

УДК: 577.32

Построение расширенных динамических контактных карт по данным молекулярно-динамических расчетов

Лихачев И.В*., Балабаев Н.К.

Институт математических проблем биологии, Российская академия наук, Пущино, Московская область, 142290, Россия

Аннотация. В работе представлены алгоритмы построения расширенных контактных карт и лент контактов для анализа динамики молекулярных систем. Проведение такого анализа возможно на базе программного комплекса TAMD – анализатора траекторий молекулярной динамики.

Ключевые слова: контактные карты, анализ траекторий молекулярной динамики, молекулярная динамика, анализ контактов белковых структур

ВВЕДЕНИЕ

Настоящая работа посвящена одному из способов анализа молекулярных структур – составлению и обработке контактных карт.

Новизна работы заключается в предоставлении возможности одновременного анализа как статических, так и динамических структур. Представлен алгоритм построения единой матрицы по всей траектории молекулярно-динамической системы, так называемой ленты контактов. Данная структура способна быстро и наглядно изобразить динамические процессы, происходящие в системе.

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. Плоское изображение является одним из преимуществ контактных карт. За счет этого их информативность крайне высока.

В статье предлагается новый подход к построению контактных карт, пригодный, в том числе, и для динамических систем. Описываются быстрые интерактивные методы интерпретации таких данных.

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 .

* ilya_lihachev@mail.ru

Выбор параметра r_{max} зависит от рода анализируемых контактов. В примерах исследуются траектории белковых молекул. Было выбрано характерное расстояние 8 Å 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 из 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.

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 в системе RGB каналов (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 цветовых каналов в равной пропорции, согласно формулам (2), дает контактную карту в градациях серого цвета. Полученную контактную карту назовем *монохромной* (рис. 1).

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

$$\begin{aligned}
 r_{i,j} &= \text{const1}, \\
 g_{i,j} &= \begin{cases} 1, & \text{если } R(i, j) \leq r_{\min}, \\ 1 - \frac{(R(i, j) - r_{\min})}{(r_{\max} - r_{\min})}, & \text{если } r_{\min} < R(i, j) \leq r_{\max}, \\ 0, & \text{если } R(i, j) > r_{\max}, \end{cases} \quad (3) \\
 b_{i,j} &= \text{const2}.
 \end{aligned}$$

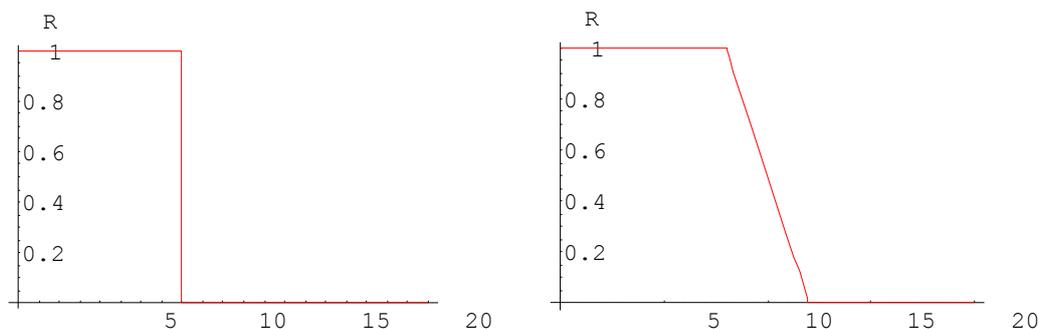


Рис. 1. Сравнение булевой (слева) и монохромной (справа) контактных карт.

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.

3. РАСШИРЕНИЕ ВОЗМОЖНОСТЕЙ КОНТАКТНЫХ КАРТ С ПОМОЩЬЮ ПОСТРОЕНИЯ ПОЛНОЦВЕТНЫХ КОНТАКТНЫХ КАРТ

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. ДИНАМИКА. ЛЕНТЫ КОНТАКТОВ

Рассмотрим динамическую молекулярную систему. Обычно математические эксперименты над такими системами записываются в виде траекторий – записи координат атомов с заданной периодичностью на устройстве хранения информации.

Развитие во времени расширенных и монохромных контактных карт можно представить в виде анимации. