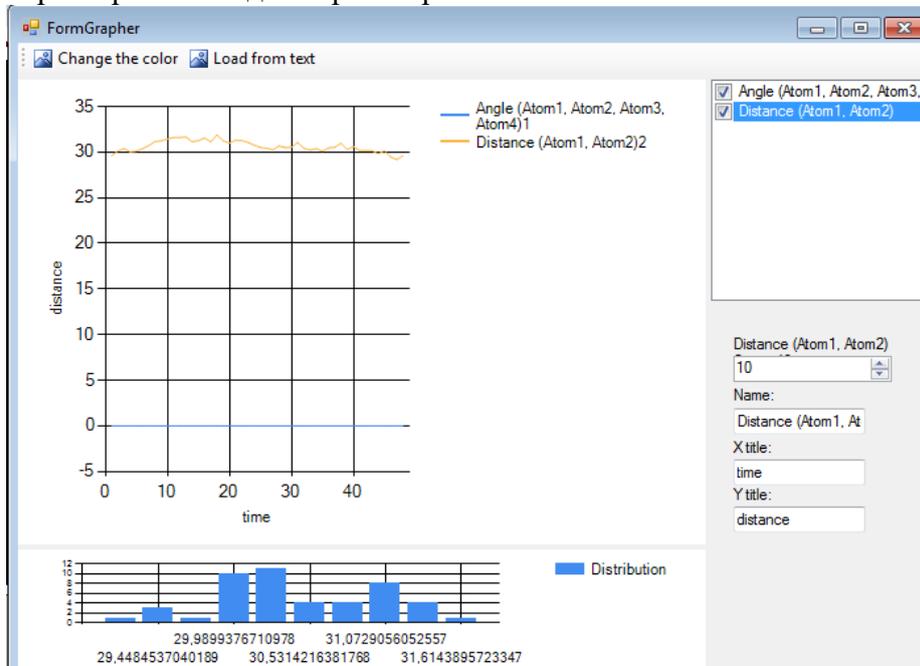
Доступны следующие функции:

1. Расстояние между двумя атомами.
2. Угол между прямыми, задаваемый тремя атомами.
3. Тензор инерции.
4. Угол между прямыми, задаваемый четырьмя атомами.
5. Угол между двумя группами атомов, рассчитываемый как угол между главными осями инерции.
6. Угол между геометрическими центрами двух групп атомов.

По окончании расчета выводится соответствующий график, см. Средство работы с графиками.

2.4 Средство работы с графиками

Данный режим работы с программой доступен с помощью меню View -> Grapher. Также средство работы с графиками появляется автоматически после расчета характеристики вдоль траектории.



Справа сверху выбирается интересующий график. Его можно отобразить на рисунке, а также скрыть, убрав галочку напротив него.

При выборе графика внизу высвечивается плотность распределения характеристики.

У текущего графика можно изменить цвет, нажав на кнопку Change the color.

Можно загрузить из текстового файла набор зависимостей при помощи кнопки Load from text.

3 Дополнительные режимы работы

3.1 Выделение групп атомов

В программе предусмотрена возможность вывода химических связей между атомами. Делается это автоматически для всех пар атомов, расстояние между которыми не превышает заданное число.

Однако для больших молекул такой возможности визуализации может оказаться недостаточно – химические связи могут показаться слишком мелкими на общем фоне. Здесь мы предлагаем устанавливать свои связи между атомами, а также выделять атомы в группы.

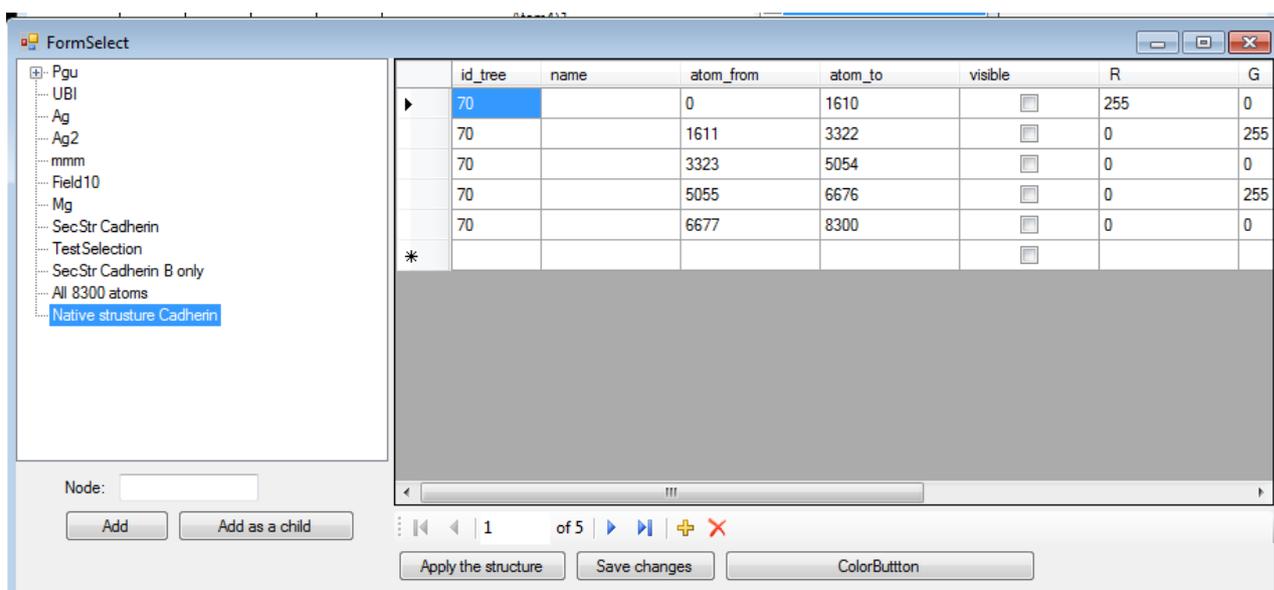


Рисунок 3. Окно групп и связей.

Записи о группировке структуры молекулы (или групп молекул) могут оказаться многочисленными и их неудобно вводить в ходе каждого эксперимента. Поэтому в программе предусмотрена возможность хранения сведений о структуре эксперимента в дереве, напоминающем структуру каталогов жесткого диска. Пользователь может создавать свою структуру дерева произвольной степени вложенности.

Для управления деревом служат кнопки Add и Add as a child.

- Add (добавить элемент). Добавляет новый элемент в дерево на том же уровне вложенности.
- Add as a Child (добавить дочерний элемент). Добавляет новый элемент в дерево, причем текущий элемент является родителем добавляемому.

Сами сведения о текущей структуре хранятся в виде таблицы.

После каждого редактирования таблицы необходимо нажимать кнопку Save Changes.

После работы над структурой пользователь должен нажать кнопку «Apply the structure» (применить структуру), чтобы записанные сведения вступили в силу.

Заполняются следующие поля таблицы:

- name – произвольное имя группы;
- atom_from – начало группы (включительно);
- atom_to – конец группы (включительно);
- visible – не используется;
- r, g, b, a – компоненты цвета в формате RGBA (красный, зеленый, синий, прозрачность). Диапазон каждого канала от 0 до 255 включительно.
- Type – тип отображения – 1 – цвета из pdb-файла, 2 – цвета из этой таблицы.

3.2 Визуализация расширенных траекторий

Для включения визуализации расширенных траекторий служит меню View options окна трехмерной визуализации, а также ряд дополнительных органов управления.

Режим включается выбором Colors by R из меню View options. При этом раскраска меняется от синего через белый к красному, в зависимости от значения расширенной компоненты.

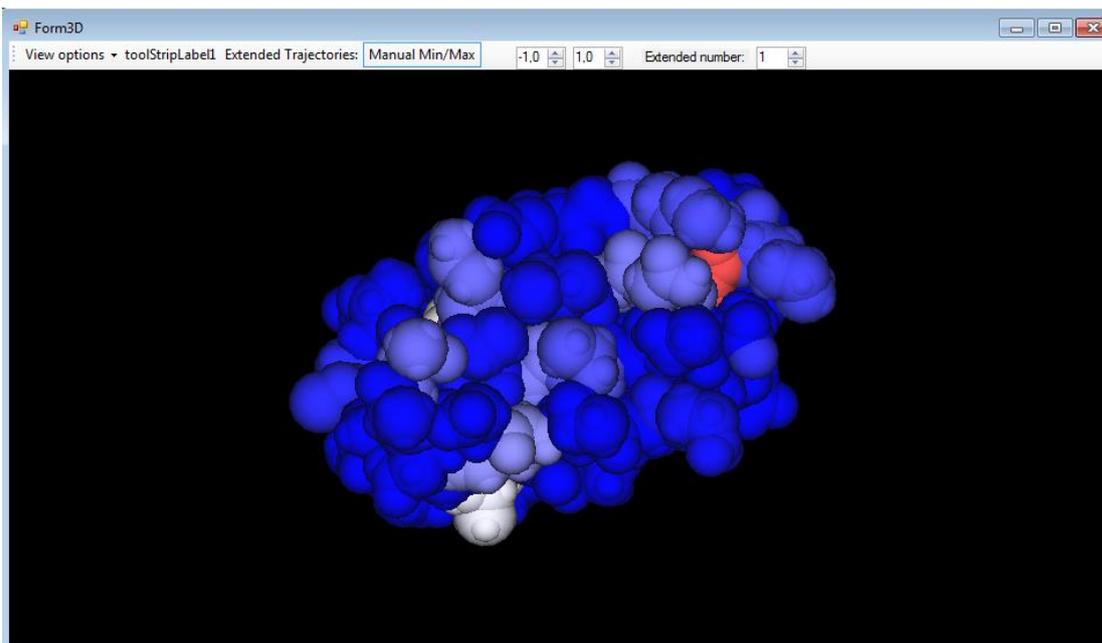


Рисунок 4. Визуализация расширенной траектории.

Нормировка осуществляется автоматически при включении режима Colors by R. Расширенные компоненты отдельных атомов могут уходить далеко от средних значений. Поэтому в программе введен режим ручной нормировки. Он включается выбором Manual Min/Max и заданием соответствующих границ.

Существует режим раздельного вывода атомов, расширенная компонента которых больше или меньше некоторого среднего значения $(R_{Min}+R_{Max})/2$. Для раздельного просмотра таких атомов необходимо убрать галочки у пунктов меню R more/less $(R_{Min}+R_{Max})/2$. Поддерживается работа с несколькими расширенными компонентами. Выбор между компонент осуществляется в окне Extended number.

3.4 Расширенные динамические контактные карты

Инструмент работы с расширенными динамическими контактными картами создан вызывается путем выбора меню View -> Contact Map.

Для работы с контактной картой необходимо сначала выбрать уровень аппаратной поддержки с помощью выпадающего списка Performance в основном окне приложения. Если необходимо, включить режим работы с группами атомов. Настроить параметры карты (Atom from-to, contact from to). Выбрать режим работы инструмента.

Отображение динамической расширенной контактной карты осуществляется при каждом смене кадра при нажатом переключателе On/Off.

Доступны следующие режимы построения контактных карт:

1. Вывод всех атомов.
2. Вывод заданного диапазона атомов (atom from-to).
3. Вывод групп атомов.
4. Вывод контактов только между C-Alpha-атомами.

Для использования вывода контактной карты по C-Alpha атомам программа просматривает таблицу параметров атомов PDB (View -> Options -> PDB) на предмет присутствия единицы в столбце CAlpha.

3.5 Квазигармонический анализ

В Анализатор траекторий молекулярной динамики включен режим анализа колебаний молекулярной системы методом квазигармонического анализа. Несмотря на похожесть с алгоритмом анализа нормальных мод, метод квазигармонического анализа позволяет производить расчеты полностью по траектории.

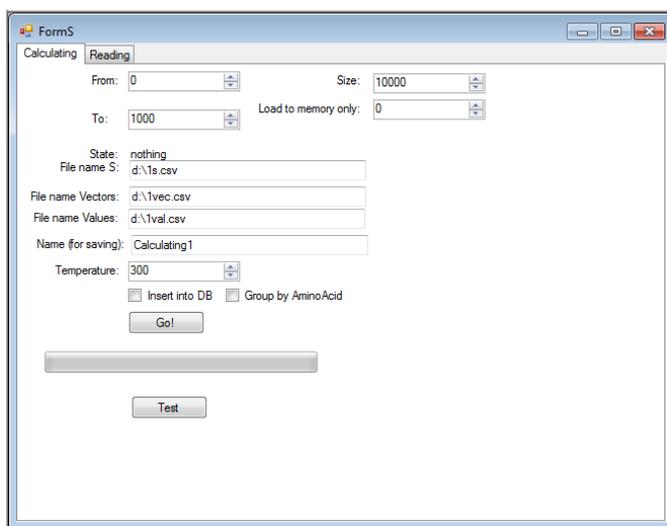


Рисунок 5. Параметры запуска квазигармонического анализа.

Вызов инструмента квазигармонического анализа осуществляется путем выбора пункта Quasiharmonic analysis меню View.

Инструмент анализа колебаний, называемый также визуализатор колебаний, состоит из двух частей:

- инструмент колебательных мод (вкладка Calculating);
- визуализатор колебаний (вкладка Reading).

Инструмент вычисления колебательных мод выполняет следующие функции:

- вычисления средних положений всех атомов по диапазону траектории, задаваемого при помощи параметров From и To;
- вычисление матрицы ковариации с поддержкой многопроцессорного режима (Performance -> Many CPU) и расчета на графическом процессоре (Performance -> CUDA);
- сохранение матрицы ковариации в текстовый файл в формате .csv (comma separated values);
- вычисление частот и колебательных мод (при помощи нахождения обратной матрицы);
- сохранение набора частот и направлений колебаний в текстовые файлы в формате .csv;
- сохранение набора колебательных мод в базу данных программы.

После открытия траектории, задания параметров программы и ввода названия вычислений алгоритм необходимо запустить на счет, нажав на кнопку Go.

После завершения всех этапов расчета можно перейти к режиму чтения частотных наборов.

Режим чтения и визуализации колебательных мод вызывается из вкладки Reading режима Quasiharmonic analysis.

Сначала необходимо выбрать название расчетов. При его выборе отображается набор частот, отсортированных по возрастанию. Загрузка всего набора данных осуществляется путем нажатия кнопки Load.

Выбрав частоту, можно получить набор направлений колебания.

После загрузки набора колебательных мод можно перейти в окно визуализации.

Появляется дополнительная панель Quasiharmonic analysis. Необходимо выбрать требуемую частоту, которую можно менять при воспроизведении колебания и нажать кнопку Play. Вектора колебаний нормированы. Чтобы колебания было лучше видно, в окне Amplitude задается амплитуда колебаний.

Остановить режим визуализации колебаний можно, нажав на клавишу Stop.

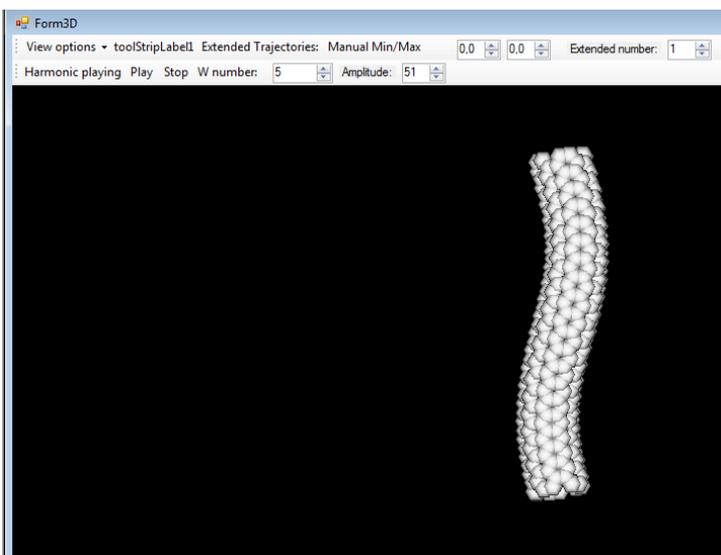


Рисунок 6. Визуализация колебаний 5-ой моды.

3.7 Визуализация траекторий с периодическими граничными условиями

Исходная система может выглядеть следующим образом:

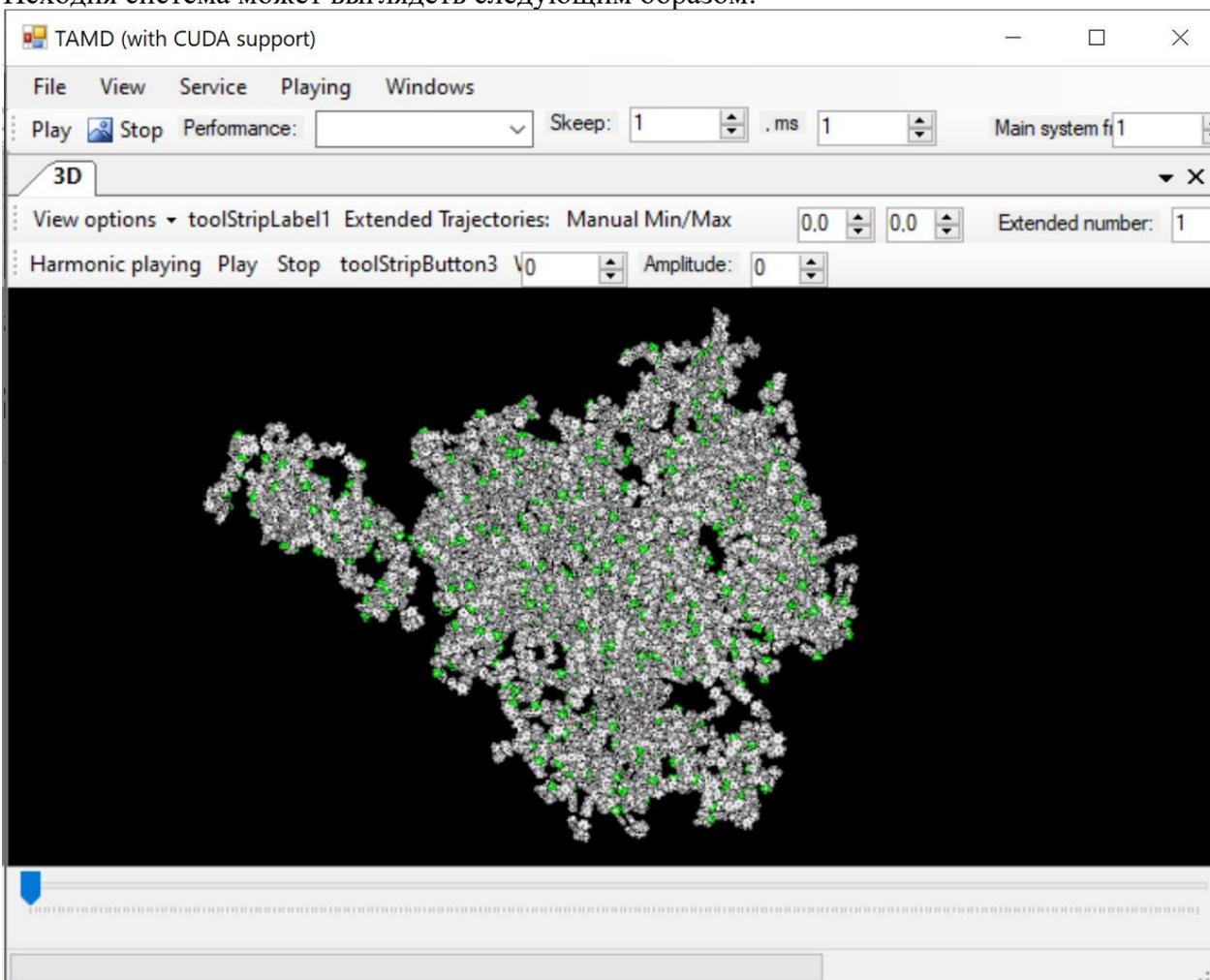


Рисунок 7. Исходная система

Для включения режима работы с периодическими граничными условиями необходимо выбрать пункт меню View->Options... (горячая клавиша F11), затем выбрать вкладку Slice.

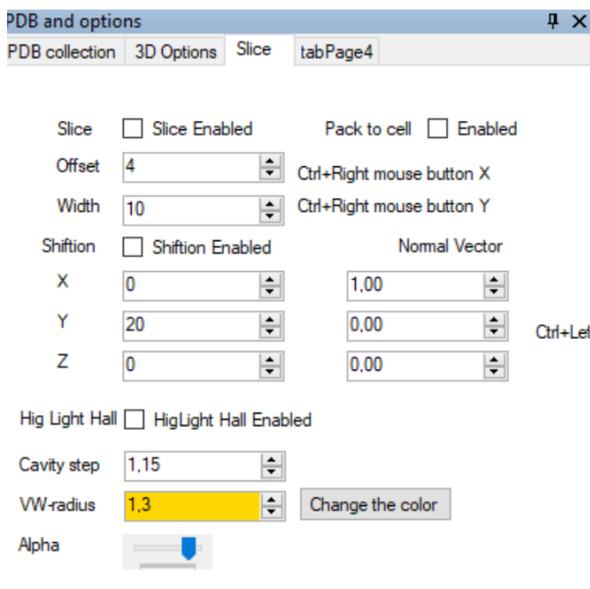


Рисунок 8. Опции работы с периодическими граничными условиями.

Доступны следующие опции:

- Pack to Cell – включение режима работы с периодическими граничными условиями (с заголовке траектории должен быть указан размер расчетной ячейки).
- Slice Enabled – вырезка слоя толщиной Width и смещением Offset.

После включения периодических граничных условий система запаковывается в прямоугольный параллелепипед:

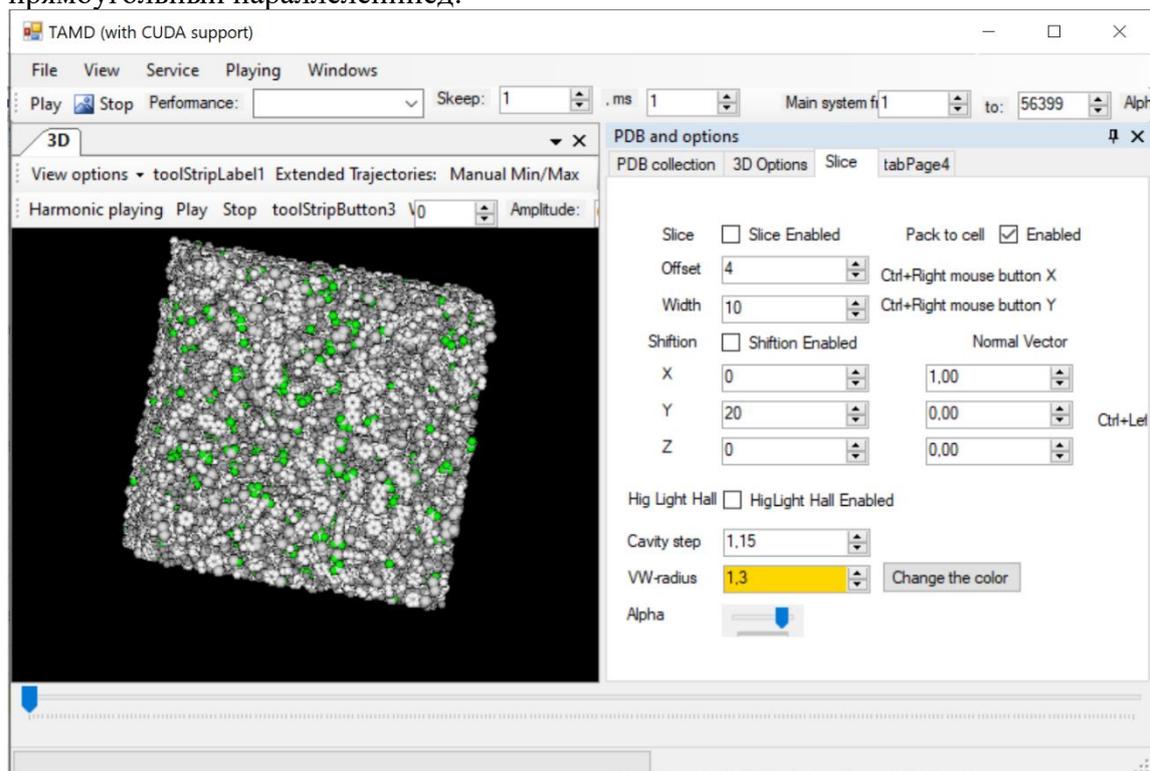


Рисунок 9. Система, запакованная в параллелепипед (включение режиме периодических граничных условий).

После включения опции вырезки слоя (Slice) удобно сканировать систему, отображая слой небольшой толщины. Для Этого необходимо нажать клавишу Control на клавиатуре и перемещать мышь при зажатой левой кнопкой:

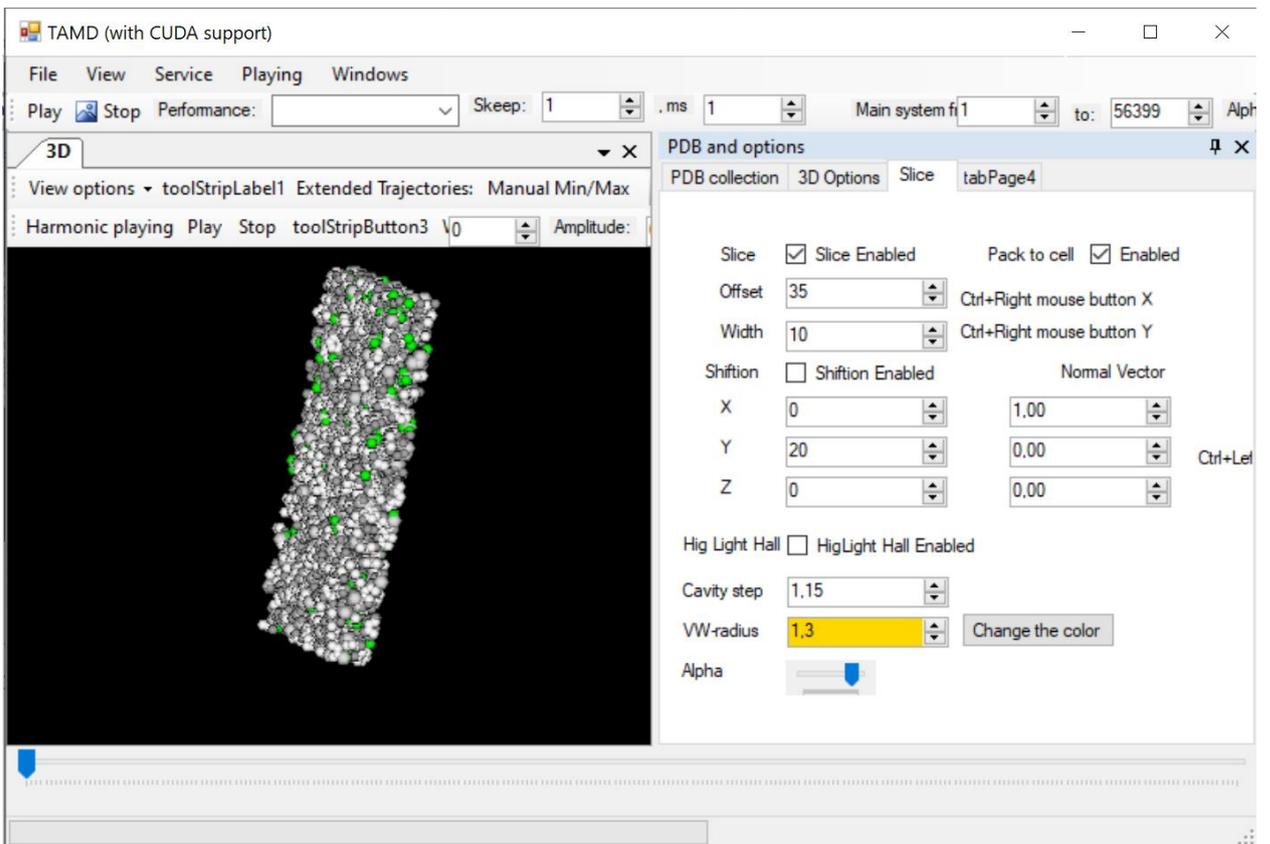


Рисунок 10. Опция вырезания слоя.

3.8 Визуализация полостей

Визуализация полостей происходит посредством заполнения пустого пространства расчетной ячейки шарами радиуса $VW\text{-radius}$ с шагом $Cavity\ step$. Чем меньше параметр $Cavity\ Step$, тем чаще будут изображаться сферы, обозначающие пустоту.

Визуализация полостей работает при включенном режиме работы с траекториями с периодическими граничными условиями (для возможности определения размеров расчетов ячейки). См. пункт 3.8.

Для включения поддержки работы с периодическими граничными условиями необходимо выбрать пункт меню `View->Options...` (горячая клавиша F11), затем выбрать вкладку `Slice`, поставить галочку `Slice Enabled`. На этой же вкладке задаются параметры $VW\text{-radius}$ и $Cavity\ step$.

Для включения режима отображения полостей необходимо выбрать из меню `View options` окна трехмерной визуализации опцию `Cavity`.

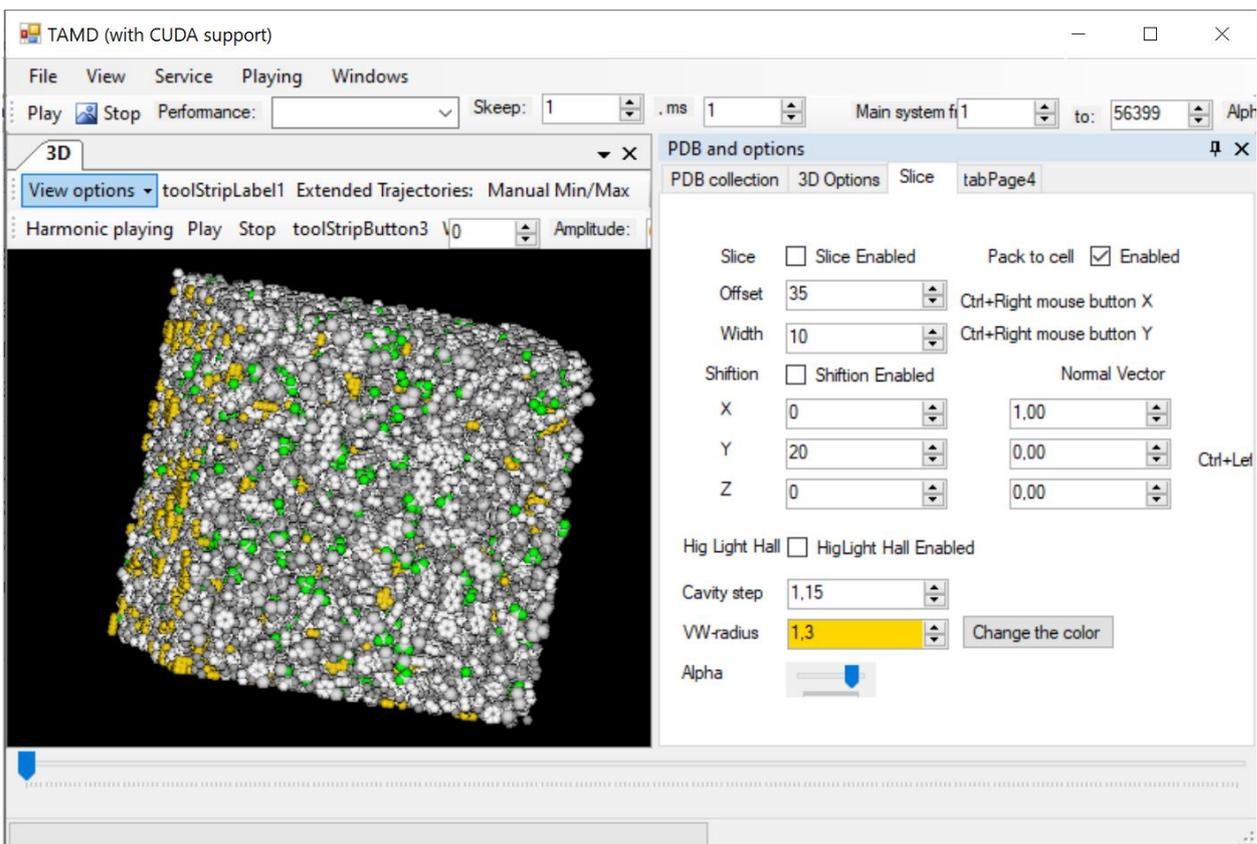


Рисунок 11. Включение режима визуализации полостей.

Полезно показывать одни лишь полости без остальных атомов системы. Для этого необходимо убрать опцию Atoms меню View options окна трехмерной визуализации.

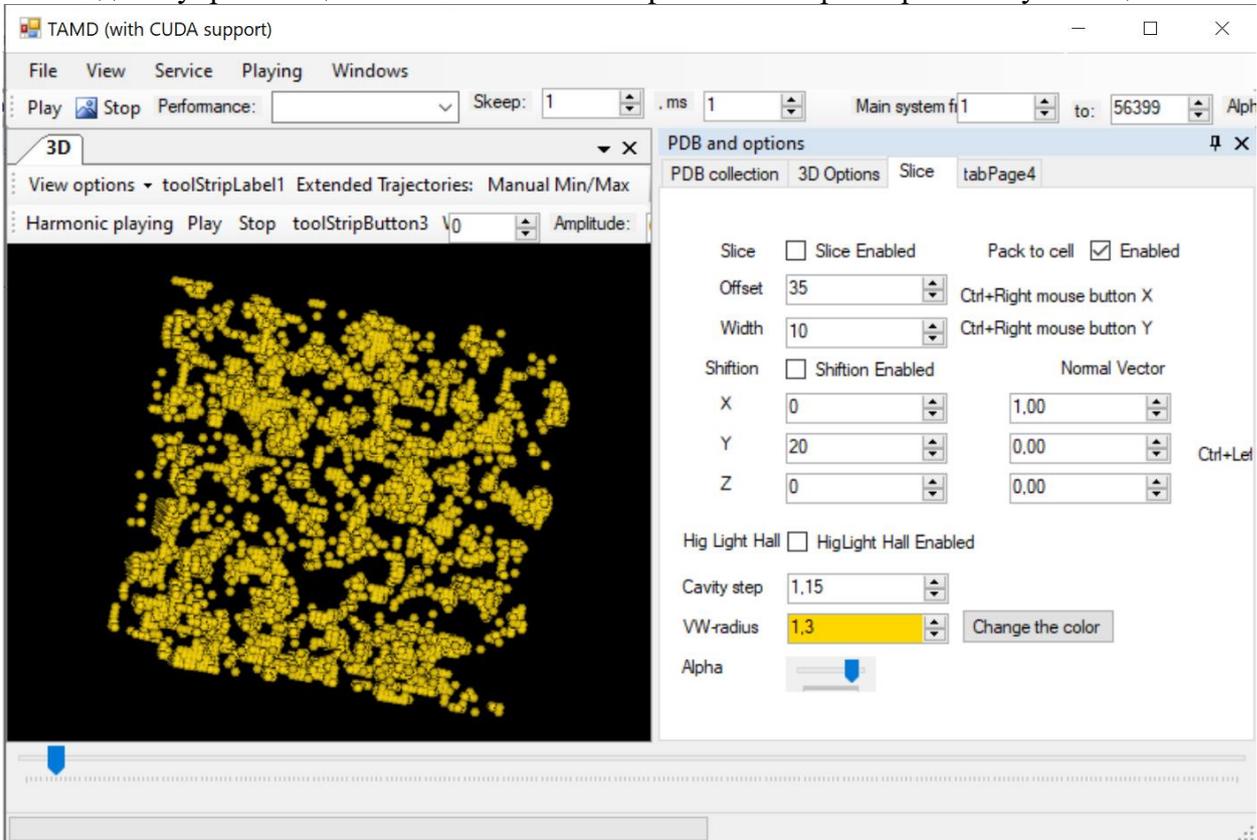


Рисунок 12. Включение режима визуализации полостей без отображения атомов системы. В режиме визуализации полостей доступны все режимы работы Анализатора, включая воспроизведение молекулярного кино.

4. Алгоритмы и их реализация

4.1 Чтение файлов траекторий молекулярной динамики

От организации чтения файлов траекторий молекулярной динамики зависит быстродействие всей программы.

4.1.1 Структура траекторного файла

Файл траектории представляет собой текстовый файл, который содержит кадры траектории. Кадры траектории содержат значения координат системы через определённые промежутки времени. Как правило, но не обязательно, кадры траектории записываются через равные промежутки времени. В этом случае порядковый номер кадра говорит о моменте времени, в который этот кадр был сделан. Анализатор ориентируется только на последовательный номер кадра.

В кадр могут быть записаны мгновенные, либо усреднённые значения координат. Выбор режима записи зависит только от программы моделирования молекулярной динамики. Для Анализатора это одни и те же траектории.

Кадр состоит из $N+2$ строк, где N – количество атомов в системе. Все кадры траектории должны содержать одинаковое количество строк. Все кадры должны занимать одинаковое количество байт дисковой памяти. Именно этот факт предоставляет возможным организовать прямой доступ к произвольному кадру траектории. Две дополнительных строки в начале кадра служат заголовком.

Если кадры состоят из различного количества строк, то такая траектория считается испорченной. Для исправления траектории служит возможность исправления (меню Service – Utils – repair).

Пример начала файла траектории:

```
Nstep= 1000 1259 1 0.50
 66.495 48.637 47.572
-15.738 -2.569 1.236
-15.466 -2.960 -0.145
-14.207 -3.821 -0.302
-13.310 -3.758 0.515
```

...

Каждый кадр должен начинаться заголовком из двух строк. Первая строка содержит маркер «Nstep=», после которого следует номер кадра и другую информацию. Номер кадра и прочая информация игнорируются. Смысл имеет только порядковый номер кадра с начала траектории.

Вторая строка заголовка содержит размер ячейки при работе с периодическими граничными условиями. Она влияет только на вывод в режиме упаковки (режим Pack to cell работы с опциями) в ячейку. В случае работы без периодических граничных условий данная строка бесполезна, но должна присутствовать. И в ней должны находиться три числа с плавающей точкой. Они нигде не используются. При получении траекторий от программного комплекса моделирования молекулярной динамики PUMA-CUDA в этих трёх числах содержится размер ячейки алгоритм сканирования по пространству.

В последующих строках содержатся координаты атомов в трёхмерном пространстве. Это три вещественных числа, разделённых через один или несколько пробелов.

Предусмотрена также работа с расширенными траекториями молекулярной динамики. В каждой строчке таких траекторий (не считая заголовков кадров), помимо координат в пространстве, содержится один или несколько дополнительных величин, относящиеся к каждому атому. Это может быть температура, вклад атома в энергию Ван-дер-Ваальса, Кулона и пр. При чтении такой траектории в память заносятся три координаты и одна из компонент энергии. Выбор компоненты осуществляется в окне трёхмерного промотра (3D), поле Extended number:

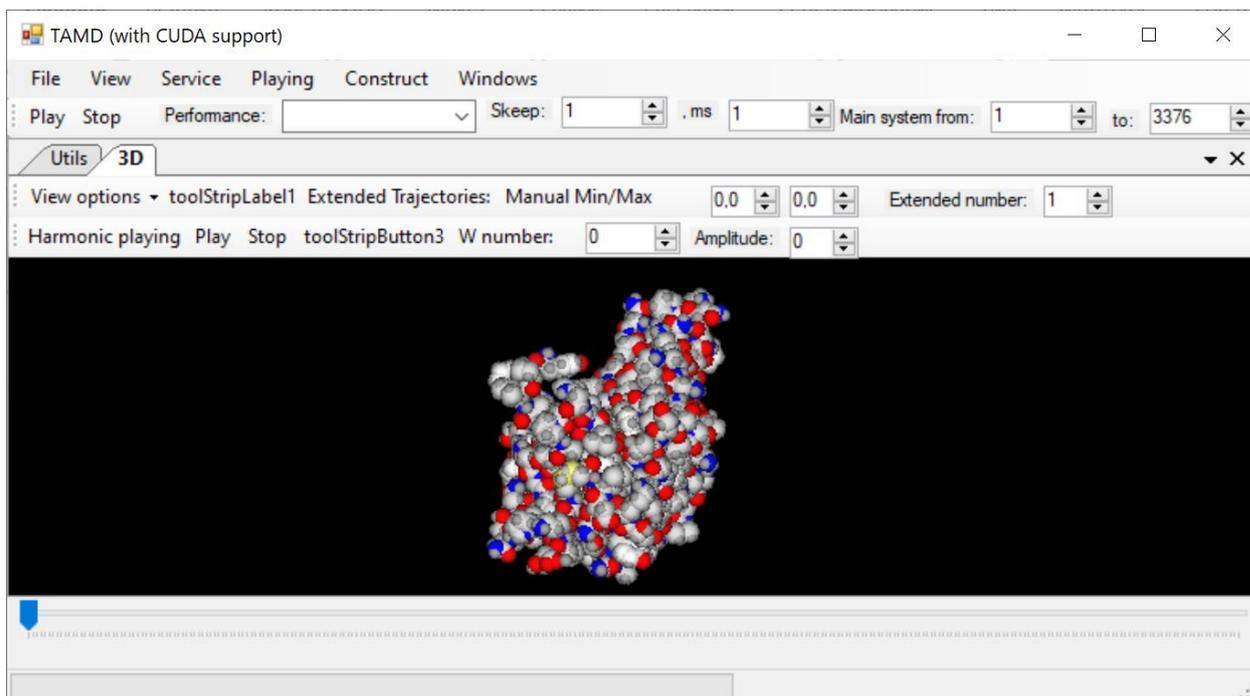


Рисунок 13. Загруженная траектория молекулярной динамики с расширенной компонентой под номером 1.

4.1.2 Буфер случайного доступа к траектории

Нередко расчет траектории занимает огромное количество времени, что приводит к необходимости работы с траекторными файлами большого объема. С другой стороны, при увеличении степени подробности даже малой траектории (уменьшении шага, через который координаты атомов записываются на диск) могут также существенно возрасти дисковые объемы данных.

Под большим объемом данных будем иметь в виду объем, превышающий физическую оперативную память ЭВМ. Подготовленный пользователь сразу спросит, а как же обстоит дело с виртуальной памятью? Поясним этот вопрос. Дело в том, что управлением памяти занимается операционная система. Напомним, что под операционной системой подразумевается аппаратно-программная начинка компьютера, которая предоставляет некий программный интерфейс для работы остальных программ, оборудования, а также взаимодействия с ними пользователя. Программа, в частности TAMM, дает процессору команды на чтение определенного адреса памяти, сама не зная о том, находится ли он в физической памяти, либо только на диске. Как известно, доступ к физической оперативной памяти (RAM) происходит намного быстрее, чем к дисковой. Использование виртуальной памяти приводит к появлению еще одной копии траектории на диске, что также неэффективно. С другой стороны, нельзя не согласиться, что описываемый алгоритм будет работать медленнее с короткой траекторией, чем алгоритм полной загрузки данных в оперативную память.

Перечислим основные приемы работы с дисковыми данными:

1. Полное чтение данных в память. Самое быстрое с точки зрения обработки данных. Однако данный способ полностью непригоден для большого объема данных;
2. Прямое чтение. Каждый кадр читается при необходимости его использования. Требуется минимум оперативной памяти. Однако данный способ неэффективен при повторном чтении данных, что постоянно происходит при просчете будущих координат атомов для определения направления их движения;
3. Буферизованное чтение.
Остановимся подробнее на алгоритме последнего (см. Табл. 1).

Здесь номер кадра говорит только о времени доступа к нему и напрямую не зависит от кадра траектории.

Представим себе очередь, организованную по принципу FIFO (first input – first output; первым пришел – первым обработан), только усовершенствованную – кадры могут меняться местами в процессе доступа к ним.

Таблица 1. Организация очереди буфера.

Кадр 1
...
Рабочий кадр
...
Кадр n-1
Кадр n

Полный алгоритм работы буфера следующий:

- Если происходит попытка чтения кадра, которого нет в буфере, то он попадает на место рабочего кадра. При этом последний кадр затирается предпоследним;
- Если происходит обращение к кадру, который есть в памяти под номером m , то он меняется местами с кадром номер $m-1$. Таким образом, затирается всегда кадр, к которому не было доступа максимально длительное время;
- В буфере есть указатель на рабочий кадр – это последний кадр, к которому был доступ. Благодаря указателю, доступ к нему происходит наиболее быстро. При создании алгоритма работы буфера разумно было предположить, что наиболее часто происходит доступ к одному и тому же кадру.

Процедура смены кадров местами не приводит к полному обмену данными между ними. Просто происходит обмен адресами памяти, указывающими на соответствующие кадры. Можно отметить, что разработанный и реализованный метод чтения данных позволяет достичь максимального быстродействия файлового ввода при ограниченном выделении оперативной памяти.

4.1.3 Чтение траекторий из баз данных

При реализации поддержки сетевых возможностей в Анализатор были добавлены функции чтения траекторий из баз данных. Они обсуждаются в главе, посвященной Trajectory Commander – средству работы с траекториями и сопутствующей информации в различных СУБД.

Настоящий раздел касается взаимодействию Trajectory Commander и буфера случайного доступа.

Класс буфера случайного доступа к траектории разделён на высокоуровневые функции управления буфером и низкоуровневые функции чтения кадра. При работе с базами данных используются те же самые высокоуровневые функции работы с буфером. При обращении к функциям чтения определённого кадра происходит вызов функций Trajectory Commander, которые считывают координаты кадра в те же самые структуры, что из функции чтения кадра с диска.

Подобная организация внутреннего API выгодна как для пользователей Анализатор, так и для программистов. Для пользователя процесс выглядит точно также, как и при работе с файлами траекторий, расположенными на диске. Более того, для программиста, занимающегося реализацией алгоритмов, нацеленных на анализ траекторий, также не имеет значений ведётся ли работа с траекторий на диске, либо она считывается по сети из базы данных.

4.2 Работа с PDB-файлом

При работе с траекторией pdb-файл (Protein Data Bank) используется для определения типов атомов и номеров аминокислотных остатков.

После выбора файла траектории возникает диалоговое окно открытия pdb-файла. От загрузки pdb-файла можно отказаться. В том случае все атомы будут одного цвета, одного размера, несуществующего «служебного» типа атомов.

Анализатор траекторий сохраняет информацию о уже когда-либо открытых парах траекторный файл – pdb-файл. Если траектория открывается из этого списка, то Анализатор спрашивает, не стоит ли использовать уже заполненный pdb-файл. Как правило, на этот вопрос лучше ответить утвердительно для более продуктивной работы.

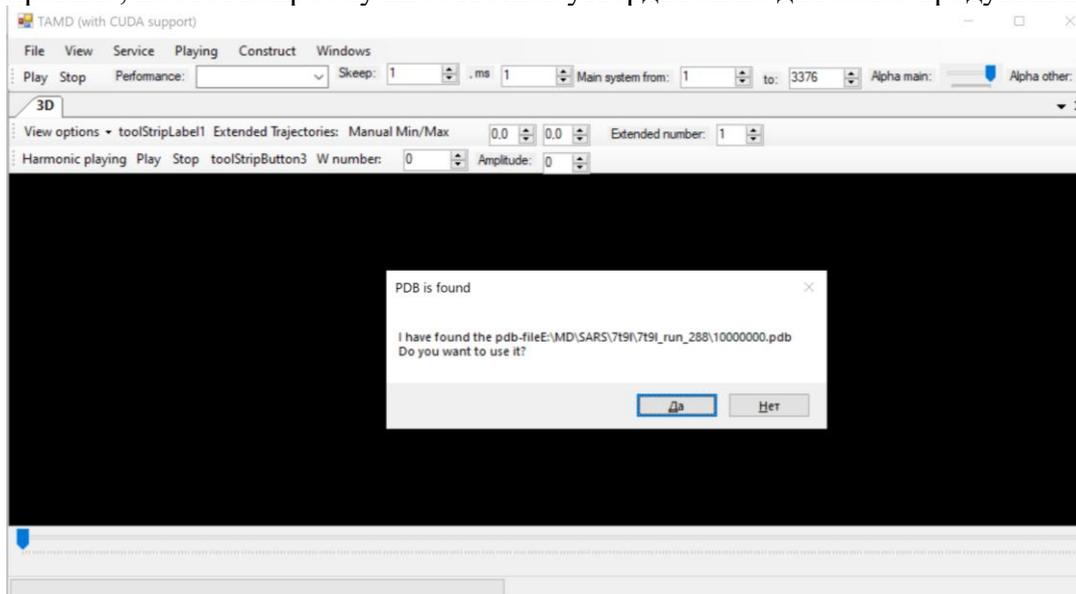


Рисунок 14. Диалог повторного открытия pdb-файла.

Анализатор траекторий содержит класс для работы с pdb-файлом. Он способен загружать файл как с диска, так и из текстовой строки.

Загрузка pdb-файла из строки сделана для чтения файла из баз данных. В них pdb-файлы хранятся как текстовые строки большой длины.

Для хранения списка пар «траектория – pdb-файл» про программе используется таблица СУБД Access Loader. При загрузке новой траектории и pdb-файла создается новая строка, заполняя поля FileNameTrj и FileNamePDB полными путями к файлам. Поле ID проставляется автоматически как первичный ключ. При загрузке траектории ведётся поиск по полю FileNameTrj. Если строчка находится, то пользователю предлагается загрузить уже запомненный pdb-файл.

PDB-файл загружается в оперативную память целиком. При чтении файла загружаются, в том числе, и координаты. Как правило, в программе везде используются координаты из текущего кадра траектории. Координаты из pdb-файла используются, к примеру, в режиме конструирования систем.

4.3 Характеристики вдоль траектории

Характеристики вдоль траектории – это функции от времени, значение которых вычисляется по координатам одного кадра.

$$Y=F(\text{frame})$$

где frame – номер кадра траектории.

Список поддерживаемых на момент написания документации функций:

1. Distance (Atom1, Atom2)
2. Angle (Atom1, Atom2, Atom3)
3. R

4. Angle (Atom1, Atom2, Atom3, Atom4)
5. Angle (Atom1, Atom2, Atom3, Atom4) by R
6. Distance (Atom1-Atom2, Atom3-A4)
7. Residue's fluctuations
8. Angle (Atom1-Atom2, Atom3-A4)
9. Angle of projections (Acid2-Acid1, Acid3-Acid4)
10. Angle (Acid2-Atom1, Acid3-A4)
11. C-alpha fluctuations
12. C-alpha fluctuations by R (all trajectory)
13. C-alpha fluctuations by R (frames a1 and a2)
14. All atoms fluctuations by R (all trajectory)
15. All atoms fluctuations by R (frames a1 and a2)
16. All atoms fluctuations by R (all trajectory SELECTED atoms)
17. All atoms fluctuations by R (frames a1 and a2 SELECTED atoms)
18. C-alpha fluctuations by R (frames from a1 to a2)
19. C-alpha pairs distances for frame a1
20. Ramachandran map (phi, psi)

Аргументами функций, как правило, служат номера атомов (atom1, atom2, Первый атом имеет порядковый номер 1.

Также некоторые функции принимают на вход номера кадров траектории. Первый кадр траектории также имеет порядковый номер 1.

Все функции выводят результат своей работы в модуль работы с двумерными графиками (TAMD-Grapher, см. соответствующую главу документации).

4.3.1 Расстояние между атомами (функция Distance)

Функция вычисляет расстояние между атомами с номерами Atom1 и Atom2 в декартовом смысле, как квадратный корень из суммы координат.

4.3.2 Угол между атомами (функция Angle от 3 атомов)

Функция вычисляет угол между векторами, образованными атомами (Atom1, Atom2) и (Atom2, Atom3) по определению угла между векторами.

Полезна для вычисления валентных углов.

4.3.3 Радиус инерции системы (функция R)

Находится геометрический центр системы M(x, y, z).

Затем для каждого атома вычисляется матрица

$$R = \begin{vmatrix} r_{xx} & r_{xy} & r_{xz} \\ r_{yx} & r_{yy} & r_{yz} \\ r_{zx} & r_{zy} & r_{zz} \end{vmatrix}$$

Где $r_{a,b} = (\text{Atom.a-M.a}) * (\text{Atom.b-M.b})$, a, b $\in \{x, y, z\}$

Находится сумма матриц R. Радиус инерции равен квадратному корню из элементов главной диагонали матрицы инерции.

$$R_{inertia} = \sqrt{(r_{xx}^2 + r_{yy}^2 + r_{zz}^2)}$$

Операция повторяется для каждого кадра траектории.

4.3.4 Угол между атомами (функция Angle от 4 атомов)

Функция вычисляет угол между векторами, образованными атомами (Atom1, Atom2) и (Atom3, Atom4). По определению угла между векторами.

Полезна для вычисления торсионных углов.

4.3.5 Угол между двумя подсистемами

Функция вычисляет угол между двумя подсистемами. Каждая подсистема задается диапазоном атомов [Atom1; Atom2] и [Atom3; Atom4]. Вектор, представляющий подсистему, рассчитывается как главная ось тензора инерции. Угол между подсистемами определяется как угол между главными осями инерции подсистем.

4.3.6 Расстояние между диапазонами атомов (функция Distance от 4 атомов)

Функция вычисляет расстояние между двумя подсистемами. Каждая подсистема задается диапазоном атомов [Atom1; Atom2] и [Atom3; Atom4]. Точка, представляющая подсистему, рассчитывается как геометрический центр соответствующей подсистемы. Расстояние между подсистемами определяется как расстояние между их геометрическими центрами.

4.3.7 Флуктуации (колебания) аминокислотных остатков (функция Residue's fluctuations)

Функция вычисляет дисперсию колебаний аминокислотных остатков.

Для каждого аминокислотного остатка на каждом шаге траектории рассчитывается его геометрический центр. Далее находится среднее значение длины вектора – центра остатков вдоль всей траектории, а также его квадрата. Дисперсия вычисляется по формуле:

$$D[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

4.3.8 Угол между векторами (Atom1-Atom2, Atom3-Atom4) (функция Angle от 4 атомов)

Численно равна функции Angle (Atom1, Atom2, Atom3, Atom4), вычисляющей угол между векторами, проходящими через атомы (Atom1, Atom2) и (Atom3, Atom4).

Реализовано для удобства физиков.

4.3.9 Угол между проекциями аминокислот

Рассчитывается угол между проекциями векторов (Acid1, Acid2) и (Acid3, Acid4), соединяющие геометрические центры аминокислотных остатков, на плоскость, нормаль к которой является вектором, соединяющим геометрические центры аминокислотных остатков Acid1 и Acid4.

Работает совместно с опцией «only acute corners». Если она выбрана и угол больше 90°, то угол вычитается из 180°.

4.3.10 Угол между четырьмя остатками (Acid2-Acid1, Acid3-A4)

Рассчитывается угол между векторами, образованными геометрическими центрами четырёх аминокислотных остатков.

Подобен расчету угла, заданному четырьмя атомами. Только в качестве вершин векторов выступают геометрические центры аминокислотных остатков.

Работает совместно с опцией «only acute corners». Если она выбрана и угол больше 90°, то угол вычитается из 180°.

4.3.11 Флуктуации C-alpha-атомов (C-alpha fluctuations)

Рассчитываются средние отклонения C-alpha-атомов от их средних положений. Средние положения определяются вдоль всей траектории. Таким образом, это двухпроходный алгоритм.

4.3.12 Флуктуации C-alpha-атомов в координатах центра тяжести по всей траектории (C-alpha fluctuations by R (all trajectory))

Рассчитываются средние отклонения C-alpha-атомов от их средних положений. В отличие от предыдущего пункта, координаты атомов переводятся в координаты относительно главных осей тензора инерции. Таким образом, алгоритм даёт независимость от перемещения и вращения белковой глобулы как целого.

Расчёт ведётся по всей траектории.

4.3.13 Флуктуации C-alpha-атомов в координатах центра тяжести между кадрами a1 и a2 (C-alpha fluctuations by R (frames a1 and a2))

Рассчитываются средние отклонения C-alpha-атомов от их средних положений.

Координаты атомов переводятся в координаты относительно главных осей тензора инерции. Таким образом, алгоритм даёт независимость от перемещения и вращения белковой глобулы как целого.

Расчёт ведётся по заданным кадрам a1 и a2.

4.3.14 Флуктуации всех атомов в координатах центра тяжести (All atoms fluctuations by R (all trajectory))

Рассчитываются средние отклонения всех атомов от их средних положений. Координаты атомов переводятся в координаты относительно главных осей тензора инерции. Таким образом, алгоритм даёт независимость от перемещения и вращения белковой глобулы как целого.

Расчёт ведётся по всей траектории.

4.3.15 Флуктуации всех атомов в координатах центра тяжести (All atoms fluctuations by R (frames a1 and a2))

Рассчитываются средние отклонения всех атомов от их средних положений. Координаты атомов переводятся в координаты относительно главных осей тензора инерции. Таким образом, алгоритм даёт независимость от перемещения и вращения белковой глобулы как целого.

Расчёт ведётся по заданным кадрам a1 и a2.

4.3.16 Флуктуации заданных аминокислотных остатков в координатах центра тяжести (All atoms fluctuations by R (all trajectory SELECTED atoms))

Рассчитываются средние отклонения заданных аминокислотных остатков от их средних положений. В отличие от предыдущего пункта, координаты атомов переводятся в координаты относительно главных осей тензора инерции. Таким образом, алгоритм даёт независимость от перемещения и вращения белковой глобулы как целого.

Расчёт ведётся по всей траектории.

Выбор аминокислотных остатков осуществляется по имени в блоке выбора типов атомов Selection. В качестве имени выступает поле Name. Редко используемая возможность.

4.3.17 Флуктуации заданных аминокислотных остатков в координатах центра тяжести (All atoms fluctuations by R (frames a1 and a2 SELECTED atoms))

Рассчитываются средние отклонения заданных аминокислотных остатков от их средних положений. В отличие от предыдущего пункта, координаты атомов переводятся в координаты относительно главных осей тензора инерции. Таким образом, алгоритм даёт независимость от перемещения и вращения белковой глобулы как целого.

Расчёт ведётся по заданным кадрам a1 и a2.

Выбор аминокислотных остатков осуществляется по имени в блоке выбора типов атомов Selection. В качестве имени выступает поле Name. Редко используемая возможность.

4.3.18 Флуктуации C-alpha-атомов в координатах центра тяжести между кадрами a1 и a2 (C-alpha fluctuations by R (frames from a1 to a2))

Рассчитываются средние отклонения C-alpha-атомов от их средних положений.

Координаты атомов переводятся в координаты относительно главных осей тензора инерции. Таким образом, алгоритм даёт независимость от перемещения и вращения белковой глобулы как целого.

Расчёт ведётся по заданным кадрам a1 и a2.

4.3.19 Расстояния между парами C-Alpha-атомов для кадра a1 (C-alpha pairs distances for frame a1)

Рассчитываются расстояния между парами C-alpha-атомов в кадре a1.

4.3.20 Карта Рамачандрана (Ramachandran map (phi, psi))

Рассчитываются карта Рамачандрана для кадра a1.

Карта Рамачандрана представляет собой зависимость углов ϕ от ψ . Эти углы образуются первыми тремя атомами аминокислотных остатков.

Угол ϕ находится по координатам 1, 2, 3 атомов текущего остатка и 1-го атома следующего остатка. Угол ψ находится по координатам 3-го атома предыдущего остатка и 1-го, 2-го, 3-го атомов текущего остатка. Если 3-го атома предыдущего остатка не существует, то берётся второй. При его отсутствии – первый.

4.4 Средство работы с графиками TAMD-Grapher

Для отображения графиков всех функций, а также других зависимостей предусмотрено средство TAMD-Grapher.

TAMD-Grapher позволяет:

- получать графические представления зависимостей;
- выдавать несколько кривых на одном графике;
- менять состав одновременно видимых графиков;
- изменять масштаб графиков;
- задавать цвет графиков;
- изменять представление графиков;
- добавлять подписи к осям;
- строить диаграмму распределения плотности величины;
- находить быстрое преобразование Фурье величины;
- сохранять графики в пиксельном представлении, а также в числовом;
- загружать и анализировать данные из внешних программ.

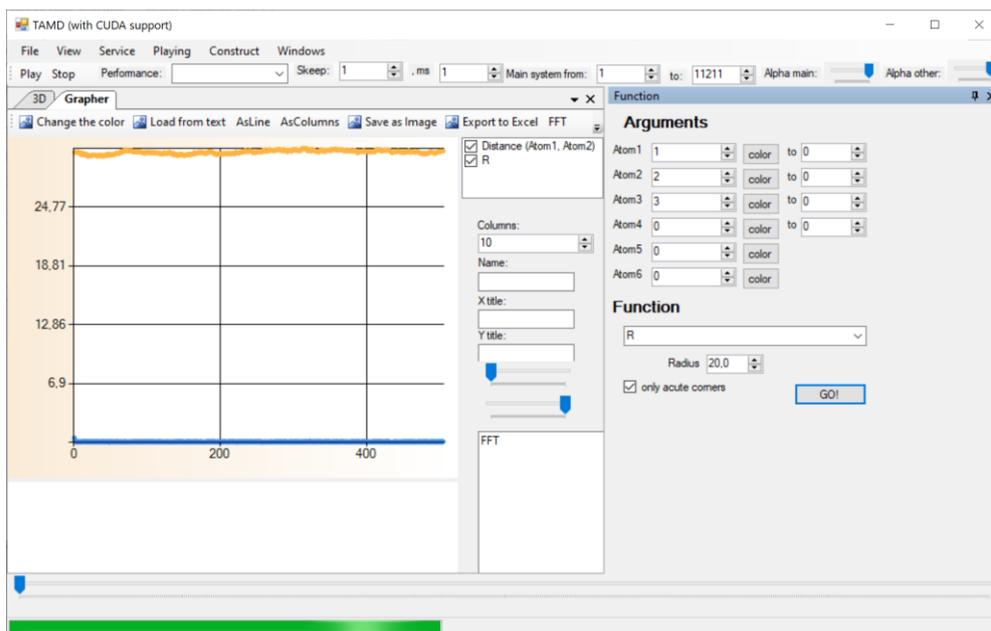


Рисунок 15. Программа TAMD с одновременным отображением средств вычисления функций (справа, окно Function) и построения графиков (слева, окно Grapher).

4.4.1 Вывод графиков

После окончания расчета вычисления любой функции данные попадают в средство работы с графиками. Окно TAMD-Grapher открывается автоматически. Также его можно открыть, воспользовавшись меню View – Grapher, либо горячей клавишей F6.

Каждая функциональная зависимость попадает не только на график, но и в список графиков, он отображен справа сверху окна TAMD-Grapher.

График можно убрать из окна представления графической информации и отобразить снова, воспользовавшись галочкой, слева от названия графика.

Выбор *текущего* графика осуществляется нажатием по названию графика.

4.4.2 Изменение масштаба графиков

После загрузки каждого нового графика масштаб графического представления подстраивается автоматически.

Для изменения масштаба служат ползунки Scale, снизу от подписей осей. Первый ползунок соответствует минимальной ординате, а второй ползунок максимальной ординате. После загрузки нового графика ползунки отводятся в минимальное и максимальное положение соответственно, что соответствует автоматически выбранному масштабу.

4.4.3 Изменение цвета графиков

После загрузки каждого нового графика кривая получает новый цвет.

Чтобы изменить цвет графика, требуемый график необходимо сделать *текущим*. Для этого необходимо щелкнуть по названию графика в списке графиков. Щелкать лучше именно по названию, не затрагивая галочку, которая отвечает за видимость графика. Далее нажимается кнопка «Change the color», которая открывает диалог выбора цвета.

4.4.4 Изменение представления графиков

После загрузки каждого нового графика почти всегда кривая отображается в виде линии. Исключение составляет построение карты Рамачандрана, у которой график отображается в виде несвязанных точек.

В TAMD-Grapher к моменту написанию текущей документации предусмотре вывод графиков в виде:

- линии;
- колонок;
- точек.

Для изменение вывода графика, требуемый график необходимо сделать *текущим*. Для этого необходимо щелкнуть по названию графика в списке графиков. Щелкать лучше именно по названию, не затрагивая галочку, которая отвечает за видимость графика. Чтобы изменить вывод текущего графика, достаточно щелкнуть по кнопке «As Line», «As Columns», либо выбрать вид графика в меню «Show as». В этом меню доступны пункты «Line», «Columns» и «Points».

4.4.5 Сохранение графика в виде рисунка

Для сохранения текущего представления графиков в виде точечного рисунка реализована кнопка «Save as image», которая экспортирует картинку в формат JPEG.

Размер окна TAMD и разрешение экрана влияют на качество изображения. Поэтому можно порекомендовать развернуть окно программы TAMD на весь экран.

4.4.6 Экспорт графика в Excel

Для сохранения текущих данных в формате электронной таблицы Excel реализована кнопка «Export to Excel», которая экспортирует картинку в программу Excel.

При экспорте в таблицу каждый график занимает один столбец. Также добавляется один столбец «Step» в начале в качестве оси абсцисс, он соответствует номеру шага. Зная шаг траектории (который обычно составляет 1 или 10 пс), легко вычислить время.

Если количество точек равно количеству аминокислотных остатков, то в конец добавляются столбец «Residue» с трехсимвольным обозначением остатка из .pdb-файла и столбец «Residue number», соответствующий номеру остатка.

Программа Excel, либо пакет MS Office, должны быть установлены на компьютере.

4.4.7 Плотность распределения

Диаграмма плотности распределения строится всякий раз при выборе конкретного графика в списке графиков. За количество столбцов диаграммы отвечает параметр Columns. При его изменении диаграмма плотности распределения пересчитывается автоматически.

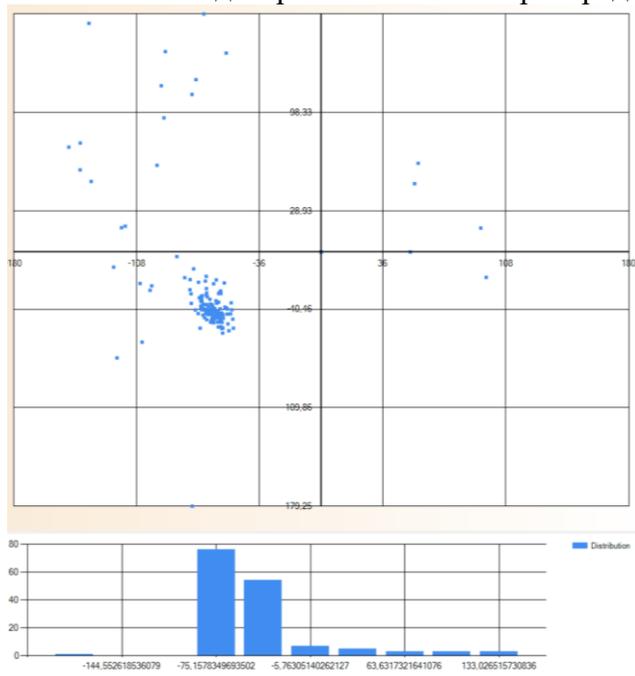


Рисунок 16. График функции (сверху) и его плотность распределения (снизу)

4.4.8 Быстрое преобразование Фурье

Построение быстрого преобразования Фурье текущего графика доступно по кнопке FFT. Результат построения отображается на диаграмме внизу окна Grapher.

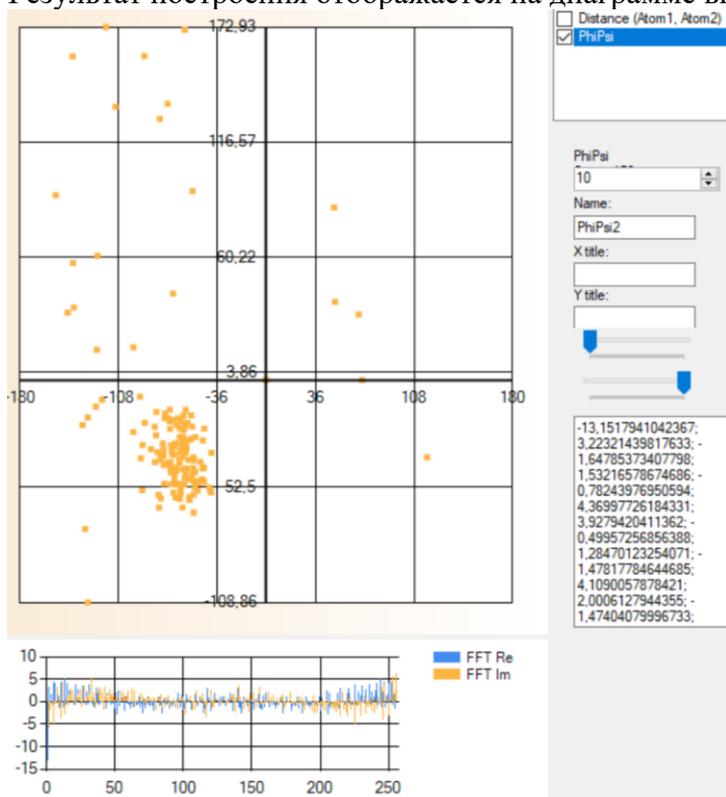


Рисунок 17. График функции (сверху) и его частотный спектр (снизу)

В правом нижнем служебном окне отображаются числа (в формате csv – comma separated values) для возможности их копирования в другие программы. В Excel они не экспортируются при экспорте всех графиков, чтобы не загромождать лишней информацией.

4.4.9 Выход на плато

Возможность служит для определения времени, через которая система выходит на плато. Как правило, речь идёт об энергии, либо о другой макроскопической величине.

Рассматриваются последние 10 точек графика и передаётся среднее их значение и дисперсия. Используя вычисленное среднее значение, определяется точка графика (момент времени), когда функция впервые подходит к этой точке на расстояние не более, чем 10 единиц измеряемой величины (граница поиска).

Результат вычислений выводится в журнал (лог), который доступен через меню Service – Utils – Log.

Количество точек усреднения и границу поиска можно менять, используя TAMD-Script – см. соответствующий раздел.

4.5 Контактные карты

Контактные карты, как инструмент анализа молекулярных структур, начали применять с 70-80 годов. Обычная контактная карта представляет собой булеву матрицу, часто изображаемую в виде двуцветного точечного рисунка. Анализ таких карт не представляет труда. Благодаря контактной карте можно быстро охарактеризовать сложную трехмерную статическую молекулярную структуру.

Авторы статьи предлагают новый подход к построению контактных карт, пригодный, в том числе, и для динамических систем. Также в статье рассказывается о быстрых интерактивных методах интерпретации таких данных.

4.5.1 Булева контактная карта

Рассмотрим молекулярную систему (группу молекул, либо часть макромолекулы), состоящую из N атомов. Будем считать, что один атом контактирует с другим, если расстояние между ними не больше некоторого заданного числа r_{\max} . Определим матрицу A размером $N \times N$ таким образом, что каждый ее элемент:

$$a_{i,j} = \begin{cases} 1, & \text{если } R(i, j) \leq r_{\max} \\ 0, & \text{если } R(i, j) > r_{\max} \end{cases}, \quad (1)$$

где $R(i, j)$ – расстояние между атомами i и j .

Выбор параметра r_{\max} зависит от рода анализируемых контактов. В примерах авторы исследуют траектории белковых молекул. Было выбрано характерное расстояние 8 \AA для подсчета контактов между атомами различных вторичных структур.

Матрицу $A = \{a_{i,j}\}$ назовем матрицей контактов или *контактной картой*.

На диагонали матрицы A будут стоять единицы.

Заметим, что такую контактную карту в виде матрицы достаточно просто построить на любом устройстве вывода информации, например текстовом, в виде, матрицы единиц и нулей. Однако текст занимает слишком много места и не подходит для случаев, когда речь идет о сотнях и тысячах атомов.

Матрицу A можно представить в виде двухцветного точечного рисунка на устройстве вывода графической информации, таком, как экран или принтер (0 – белый цвет, 1 – черный).

В то же время возможности современных устройств вывода графической информации (равно как и органов зрения человека) намного превосходят 2 цвета. Предложим алгоритмы построения контактных карт, более полно использующие возможности человеческого зрения.

4.5.2 Построение монохромной контактной карты

Изображение на современных цветных устройствах вывода графической информации (монитор, принтер) представляется в виде матрицы цветных пикселей – наиболее мелких единиц изображения. Каждый пиксель имеет свой цвет, который кодируется определенным образом в виде целого числа, в зависимости от принятой системы кодирования цветов. Разберем построение контактной карты системе RGB (красный, зеленый, синий). В ней каждый цвет записывается в виде яркости трех его составляющих: красного, зеленого и синего (r, g, b). Каждая составляющая принимает значения в диапазоне от 0 до 1. Значение $(0, 0, 0)$ соответствует черному цвету, $(1, 1, 1)$ белому, $(1, 0, 0)$ красному и т.д.

Для построения точечного рисунка введем закон, по которому одновременно будут изменяться яркости всех каналов.

Каждый элемент матрицы A представим в виде вектора, компонентами которого являются яркости красного (r), зеленого (g) и синего (b) каналов:

$$a_{i,j} = (r_{i,j}, g_{i,j}, b_{i,j}), i, j = 1, \dots, n$$

Для удобства примем, что яркость каждого канала меняется в диапазоне $[0, 1]$. Тогда

$$r_{i,j} = g_{i,j} = b_{i,j} = \begin{cases} 1, & \text{если } R(i, j) \leq r_{\min} \\ 1 - \frac{(R(i, j) - r_{\min})}{(r_{\max} - r_{\min})}, & \text{если } r_{\min} < R(i, j) \leq r_{\max} \\ 0, & \text{если } R(i, j) > r_{\max} \end{cases}, \quad (2)$$

$r_{\min} < r_{\max}$ – наперед заданные положительные числа. От выбора этих параметров зависит чувствительность контактной карты.

Выбор яркости цветowych каналов в равной пропорции, согласно формулам (2), дает контактную карту в градациях серого цвета. Полученную контактную карту назовем *монохромной*.

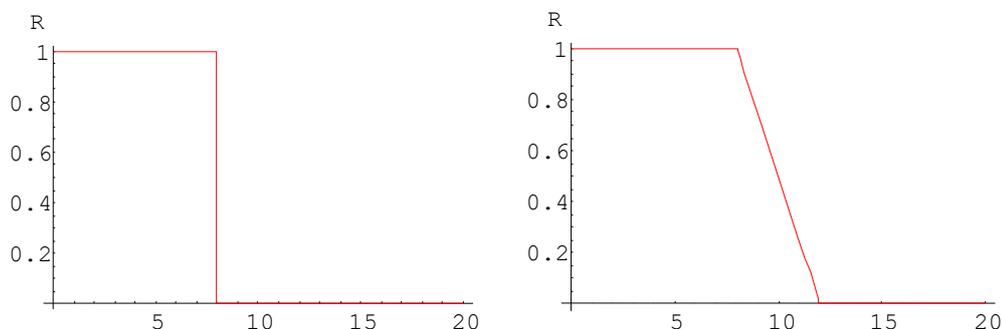


Рисунок 18. Сравнение булевой (слева) и монохромной (справа) контактных карт

Для построения контактной карты будем использовать не все три канала, а только один канал – зеленый.

$$r_{i,j} = const1$$

$$g_{i,j} = \begin{cases} 1, & \text{если } R(i, j) \leq r_{\min} \\ 1 - \frac{(R(i, j) - r_{\min})}{(r_{\max} - r_{\min})}, & \text{если } r_{\min} < R(i, j) \leq r_{\max} \\ 0, & \text{если } R(i, j) > r_{\max} \end{cases}, \quad (3)$$

$$b_{i,j} = const2$$

Таким образом мы получили контактную карту в градациях зеленого цвета (при нулевых константах). В системе (3) константы const1 и const2 задаются предварительно до построения точечного рисунка, что будет использовано в следующем разделе.

4.5.3 Расширение возможностей контактных карт при анализе структурированных систем при помощи построения полноцветной контактной карты

Под *структурированной системой* понимается такая молекулярная система, в которой можно выделить по порядковым номерам группы подряд идущих атомов согласно некоторому смысловому критерию. К примеру, это может быть отнесение атомов белковой молекулы к элементам вторичной структуры.

Заметим, что часто неструктурированную систему можно представить в виде структурированной при помощи перенумерации атомов.

Согласно формуле (3) мы имеем 2 свободных канала (каналы красного и синего).

Цвет элемента структуры заранее задается для каждого значимого с точки зрения проводимого эксперимента элемента структуры.

Для обозначения каждого элемента структуры нам понадобится два числа: первый атом и последний (s_i, e_i), $1 \leq i \leq s$, $s < n$, i – номер элемента структуры, s – количество элементов структуры. Будем отмечать эти числа сразу на двух осях. Если обозначить область между ними: $s_i \leq x \leq e_i$, $s_i \leq y \leq e_i$, то мы получим квадрат, у которого одна из диагоналей совпадает с главной диагональю контактной карты. Полученная область подвергается заливке (установке всех пикселей, входящих в заданную область) в цвет, заранее сопоставленный группе. При этом будем использовать только свободные красный и синий каналы.

Теперь нам потребуется указать на карте области, отвечающие за контакт двух структур i и j . Для этого потребуется залить два симметричных относительно главной диагонали участка: $s_i \leq x \leq e_i$, $s_j \leq y \leq e_j$ и $s_j \leq x \leq e_j$, $s_i \leq y \leq e_i$, $1 \leq i \leq s$, $1 \leq j \leq s$. Для цвета будем по-прежнему использовать только красный и синий каналы. Перед построением карты задается только цвет для каждой группы. Для определения цвета области контакта между группами будем использовать побитовое сложение. С одной стороны, это простая операция, которая требует минимум машинного времени, а с другой стороны получаемые цвета становятся логически понятными.

Таким образом мы получили расширенную контактную карту, обладающую следующими свойствами:

- расширенная контактная карта по-прежнему несет информацию о контактах всех атомов, пригодную для быстрой интерпретации;
- имеется возможность визуального выделения контактов между группами.

4.5.5 Программная реализация

Траектория молекулярной динамики белка получена при помощи программного комплекса PUMA [1].

Изложенный теоретический материал о контактных картах реализован в программно-алгоритмическом комплексе «Анализатор траекторий молекулярной динамики» (TAMD) [2]. Анализатор обладает следующими возможностями по построению контактных карт:

- построение обычных монохромных и расширенных контактных карт по любому кадру траектории;
- вывод контактных карт в динамике синхронно с выводом интерактивного трехмерного молекулярного кино;
- вывод информации по атомам на контактной карте с быстрым переходом на расположение атома в пространстве;
- экспорт динамики контактов в виде последовательности графических файлов (image sequence).

4.5.6 Пример использования

Методику исследования динамической белковой структуры при помощи динамических расширенных контактных карт и лент контактов рассмотрим на примере В1 домена иммуноглобулина связывающего белка L (в дальнейшем 2PTL) [3]. Средствами программного комплекса по моделированию молекулярной динамики был проведен математический эксперимент по растяжению белковой глобулы под действием постоянных сил, приложенных к концам макромолекулы. Белок состоит из 2-х бета-шпилек, одной альфа-спирали и хвоста.

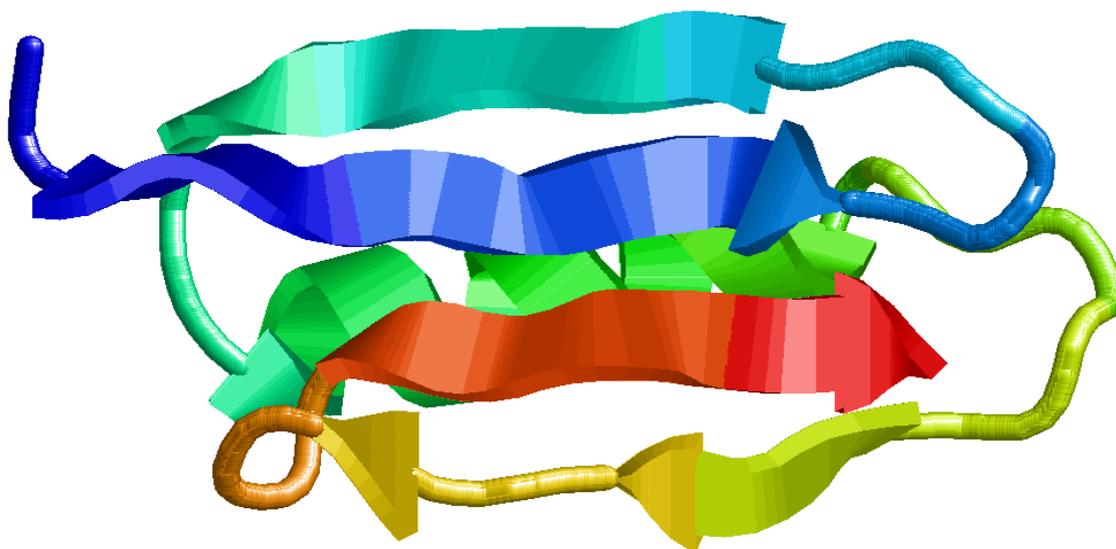


Рисунок 19. Изображение структуры белка 2 ptl.

4.5.7 Методика проведения исследования. Интерпретация результатов

Сначала построим монохромную контактную карту белковой глобулы до начала растяжения.

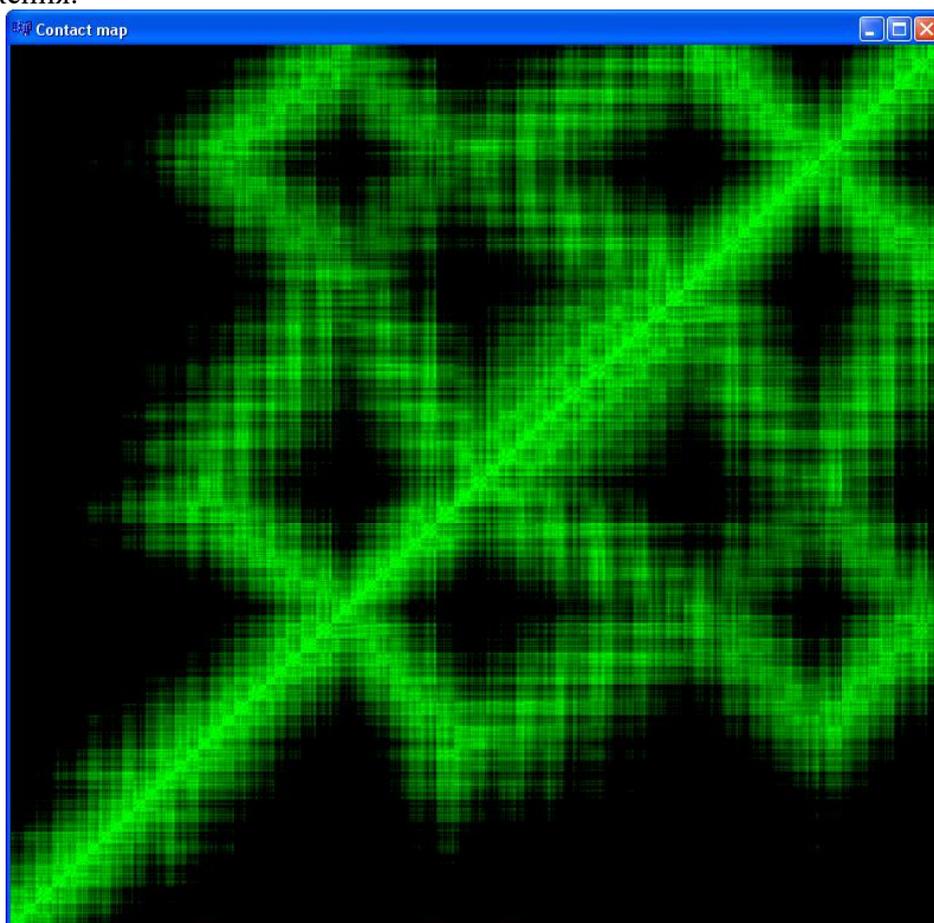


Рисунок 20. Контактная карта белка. Начало эксперимента.

Для построения карты мы использовали систему координат, в которой элемент матрицы $a_{1,1}$ расположен в левом нижнем углу окна.

Четко видна главная диагональ. Толщина главной диагонали говорит о близком расположении в пространстве соседних через 1, через 2 и т.д. атомов.

Также наблюдаются отрезки, параллельные и перпендикулярные главной диагонали. Они говорят о существовании контактов между целыми участками цепи. Чем дальше отрезок от диагонали, тем дальше по белковой цепи контактируемые участки расположены друг от друга. Параллельность либо перпендикулярность говорит о том, в каком направлении осуществляется контакт. Параллельность отрезка главной диагонали означает, что начало одного участка цепи контактирует с началом другого. Перпендикулярность – начало одного с концом другого.

При помощи интерактивного программного инструмента легко можно узнать на карте, какие атомы отвечают за тот или иной пиксель. Данный тип анализа является достаточно подробным. При возникновении потребности в оценке динамики контактов всей траектории рекомендуется воспользоваться лентами контактов.

4.5.8 Динамика

Как видно из названия, программный комплекс «Анализатор траекторий молекулярной динамики» создавался изначально для анализа молекулярных структур во времени.

В программе реализован вывод контактной карты в виде непрерывной анимации совместно с молекулярным кино.

4.5.9 Ускорение вывода контактных карт

Заметим, что каждый элемент матрицы контактной карты может быть рассчитан независимо от остальных. Благодаря этому факт возможно использование технологий параллельного программирования для ускорения построения контактных карт. В TAMD при построении контактной карты применяются следующие технологии параллельного программирования:

1. Многопроцессорная обработка.
2. Использование графических ускорителей, работающих по технологии CUDA

Выбор технологии ускорения расчетов осуществляется из выпадающего списка Performance главного окна. Необходимо выбрать Many CPUs для многопроцессорной обработки и CUDA для расчета на видеоускорителях.

4.5.10 Ленты контактов

Лента контактов представляет сбор графики зависимостей количества контактов между всеми подструктурами.

Для вычисления ленты контактов используется пункт «Type of contacts» меню View. Для начала необходимо разметить структуру. Делается это при помощи средства выделения подсистем. Для этого можно воспользоваться пунктом «Selection» меню «View», либо горячей клавишей F9.

В качестве примера выделено две подсистемы: первая будет изображаться зелёным цветом, а вторая красным:

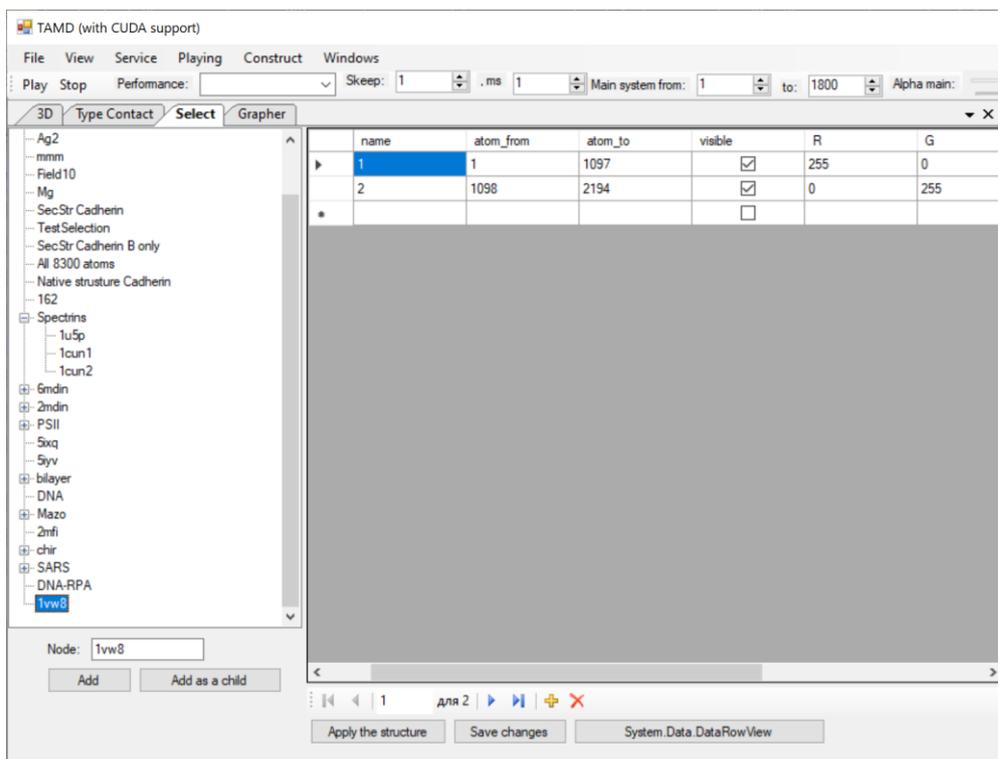


Рисунок 21. Выделение двух подсистем в окне Selection. После сохранения информации о подсистемах по кнопке «Save changes» и принятия изменений по кнопке «Apply the structure» в окне трёхмерного просмотра будут изображаться две подструктуры:

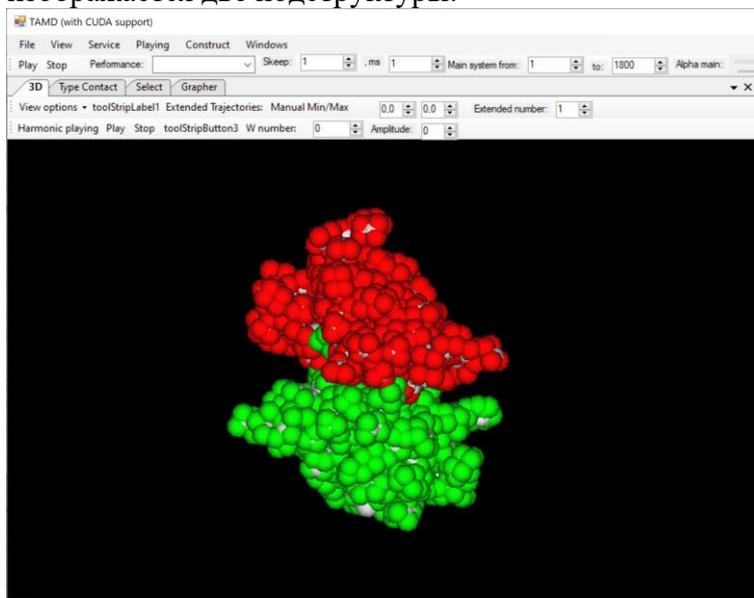


Рисунок 22. Вывод двух подсистем в окне вывода трехмерного изображения. Затем необходимо выбрать пункт «Type of contacts» меню View. Лента контактов строится по всей траектории, ли по интересующей её части.

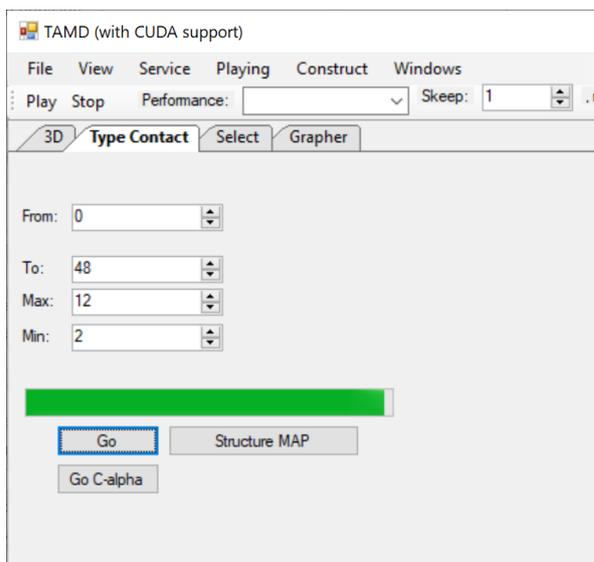


Рисунок 23. Средство построения лент контактов.

Значения полей следующие:

- From – начальный кадр траектории для построение ленты контактов.
- To – конечный кадр траектории. При открытии траектории на поле накладывается ограничение сверху, равное количеству кадров.
- Min – нижняя граница контакта.
- Max – верхняя граница контакта.

Программа просматривает контакты между всеми атомами выбранных структур, а также между структурами. Если расстояние между атомами меньше Min, то расстоянию приписывается 1 (контакт), если больше Max, то 0 (атомы не взаимодействуют). Промежуточный контакт от Min до Max соответствует линейно убывающей величине контакта от 1 до 0:

$$d = 1 - \frac{d - Min}{Max - Min}$$

Где d – величина контакта.

Для каждой структуры суммируются все величины d , указывающие суммарное количество контактов.

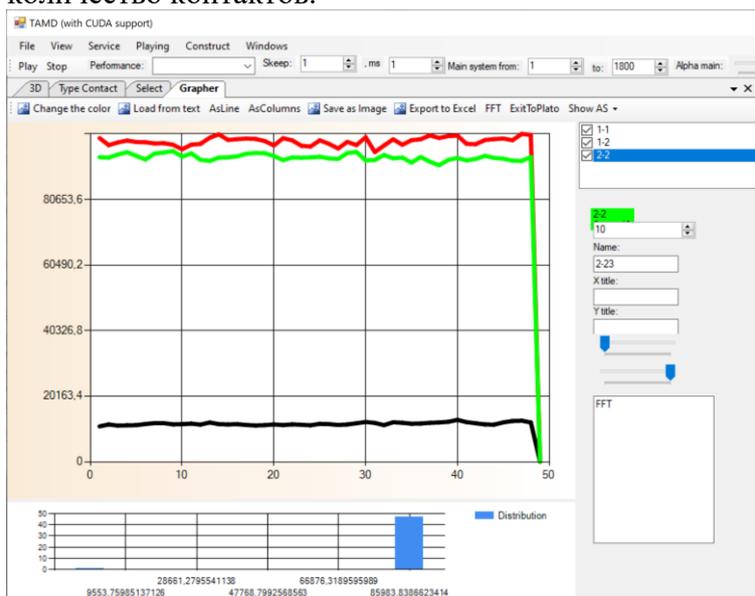


Рисунок 24. Вывод лент контактов в окне Grapher.

Как видно из Рисунок 24, построены три «ленты» – это графики расстояний внутри структур, а также между структурами:

- красная кривая – контакты внутри первой структуры;

- черная кривая – контакты между первой и второй структурами;
- зелёная кривая – графики внутри второй структуры.

Таким образом контактная карта даёт быстрое представление о динамике контактов между структурами в течении МД-эксперимента.

Работая со средством построения графиков «Grapher», удобно отключать контакты внутри структур и оставлять только между структурами, т.к. обычно их масштаб разный.

4.5.10 О перспективах программы TAMD

Средство работы с контактными картами, а также его ускорение с использованием технологий параллельного программирования было сделано в районе 2010 года. С тех пор работа с многопроцессорными системами в .NET практически не изменилась. А для технологии CUDA многое поменялось. NVIDIA CUDA ориентировано, в первую очередь, на язык C++, а Анализатор траекторий написан на C#. Для этого языка в районе 2010 года существовала единственная библиотека CUDA.NET, которая к настоящему моменту времени признана устаревшей. По всей вероятности, при потребности в написании вычислительно ресурсоемких возможностей будут использоваться новые библиотеки. Учитывая текущий год написания документации (2023), это будет ManagedCuda или Cudafy.

Пользователи не пострадают. По сравнению с 2010 годом, количество ядер на центральных процессорах значительно увеличилось, как и возросла производительность каждого ядра. Поэтому скорость вычисления контактной карты в многопроцессорном режиме весьма высока, даже для больших систем.

4.6 Коэффициент вариации молекулярной динамики. Распознавание процессов на исследуемой системе без предварительных сведений

В данном разделе работы предлагается алгоритм получения новой характеристики молекулярной системы, которая позволяет говорить о поведении системы в целом. Коэффициент формы был взят из обобщенного спектрально-аналитического метода (OSAM) [3].

Траектория делится на временные промежутки – окна. Осуществляется переход к траектории в торсионных углах. Каждый торсионный угол определяется положением четырех атомов. У каждого торсионного угла на промежутке окна определяется коэффициент формы. Таким образом на каждом окне получается вектор коэффициентов формы. У каждого такого набора коэффициентов находится дисперсия. Полученная характеристика называется коэффициентом вариации.

Коэффициент вариации позволяет выделить места перестройки структуры макромолекулы. Дополнительную информацию о локализации изменений дают мгновенные значения коэффициентов формы на интервале.

4.6.1 Использование коэффициента формы при анализе траекторий молекулярной системы

Под коэффициентом формы в общем случае понимается следующая характеристика процесса:

$$K = \frac{\int f(t)\gamma(t)dt}{\int f(t)\eta(t)dt}, \quad (4.6.1)$$

где $f(t)$ - исследуемый процесс, $\gamma(t)$ и $\eta(t)$ - некоторые функции.

Для анализа траектории в качестве функций $\gamma(t)$ и $\eta(t)$ будем использовать полиномы первой степени. Пусть процесс $f(t)$ определен на интервале $t \in [0; T]$. Зададим функции $\gamma(t)$ и $\eta(t)$ в следующем виде:

$$\begin{aligned} \gamma(t) &= T - t \\ \eta(t) &= t \end{aligned} \quad (4.6.2)$$

Подставляя (2) в (1), получим:

$$K_1 = \frac{\int_0^T f(t)(T - t)dt}{\int_0^T f(t)tdt} \quad (4.6.3)$$

Смысл коэффициента формы таков, что

$$\begin{cases} K_1 > 1; \text{ для убывающих процессов} \\ K_1 \approx 1; \text{ для стационарных процессов} \\ K_1 < 1; \text{ для возрастающих процессов} \end{cases} \quad (4.6.4)$$

2.5.2 Алгоритм

Пусть имеется траектория молекулярной системы, записанная в трехмерных пространственных координатах. Перейдем к траектории в торсионных углах. Каждый торсионный угол определяется положением четырех атомов. Информацию о торсионных углах TAMД получает из соответствующего структурного файла, созданного программным комплексом PUMA.

Переход к торсионным углам исключает зависимость от поступательного и вращательного движений системы как целого. Нас будет интересовать динамика процесса. Вся траектория длиной $l=mT$ (m и T – натуральные числа) делится на участки длиной T по временной шкале (окна). Ширина окна T является одним из параметров. Выбор T может зависеть от длины траектории и от степени детализации проводимого исследования.

В каждом окне $j=1, \dots, m$ у каждого торсионного угла $\varphi_i^{(j)}(t)$ рассчитывается коэффициент формы:

$$K_i^{(j)}(\varphi_i^{(j)}(t)) = \frac{\int_{(j-1)T}^{jT} \varphi_i^{(j)}(t)(T - t)dt}{\int_{(j-1)T}^{jT} \varphi_i^{(j)}(t)tdt}, \quad i = 1, \dots, n; \quad j = 1, \dots, m \quad (4.6.5)$$

где n – количество торсионных углов в молекуле (молекулярной системе).

Таким образом получится набор коэффициентов формы

$$K^{(j)} = \{K_i^{(j)}, i = 1, \dots, n\} \quad (4.6.6)$$

Мы будем находить дисперсию каждого набора коэффициентов формы в каждом окне. Это и будет новая характеристика молекулярной системы.

Необходимо отличать процессы динамической активности от стационарных, то есть области возрастания или убывания углов. При этом не важно направления изменения

торсионного угла, а лишь факт нестационарности процесса. Будем искать среднеквадратичное отклонение от единицы следующей величины:

$$K_{1+} = \begin{cases} K_1; & \text{если } K_1 \geq 1 \\ \frac{1}{K_1}; & \text{если } 0 < K_1 < 1 \end{cases} \quad (4.6.7)$$

Полученную характеристику назовем *коэффициентом вариации молекулярной динамики*.

4.6.3 Исследование protein L при помощи коэффициента вариации молекулярной динамики

Рассматривается практический пример использования коэффициента вариации.

Исследования проводились на траектории молекулы protein L, которую растягивали с постоянной силой 1400 пН за концы на протяжении всего эксперимента.

Зависимость между концами цепи от времени приведена на рис. 16. Будем сравнивать поведение этой величины с изменением коэффициента вариации вдоль траектории молекулярной системы.

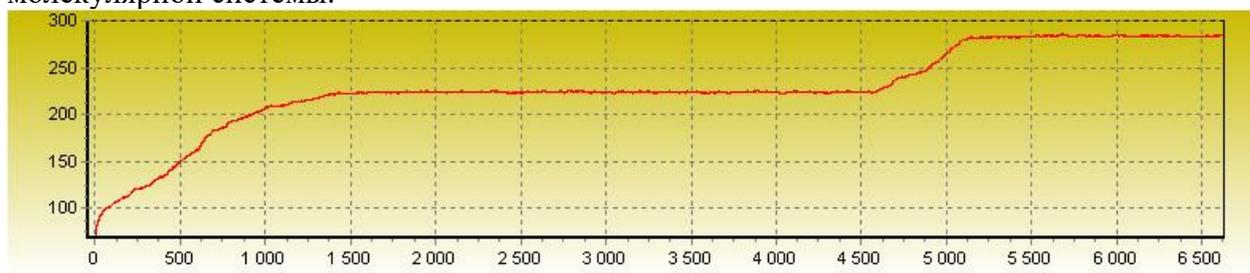


Рисунок 25. Изменение расстояния между концами макромолекулы при растяжении ее за концы постоянной силой 1400 пН.

На интервале [0; 1500] наблюдается участок интенсивного растяжения молекулы. Далее на участке [1500; 4500] расстояние между концами практически не меняется. После этого опять наблюдается интенсивное растяжение [4500; 5200], которое плавно прекращается и далее молекула ведет себя, как натянутая струна практически без растяжения.

На рисунке 17 показано изменение коэффициента вариации в процессе растяжения цепи. Наблюдается два четких локальных максимума: один в начале эксперимента и второй на промежутке [4800; 5100].

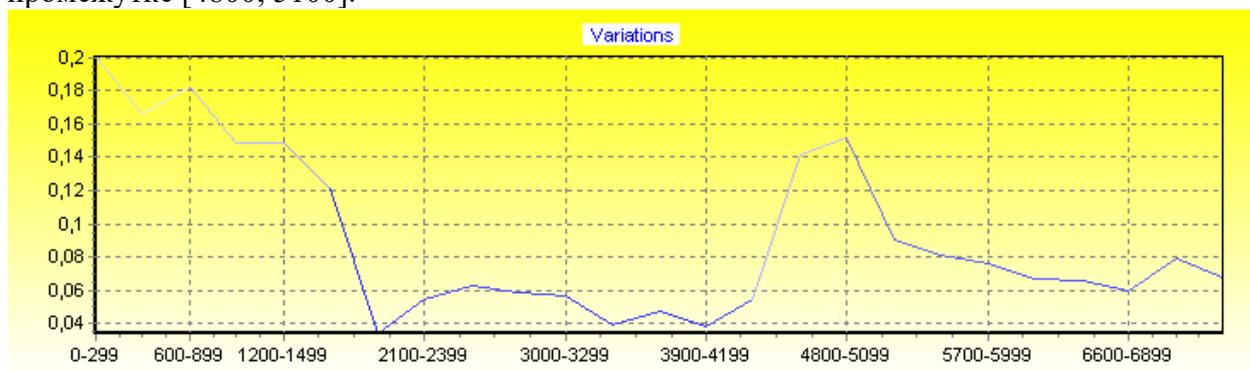


Рисунок 26. Коэффициент вариации.

Рисунок 17 демонстрирует корреляцию измеренной величины с изменением расстояния между концами и физическими процессами, которые происходят в системе. Из этого следует, что характеристика позволяет выделить места перестройки структуры макромолекулы. Дополнительную информацию о локализации изменений дают значения коэффициентов формы на заданном интервале.

4.6.4 Физический смысл коэффициента вариации. Обоснование

Итак, мы убедились, что поведение коэффициента вариации коррелирует с априорными сведениями о молекулярной системе.

Сравним наборы коэффициентов формы на разных временных интервалах.

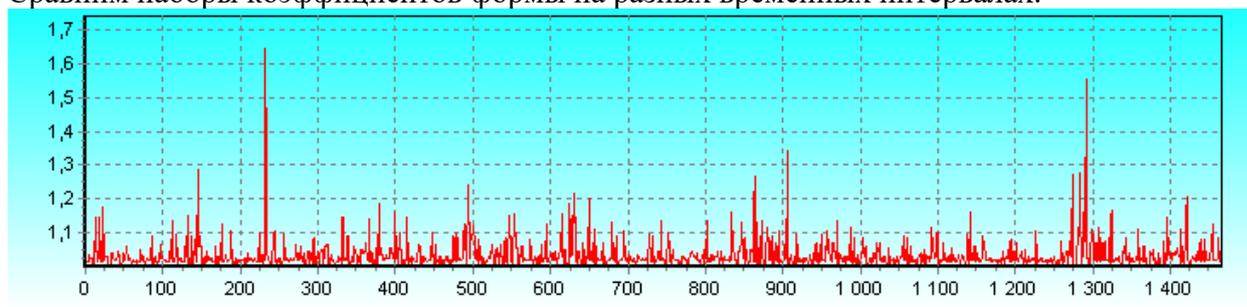


Рисунок 27. Мгновенные значения коэффициента формы, дающие малый коэффициент вариации на интервале слабых структурных изменений.

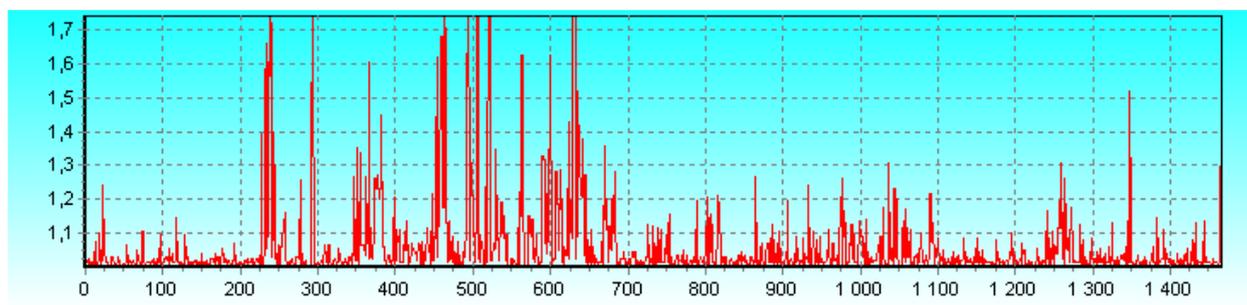


Рисунок 28. Мгновенные значения коэффициента формы, дающие большой коэффициент вариации на интервале сильных структурных изменений.

Сравнивая рисунки 18 и 19 можно заметить, что коэффициенты формы больше отличаются от единицы во время структурных изменений.

При разворачивании, равно как и при любых других видах изменения белковой структуры, некоторые торсионные углы у молекулы заметно меняют свое значение. Большинство углов представляют собой стохастический процесс. Среднее значение коэффициентов формы практически не меняется. Подвижные торсионные углы увеличивают дисперсию и коэффициент вариации.

Таким образом, коэффициент вариации дает качественное представление о процессе структурных изменений в молекулярной системе. Его составляющие – коэффициенты формы – способны ответить на вопрос локализации (внутри белковой цепи) этих изменений.

4.6.5 Реализация и возможности расчета

Для работы с коэффициентом вариации молекулярной динамики предусмотрено окно «K1», вызываемое из меню View одноимённой командой.

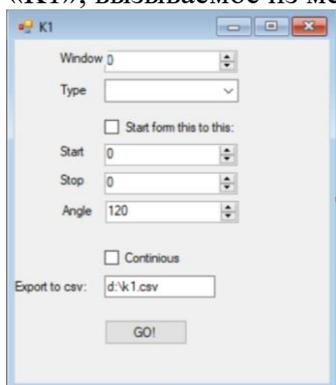


Рисунок 29. Окно работы с коэффициентом вариации молекулярной динамики
Поле Window служит для задания окна расчета. Типичные значения 50-100 шагов траектории.

В поле Type выбирается один из алгоритмов расчета коэффициента вариации (см. ниже).
Галочка Start from this to this ставится в том случае, если необходим расчет по отрезку траектории.

Поля Start и Stop задают границу отрезка траектории. Работают только в том случае, если установлен режим работы с отрезком.

Галочка Continuous включает режим непрерывного движения окна. В противном случае коэффициент вариации рассчитывается каждый window шагов траектории.

Поле Export to csv задаёт имя .csv-файла (comma separated values) для экспорта, который может быть использовать при анализе во внешних приложениях.

4.6.5.1 Алгоритмы расчета

В Анализаторе траекторий реализовано несколько алгоритмов расчета коэффициента вариации молекулярной динамики. Все они основаны на формулах (4.6.5), (4.6.6) и (4.6.7). Разницы состоит лишь в том, какая используется функция $\varphi(t)$.

На текущий момент времени реализовано шесть функций:

1. One and next distance. Функция отслеживает текущую и следующую позицию атомов.
2. All relation distance. Длины связей между выбранными парами атомами. Выбор осуществляется в окне выбора структур Selection... (F9). Каждая строчка в окне выбора соответствует одному атому.
3. Angles by C-alpha. Углы, образованные отрезками, соединяющие C-alpha-атомы.
4. H-bonds (main chain). Водородные связи, образованные главной цепью.
5. H-bonds (residues). Водородные связи, образованные аминокислотными остатками.
6. All H-bonds. Все водородные связи.

Под водородной связью понимается связь, образованная при помощи атомов водорода, кислорода и азота. Как известно, в .pdb-файлах одному и тому же химическому элементу может соответствовать несколько названий атомов. Именно по этой причине для обозначения этих атомов используется таблица PDB Анализатора траекторий, вывод которой возможен по горячей клавише F11, либо воспользовавшись меню View – Options:

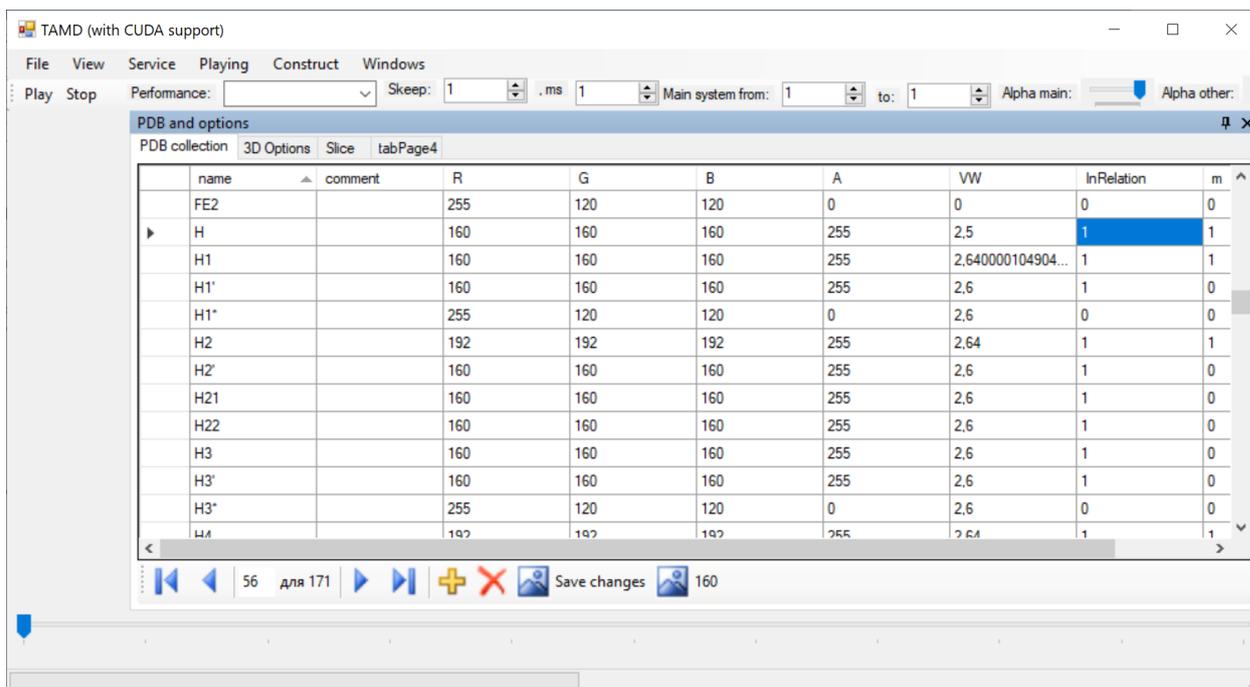


Рисунок 30. Окно вывода опций Анализатора траекторий. Вкладка PDB Collection. Атомы водорода отмечаются цифрой «1» в поле InRelation таблицы PDB Collection (таблица pdb вспомогательной базы данных). Атомы кислорода отмечаются цифрой «2». Атомы азота отмечаются цифрой «3».

4.6.5.2 Ускорение расчета за счет параллельного программирования

Заметим, что коэффициент формы для каждого атома может быть вычислен независимо от остальных. Поэтому набор коэффициентов формы можно найти параллельно. Параллельное нахождение набора коэффициентов формы доступно при выборе значения Many CPUs выпадающего списка Performance Type панели инструментов Анализатора траекторий.

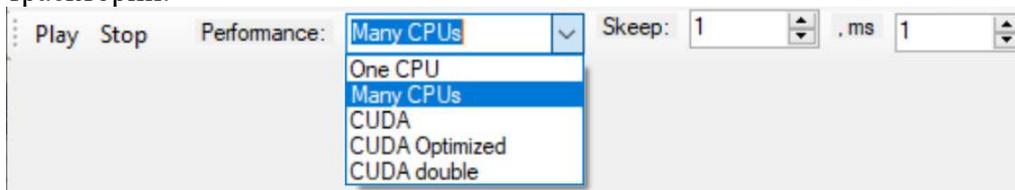


Рисунок 31. Выбор Performance Type на панели инструментов Анализатора траекторий.

TAMD, в основном, написан на языке программирования C#. В данном случае в качестве конкретной реализации технологии параллельного программирования используется средство организации параллельных циклов библиотеки .NET Parallel.For (полное имя System.Threading.Tasks.Parallel.For).

4.7 Матрица среднеквадратичных отклонений атомов

Расчет и вывод матрицы среднеквадратичных отклонений атомов доступен выбором пункта RMS из меню View.

Основной алгоритм работы заключается в вычислении отклонения координат от их средних значений. Средние значения находятся вдоль всей траектории.

Алгоритм схож с поиском отклонений C-alpha-атомов, который доступен из меню Functions. Поиск отклонений C-alpha-атомов предоставляем график отклонений атомов между двумя кадрами. Матрица отклонений, в свою очередь, строится по всей траектории, а сравнение идёт со средними положения атомов.

При этом алгоритм построения матрицы содержит ряд дополнительных возможностей, например вывод в режиме реального времени и построение по расширенным траекториям молекулярной динамики.

4.7.1 Построение матрицы среднеквадратичных отклонений атомов

Для построения матрицы среднеквадратичных отклонений атомов необходимо:

1. Загрузить траекторию, желательно с rdb-файлом.
2. Выбрать меню View, пункт RMS.
3. Установить режим из выпадающего списка.
4. Установить дополнительные опции, если требуется.
5. Выбрать номер дополнительной компоненты расширенной траектории молекулярной динамики, если требуется.
6. Нажать кнопку Go.

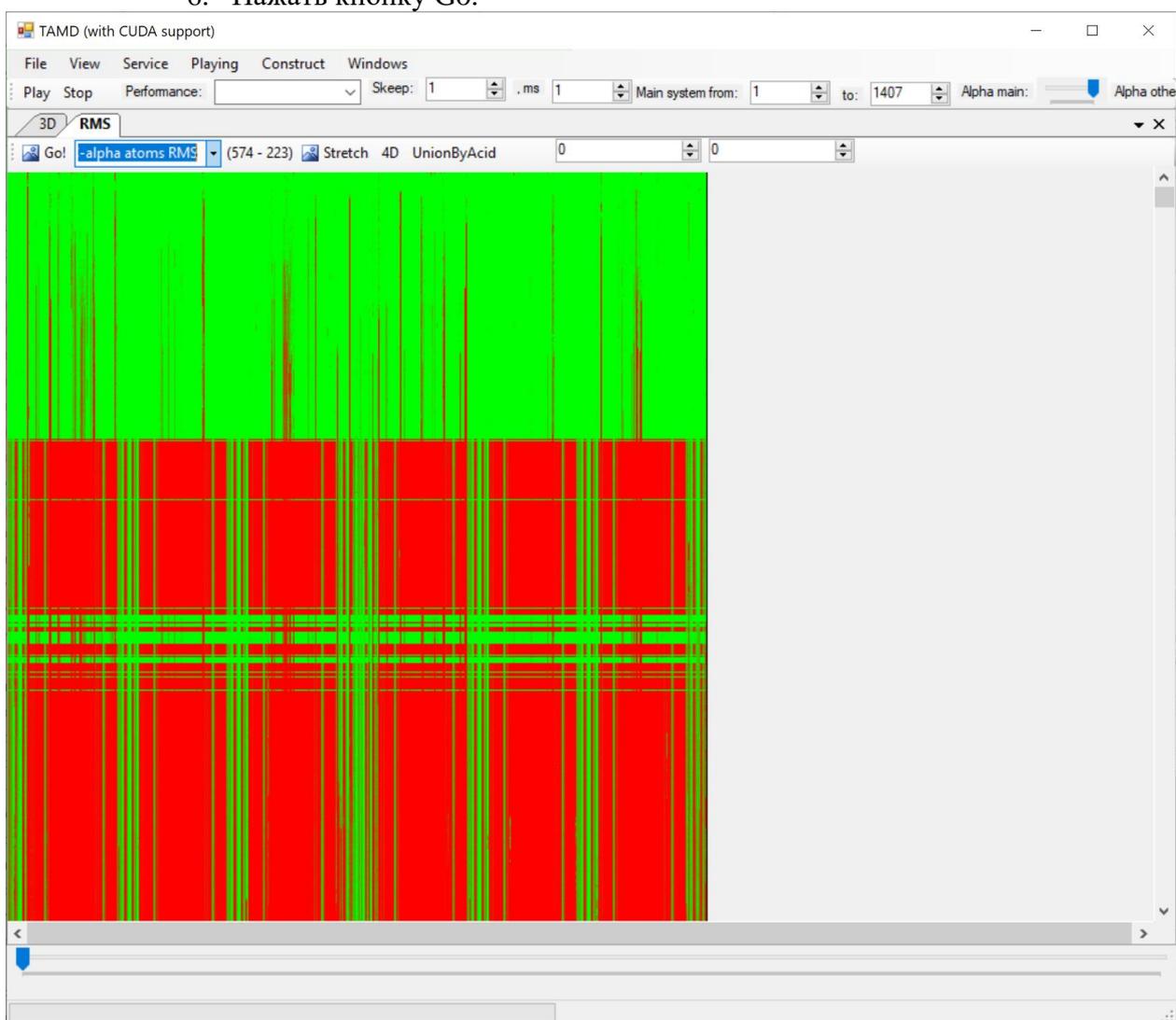


Рисунок 32. Построенная матрица среднеквадратичных отклонений C-alpha-атомов.

4.7.2 Режимы построения матрицы среднеквадратичных отклонений

Режимы построения матрицы среднеквадратичных отклонений атомов выбираются при помощи выпадающего списка на панели инструментов окна RMS.

Доступны следующие режимы:

1. All atoms RMS. Построение матрицы по всем атомам.
2. C-alpha atoms RMS. Построение матрицы только по C-alpha-атомам.

3. All atoms RMS + Structure. Построение матрицы по всем атомам, диапазон которых попадает в указанные в окне Selection интервалы.
4. C-alpha atoms RMS + Structure. Построение матрицы только по C-alpha-атомам, диапазон которых попадает в указанные в окне Selection интервалы.
5. All atoms RMS + Structure + Rij. Построение матрицы по всем атомам, диапазон которых попадает в указанные в окне Selection интервалы. Приведение к координатам тензора инерции.
6. C-alpha atoms RMS + Structure + Rij. Построение матрицы только по C-alpha-атомам, диапазон которых попадает в указанные в окне Selection интервалы. Приведение к координатам тензора инерции.

О работе в системе координат, связанной с тензором инерции, читать в главе 4.8.

4.7.2 Вывод матрицы среднеквадратичных отклонений в режиме реального времени

Анализатор траекторий выводит матрицу отклонений в режиме реального времени. Это достигается за счет прямого редактирования картинка как массива неуправляемой (unmanaged) памяти.

В C# это считается небезопасным кодом. Однако авторы программы, не считают это плохим стилем программирования. Анализатор траекторий – это, в первую очередь, вычислительная программа, которая не может быть использована злоумышленниками для доступа, к примеру, к персональным данным. Разумеется, она содержит многие системные функции, например, при работе с базами данных. Но там используется управляемый код. И он не спасает от перехвата паролей.

Использование неуправляемой памяти гораздо быстрее, чем вызов функций рисования на холсте (canvas). Прямой доступ к элементу массива гораздо быстрее вызова любой функции. Поэтому применение программирования с использованием неуправляемой памяти оправдано.

4.8 Тензор инерции и его применение в Анализаторе траекторий

Задача вычисления тензора инерции – переход к координатам, независимым от перемещения в пространстве и вращения системы как целого.

Матрица тензора инерции имеем вид:

$$J = \begin{vmatrix} J_{xx} & J_{xy} & J_{xz} \\ J_{yx} & J_{yy} & J_{yz} \\ J_{zx} & J_{zy} & J_{zz} \end{vmatrix}$$

Где каждый элемент её вычисляется по формуле:

$$J_{\alpha\beta} = \int_V (\alpha - \bar{\alpha})(\beta - \bar{\beta})dV$$

Где $\alpha, \beta \in \{x, y, z\}$

После нахождения матрицы тензора инерции вычисляются её собственные значения и собственные вектора. Собственные вектора и будут новой системой координат.

Для поиска собственных значений и собственных векторов матрицы используется библиотека `alglib`.

Однако, в математике порядок собственных векторов не определён. Для алгоритмов он тоже не важен. Но важно сохранять одинаковый порядок собственных векторов. Иначе оси будут меняться местами. Ввиду этого в программе TAMM предусмотрена сортировка собственных векторов по убыванию собственных значений.

Следующая особенность вычислений – собственные вектора определены с точностью до множителя. Первый этап решения этой проблемы – нормировать собственный вектор, чтобы его длина составляла единицу.

Теперь собственный вектор будет определён с точностью до знака. Для решения этой проблемы запоминается предыдущее значение собственных векторов. В том случае, если скалярное произведение текущего собственного вектора на предыдущее его значение будет отрицательным, но необходимо каждый его компонент умножить на -1.

4.8.1 Тензор инерции в функциях Анализатор траекторий

Приведение координат к независимой от вращения и перемещения системе координат применяется в функциях нахождения среднеквадратичных отклонений координат, а также при построении одноимённой матрицы.

4.8.2 Визуализация системы в координатах тензора инерции

В Анализатор траекторий включена возможность вывода молекулярного кино в координатах тензора инерции. Для этого достаточно установить галочку «Rotate By Inertia Tensor» в окне PDB and Options (доступной из меню View, либо по горячей клавише F12), на вкладке Slice.

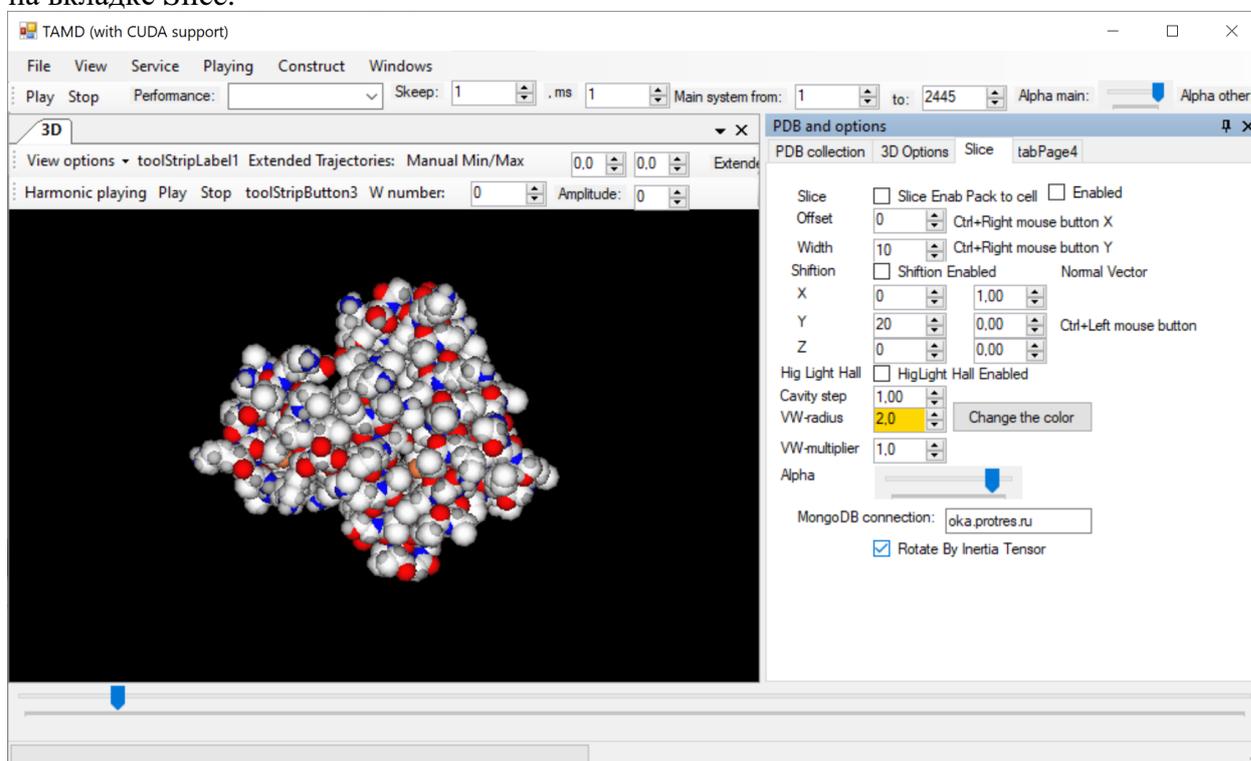


Рисунок 33. Воспроизведение молекулярного кино в координатах тензора инерции.

4.9 Работа с подсистемами

Для выделения подсистем в Анализаторе траекторий предусмотрено средство работы с подсистемами. Оно вызывается из меню View, Selection. Также оно доступно по горячей клавише F9.

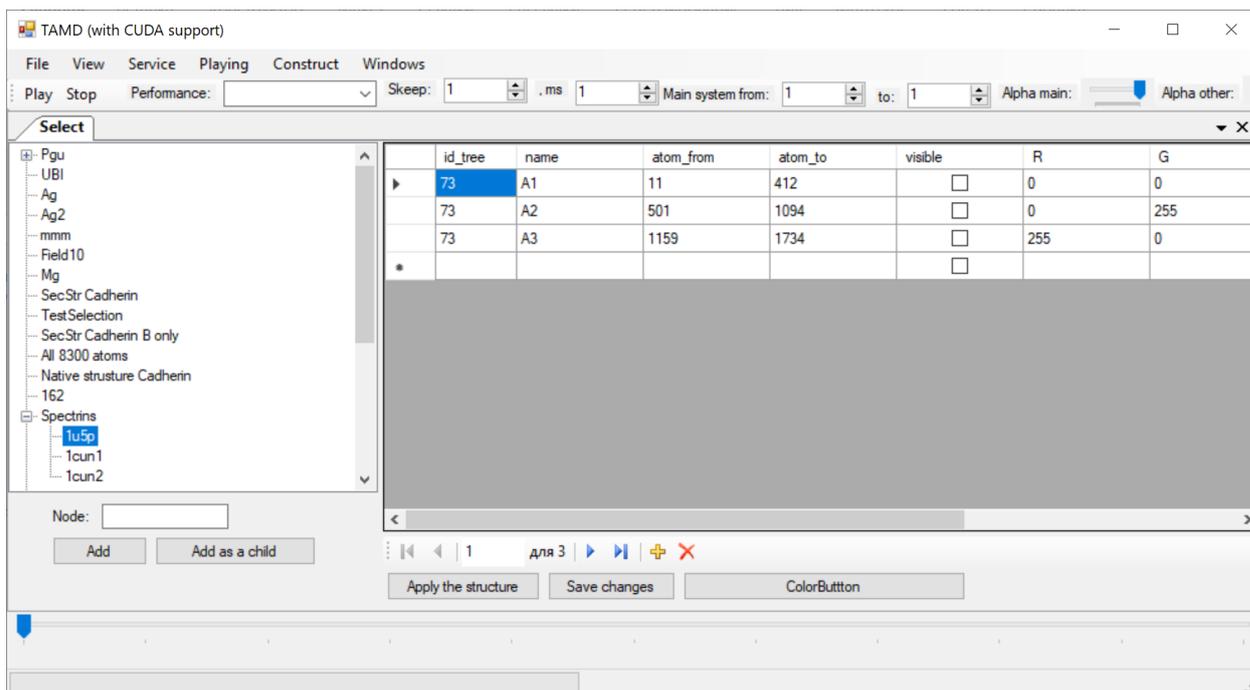


Рисунок 34. Средство работы с подсистемами.

Под подсистемой подразумевается диапазон атомов. Для каждого объекта можно хранить неограниченный набор подсистем.

Каждому объекту может соответствовать свой набор подсистем. Обычно некоторой группе молекулярно-динамических экспериментов соответствует один набор подсистем. Для хранения набора подсистем в Анализаторе траекторий каждому набору присваивается имя.

Подсистема может быть видима и невидима. Подразумевается, что сначала осуществляется разметка системы на избыточное количество подсистем. Затем высвечиваются необходимые подсистемы.

Подсистеме присваивается свой цвет в формате RGB (красный, зелёный, синий).

Подсистема может изображаться как обычными цветами атомов, так и одним выбранным цветом.

4.9.1 Заполнение списка подсистем

Сначала для подсистемы необходимо выбрать имя. Имя указывается в поле Node. Затем нажимается кнопка Add, чтобы добавить имя в корень имён подсистем. Либо нажимается кнопка Add child, чтобы добавить имя в подчинении другому имени (подобно подкаталогам в файловой системе). При этом имя родителя должно быть выбрано.

Для работы со списком диапазонов атомов в подсистеме предусмотрен набор кнопок навигации:

- По кнопке «+» добавляется новая строка. При этом поле id_tree заполняется автоматически.
- Кнопка «X» удаляет строку.
- По кнопке «|<» осуществляется переход на первую строчку.
- По кнопке «<<» осуществляется переход на предыдущую строчку.
- По кнопке «>>» осуществляется переход на следующую строчку.
- По кнопке «>|» осуществляется переход на последнюю строчку.
- По кнопке «ColorButton» осуществляется выбор цвета.
- По кнопке Save changes осуществляется сохранение изменений в служебной базе данных.

Поля имеют следующие значения:

- 1) `id_tree` – идентификатор имени. Служебное поле. Нет необходимости его использовать, т.к. оно заполняется автоматически при добавлении новой строки. Однако было принято решение оставить это поле на интерфейсе для легкого ручного редактирования списка подсистем напрямую в базе данных.
- 2) `name` – имя подсистемы. Поле служит для удобства пользователя и не влияет на отображение.
- 3) `atom_from` – начальный номер атома подсистемы. Номера атомов начинаются с 1, подобно порядку атомов в `pdb`-файле.
- 4) `atom_to` – конечный номер атома подсистемы, включительно. Номера атомов начинаются с 1, подобно порядку атомов в `pdb`-файле.
- 5) `visible` – флажок видимости подсистемы.
- 6) `R` – значение красного канала цвета подсистемы. Имеет значение от 0 до 255.
- 7) `G` – значение зелёного канала цвета подсистемы. Имеет значение от 0 до 255.
- 8) `B` – значение синего канала цвета подсистемы. Имеет значение от 0 до 255.
- 9) `A` – значение альфа-канала цвета подсистемы (прозрачности). Имеет значение от 0 до 255.
- 10) `type` – тип вывода раскраски атомов:
 - a) 1 – означает вывод атомов в соответствии со стандартной раскраской.
 - b) 2 – означает вывод атомов одним цветом в соответствии с полями `RGB`.

Цвет подсистемы можно указывать как вручную заполняя поля `R`, `G`, `B`, так и с использованием кнопки `ColorButton`.

После заполнения списка диапазонов атомов необходимо нажимать кнопку `Save changes` для сохранения данных в служебной базе данных.

4.9.2 Использование подсистем

Для использования заполненного списка подсистем в Анализаторе траекторий необходимо сначала выбрать название этой подсистемы в окне `Select`, а затем нажать кнопку `Apply the structure`.

Подсистемы используются для цветовой окраски в окне визуализации, а также при расчете характеристик. Подсистемы поддерживаются при работе с контактными картами и при расчете функций вдоль траектории. Списков режимов (функций), как правило включает функции работы со всей системой. Название той же функции с ключевым словом `select` означает использование функции с учетом подсистем.

4.9.3 Хранение подсистем в СУБД

Сведения о подсистемах сохраняются в служебной СУБД, тем самым они доступны во время всего периода эксплуатации программы.

Хранение подсистем (групп атомов) в СУБД занимает две таблицы:

1. Таблица `tree` хранит сведения о названиях подсистем в виде иерархического списка.
2. Таблица `groups` хранит сведения о непосредственно самих группах атомов.

Таблицы связаны соотношением один-ко-многим. Поле `id_tree` таблицы `groups` ссылается на поле `id` таблицы `tree`, как это показано на следующем рисунке:

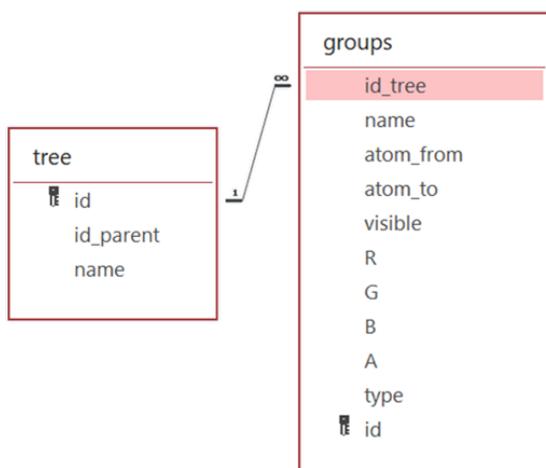


Рисунок 35. Схема хранения информации о подсистемах

4.9.3.1 Структура таблиц хранения информации о подсистемах СУБД

Таблица tree имеет следующую структуру:

№	Имя поля	Тип поля
1	id	целое
2	id_parent	целое
3	name	строка

Таблица 4.9.1 Структура таблицы tree.
 Назначение
 Первичный ключ
 Ссылка на родительскую группу
 Наименование набора групп атомов

Таблица groups имеет следующую структуру:

№	Имя поля	Тип поля
1	id	целое
2	id_tree	целое
3	name	строка
	atom_from	целое
	atom_to	целое
	visible	булево
	R	целое
	G	целое

Таблица 4.9.1 Структура таблицы groups.
 Назначение
 Первичный ключ
 Идентификатор имени. Внешний ключ. Ссылка на поле id таблицы tree
 Имя подсистемы. Поле служит для удобства пользователя и не влияет на отображение
 начальный номер атома подсистемы. Номера атомов начинаются с 1, подобно порядку атомов в pdb-файле
 конечный номер атома подсистемы, включительно.
 Номера атомов начинаются с 1, подобно порядку атомов в pdb-файле
 флажок видимости подсистемы
 значение красного канала цвета подсистемы. Имеет значение от 0 до 255
 значение зелёного канала цвета подсистемы. Имеет значение от 0 до 255.

B	целое	значение синего канала цвета подсистемы. Имеет значение от 0 до 255
A	целое	значение альфа-канала цвета подсистемы (прозрачности). Имеет значение от 0 до 255.
type	целое	тип вывода раскраски атомов: 1 – означает вывод атомов в соответствии со стандартной раскраской. 2 – означает вывод атомов одним цветом в соответствии с полями RGB

Крайне рекомендуется давать осмысленные имена в поле name таблицы tree – это имена набора групп атомов, которые будут использоваться всякий раз при вызове сохранённого выделения. Напротив, имя конкретной группе атомов можно не давать, оставляя поле name таблицы groups пустым.

4.9.3.2 Реализация. Хранение древовидной информации в СУБД

Для хранения информации о названиях наборов групп атомов в древовидной форме можно было бы использовать несколько связанных таблиц. На каждый уровень иерархии по таблице.

Вместо этого представляется универсальное решение хранения иерархической информации неограниченного уровня вложенности в одной таблице.

Каждая строка таблицы tree имеет ссылку на родителя в качестве поля id_parent, который содержит значение поля id родительского элемента. Родительский элемент, в свою очередь, также ссылается на своего родителя и так далее до корневого уровня. Корневые элементы не имеют родителя, поле id_parent у них содержит значение 0.

4.9.3.3 Реализация вывода иерархического списка в древовидном виде

Вывод иерархического списка начинается с поиска корневых элементов. Их легко найти, используя запрос:

```
SELECT * FROM tree WHERE id_parent=0
```

При потребности вывода дочерних элементов, принадлежащих определённому родителю – при нажатии символа «+», раскрывающего дочерние элементы – для поиска подчинённых элементов выполняется аналогичный запрос. Только вместо 0 используется значение поля id конкретного родительского элемента.

Таким образом, из таблицы tree служебной СУБД загружается только минимум информации, необходимый для отображения данных на экране в древовидной форме.

Достигается экономия памяти.

Из таблицы groups загружается также минимум информации – только строки, соответствующие данному набору данных.

4.10 Опции программы

Опции программы доступные выбором пункта Options... меню View. Также можно использовать горячую клавишу F11.

Опции следует воспринимать именно как оперативные опции отображения, а не как настройки, которые нечасто редактируются при эксплуатации программы.

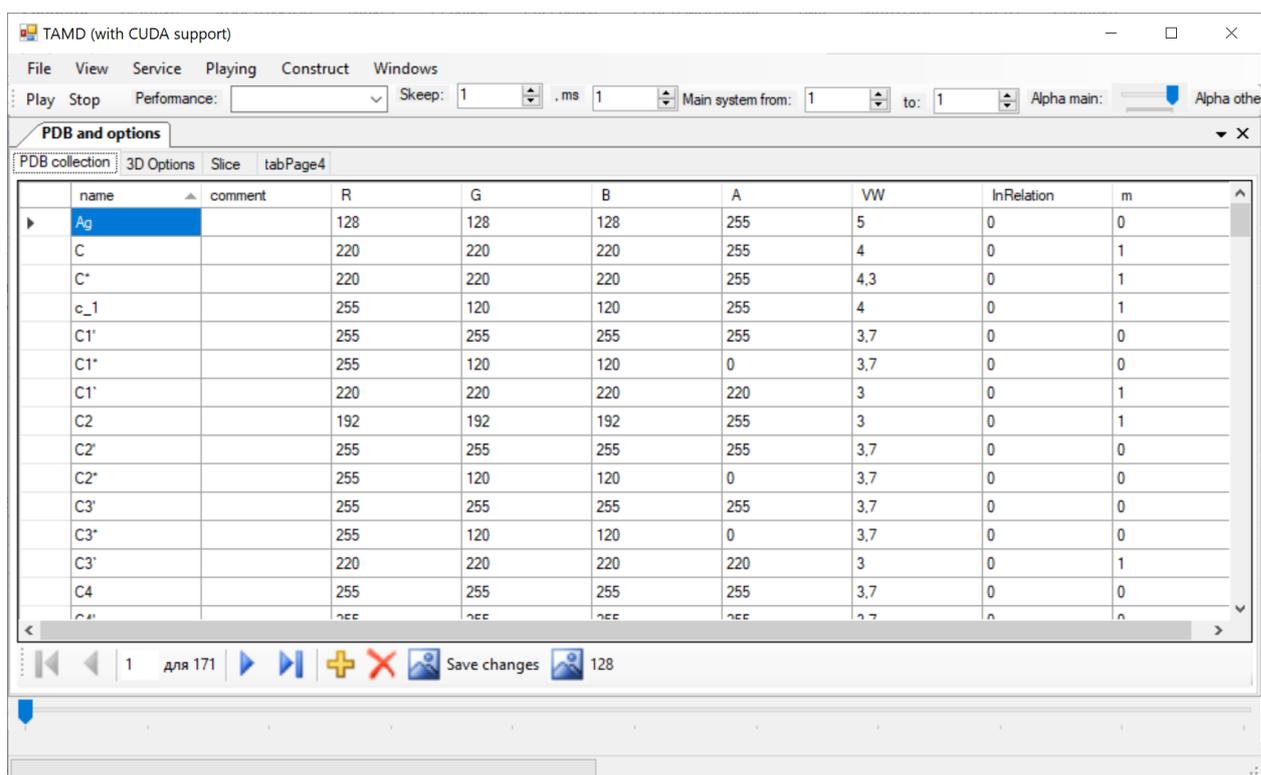


Рисунок 36. Настройки программы TAMD. Вкладка коллекции типов атомов (PDB collection)

4.10.1 Коллекция типов атомов

Коллекция типов атомов доступна на вкладке PDB collection.

Коллекция атомов содержит информацию о названиях атомов, применяемых в Protein Data Bank и в биоинформатике. Одному и тому же химическому элементу соответствует несколько названий атомов из PDB. Название зависит от положения атома в аминокислоте. В молекулярной динамике именно названию атома соответствуют определённые параметры силового поля.

При открытии PDB-файла Анализатор производит сопоставление по имени атома из этой таблицы, определяя цвет отображения и другие параметры.

Переход по строчкам осуществляется как мышью, так и при помощи панели навигации внизу окна. После изменения значений необходимо нажимать кнопку Save changes.

Список полей и их значений в программе TAMD:

- name – имя атома из файла PDB. Именно по этому полю ведётся сопоставление при открытии pdb-файла. Регистр символов при поиске не учитывается. Но он учитывается при записи pdb-файла.
- comment – комментарий. Поле введено для удобства пользователя. Не несёт смысловой нагрузки.
- R, G, B, A – цвет атома в формате красный, зелёный, синий, прозрачность (alpha-канал). Каждому каналу соответствует число от 0 до 255. Все нули соответствуют черному цвету, а максимальные значения каналов – белому.
- VW – радиус Ван-дер-Ваальса. Это радиус отображения атомов. Задаётся в ангстремах [Å].
- InRelation – флаг участия атомов в связи. Используется для опции вывода связей C-alpha-атомов в окне трехмерного вывода. Принимает значения 0 (не участвует) или 1 (участвует в отображении связей). По умолчанию принимает значение 0.
- m – масса атома в а.е.м. Используется для квазигармонического анализа.

При выводе связей C-alpha-атомов в окне трехмерного вывода необходимо проставить значение 1 у всех C-alpha-атомов. Получится грубая модель пептидной цепи. Помимо этого, можно изобразить любую цепь, выделяя таким образом необходимые атомы. Например, обозначить спиралевидные структуры, нанотрубки и пр.

4.10.2 Опции отображения трехмерной структуры

Опции отображения трехмерной структуры доступны на вкладке 3D Options. На ней собраны основные настройки отображения, а также параметры записи этого изображения в видео-файл (.avi-файл).

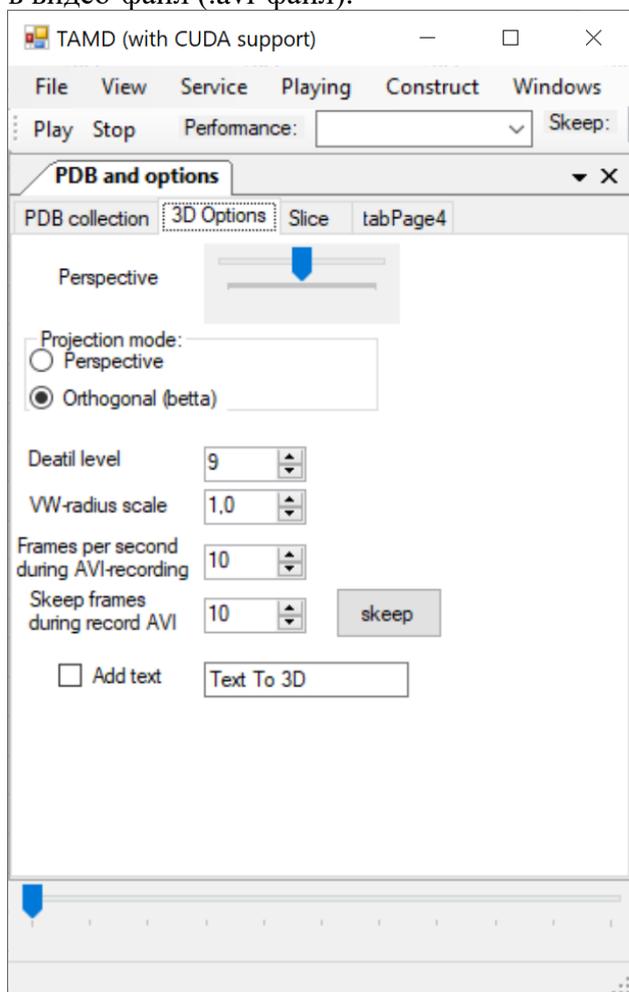


Рисунок 37. Вкладка выбора опций трехмерной визуализации.

В окне доступны следующие опции:

- Perspective – угол перспективной проекции.
- Projection mode: – режим проекции:
 - Perspective – перспективная проекция. Эта проекция соответствует человеческому глазу. Камера (наблюдатель) находится перед объектом на некотором расстоянии. Чем объект ближе, тем он больше. Таким образом, параллельные линии, которые перпендикулярны наблюдателю, выглядят под некоторым углом. Если бы они были бы бесконечны, то они бы сходились в одну точку на горизонте (в точке, удалённой на бесконечность от наблюдателя).
 - Orthogonal – ортогональная проекция. Эта проекция используется в черчении, где необходимо четко соблюдать размеры. Все параллельные линии выглядят параллельно. Может быть полезна при отображении структур, выполненных в форме прямоугольного параллелепипеда. Помогает проверять параллельность структур.

- Detail level – уровень детализации. Чем это число выше, тем плавнее смотрятся сферы и цилиндры. И наоборот: низкое значение уровня детализации помогает быстрее отображать сложные объекты, быстрее проводить анимацию. Конкретное значения этого параметра соответствует количеству граней многогранника, который изображает сферу (атом).
- VW-radius scale – коэффициент масштабирования радиусов атомов. Если задать 2, то из этому получатся в 2 раза больше. Если 0,5, то в 2 раза меньше. Исходный радиус атома соответствует значению VW вкладки PDB collection.

Опции записи в видеофайл (avi-файл)

- Frames per second – количество кадров в секунду. Для объяснения этого параметра лучше представить, что воспроизведение молекулярного кино зависит только лишь от мощности вычислительной установки. При записи avi-файла Анализатор генерирует кадры с одной частотой, а этот параметр будет определять частоту, с которой кадры молекулярного кино будут воспроизводиться при проигрывании готового avi-файла.
- Skip frames during record AVI – количество кадров, которые пропускаются (записываются) при нажатии кнопки Skip.
- Кнопка Skip. При нажатии этой кнопки в avi-файл записывается число одинаковых кадров, указанное в предыдущем параметре. В момент воспроизведения картинка зависает на определенное время, которое легко вычислить, разделив количество пропущенных кадров на частоту записи кадров. Это может быть полезно, чтобы акцентировать внимание на определенное место вычислительного эксперимента. Например, чтобы зрители успели рассмотреть структуру или прочитать надпись.
- Галочка «Add text» – включает режим добавления текста в видеофайл из текстового поля рядом с этой галочкой. На трехмерной визуализации при этом ничего не происходит. Текст накладывается непосредственно в сам видеофайл.

4.10.3 Опции отображения разреза

Опции работы с разрезом структуры доступны на вкладке Slice.

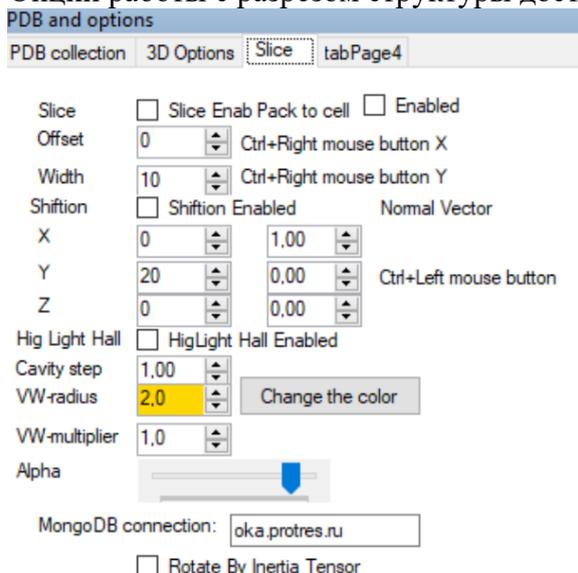


Рисунок 38. Вкладка выбора опций отображения разреза и прочих опций вывода. Предполагается, что существует две параллельных плоскости. Отображается все атомы, которые расположены между двумя плоскостями. Режим включается галочкой Slice.

Опцию рекомендуется использовать совместно с выводом в режиме упаковки в расчётную ячейку периодических граничных условий (галочка «Pack to cell»).

Отсчет ведётся от угла ячейки периодических граничных условий с координатами (0, 0, 0). Первая плоскость задаётся нормальным вектором, который по умолчанию равен (1, 0, 0) и изначально проходит через точку (0, 0, 0). По умолчанию первая плоскость совпадает с осью OX, иными словами является плоскостью YOZ.

За сдвиг первой плоскости относительно начала расчётной ячейки отвечает параметр Offset. К примеру, если вектор нормали равен (0, 1, 0), а Offset=5, то первая плоскость будет проходить через точку с координатами (0, 5, 0).

Вторая плоскость отстоит от первой на расстоянии, задаваемым параметром Width. Width – это ширина отображаемого промежутка атомов.

Отображению подлежат все атомы, находящиеся между этими двумя плоскостями.

Алгоритмически это записывается в следующем виде:

- Нормализуется вектор нормали (делим все его координаты на длину вектора).
- Нормализованный вектор нормали умножаем на параметр Offset.
- Находится скалярное произведение между координатами атомами и полученного вектора.
- Если оно меньше нуля, то атом не отображается (атом слева от плоскости). Иначе – проверяем взаимное расположение со второй плоскостью.
- Вторая плоскость задаётся почти также, учитывая две особенности. Она отстоит на расстоянии Width от первой по направлению вектора нормали. При проверке скалярного произведения атом подлежит отображению в том случае, если скалярное произведение меньше нуля.

4.10.4 Опции отображения полостей

Опции отображения полостей также доступны на вкладке Slice окна PDB and options.

Рассматривается область пространства, ограниченная прямоугольным параллелепипедом. Координаты этого прямоугольного параллелепипеда совпадают с размерами расчётной ячейки. Пространство заполняется виртуальными частицами с шагом Cavity Step [Å] таким образом, чтобы эти виртуальные атомы не пересекались с реальными атомами системы.

Размер виртуальных частиц определяется параметром VW-radius.

Цвет виртуальных частиц задается при помощи кнопки Change the color, рядом с полем VW-radius.

Прозрачность виртуальных частиц задаётся при помощи ползунка Alpha.

Режим отображения полостей включается при выборе пункта Cavity меню View options окна 3D.

Рекомендуется убирать отображение атомов при отображении полостей. Это делается путём снятия галочки Atoms (и других галочек, если они были установлены) меню View options окна 3D.

4.10.5 Прочие опции отображения и опции программы

Прочие опции также доступны на вкладке Slice окна PDB and options.

Значение VW-multiplier задаёт множитель радиуса вывода атомов. Изначально атомы отображаются с радиусом, заданным в поле VW (согласно их типу) таблицы PDB Collection, доступной из окна PDB and options. Значение радиуса умножается на множитель. Таким образом, чтобы получить вдвое больший радиус, необходимо задать значение 2 в поле VW-radius.

Поле MongoDB connection задаёт имя сервера, которое используется при загрузке данных их СУБД MongoDB. Устаревшая, но работающая возможность. Вместо этого удобнее работать с несколькими СУБД в окне Trajectory Commander.

4.11 Анализ расширенных компонент

Анализатор траекторий поддерживает расширенные траектории молекулярной динамики. В расширенных траекториях каждому атому, помимо трех его координат, добавляется одно и несколько чисел, соответствующих какой-либо характеристике, рассчитываемой в программе моделирования молекулярной динамики (такой, как PUMA или PUMA-CUDA). Выбор расширенной компоненты траектории осуществляется в окне 3D изменением значения Extended number. Этот выбор влияет на весь Анализатор траекторий. TAMM использует буфер случайного доступа к траектории. От выбора номера расширенной компоненты зависит, какая расширенная компонента траектории будет сохранена в буфер, расположенный в оперативной памяти для быстрого доступа к последним прочитанным кадрам.

4.11.1 Визуализация расширенных компонент траектории на молекулярном кино

Включение визуализации расширенных компонент осуществляется в меню View options окна 3D.

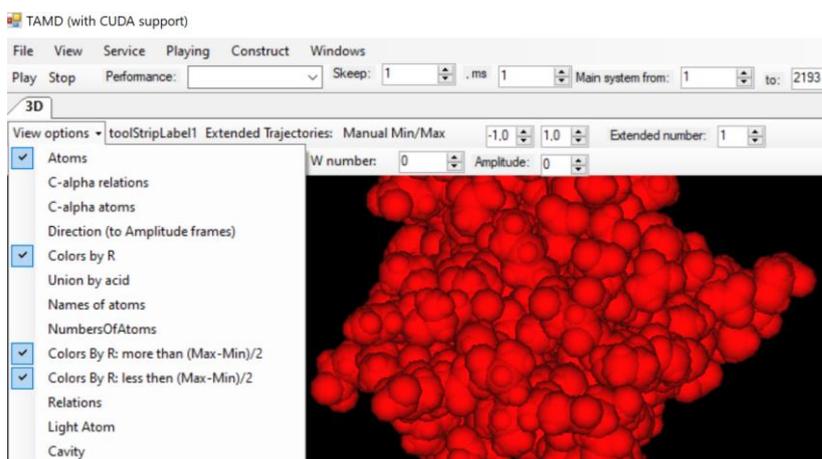


Рисунок 39. Меню View options окна 3D в включенном отображении расширенных компонент.

Для включения режима достаточно выбрать пункт «Colors By R». Под R подразумевается расширенная координата. Её номер соответствует полю Extended number. При этом возможно автоматическое определение границ цветов значения расширенной компоненты (по умолчанию), либо ручное их задание. Ручное задание цвета включается нажатием кнопки «Manual Min/Max». В двух полях рядом с этой кнопкой задаётся значения, которые принимаются за минимальное и максимальное. Если расширенная компонента меньше минимального значения, то она изображается зеленым цветом, а если больше максимального, то красным. Если её значение располагается между минимальным и максимальным, то происходит постепенная смена цветов.

Алгоритм расчёта цвета.

Если расширенная меньше половин

$$Color(R, G, B) = \begin{cases} (1,0,0), \text{ если } r \geq RMax \\ \left(\begin{array}{c} 1, \\ 1 - 2 \frac{r - (RMax + RMin)/2}{RMax - RMin}, \\ 1 - 2 \frac{r - (RMax + RMin)/2}{RMax - RMin} \end{array} \right), \text{ если } \frac{RMax + RMin}{2} \leq r < RMax \\ \left(\begin{array}{c} 2 \frac{r - RMin}{RMax - RMin}, \\ r - RMin \\ 2 \frac{r - RMin}{RMax - RMin} \\ 1 \end{array} \right), \text{ если } RMin < r \leq \frac{RMax + RMin}{2} \\ (0,1,0), \text{ если } r \leq RMin \end{cases}$$

Здесь:

Color(R,G,B) – трехкомпонентное изображение цвета. Каждый канал принимает значения от 0 до 1. (0, 0, 0) соответствует черному цвету, (1, 1, 1) – белому;

r – текущее значение расширенной компоненты;

RMin и RMax – нижняя и верхняя границы градиентного цвета изображения расширенных компонент

Смысл формулы следующий. Если расширенная компонента больше верхней границы $r \geq RMax$, то атом изображается красным. Если расширенная компонента меньше нижней границы $r \leq RMin$, то атом изображается зелёным (0, 1, 0). Если между ними $RMin < r < RMax$, то необходимо вычислить пропорцию цветов. Этот интервал разделяется на два: сначала убывает зелёный канал, затем увеличивается красный канал.

4.11.2 Анализ расширенных компонент вдоль всей траектории

В Анализаторе имеется возможность проводить анализ расширенных компонент вдоль всей траектории путем вывода информации на графики, а также в текстовый файл в формате .csv.

Анализатор рассчитывает статистику значений расширенных компонент вдоль траектории, группируя их по аминокислотным остаткам. Для каждой группы находится математическое ожидание (среднее значение), дисперсия и процент кадров траектории, в которых значение расширенной компоненты было положительным или отрицательным. Для включения режима анализа расширенных компонент необходимо выбрать пункт «Extended Component Analyzer» меню «View».

Перед началом анализа расширенных компонент необходимо установить номер интересующей расширенной компоненты траектории в окне «3D», поле «Extended number».

На первой вкладке задаётся режим расчета:

- Поля From и To определяют границу траектории в кадрах, по которым идёт сбор статистики.
- Поле Size содержит размер анализируемой системы в атомах. Как правило, используется для исключения атомов растворителя.
- Поле Filename содержит имя файла вывода информации. Вне зависимости от расширения файла, будут записаны данные в формате CSV (comma separated values).
- Кнопка Go включает режим расчета.

Процент готовности счета будет изображаться на ProgressBar.

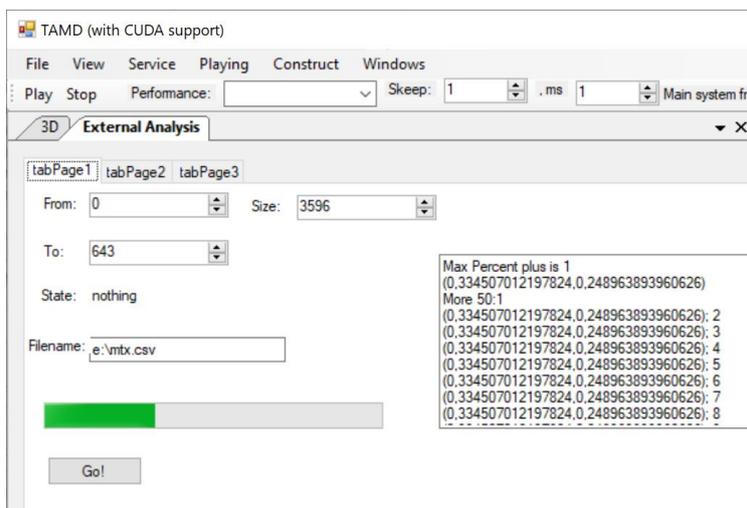


Рисунок 40. Окно анализа расширенных компонент. Выбор режима. По окончании расчета будет выведена информация в текстовый файл, на график в окно Grapher, а также в таблицу на второй вкладке.

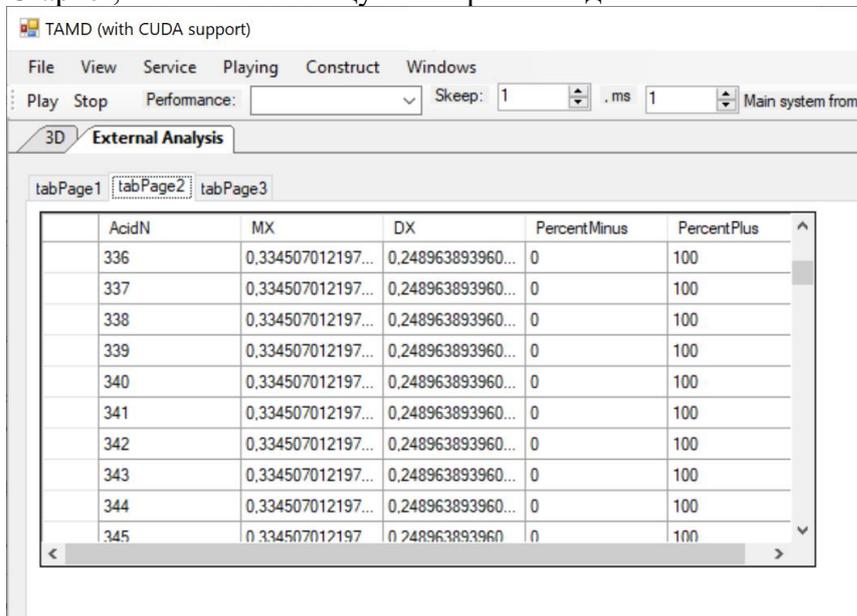


Рисунок 41. Окно анализа расширенных компонент. Вывод статистики. Значения полей в таблицы вывода статистической информации следующие:

- AcidN – номер аминокислотного остатка, по порядку.
- MX – математической ожидание (среднее значение).
- DX – дисперсия.
- PercentMinus – процент кадров траектории, в которых сумма расширенных компонент у атомов данной траектории была отрицательной.
- PercentPlus – процент кадров траектории, в которых сумма расширенных компонент у атомов данной траектории была положительной.

4.12 Восстановление траекторий

В ходе проведения МД-эксперимента на суперкомпьютерах в результате многократной остановки и продолжения расчетов, а также передачи файлов траекторий по каналам связи возможны повреждения файлов траекторий молекулярной динамики.

4.12.1 Объяснение работы с повреждённой траекторией

При работе с поврежденной траекторией средствами Анализатора возникают сообщения об ошибках при попытке чтения кадров, записанных не полностью, а также расположенных в ошибочных позициях.

В Анализаторе используется буфер случайного доступа к траектории. При открытии файла траектории вначале считывается только первый кадр. Остальные кадры считываются только по мере необходимости.

Повреждение одного кадра, как правило, означает, что он записан не полностью. Буфер случайного доступа к траектории для доступа к произвольному кадру, ровно как и при последовательном считывании траектории, вычисляет начало кадра по его длине (размеру кадра на диске) и номеру. У обычной траектории все кадры имеют один и тот же размер на диске. Повреждение одного кадра ведёт к тому, что начальные позиции всех последующих кадров находятся не на своих предполагаемых местах. Разумеется, это приводит к ошибке чтения траектории.

4.12.2 Признаки поврежденной траекторий

Вычисление большинства характеристик основано на последовательном чтении траектории. Ошибка любого алгоритма работы до его завершения, как правило, свидетельствует о поврежденном файле траектории.

Чтобы быть полностью уверенным, можно запустить воспроизведение траектории. При попытке считывания ошибочного кадра появится сообщение об ошибке.

Для быстрой проверки траектории можно вручную подвигать ползунок позиционирования траектории с открытым окном вывода трехмерной информации. И даже быстро передвинуть его на конец траектории. Как правило, поврежденный кадр в середине траектории приводит к нечитаемости всех последующих кадров. Это один из скорейших способов проверки траектории.

4.12.3 Алгоритм восстановления траекторий

Потеря одного или нескольких кадров, как правило, несущественна.

Средство восстановления траекторий основано на попытке чтения всей траектории построчно. В кадре должно быть минимум ровно то количество атомов, которое указывается до начала работы алгоритма. Если атомов больше, то лишние из них выбрасываются. Скорее всего, это будет следующий кадр траектории, оказавшийся без заголовка. Такой кадр бесполезен. Если атомов в кадре меньше указанного количества, то такой кадр также выбрасывается. Он также бесполезен.

Помимо подсчёта количества атомов в кадре, делается попытка прочитать три числа – три координаты каждого атома. Если она заканчивается неудачей хотя бы для одного атома, то весь кадр целиком считается испорченным.

Благодаря заданию количества атомов, средство восстановления можно использовать для обрезания кадров траектории. Таким образом можно удавить растворитель (воду) из каждого кадра, задав количество атомов белка. Это уменьшит размер траектории.

Также можно проредить траекторию, оставляя каждый i -ый кадр. Это также уменьшит размер траектории в i раз.

4.12.4 Порядок восстановления файлов траекторий

Средства для восстановления файлов траекторий молекулярной динамики находится в окне утилит, доступном из меню «Service», пункт «Utils», вкладка «Repair», как показано на следующем рисунке:

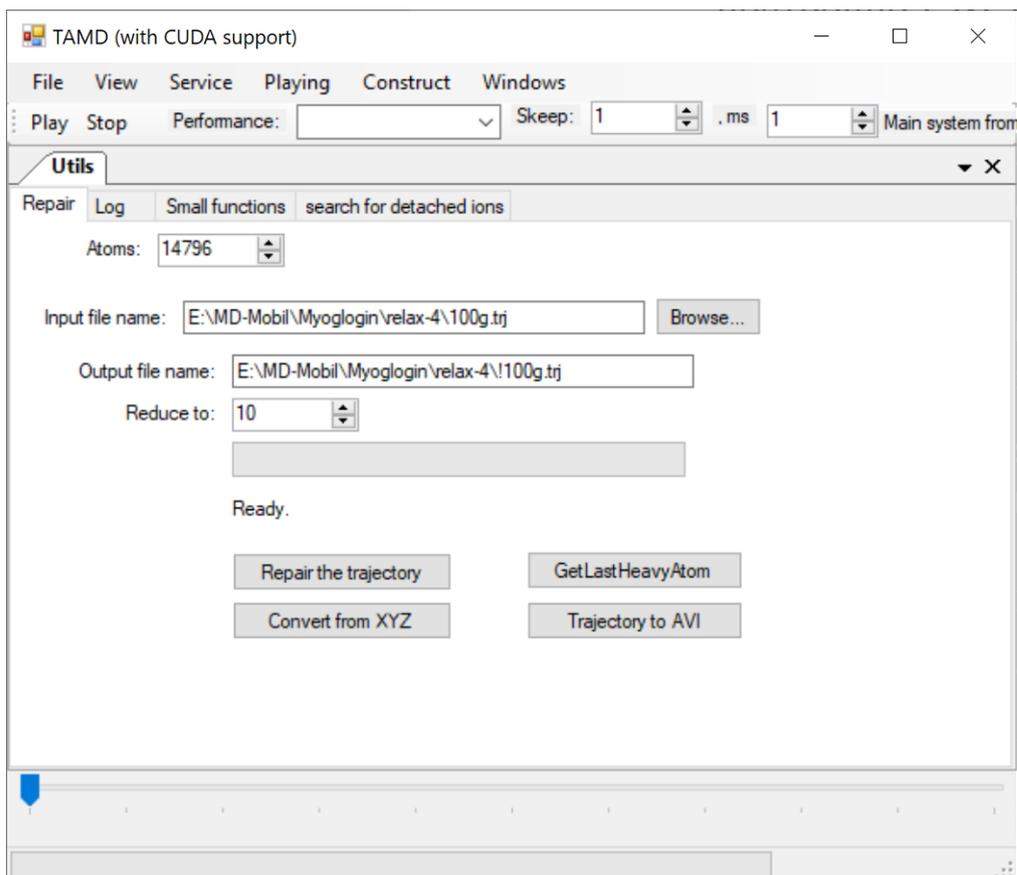


Рисунок 42. Окно восстановления файлов траекторий.

Порядок восстановления файлов траекторий молекулярной динамики следующий:

1. Заносится полное имя файла (с путём) в поле «Input file name». Также можно воспользоваться диалогом выбора файла, доступного по кнопке «Browse».
2. Имя выходного файла заносится в поле «Output file name». Если используется кнопка «Browse», то имя выходного файла устанавливается таким же, как у входного. Разумеется, необходимо его изменить. Хорошей практикой считается добавление в имени выходного файла одного и того же префикса. В данном случае используется восклицательный знак. Подобная практика позволит в будущем легко отделять исходные файлы от восстановленных. Исходные файлы, как правило, удаляются послед процесса восстановления.
3. После выбора файла, используя кнопку «Browse», программа автоматически определяет количество атомов по первому кадру траектории и заносит его в поле Atoms. Пользователь может либо согласиться с этим значением, оставив его без изменений, либо поставить своё значение.
4. Также количество атомов можно уменьшить до количества атомов белка по команде «Get last heavy atom». Она найдёт первый атом воды, но делает это по .pdb-файлу, если траектория открыта.
5. Для прореживания кадров траектории служит поле «Reduce». При этом будет записываться каждый i -ый кадр, уменьшая размер файла траектории в i раз. Например, значение 1 означает, что необходимы все кадры траектории, а значение 10 – что необходимо оставить только каждый 10-ый кадр. По умолчанию в поле присутствует значение 1.
6. Процесс восстановления запускается по команде «Repair the trajectory». При этом на экране будет отображаться сообщение «Working» на синем фоне и будет изменяться положение индикатора статуса процесса.

7. По окончании процесса восстановления отобразится надпись «Ready». При правильном задании параметров работы и после успешного восстановления файлов траекторий исходные файлы можно удалить.

4.13 Статистическая обработка энергетических траекторий и прочих файлов с данными

Под энергетическими траекториями понимаются зависимости макроскопических и других величин от времени, записываемых в течении МД-эксперимента.

Анализатор Траекторий содержит функции массовой статистической обработки энергетических траекторий. Под массовой обработкой подразумевается сбор статистики по всем заданным файлам.

Помимо энергетических траектории, блок Статистической обработки может использоваться для статистического анализа любых других файлов с данными схожего формата.

Эти данные записываются в так называемые «файлы с данными», которые представляют собой текстовые файлы, состоящие из фиксированного набора колонок. Каждая колонка имеет свой заголовок в первой строчке файла. Значения располагаются строго под заголовком построчно. Все значения отделяются друг от друга пробелами.

	Time	Eqq	Eww	EV12	EV13	EV14
1	0.00000	-99134.11350	13477.73061	11417.31858	5307.73027	1309.88603
2	0.10000	-99317.39212	13576.43857	8826.82347	4901.25129	1305.75245
3	0.20000	-99201.68604	13503.10606	8873.58651	4709.76876	1333.19225
4	0.30000	-99016.08503	13445.15010	8932.31611	4770.16548	1333.74401
5	0.40000	-98957.68771	13222.40291	8864.63519	4731.05864	1354.25491
6	0.50000	-99255.51871	13477.12353	8860.01336	4857.66466	1306.57935
7	0.60000	-99032.11647	13273.04161	8860.29631	5021.17239	1335.81248
8	0.70000	-98797.83402	13288.23411	8925.94840	4792.82822	1326.28202
9	0.80000	-98952.24978	13502.82998	8827.86461	4847.32295	1317.23483
10	0.90000	-99074.24991	13458.54891	8899.93330	4852.87641	1333.06943
11	1.00000	-98844.62889	13266.60501	8835.05914	4750.75926	1320.16634
12	1.10000	-98781.17698	13262.02735	9011.00988	4736.57869	1335.79034
13	1.20000	-98945.79951	13261.77428	8964.53451	4841.80549	1333.48442
14	1.30000	-98845.18425	13113.55925	8889.00008	4736.16481	1318.29781
15	1.40000	-99233.38252	13472.95400	8907.90154	4828.23584	1351.56142
16	1.50000	-99542.73803	13675.61235	8909.57226	4788.34749	1338.50644
17	1.60000	-99432.66029	13647.48733	8844.84234	4823.29832	1312.93832
18	1.70000	-99625.02417	13690.30492	8908.56763	4820.74358	1310.85365
19	1.80000	-99108.22598	13261.78555	8846.66719	4911.43317	1316.96133
20						

Рисунок 43. Фрагмент файла энергетической траектории молекулярной динамики.

Анализатор траекторий позволяет находить следующие характеристики временных зависимостей (предполагается, что данные представляют собой зависимости от времени, но под временем может подразумеваться и любая координата, записанная в первом столбце):

- максимальное значение (столбец Max);
- минимальное значение (столбец Min);
- точка максимума (столбец MaxAt) – значение абсциссы (времени), при котором функция принимает максимальное значение;
- точка минимума (столбец MinAt) – значение абсциссы (времени), при котором функция принимает минимальное значение;
- среднее значение функции (столбец Mid);
- усреднение исходных данных методом скользящего среднего (параметр Mid by);
- задание верхних и нижних границ анализа (поля Find Min From и Max time).

Как правило, в энергетических траекториях записываются сильно флуктуирующие величины. Первоначально требуется их усреднить. Для этого выбран метод скользящего среднего, где каждая точка усреднённых данных рассчитывается по временному отрезку

исходных данных, а все отрезки отстоят друг от друга ровно на один дискретный момент времени – частоту замера исходных данных.

$$\langle Y_i \rangle = \frac{\sum_{j=i}^{i+w} Y_j}{w}, i \in [1; N-w]$$

Где:

- $\langle Y_i \rangle$ – усредненное значение;
- Y_i – исходное значение;
- w – ширина окна усреднения (window);
- N – размер исходных данных.

Концы зависимостей могут сильно выбиваться от общего тренда. И их целесообразно бывает обрезать. Для этих целей введена возможность задания верхних и нижних границ анализа (поля «Find Min From» и «Max time»).

4.13.1 Порядок проведения анализа

Для открытия окна статистического анализа необходимо выбрать пункт «Statistic» меню «Service».

Вид Анализатора в процессе статистического анализа показан на следующем рисунке:

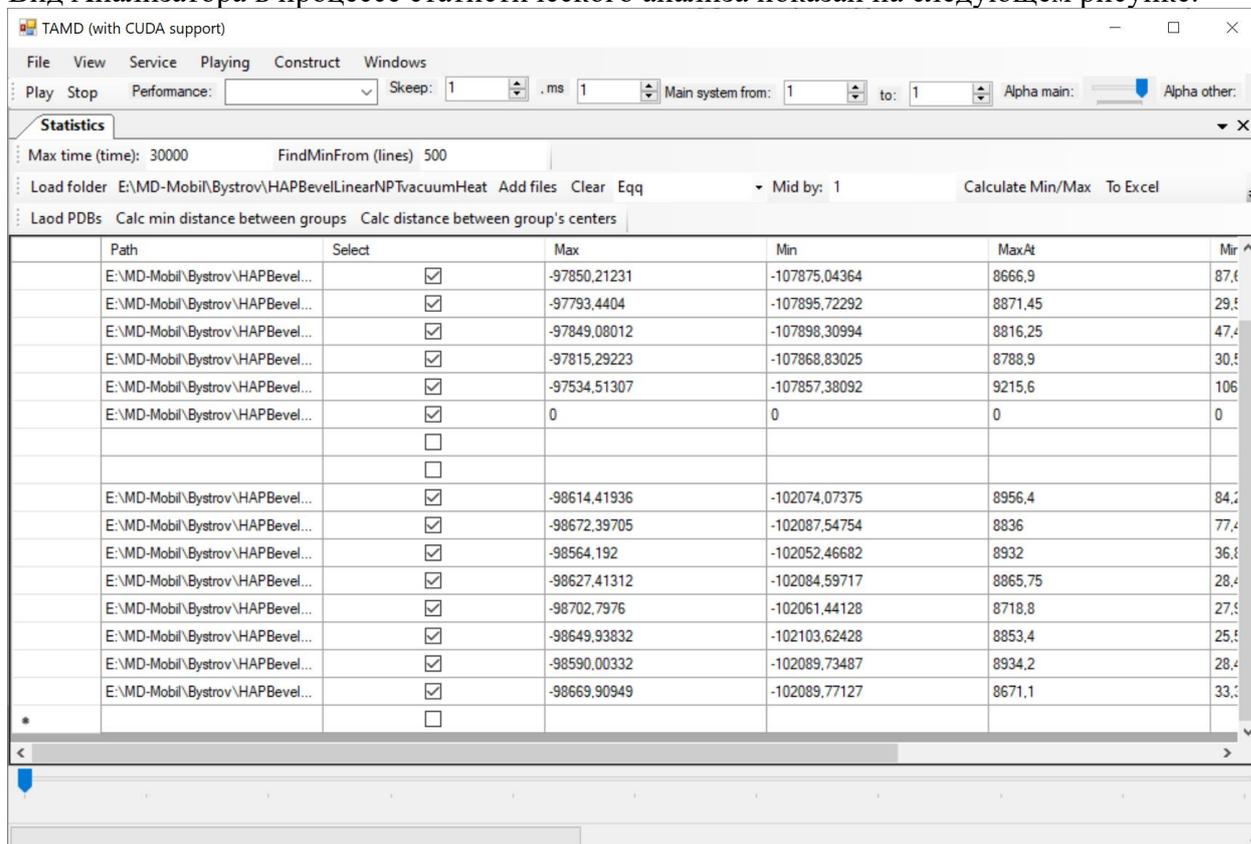


Рисунок 44. Статистический анализ энергетических траекторий молекулярной динамики. В отличие от большинства возможностей Анализатора траекторий, блок статистической обработки работает независимо от механизмов работы с классическими трехмерными траекториями молекулярной динамики.

Порядок проведения статистического анализа энергетических траекторий молекулярной динамики следующий:

1. Открыть окно статистической обработки данных (меню «Service», пункт «Statistic»).
2. Выбрать директорию, в которой расположены файлы энергетических траекторий (файлы с расширением .txt), при помощи кнопки Load Folder.

3. Нажать кнопку «Add files». При этом в таблицу будут добавлены все файлы с расширением .txt.
4. При необходимости, убрать галочки «Select» с ненужных файлов. Галочка «Select» ставится по умолчанию. Она означает, что анализ будет проведён над конкретным файлом.
5. Если подразумевается обработка файлов, расположенных в различных каталогах, то повторить пункты 2-4.
6. Выбрать столбец для проведения анализа (поле Eqq на рисунке). Изначально поле с выпадающим списком после кнопки «Clear» пустое. Значения в выпадающем заполняются после нажатия на кнопку «Add files». Перед началом анализа необходимо выбрать какое-либо поле.
7. Выбрать остальные параметры анализа: поле выбора ширины окна усреднения «Mid by» (по умолчанию 25 строчек данных, если усреднение не требуется, то выбрать 1), поля выбора границ «Find Min From» и «Max time».
8. Для проведения статической обработки файлов нажать на кнопку «Calculate Min/Max». Процесс может занять некоторое количество времени. Таблица статической обработки будет постепенно заполняться. Как только заполняется последняя строчка, процесс обработки закончен.
9. Для сохранения статистической информации предусмотрен экспорт в электронную таблицу Excel. Для этого достаточно нажать кнопку «To Excel». При этом в таблицу войдут все поля, кроме «Select», которое не несёт смысловой информации.

При экспорте в Excel результатов обработки нескольких директорий между каталогами делается пропуск в виде одной строки для повышения удобочитаемости данных. Вид таблицы показан на следующем рисунке. Для работы возможности требуется, чтобы на компьютере была установлена программа MS Excel.

File name	Max	Min	MaxAt	MinAt	Mid
E:\MD-Mobil\Bystrov\HAPBevelLinearNPT\100ww.txt	-97786,2	-107945	8810	50,8	-102790
E:\MD-Mobil\Bystrov\HAPBevelLinearNPT\101ww.txt	-98112,9	-107867	8507	43,7	-102895
E:\MD-Mobil\Bystrov\HAPBevelLinearNPT\102ww.txt	-97669	-107920	8989,15	48,05	-102697
E:\MD-Mobil\Bystrov\HAPBevelLinearNPT\103ww.txt	-97850,2	-107875	8666,9	87,6	-102797
E:\MD-Mobil\Bystrov\HAPBevelLinearNPT\104ww.txt	-97793,4	-107896	8871,45	29,5	-102773
E:\MD-Mobil\Bystrov\HAPBevelLinearNPT\105ww.txt	-97849,1	-107898	8816,25	47,45	-102798
E:\MD-Mobil\Bystrov\HAPBevelLinearNPT\106ww.txt	-97815,3	-107869	8788,9	30,5	-102780
E:\MD-Mobil\Bystrov\HAPBevelLinearNPT\107ww.txt	-97534,5	-107857	9215,6	106,55	-102608
E:\MD-Mobil\Bystrov\HAPBevelLinearNPT\ww100.txt	0	0	0	0	0
E:\MD-Mobil\Bystrov\HAPBevelLinearNPTvacuumHeat\100ww.txt	-98614,4	-102074	8956,4	84,2	-100520
E:\MD-Mobil\Bystrov\HAPBevelLinearNPTvacuumHeat\101ww.txt	-98672,4	-102088	8836	77,45	-100532
E:\MD-Mobil\Bystrov\HAPBevelLinearNPTvacuumHeat\102ww.txt	-98564,2	-102052	8932	36,8	-100525
E:\MD-Mobil\Bystrov\HAPBevelLinearNPTvacuumHeat\103ww.txt	-98627,4	-102085	8865,75	28,45	-100523
E:\MD-Mobil\Bystrov\HAPBevelLinearNPTvacuumHeat\104ww.txt	-98702,8	-102061	8718,8	27,9	-100531
E:\MD-Mobil\Bystrov\HAPBevelLinearNPTvacuumHeat\105ww.txt	-98649,9	-102104	8853,4	25,55	-100533
E:\MD-Mobil\Bystrov\HAPBevelLinearNPTvacuumHeat\106ww.txt	-98590	-102090	8934,2	28,4	-100523
E:\MD-Mobil\Bystrov\HAPBevelLinearNPTvacuumHeat\107ww.txt	-98669,9	-102090	8671,1	33,35	-100526

Рисунок 45. Экспорт результатов статистической обработки энергетических траекторий молекулярной динамики в формат Excel.

При желании эту таблицу легко отформатировать. При составлении настоящей документации мы расширили колонку с именем файла, дважды щелкнув мышью на между столбцами A и B в программе MS Excel.

Все значения будут записаны в единицах измерения исходных данных. Точки минимума и максимума вычисляются с использованием шкалы времени, роль которой играет первый столбец исходных данных. Иными словами, в столбцах «MinAt» и «MaxAt» будет время, а не номера строк.

Решение трудностей.

Для проведения статистического анализа необходимо выбрать столбец для анализа из выпадающего списка после кнопки «Clear». Этот список заполняется при нажатии на кнопку «Add Files». При этом программа считывает заголовок первого файла. Если список полей, то, по всей вероятности, первый файл не содержит правильного заголовка. Подразумевается, что наименования столбцов в заголовке, как и все строки в файле, отделены друг от друга пробелом. При этом несколько пробелов воспринимаются как один. Таким образом, в наименование любого поля не может содержать пробелы.

4.14 Запись видео

В Анализаторе существует несколько способов записать видео трёхмерной проекции:

1. Интерактивный.
2. Автоматический

В интерактивном режиме записываются все отображаемые кадры. Пользователь может включить воспроизведение, настроить режимы отображения, остановить

воспроизведение, повернуть молекулу под нужным ракурсом, приблизить её и т.д. Все эти кадры войдут в видеофайл.

В автоматическом режиме запишется всё молекулярное кино от начала и до конца. В качестве выходного файла используется несжатый файл формата .avi. Выбор сделан ввиду отсутствия привязки к каким-либо видеокодекам, установленным в системе. Рекомендуется перекодировать этот файл для дальнейшего распространения. Разрешение видеофайла берётся равным разрешению окна трёхмерной проекции. В режиме записи нельзя изменять размеры окна Анализатора, т.к. один видеофайл подразумевает одно разрешение. Однако можно записать один видеофайл, затем изменить размеры окна Анализатора и записать другой видеофайл.

Перед записью видеофайла рекомендуется подумать о желаемом разрешении и его пропорциях. Каким именно образом отображаемый объект будет лучше смотреться на экране?

Количество кадров в секунду при записи видеофайла по умолчанию равно 10. Оно может быть изменено из меню View, пункта Options, вкладки 3D Options выбором требуемого значения в поле «Frames per second during record AVI».

Скорость воспроизведения на экране не имеет никакого отношения к скорости будущего видеофайла. Поле «ms» главной панели инструментов не влияет на процесс записи.

Определённые действия можно сделать быстро, некоторые медленно, требующие от пользователя многократного переключения опций. Выходной видеофайл будет иметь скорость воспроизведения, указанно в поле «Frames per second during record AVI».

Учитывая вышесказанное, для повышения качества видео рекомендуется увеличивать поле «Detail level» из меню View, пункта Options, вкладки 3D Options.

4.14.1 Запись видео в интерактивном режиме

Для записи видео в интерактивном режиме необходимо приготовить файл траектории для записи молекулярного кино, а также выбрать режимы отображения молекулярной системы.

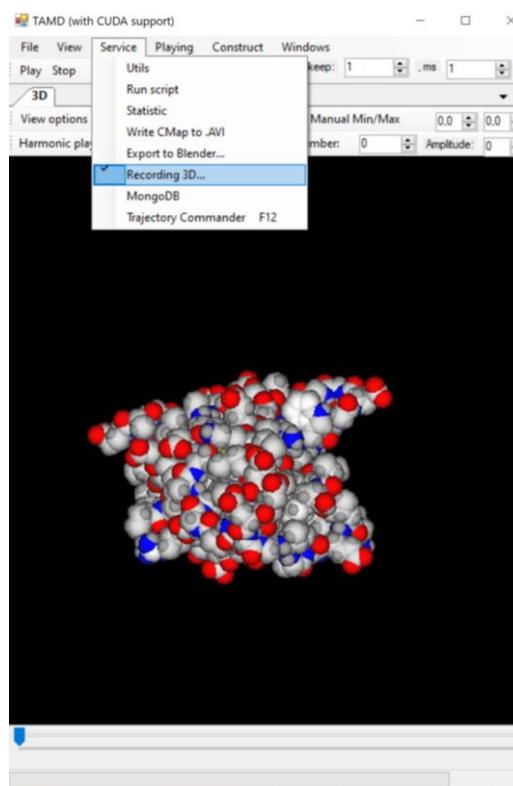


Рисунок 46. Включение записи видео в интерактивном режиме. Последовательность действий для записи видео может быть следующей:

1. Включение режима записи происходит выбором пункта «Recording 3D...» из меню «Service». При этом Анализатор траекторий запросит имя сохраняемого .avi-файла. Напротив пункта «Recording 3D...» отобразится галочка, которая будет свидетельствовать о включенном режиме записи.
2. Включить воспроизведение. Также можно производить любые действия по выбору режимов, а также масштабированию или перемещению молекулы в пространстве. Все отображаемые кадры будут записаны.
3. Для записи паузы, акцентирования внимания на определённом кадре используется кнопка Skip, которая доступна из меню View, пункт Options, вкладка 3D Options. При этом в видео запишется несколько одинаковых кадров, количество которых указывается в поле «Skip frames during record AVI».
4. Для окончания записи в файл необходимо повторно выбрать пункт «Recording 3D...» из меню «Service». При этом галочка напротив пункта «Recording 3D...» исчезнет. Видеофайл закроется.

Открытие дополнительных окон Анализатора, таких, как выбор параметров, в режиме записи может изменить разрешение окна трехмерной проекции. Это приведёт к ошибке записи видео. Рекомендуется открывать окна Анализатора до начала записи.

4.14.2 Наложение текста при записи видео

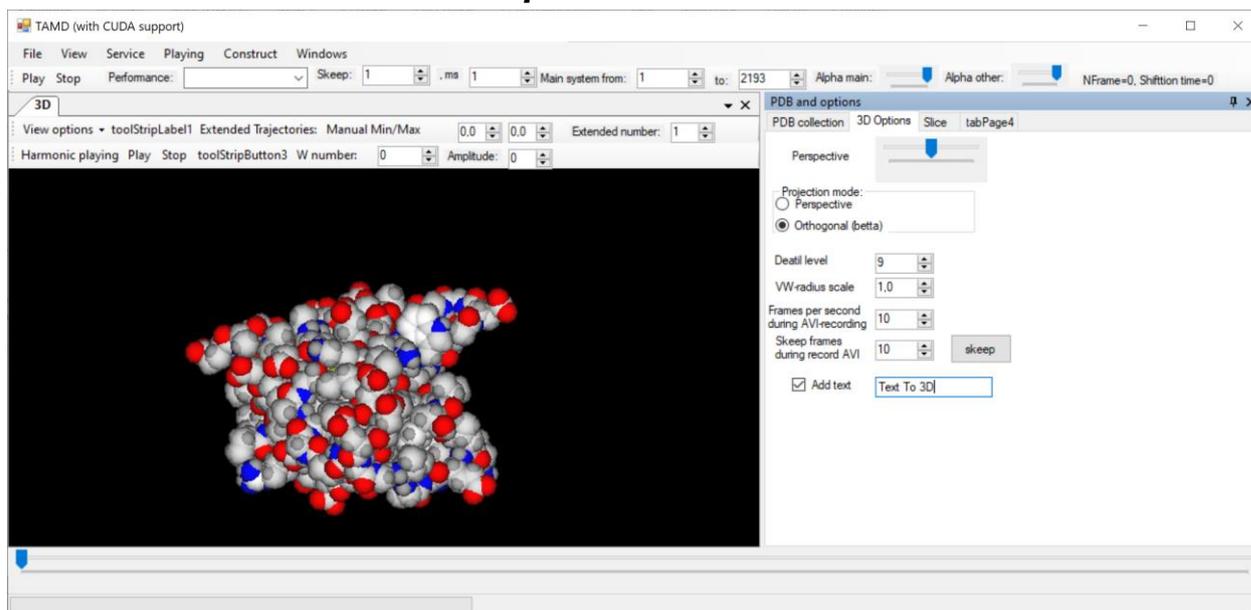


Рисунок 47. Наложение текста при записи видео.

Для наложения текста на видео используется меню View, пункт Options, вкладку 3D Options. Наложение текста включается галочкой Add text. При этом поверх видео записывается надпись, вводимая в текстовом поле справа от галочки Add text.

4.14.3 Запись видео в автоматическом режиме

Анализатор траекторий позволяет записать видео большой траектории в автоматическом режиме без участия пользователя в процессе записи.

Для этого необходимо сначала открыть файл траектории и выбрать необходимые режимы записи, а также позаботиться о желаемом разрешении записи, изменяя размеры окна Анализатора траекторий.

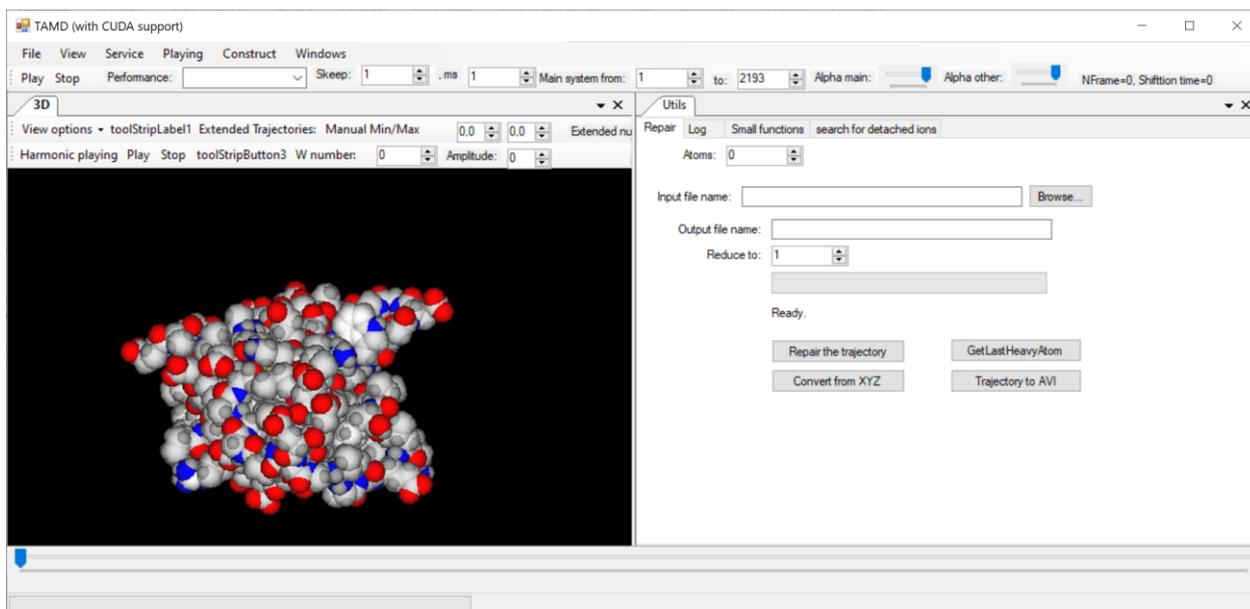


Рисунок 48. Запись молекулярного кино в видеофайл в автоматическом режиме. Для записи молекулярного кино всей траектории в автоматической режиме необходимо из меню Service, выбрать пункт Utils. Включение режима автоматической записи осуществляется нажатием кнопки Trajectory to AVI. Анализатор траекторий запросит имя видеофайла и произведёт запись всех кадров траектории. Во время записи работает также опция наложения текста на видео. Изменить текст, также другие опции отображения в процессе автоматической записи весьма сложно. Если в этом есть потребность, то рекомендуется перейти к интерактивной записи видео молекулярного кино.

4.15 Построение твердотельных моделей

Анализатор траекторий позволяет подготавливать .pdb-файлы для печати на 3D-принтерах.

Для печати белка, либо другой макромолекулы, необходимо подготовить её трехмерную модель. Затем эта модель используется программой трехмерной печати, такой как Cura Ultimaker или Repetier host, чтобы составить программу для 3D-принтера на языке G-CODE (язык программирования устройств с числовым программным управлением (ЧПУ)).

Распространённым форматом для хранения трехмерных моделей является .STL. Это формат файла, широко используемый для хранения трёхмерных моделей объектов для использования в аддитивных технологиях. Информация об объекте хранится как список треугольных граней, которые описывают его поверхность, и их нормалей.).

В свою очередь, чтобы получить модель в формате STL, целесообразно использовать программу трёхмерного моделирования, такую как Blender. С другой стороны, можно написать прямой экспорт в формат STL, но использование Blender значительно снижает трудоёмкость задачи. Выбор программы Blender обусловлен двумя причинами:

1. Blender – бесплатная программа трёхмерного моделирования с открытым исходным кодом.
2. Blender поддерживает управление при помощи программ, написанных на языке Python.
3. Blender обладает богатыми возможностями по редактированию трёхмерных объектов.

После получения трёхмерной модели белка в Blender можно отредактировать модель. Например, построив подставку, чтобы модель красиво смотрелась на столе.

Для экспорта модели в Blender можно написать скрипт на языке Python, который бы содержал команды по воссозданию модели белка в программе трёхмерного моделирования.

Весь процесс получения трёхмерной модели показан на следующем рисунке.

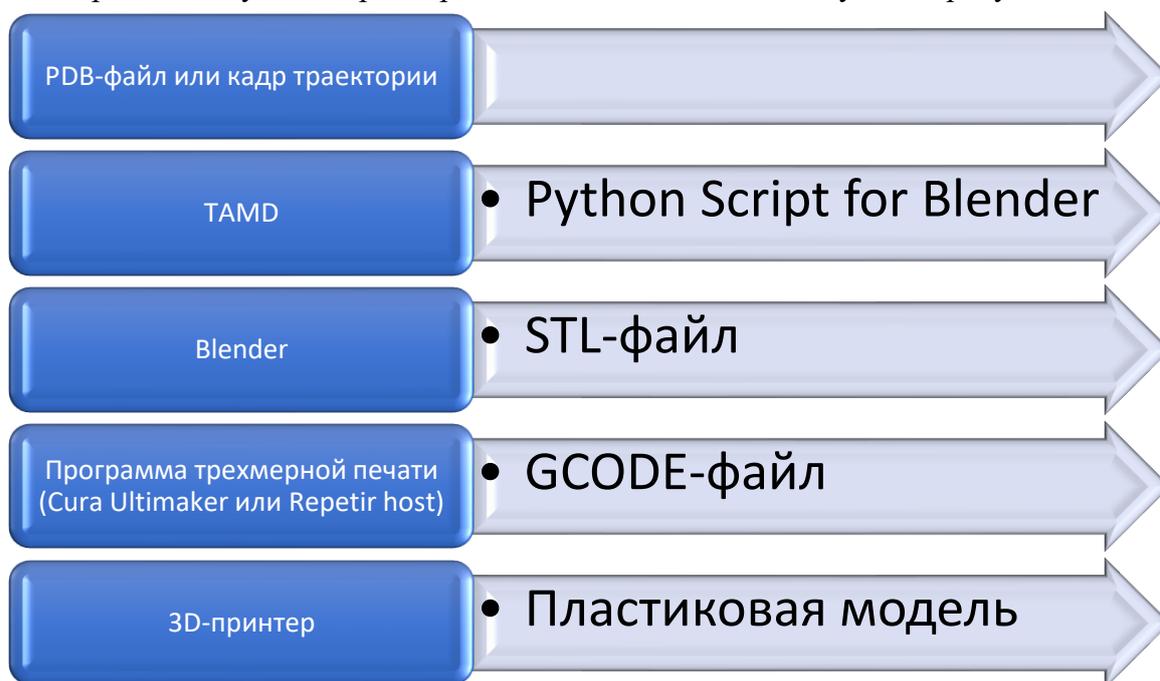


Рисунок 49. Процесс получения трехмерной модели белка.

4.15.1 Описание применяемых конструкций Python for Blender

Анализатор траекторий экспортирует программу на языке Python в виде текстового файла. В этой программе содержатся команды по построению графических примитивов, из которых состоит модель.

Атомы изображаются в виде сфер. Большую часть программы занимает повторение команды `primitive_ico_sphere_add` с различными параметрами:

```
bpy.ops.mesh.primitive_ico_sphere_add(subdivisions=4,radius=0.370000004768372, location=(-1.3283,-5.377,10.507))
```

Значения параметров:

- `subdivisions` – детализация сферы. Сфера состоит из многогранников. Большее значение приводит к большему количеству многогранников на одну сферу. При этом время, затрачиваемое на воссоздание модели увеличивается, как и размер дискового прорастав, необходимого для хранения stl-файла.
- `radius` – значение радиуса атома. Берётся из таблицы Colors, поля VW. Значение радиуса делится на делитель, указываемый в поле VW devision.
- `location=(x, y, z)` – трёхмерные координаты положения атома в пространстве. Берётся либо из PDB-файла, либо из кадра траектории.

Используемый материал и цвет.

Для отображения разноцветных сфер необходимо задать им свой материал. Материал характеризую такие параметры как цвет и прозрачность. Ниже расположен пример исходного кода для Blender, который устанавливает для сферы некий материал:

```
import bpy

def makeMaterial(diffuse, alpha):
    mat = bpy.data.materials.new('materialchik')
    mat.diffuse_color = diffuse
```

```
mat.diffuse_shader = 'LAMBERT'  
mat.diffuse_intensity = 1.0  
mat.specular_color = (1,1,1)  
mat.specular_shader = 'COOKTORR'  
mat.specular_intensity = 0.5  
mat.alpha = alpha  
mat.ambient = 1  
return mat
```

```
def setMaterial(ob, mat):  
    me1 = ob.data  
    me1.materials.append(mat)
```

Приведём описания лишь требуемых параметров:

- `mat.specular_color = (1,1,1)` – значение цвета в формате (R, G, B). Каждый канал характеризуется вещественным числом от 0 до 1. Цвет с кодом (1, 1, 1) соответствует белому цвету.
- `mat.alpha` – прозрачность. Вещественное число от 0 до 1. По умолчанию равно 1, атом полностью непрозрачен.
- `mat = bpy.data.materials.new('materialchik')`. 'materialchik' – это вымышленное название применяемого материала. В программе для каждого атома используются материалы вида `materialchik1`, `materialchik2` и т.д.

Значения остальных параметров устанавливаются в соответствии с эстетическими взглядами и не несут смыслового оттенка.

4.15.2 Порядок экспорта моделей в формат Python for Blender

Для экспорта моделей белков и макромолекул в формат Python for Blender с целью подготовки трёхмерных моделей необходимо выбрать пункт «Export to Blender» меню «Service».

Типичный вид программы показан на следующем рисунке.

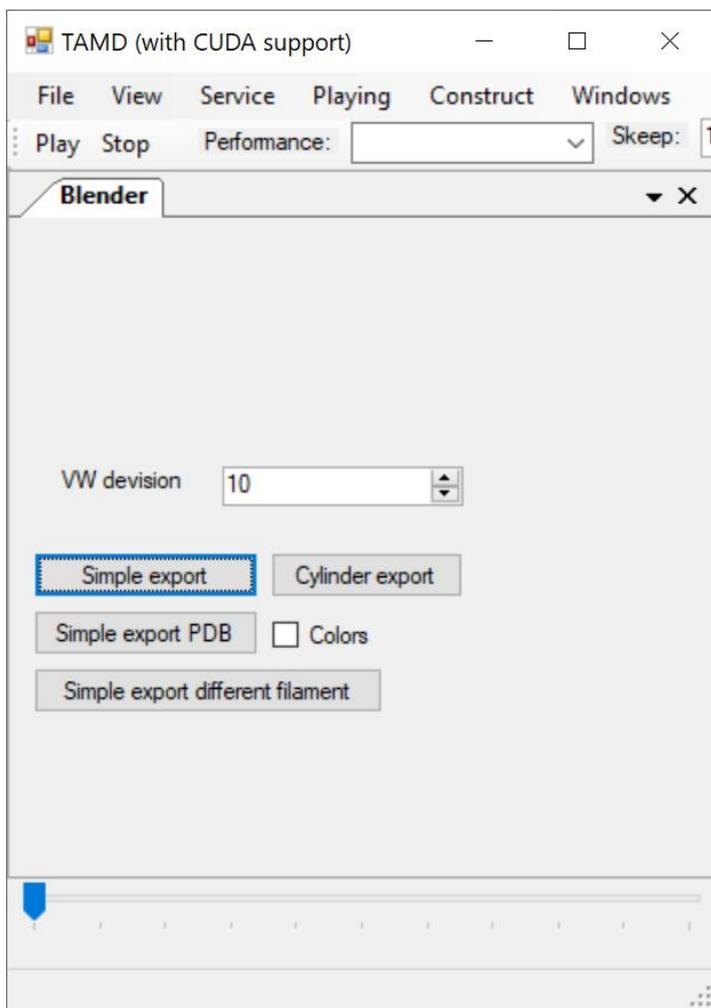


Рисунок 50. Экспорт координат атомов в формат Python for Blender.

Значения параметров и кнопок:

- VW division – делитель радиуса атомов. При экспорте атомов в сферы в качестве радиуса используется поле VW таблицы «colors» основной базы данных Анализатора траекторий, делённое на значение из поля VW division.
- Colors – если выбран этот флажок, то атомы рисуются каждый своим цветом. Компоненты цвета ищутся из полей R, G, B таблицы «colors» основной базы данных Анализатора траекторий. Это поле не имеет прямого отношения к цветной печати. См. ниже.
- Simple export – экспорт текущего кадра траектории в формат Python for Blender. Предполагается, что файл траектории открыт и стоит на требуемом кадре.
- Simple export PDB – экспорт вбираемого PDB-файла в формат Python for Blender. При этом открывается диалоговое окно выбора PDB-файла.
- Simple export different filament – экспорт вбираемого PDB-файла в формат Python for Blender. При этом открывается диалоговое окно выбора PDB-файла.

При использовании всех команд также открывается окно сохранения выходного файла.

4.15.2 Многоцветная печать трёхмерных моделей

Если 3D-принтер оборудован несколькими печатающими головками – экструдерами, либо его головка принимает несколько цветов, то на таком устройстве возможна цветная печать. Предполагается, что количество цветов ограничено и равно 3. Это не принципиальное ограничение, а лишь ограничение текущей реализации. Так или иначе,

предполагается, что количество цветов меньше, чем количество используемых типов атомов.

Каждому типу атому ставится в соответствии свой цвет, точнее, свой номер печатающей головки, либо номер филамента.

Для этих целей в таблице «colors» основной базы данных Анализатора траекторий используется поле «filament».

name	comment	R	G	B	A	VW	InRelation	m	CAlpha	filament	Щелкнит
Ag		128	128	128	255	5	0	0	0	0	2
C		220	220	220	255	4	0	1	1	1	1
C*		220	220	220	255	4,3	0	1	1	1	1
c_1		255	120	120	255	4	0	1	0	1	1
C1'		255	255	255	255	3,7	0	0	0	0	1
C1*		255	120	120	0	3,7	0	0	0	0	1
C1'		220	220	220	220	3	0	1	0	0	1
C2		192	192	192	255	3	0	1	0	0	1
C2'		255	255	255	255	3,7	0	0	0	0	1
C2*		255	120	120	0	3,7	0	0	0	0	1
C3'		255	255	255	255	3,7	0	0	0	0	1
C3*		255	120	120	0	3,7	0	0	0	0	1
C3'		220	220	220	220	3	0	1	0	0	1
C4		255	255	255	255	3,7	0	0	0	0	1
C4'		255	255	255	255	3,7	0	0	0	0	1
C4*		255	120	120	0	3,7	0	0	0	0	1
C5		255	255	255	255	3,7	0	0	0	0	1
C5'		255	255	255	255	3,7	0	0	0	0	1
C5*		255	120	120	0	3,7	0	0	0	0	1
C5M		255	120	120	0	3,7	0	0	0	0	1
C6		255	255	255	255	3,7	0	0	0	0	1
C7		255	255	255	255	3,7	0	0	0	0	1
C8		255	0	255	255	3,7	0	0	0	0	1
CA		255	255	255	255	4,34999999046	0	1	0	0	1
CB		255	255	255	255	4,3	0	1	0	0	1
CC		220	220	220	255	4,3	0	1	0	0	1
CD		255	255	255	255	4,3	0	1	0	0	1
CD1		255	120	120	0	0	0	0	0	0	0
CD2		255	120	120	0	0	0	0	0	0	0
CE		255	255	255	255	4,3	0	1	0	0	1

Рисунок 51. Таблица «Colors» основной базы данных Анализатор траекторий.

Возможные значения:

- «0» – для первого цвета;
- «1» – для второго цвета;
- «2» – для третьего цвета.

Например, тяжелые атомы могут иметь цвета 0 и 1, а водороды – цвет 2.

Один из способов подготовки моделей для цветной печати заключается в разбиении модели на несколько частей. Каждой модели при создании файла G-CODE, понятного для 3D-принтера, назначается свой номер экструдера, в который подаётся филамент определённого цвета.

Для трехцветной печати используется команда «Simple export different filament» окна Blender Анализатора траектории. После выбора PDB-файла программа запросит имя выходного файла для скрипта Python for Blender. При это она сохранит сразу три файла: по файлу на каждый цвет. Названия файлов будут оканчиваться на суффиксы «_0», «_1» и «_2» соответственно.

4.15.3 Исполнение скриптов Python в программе Blender

После экспорта координат атомов в формат Python for Blender необходимо выполнить этот файл в программе трёхмерного моделирования Blender.

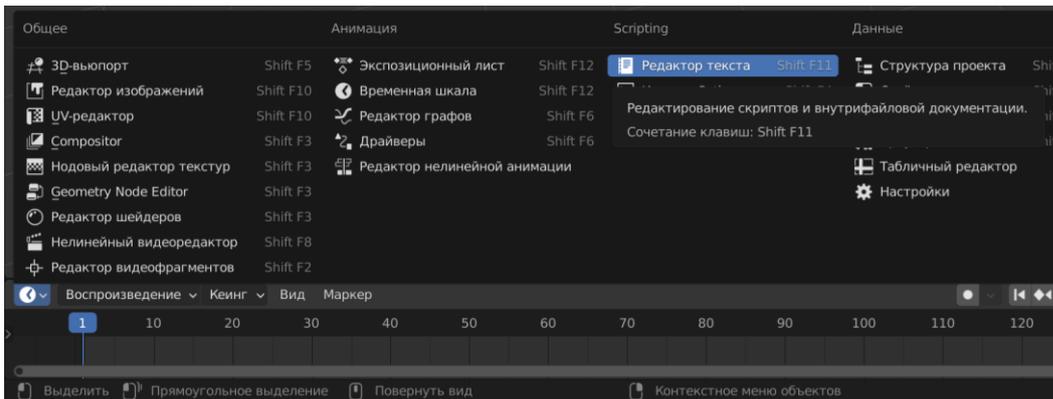


Рисунок 52. Открытие текстового редактора в программе Blender.

Для начала необходимо открыть текстовый редактор программы Blender. Это можно сделать при помощи кнопки с изображением часов нижней панели инструментов, которая отвечает за временную шкалу воспроизведения анимации. Как известно, подготовка трехмерных моделей не подразумевает их анимацию. Поэтому временная шкала анимации бесполезна.

Далее необходимо создать новый скрипт кнопкой «Создать» и вставить в него скрипт Python for Blender, полученный из Анализатора траекторий. Также можно воспользоваться кнопкой открытия файла.

Для отрисовки моделей скрипт нужно выполнить. Для этих целей служит кнопка «Выполнить» с изображением стрелки, похожей на команду воспроизведения.

При желании модель можно отредактировать. Например, можно добавить к ней подставку.

Для дальнейшей работы необходимо экспортировать модель в формат .STL. Для этого можно воспользоваться меню Файл -> Экспортировать -> STL.

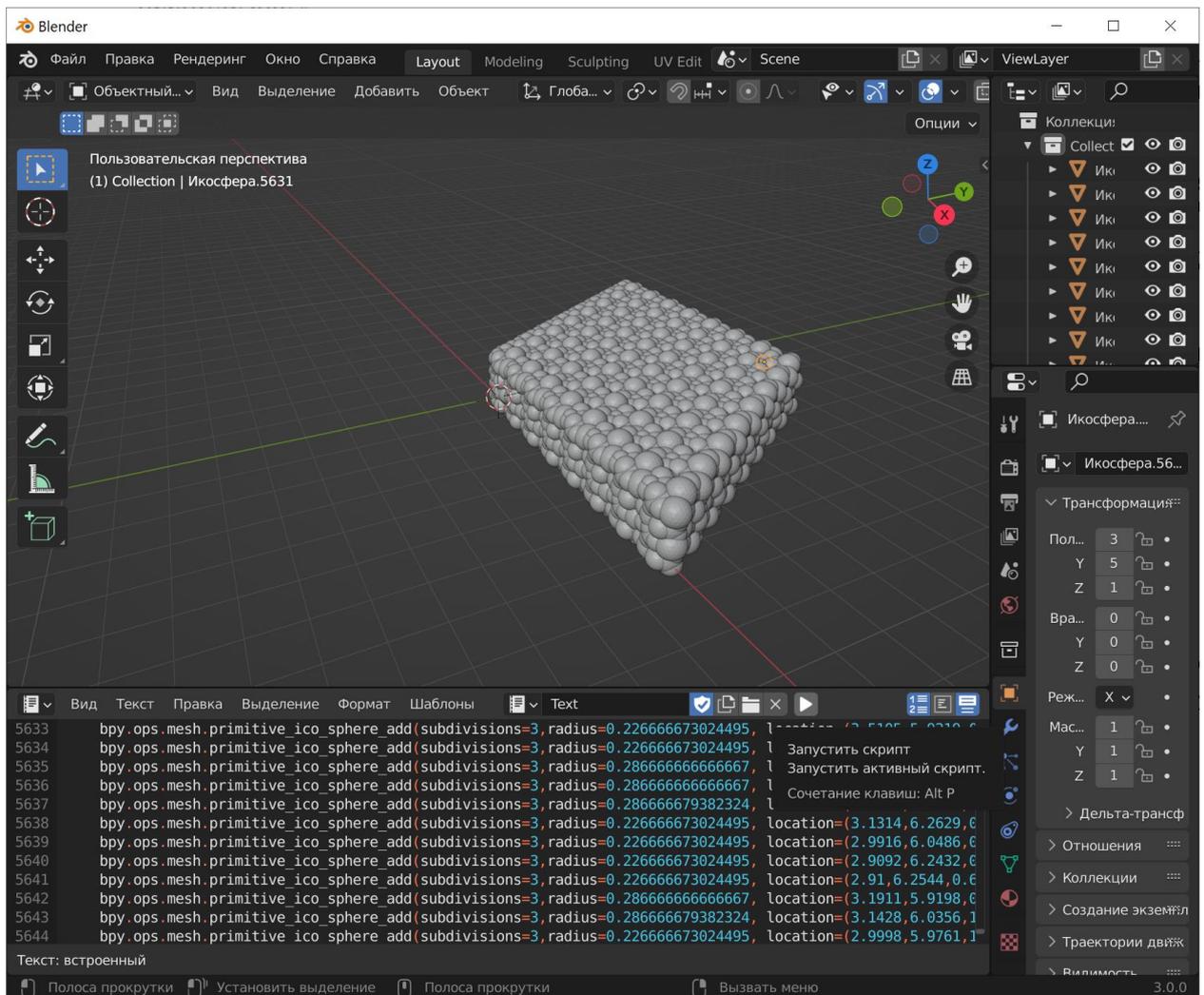


Рисунок 53. Окно программы Blender с созданной моделью и программой на языке Python.

4.15.4 Подготовка моделей в формате G-CODE

Предполагается, что создана модель в формате STL.

Для печати модели необходимо преобразовать её в формат, понимаемый 3D-принтером – формат G-CODE. Для этих целей существуют программы нарезки модели на слои, такие как Ultimaker Cura и Repetir host.

Покажем процесс подготовки файлы G-CODE на базе программы Ultimaker Cura.

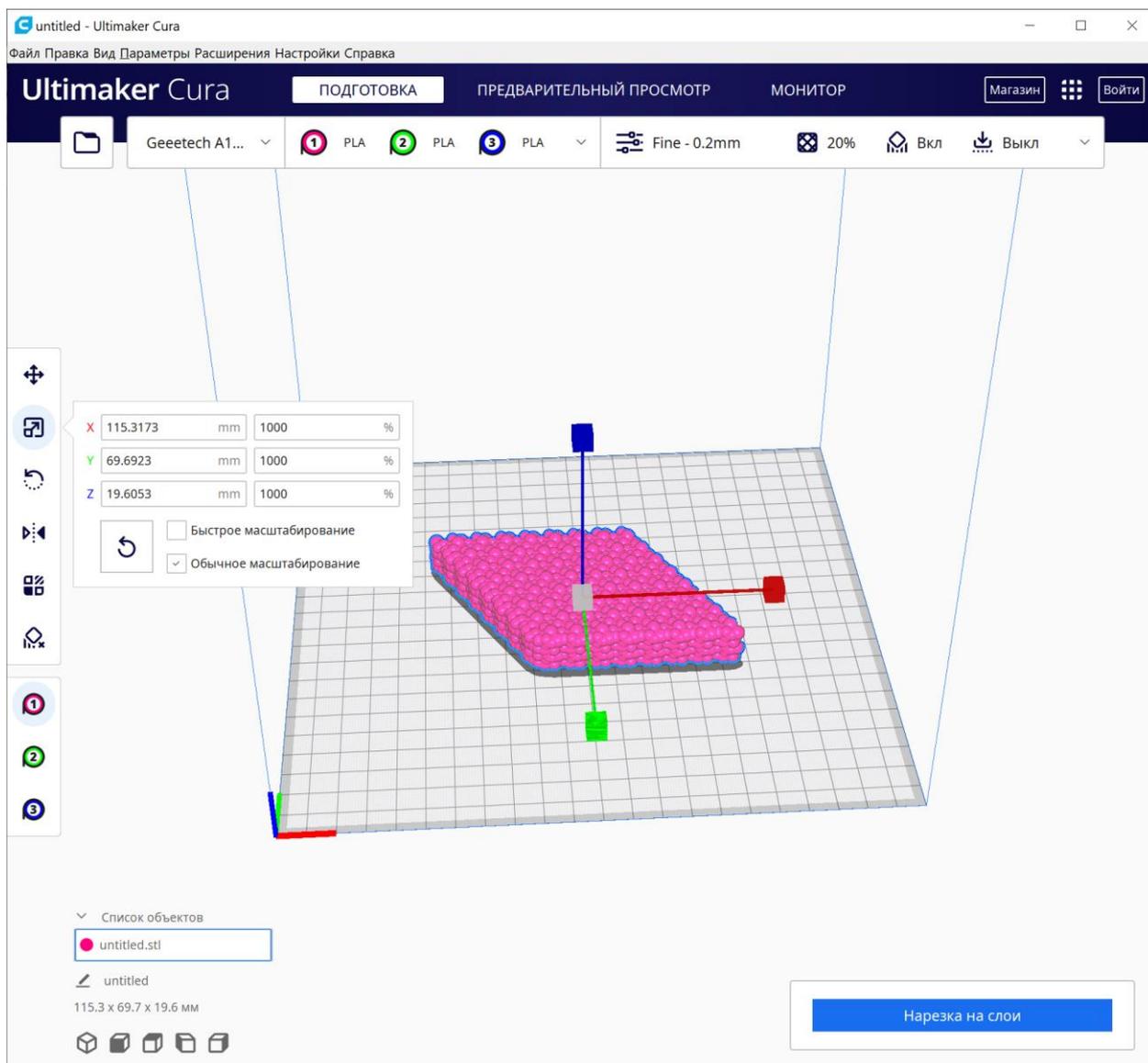


Рисунок 54. Программа Ultimaker Cura с загруженной моделью.

Примерный порядок работы с программой Ultimaker Cura может быть следующим:

1. Загрузить модель в формате .STL.
2. Масштабировать модель. Начальный масштаб модели таков, что 1 мм советует 1 Å. Типично требуется увеличить модель в 10 раз, задав масштаб 1000% для всех осей. Также можно задать масштаб под всю доступную для печати область пространства.
3. Установить многочисленные параметры печати. Этот материал относится к программе Ultimaker Cura и конкретному используемому 3D-принтеру и выходит за рамки настоящего Руководства.
4. Нарезать модель на слои соответствующей кнопкой.
5. Сохранить модель в формате G-CODE.

Особенности цветной печати

Для цветной печати необходимо последовательно загрузить все три STL-файлы модели.

При масштабировании всем моделям назначается один и тот масштаб.

Для каждой модели выбирается свой экструдер, соответствующий желаемому цвету.

Поддержки

При печати моделей строятся поддержки. Они необходимы для свисающих деталей.

Крайне желательно повернуть модель в пространстве таким образом, чтобы свисающих частей было как можно меньше.

При многоцветной печати необходимо позаботиться о выборе экструдера для поддержек.

Для поддержек также можно использовать растворимый материал, такой как PVA. Работа с этим материалом возможна на двух и трехцветных принтерах (возможно и большее количество цветов, что не принципиально).

При работе с трехцветным принтером можно поделить модель не на три цвета, а на два.

При этом в третий экструдер вставляется растворимый материал для печати поддержек.

4.16 Конструирование системы для поиска наилучшего расположения двух подсистем

Задача поиска наилучшего расположения двух молекул решается методом перебора.

Слово полный перебор здесь неуместно, потому что мы имеем дело с дискретными величинами, дискретным приращением угла. Но при достаточном количестве вычислительных мощностей задача близка к полному перебору.

В общем случае имеются две разных молекулы или две подсистемы, не связанные между собой валентными связями.

Для поиска наилучшего взаимного расположения зафиксируем в пространстве одну молекулу. Вторую будет находиться на некотором расстоянии от первой, сферически с ней не перекрываясь. Введём полярные координаты.

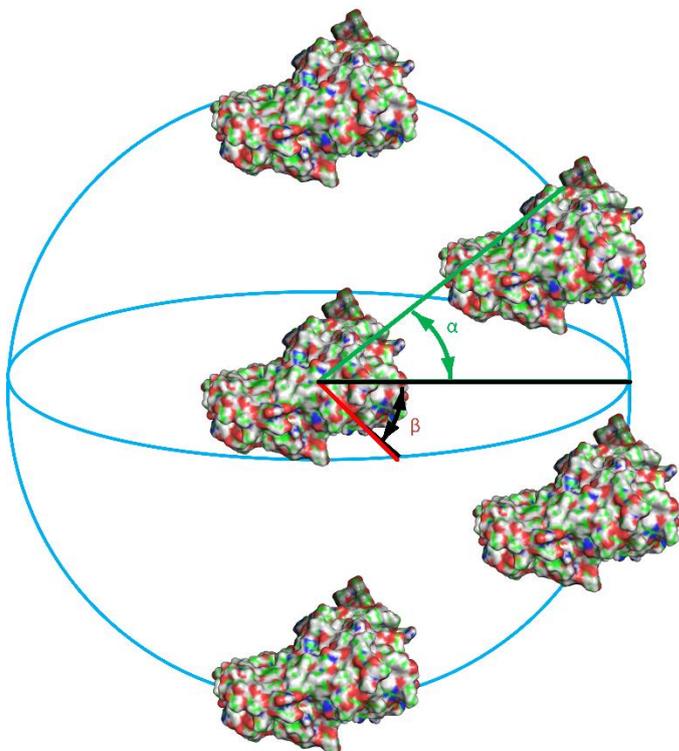


Рисунок 55. Взаимное расположение двух молекул актина

Будем варьировать углы α и β таким образом, чтобы вторая молекула описала полусферу относительно первой. В общем случае надо описать полную сферу. В случае, если обе молекулы идентичны, достаточно будет полусферы. Таким образом, угол $\alpha=180^\circ$ будет аналогичен углу $\alpha=0^\circ$.

Чтобы учесть все варианты (с оговоркой на дискретный шаг) взаимного расположения, необходимо учесть также собственные степени свободы второй молекулы актина. В полярных координатах это будет также два угла.

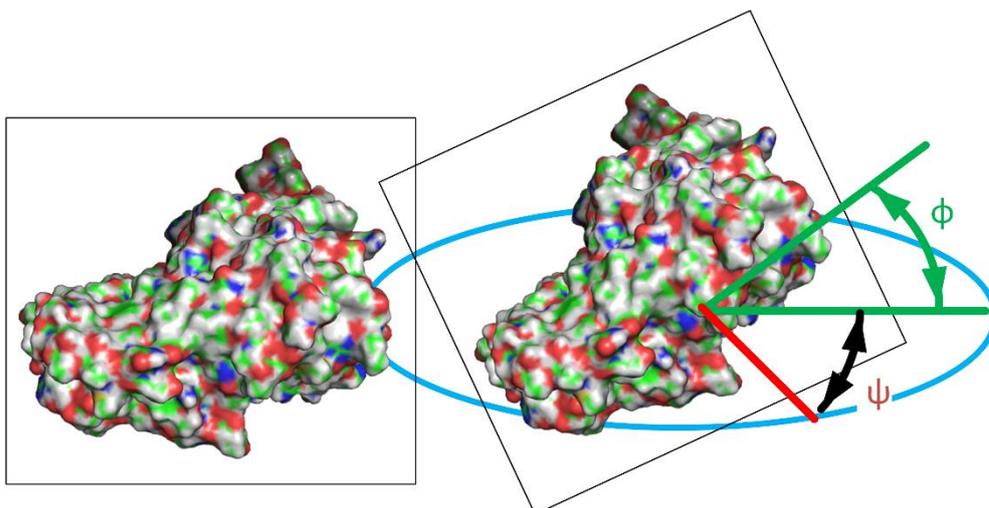


Рисунок 56. Варьирование собственными степенями свободы второй молекулы актина. Далее при помощи аффинных преобразований вторая молекула актина перемещается к первой таким образом, чтобы минимальное расстояние между ними составило 4 Å. В конкретных экспериментах углы $\alpha=\beta=\varphi=\omega=45^\circ$. Число отсчетов углов α и β , отвечающих за расположение составило 4 ($4 \times 45^\circ = 180^\circ$), а для углов φ и ω , отвечающих за поворот второй молекулы 8 ($8 \times 45^\circ = 360^\circ$). Таким образом, общее количество образцов составило $4 \times 4 \times 8 \times 8 = 1024$.

4.16.1 Порядок работы с TAMM

Для конструирования всевозможных вариантов расположения двух подсистем с помощью Анализатора траекторий необходимо открыть меню «Construct» и выбрать пункт «Two molecules».

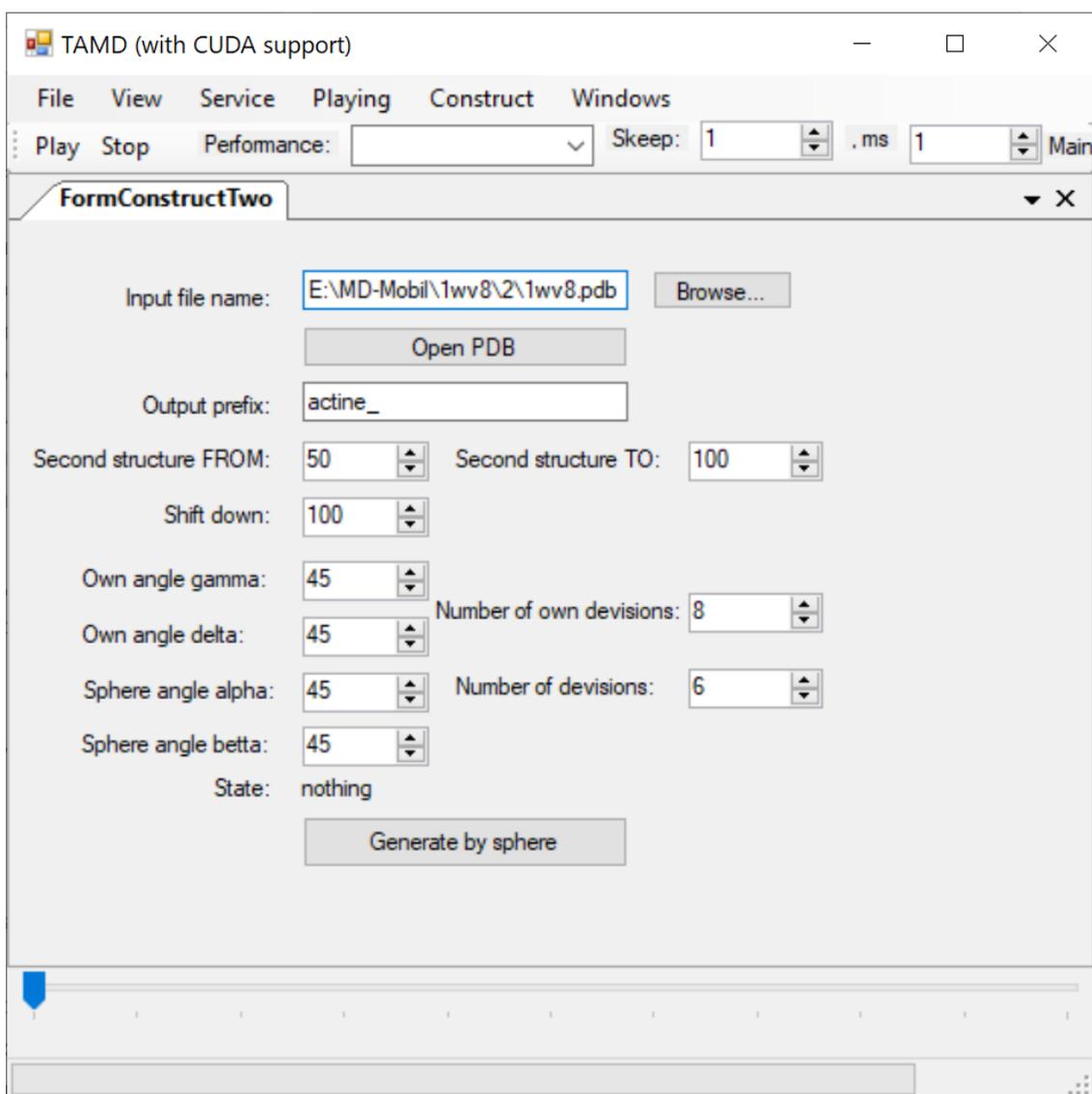


Рисунок 57. Окно Анализатора траекторий в режиме конструирования системы. Для начала необходимо выбрать исходный pdb-файл. Для этого можно ввести полное имя файла в поле «Input file name», либо воспользоваться кнопкой «Browse», которая откроет диалог выбора файла.

Далее необходимо загрузить pdb-файл при помощи кнопки «Open PDB». После открытия файла раз блокируются диапазоны полей второй подсистемы.

После этого можно установить все параметры конструирования:

- Output prefix – начальный префикс всех генерируемых (выходных) pdb-файлов.
- Secondary structure from – первый атом второй подсистемы, включительно
- Secondary structure to – последний атом второй подсистемы, включительно
- Shift down – начальная позиция второй системы относительно первой.
- Own angle gamma и Own angle delta – углы вращения второй подсистемы вокруг её центра масс.
- Number of own divisions – количество отсчетов по каждому измерению (с периодом gamma и delta градусов).
- Sphere angle alpha и Sphere angle betta – углы вращения второй системы относительно первой.
- Number of divisions – количество отсчетов по каждому измерению (с периодом alpha и betta градусов).
- State – статус генерации молекул. Nothing означает бездействие.

Для генерации набора молекул используется кнопка «Generate by sphere». Всего будет сгенерировано (записано) $(\text{Number of own divisions})^2 * (\text{Number of own divisions})^2$ pdb-файлов. Квадрат следует из того, что вращение происходит сразу по двум углам. И надо перебрать взаимное сочетание всех пар углов. Во время генерации набора файлов будет изменяться значение поля State, которое будет указывать на название текущей системы.

4.16.2 Особенности реализации

Задача решается в несколько этапов:

1. Загрузка двух подсистем как отдельных экземпляров класса PDB.
2. Перемещение каждой подсистемы в её центр масс.
 - на данном этапе две системы будут перекрываться.
3. Перемещение второй подсистемы вниз на «Shift down» ангстрем.
 - Значение поля «Shift down» по умолчанию равно 100, что подойдёт для большинства систем. Главное, чтобы это значение превосходило сумму расстояния от центра масс каждой системы до её любого атома. Иными словами, это поле должно быть больше суммы радиусов первой и второй подсистем.
 - Направление вниз выбрано случайно. Но служит для удобства выбора углов. Достаточно сместить вторую подсистему в любом направлении, чтобы алгоритм работал.
4. Перебор всех сочетаний значений углов alpha, betta, gamma, delta, учитывая количество создаваемых подсистем.
 - Для каждого сочетания углов создаётся своя система, состоящая из первой подсистемы в своём центре масс и вторая подсистема на расстоянии Shift down в позиции, задаваемой углами взаимного расположения, а также углами собственного вращения второй подсистемы.
 - Системы начинают сближаться с шагом 0.1 Å до тех пор, пока расстояние между любой парой атомов из разных подсистем не достигнет 4 Å.
5. Записывается выходной файл и процесс повторяется, начиная с пункта 3.

Для ускорения расчетов процесс поиска минимального расстояния между двумя подсистемами реализован при помощи параллельного программирования. Необходимо перебрать все пары атомов. Первый атом принадлежит первой подсистеме, а второй – второй. Задача решается при помощи двух вложенных циклов. Для использования мощностей современных многоядерных процессоров внешний цикл запускается параллельно на всех ядрах центрального процессора. При этом надо узнать, на какой итерации внешнего цикла найдётся минимальное расстояние. Вводится вспомогательный массив, размерность которого совпадает с размеров первой подсистемы или внешнего цикла. В него записывается минимальное расстояние от текущего атома первой подсистемы до ближайшего атома второй. Задача имеет сложность $O(N^2)$. После работы параллельного цикла с вложенным остаётся найти минимальное значение вспомогательного массива – решить задачу сложности $O(N)$, которое и укажет на минимальное расстояние между подсистемами.

5. TAMD-Script. Сценарный анализ траекторий молекулярной динамики

5.1 Назначение

Поддержка Анализатором скриптового языка переводит TAMD в класс программ-платформ, позволяя производить не только интерактивный анализ траекторий, но и получение группы характеристик по сценарию анализа траекторий.

5.2 Поддержка сценариев анализа траекторий

В настоящее время все больше производителей программного обеспечения оставляют возможность пользователям своих продуктов ровно как и сторонним разработчикам использовать программный инструментарий для управления своим ПО. У пользователей появляется гибкий механизм автоматизации рутинных операций. Сторонние разработчики, в свою очередь, выступают как создатели всевозможных дополнений, расширяющих возможность исходного продукта. С экономической точки зрения, остаются в выигрыше абсолютно все перечисленные лица. Программный продукт – платформа – приобретает большую популярность, позволяя решать более широкий класс задач, либо пользование им становится более удобным и быстрым.

Ярким примером программ-платформ является пакет MS Office с поддержкой VBA (Visual Basic for Applications). Также VBA поддерживается САПР AutoCAD. Последним пакетом также традиционно поддерживается AutoLisp – адаптированная версия языка Lisp. Фирма 1С пошла по другому пути: разработала свой собственный язык для написания программ для бизнеса, используя свой собственный механизм организации и доступа к СУБД.

TAMD – Анализатор траекторий молекулярной динамики содержит ряд программных средств для работы с файлами траекторий, получения по ним набора характеристик и вывода этих характеристик на экран компьютера, а также в графические, текстовые и excel-файлы. Для доступа к описанным инструментами традиционно используется GUI (графический интерфейс пользователя).

Для перехода Анализатора в класс программы-платформы реализована возможность выполнения скриптов - программ на языке Python. Программе-скрипту доступен программный интерфейс Анализатора (TAMD API), посредством которого происходит управление Анализатором и двухсторонний обмен информацией между скриптом и программой-платформой.

Поддержка выполнения скриптов предоставляет широкие возможности по автоматизации рутинных действий по анализу траекторий. Весьма часто имеется набор файлов траекторий молекулярной динамики, которые обрабатываются похожим образом. Например, требуется восстановить траекторные файлы (из-за возникновения ошибок ввода/вывода при многократном использовании суперкомпьютеров), построить ряд скриншотов с расширенной динамической контактной карты, составить графики количества контактов и пр. Эта работу довольно просто выполнять в интерактивном режиме. Однако при большом количестве данных работа становится рутинной. Внимание научного сотрудника переводится с сути экспериментальных данных на индексы файлов и построение изображений. Достаточно небольших знаний программирования, чтобы существенно снизить трудоемкость рутинной работы за счет описания своих действия на скриптовом языке.

Скриптовый (или сценарный) язык (англ. scripting language, в русской литературе принято название язык сценариев) — высокоуровневый язык программирования для написания сценариев – кратких описаний действий, выполняемых системой. Разница между программами и сценариями довольно размыта. Сценарий — это программа, имеющая дело с готовыми программными компонентами.

TAMD-скрипт представляет собой текстовый файл, содержащий программу на языке Python. Для его запуска достаточно выполнить команду меню Service -> Run script... Анализатор передает в программу-скрипт глобальную переменную-объект TAMD, являющуюся экземпляром класса TAMDScript. Управление Анализатором сводится к вызову функций объекта TAMD. Помимо явного управления TAMD, в скрипте доступны все возможности языка программирования Python.

5.3 БАЗОВЫЕ КОМАНДЫ TAMD-SCRIPT

Непосредственное управление Анализатором осуществляется путем вызова методов объекта TAMD. На данный момент пользователю доступен следующий набор функций TAMD-Script (методов объекта TAMD).

Для загрузки траектории достаточно выполнить команду:

```
void Load(string Trj, string Pdb)
```

Загрузка траектории Trj и соответствующего ей pdb-файла с именем Pdb.

```
void ContactGraphC(double Min, double Max, int First, int Last)
```

Построить контактную карту только по C-alpha атомам с критерием контакта от Min до Max ($Min \leq Max$) по кадрам траектории с First по Last. По окончании работы функции откроется средства работы с графиками.

```
void SaveGraphToJpeg(string FileName)
```

Сохранить текущий график из средства работы с графиками в jpeg-файл FileName.

```
void SaveGraphToExcel(string FileName)
```

Сохранить текущий график из средства работы с графиками в таблицу Excel под именем FileName. Один график – это один столбец данных. Первая строка – заголовок согласно легенде.

```
void GotoFrame(int Frame)
```

Перейти к кадру траектории под номером Frame. Отсчет кадров траектории идет в нуля.

```
void SaveCMapToJpeg(string FileName)
```

Сохранить текущий кадр расширенной динамической контактной карты в jpeg-файл FileName.

```
int CalculateAtoms(string FileName)
```

Вычислить количество атомов в первом кадре траектории по имени FileName. Функция возвращает количество атомов.

```
void Repair(string FileName1, string FileName2, int aim, int reduce)
```

Восстановить траекторию из файла FileName1. Результат записать в файл FileName2. Записывать только первые aim атомов каждого reduce кадра траектории. Данная функция применяется после незапланированных остановок мд-экспериментов за вычислительных кластерах. Также функцией удобно убрать воду из траекторного файла, либо сжать траекторный файл, записывая, к примеру каждый 10-ый кадр (параметр reduce).

6. Методика работы с программой

6.3 Возможности TAMD-Script

6.3.1 Обработка всех файлов в директории

Основное применение TAMD-Script – однотипная обработка экспериментов. Для однотипной обработки всех файлов в директории необходимо получить список всех папок, а также список файлов интересующего типа. Например, это могут быть файлы траекторий *.trj.

Для получения списка файлов рекомендуется использовать функцию следующего вида:

```
def getsubs(dir):
    dirs = []
    files = []
    for dirname, dirnames, filenames in os.walk(dir):
        dirs.append(dirname)
        for subdirname in dirnames:
            dirs.append(subdirname)
        for filename in filenames:
            files.append(os.path.join(dirname, filename))
    return dirs, files
```

Эта функция возвращает список файлов (files) и папок (dirs).

Цикл перебора всех траекторных файлов может выглядеть следующим образом:

```
MainPath="e:\\MD\\Myoglogin\\"
subdirs, files=getsubs(MainPath)
for SubDir in subdirs:
    Path=MainPath+SubDir
    files = os.listdir(os.path.join(MainPath, SubDir))
    ListE = filter(lambda x: x.endswith('.trj'), files)
    Path=Path+"\\"
    j=0
    while j < len(ListE):
        file1=Path+ListE[j]
```

Здесь MainPath – основной путь, по которому предполагается обработка. File1 – полное имя текущего файла.

6.3.2 Расчет отклонений C-alpha-атомов вдоль всей траектории

В Анализаторе траекторий есть возможность получения отклонений C-alpha-атомов. Для этого необходимо загрузить траекторию (выбрав пункт Open меню File), затем воспользовавшись возможностью работы с функциями (меню View, пункт Fuctions).

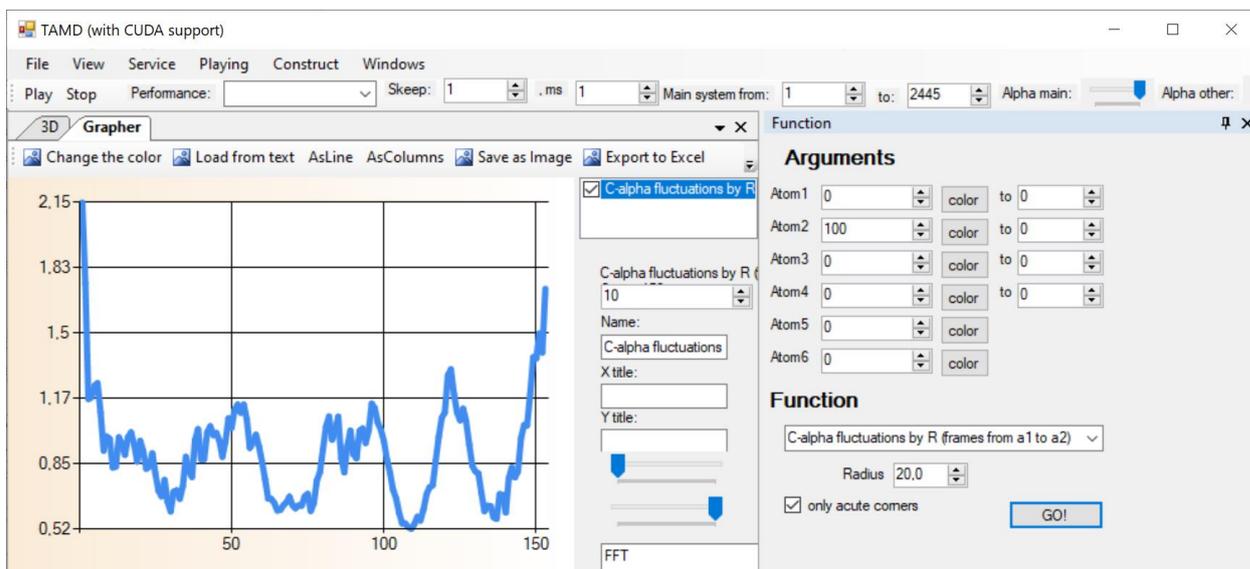


Рисунок 58. Расчет отклонений C-alpha-атомов между нулевым и 100-ым кадром траектории.

Как правило, подобный расчет необходимо проводить не единожды, а вдоль всей траектории с некоторым шагом.

Для построения отклонений C-alpha-атомов вдоль всей траектории можно воспользоваться следующим скриптом:

```
subdirs, files=getsubs(MainPath)
for SubDir in subdirs:
    Path=MainPath+SubDir
    files = os.listdir(os.path.join(MainPath, SubDir))
    ListE = filter(x.endswith('.trj'), files)
    Path=Path+"\"
    j=0
    while j < len(ListE):
        file1=Path+ListE[j]
        TAMM.Load(file1, Path+"1000000.pdb")
        for i in range(0, Size-100, 100):
            TAMM.GraphSimpleGraph(17, i, i+100, 0, 0, 0, 0)
            TAMM.GraphSetAutoAxis()
            TAMM.GraphSetXFormat("#.")
            TAMM.GraphSetXMin(0)
            TAMM.GraphSetYMin(0)
            TAMM.GraphSetToColumns()
            TAMM.SaveGraphToJpeg(Path+"TXT\\TXTd-
"+ListE[j]+str(i)+"-"+str(i+100)+"_Line.jpg")
            TAMM.SaveGraphToExcel(Path+"TXT\\TXT-
"+ListE[j]+str(i)+"-"+str(i+100)+".xls")
            TAMM.SaveGraphToCSV(Path+"TXT\\TXT-
"+ListE[j]+str(i)+"-"+str(i+100)+".csv")
            TAMM.ClearGraph()
        j+=1
```

Здесь используется функция получения списка файлов и каталогов getsubs.

Далее осуществляется обход всех траекторных файлов.

Для каждого файла строятся отклонения C-alpha-атомов при помощи функции TAMM.GraphSimpleGraph. В данном случае нас интересует 17-я функция по счету. Функция вызывается для 0-100 кадров, затем для 100-200 и т.д. до конца траектории.

После построения характеристики, график разумно сохранить. В данном примере данные сохраняются в графический формат функцией `TAMD.SaveGraphToJpeg`. Затем в Excel функцией `TAMD.SaveGraphToExcel`. А также в .csv-формат (comma separated values) функцией `TAMD.SaveGraphToCSV`.

Для использования возможности сохранения в Excel на компьютере должен быть установлен пакет MS Office. Возможность сохранения данных в формат .csv работает без дополнительных программ.

Для оформления графика используются функции:

- `TAMD.GraphSetAutoAxis()` – автоматический выбор масштаба для всех осей;
- `TAMD.GraphSetXFormat("#.")` – установка формата значений по оси X (0 знаков после запятой);
- `TAMD.GraphSetXMin(0)` – установка нуля в качестве минимального значения по оси X;
- `TAMD.GraphSetYMin(0)` – установка нуля в качестве минимального значения по оси X. Иначе график начинался бы, к примеру, с 0,52, как это показано на рисунке;
- `TAMD.GraphSetToColumns()` – выбор отображения графика в виде столбцов.

В конце цикла используется функция `TAMD.ClearGraph()` для очистки графика. В противном случае каждый следующий график накладывался бы на предыдущий. В итоге мы получаем папку TXT с набором графиков в форматах JPEG, .xls и .csv, составленных через каждый 100 шагов траектории, что составляет 1 нс вычислительного эксперимента.

6.3.3 Расчет матрицы отклонений C-alpha-атомов

В Анализаторе траекторий при помощи графического интерфейса есть возможность получения графика отклонений C-alpha-атомов между конкретными кадрами траектории. При помощи TAMD-Script можно получить последовательность таких графиков для всей траектории.

Получение матрицы отклонений C-alpha-атомов также возможно при помощи TAMD-Script. Достаточно лишь незначительно модифицировать программу из предыдущего параграфа:

```
subdirs, files=getsubs(MainPath)
for SubDir in subdirs:
    Path=MainPath+SubDir
    files = os.listdir(os.path.join(MainPath, SubDir))
    ListE = filter(x.endswith('.trj'), files)
    Path=Path+"\\"
    j=0
    while j < len(ListE):
        file1=Path+ListE[j]
        TAMD.Load(file1, Path+"10000000.pdb")
        for i in range(0, Size-100, 100):
            TAMD.GraphSimpleGraph(17, i, i+100, 0, 0, 0, 0)
            TAMD.GraphSetAutoAxis()
            TAMD.GraphSetXFormat("#.")
            TAMD.GraphSetXMin(0)
            TAMD.GraphSetYMin(0)
            TAMD.GraphSetToColumns()
```

```

TAMD.SaveGraphToJpeg(Path+"TXT\\TXTd-
"+ListE[j]+str(i)+"-"+str(i+100)+"_Line.jpg")
-----
TAMD.SaveGraphToExcel(Path+"TXT\\TXT-
"+ListE[j]+str(i)+"-"+str(i+100)+".xls")
-----
TAMD.SaveGraphToCSV(Path+"TXT\\TXT-
"+ListE[j]+str(i)+"-"+str(i+100)+".csv")
TAMD.ClearGraph()
TAMD.SaveGraphToCSV(Path+"TXT\\TXT-"+ListE[j]+".csv")
j+=1

```

Идея модификации программы состоит в том, что нам не нужно экспортировать график после каждого его построения. Также не нужно и очищать график. Таким образом мы получим целый набор графиков для одной траектории.

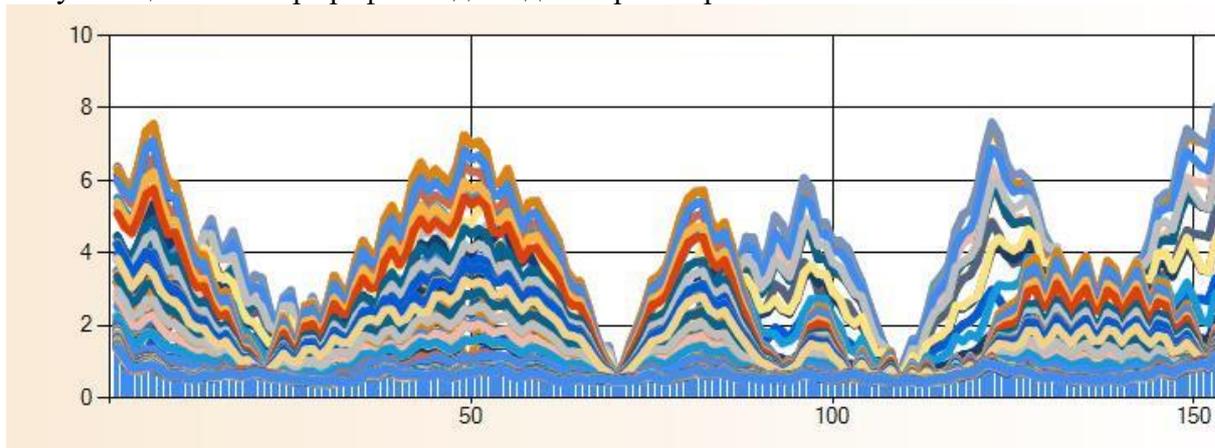


Рисунок 59. Набор графиков отклонений C-alpha-атомов для всей траектории, изображенных на одном рисунке.

При экспорте такого набора графиков в формат .csv получается матрица $N \times T$, где N - количество C-alpha-атомов (почти равно количеству аминокислотных остатков), T - количество пар сравниваемых кадров. Разумеется, к матрице добавляются заголовки названий графиков, а также отдельный столбец, соответствующий номеру аминокислотного остатка.

Количество C-alpha-атомов чуть меньше или равно количеству аминокислотных остатков ввиду того, что остатки меньше чем из 3 атомов не рассматриваются. Как правило, это отдельные ионы (Ca, Mg, Cl...) или начальные и конечные ионы H- и -OH.

6.3.4 Визуализация матрицы отклонений C-alpha-атомов средствами внешней программы

Анализатор траекторий экспортирует полученные характеристики, ровно как и матрицу отклонений C-alpha-атомов в формат .csv, для дальнейшего анализа и визуализации в сторонних программных продуктах.

Например, удобно использовать короткую программу на языке MATLAB для визуализации матрицы отклонений:

```

A=importdata("e:\MD-Mobil\100.csv");
A=A.data;
contourf(A, '.', 'linestyle', 'none');
xlabel("Time, [*100ps]");
ylabel("Residue");
colorbar;

```

Сначала программа считывает .csv-файл. Т.к. файл содержит заголовки столбцов, то сами данные сохраняются в матрице data структуры A.

Далее строится заполненный двумерный контурный график. У графика подписываются название осей, а также строится шкала, где показано соответствие между цветом и количеством контактов.

Изображение матрицы отклонений C-alpha-атомов можно увидеть на следующем рисунке:

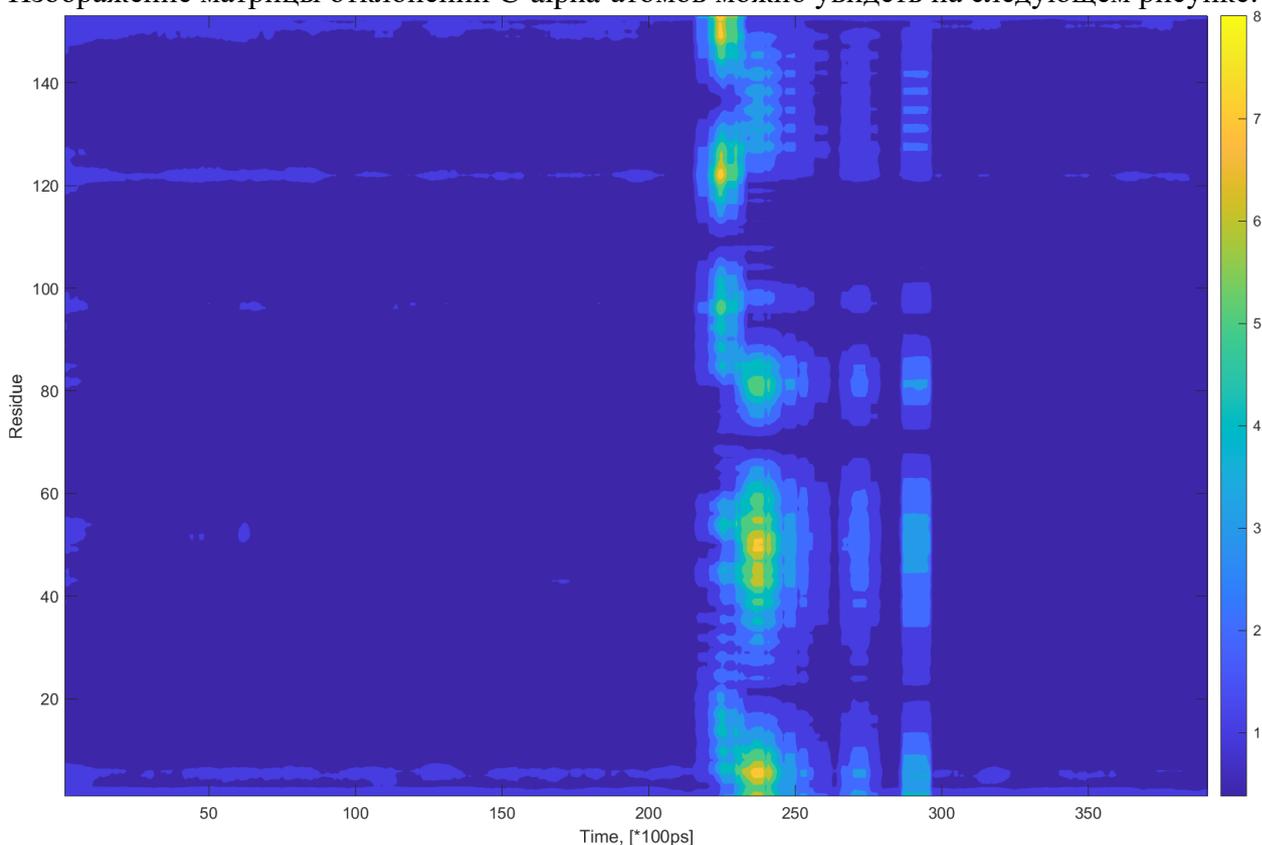


Рисунок 60. Изображение матрицы отклонений C-alpha-атомов.

Если сопоставлять набор графиков и изображение матрицы отклонений C-alpha-атомов, то видно, что максимумы графиков соответствуют более тёплым цветам на изображении матрицы отклонений.

6.3.5 Средство визуализации и интерактивного анализа матрицы отклонений C-alpha-атомов

Анализатор траекторий располагает собственным средством работы с двумерными картами – средством визуализации матриц.

Для того, чтобы рассчитать и визуализировать матрицу отклонений C-alpha-атомов, её необходимо построить. Делается это из пункта RMS by tensor of inertia меню View.

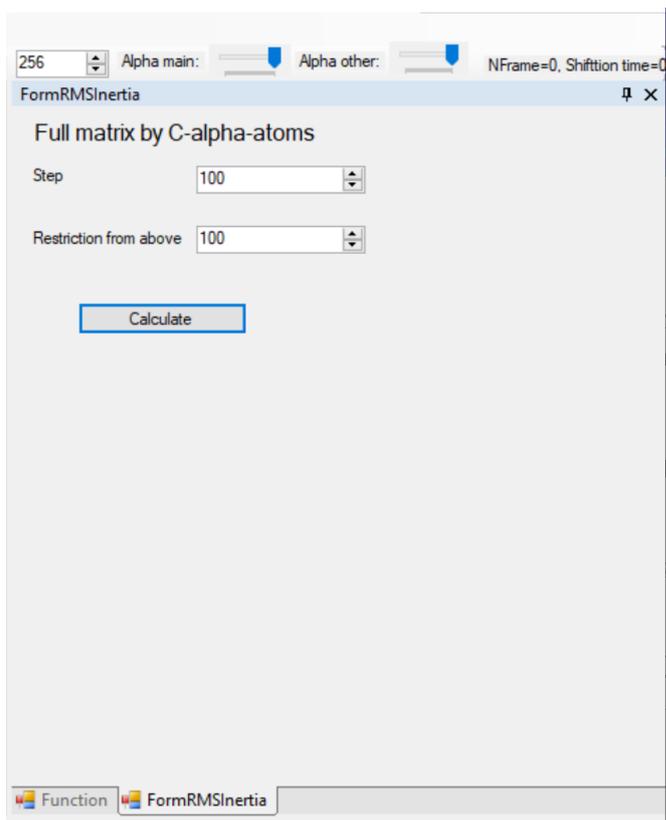


Рисунок 61. Окно построения матрицы отклонений C-Alpha-атомов по тензору инерции. Параметры имеют следующие значения:

- Step – ширина окна (в кадрах) определения матрицы флуктуаций;
- Restriction from above – ограничение флуктуации.

Если в траектории N кадров, то матрица флуктуаций будет содержать N-Step столбцов. Каждый i-ый столбец матрицы будет показывать смещения между кадрами i и i+Step. Для визуализации полученной матрицы флуктуаций можно воспользоваться пунктом Grapher Map меню View.

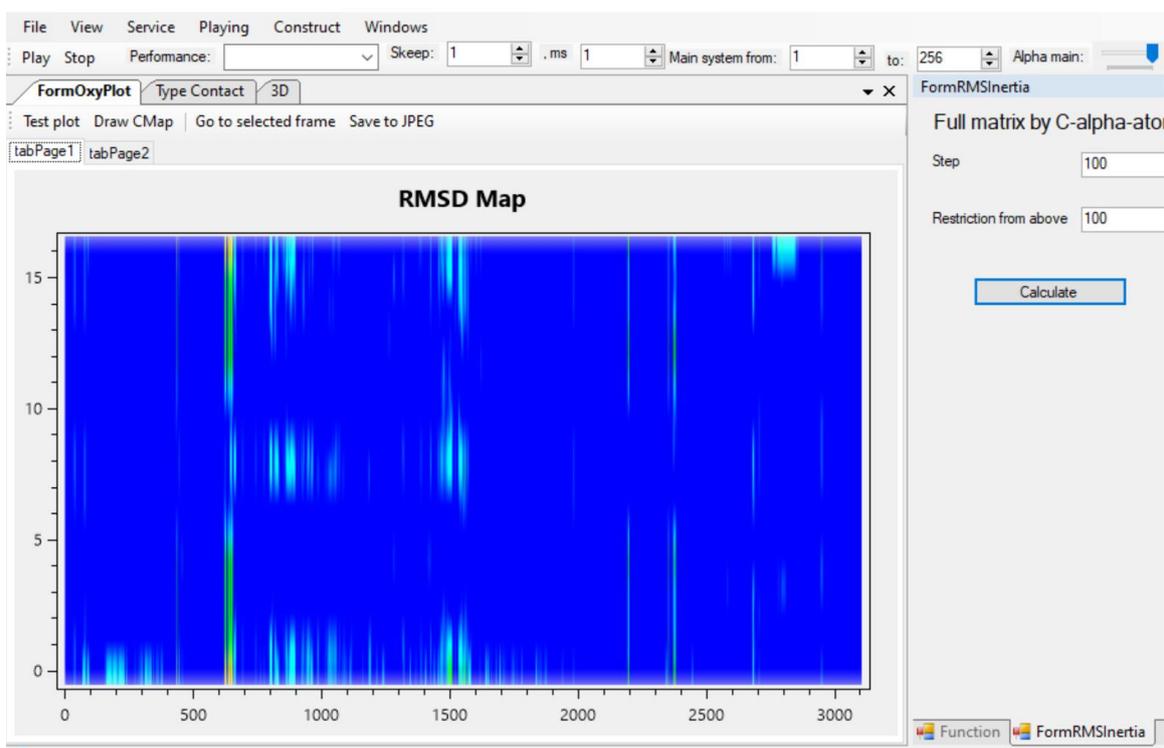


Рисунок 62. Средство визуализации матриц (слева) совместно со средством построения матрицы флуктуаций (справа).

Шкала абсцисс означает номер первого кадра (в паре i и $i+Step$), а шкала ординат означает отклонение.

Графическое изображение матрицы **интерактивно**:

- График можно масштабировать посредством вращения колёсика мыши.
- График можно перемещать посредством зажатой правой клавишей мыши. Особенно это полезно для увеличенных графиков.
- На графике можно выделять точку левой клавишей мыши для того, чтобы узнать её точные координаты, а также величину отклонения.
- Выделив точку, можно быстро перейти на соответствующий кадр траектории, воспользовавшись кнопкой *Go to selected frame*.
- Графическое изображение матрицы можно сохранить в графический файл формата JPEG, воспользовавшись кнопкой *Save to JPEG*.
- Можно также увидеть графическое изображение вектора-столбца (отклонением атомов между текущим кадром и кадром, отстоящим от текущего на *Step* кадров), перейдя на вторую вкладку средства визуализации матриц.

Интерактивное выделение точки на двумерном графике с масштабированием можно увидеть на следующей рисунке:

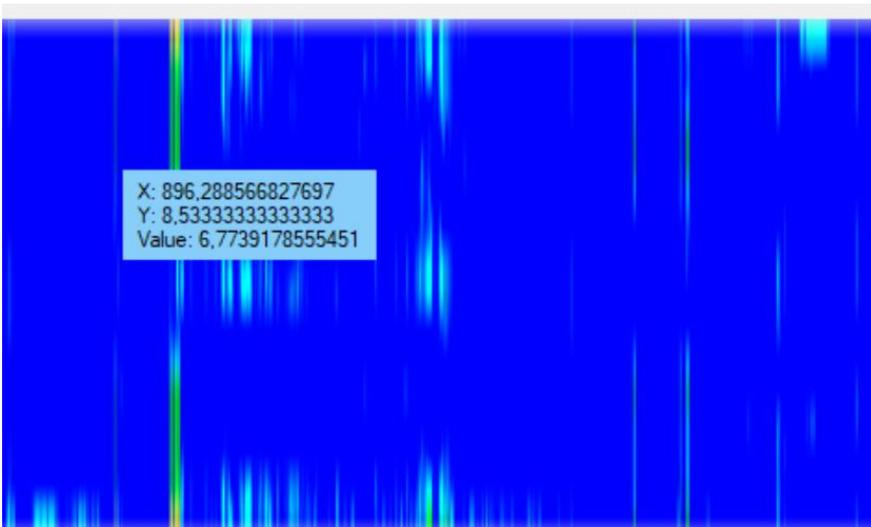


Рисунок 63. Средство визуализации с выделенной точкой.

Ниже представлено графическое изображение вектора-столбца, которой соответствует выделенной точке:

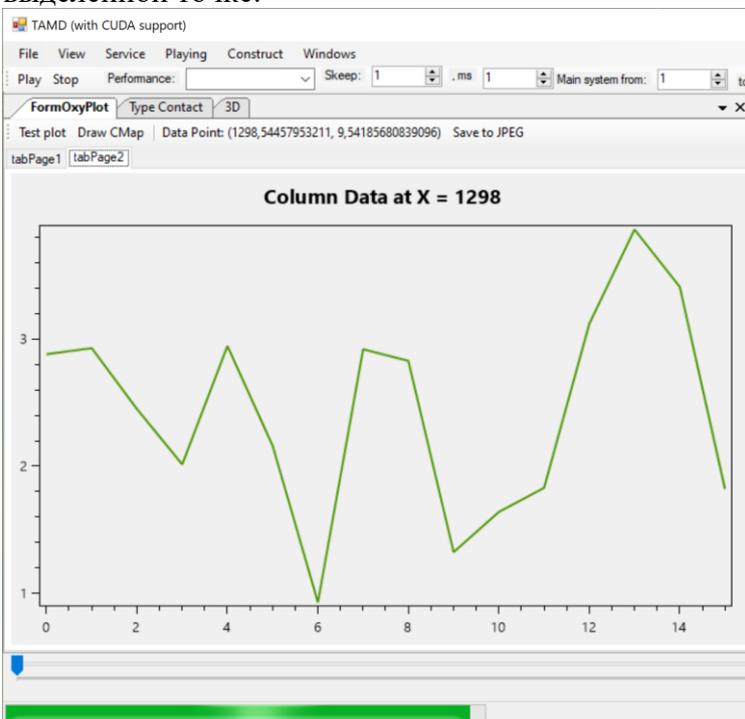


Рисунок 64. Средство визуализации матриц в режиме показа выбранного вектора.

6.3.5 Сценарий анализа лево- и правозакрученных альфаспиральных белков в опытах их разворачивания под действием постоянной силы, приложенной к концам

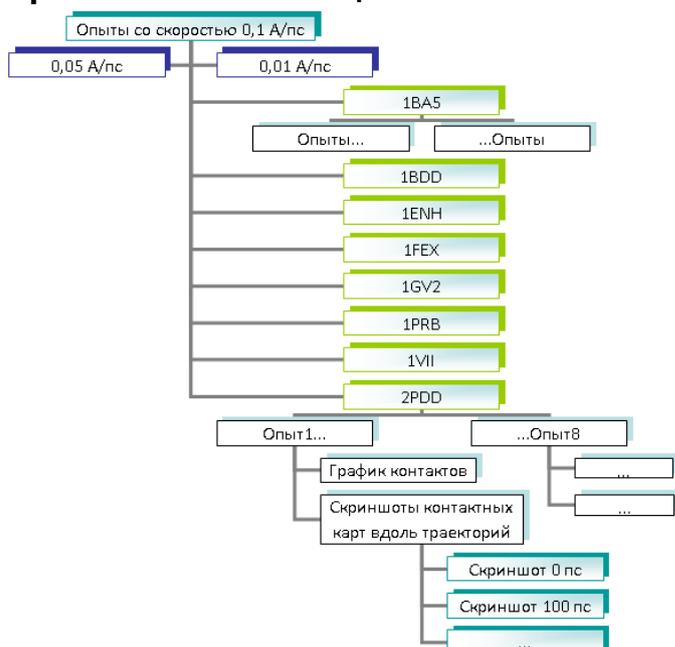


Рисунок 65. Сценарий анализа лево- и правозакрученных альфаспиральных белков в опытах их разворачивания под действием постоянной силы, приложенной к концам.

Для исследования механических характеристик лево- и правозакрученных белков были проведены эксперименты молекулярной динамики по разворачиванию белковых глобул под действием внешней силы, приложенной к концам молекулы, обеспечивающей постоянную скорость растяжения. Всего было исследовано 8 белков — 1BA5, 1BDD, 1ENH, 1GV2, 1PRB, 1VII и 2PDD [**Ошибка! Источник ссылки не найден.**]. Было выбрано по три значения скорости растяжения с каждым белком: 0,1, 0,05 и 0,01 А/пс. Каждый опыт по растяжению проводился по 8 раз. Итого: $8 \times 3 \times 8 = 132$ эксперимента молекулярной динамики или 132 траектории. См. рисунок 1.

После проведения мд-экспериментов на суперкомпьютере ИМПБ РАН необходимо выполнить анализ полученных данных. Для этого сначала следует восстановить поврежденные траектории, если повреждения имели место. Затем по каждой траектории построить график контактов вторичных структур, который совместно со скриншотами контактных карт в ключевых кадрах траектории позволит определить последовательность разворачивания вторичных структур.

Предполагается, что файлы расчетов молекулярной динамики находятся в каталоге Path. Первая траектория имеет имя 100g.trj, а имя соответствующего ей pdb-файла 10000000.pdb. Всего 8 траекторий.

Пример программы по восстановлению траекторий, содержащих исходное количество атомов и сжатию траектории в 10 раз:

```
Path="n:\..."
aim=TAMD.CalculateAtoms(Path+"\\100g.trj")
i=100
while i < 108:
    TAMD.Repair(Path+str(i)+"g.trj",Path+"!" +str(i)+"g.trj",aim,10)
    i += 1
```

Пример программы по сохранению заданных кадров расширенной динамической контактной карты:

```
# Переход к 0 кадру
TAMD.GotoFrame(0)
# Сохранение контактной карты в файл
TAMD.SaveCMapToJpeg(FirstPath+"001.jpg")
# Сохраняем остальные кадры
TAMD.GotoFrame(50)
TAMD.SaveCMapToJpeg(FirstPath+"050.jpg")
TAMD.GotoFrame(100)
TAMD.SaveCMapToJpeg(FirstPath+"100.jpg")
TAMD.GotoFrame(150)
TAMD.SaveCMapToJpeg(FirstPath+"150.jpg")
```

Пример программы по построению графиков контактов по C-alpha атомам:

```
Path="n:\..."
i=100
while i < 108:
    # Загружаем траекторию:
    TAMD.Load(Path+"\"+str(i)+"g.trj", Path+"\\10000000.pdb")
    # Строим график количества контактов
    TAMD.ContactGraphC(6,6,0,500)
    # Сохраняем график в файл
    TAMD.SaveGraphToJpeg(Path+"CG"+"-"+str(i-100)+".jpg")
    i += 1
```

РЕЗУЛЬТАТЫ И ВЫВОДЫ

Результатом выполнения данной программы-скрипта будет каталог графических файлов, соответствующий заданному набору траекторий молекулярной динамики. Файлы графиков контактов начинаются с префикса CG (contact graph), содержат номер эксперимента и оканчиваются на расширение .jpg. Снимки контактных карт начинаются с префикса, задаваемого в переменной FirstPath, содержат временной маркер, соответствующий номеру кадра траектории, и заканчиваются на расширение .jpg.

Благодаря проведению анализа траекторий молекулярной динамики опытов разворачивания 8 трехспиральных белков установлено, что правозакрученные трехспиральные белки обладают большей механической стабильностью по сравнению с левозакрученными образцами.

Применение сценариев анализа траекторий позволяет резко снизить трудоёмкость работы по анализу проведенных экспериментов. Сценарии полностью избавляют пользователя Анализатора траекторий молекулярной динамики от рутинных действий при работе с большими объемами информации.

Приложение 1. Минимальные требования к системе

ОС: Windows XP/Vista/7/8. 32/64-битные версии.

Процессор: не ниже Pentium-IV.

Оперативная память: Windows XP 1 GB, остальные ОС 2 GB.

Дисковая подсистема: 50 MB + пространство для анализируемых данных.

Видеокарта: совместимая с OpenGL.

Дополнительно: .NET Framework 4 (поставляется совместно с программой).

Приложение 2. Установка программы

Дистрибутив представляет собой обычный инсталлятор windows-приложения. При отсутствии библиотеку .NET Framework 4 можно скачать как по требованию программы-установщика (требуется подключение к интернет), так и отдельного исполняемого файла.

Программе требуются права на запись файла store.mdb, расположенного папке с установленной программой. Для некорпоративных пользователей, обычно, легче запускать программу от имени администратора.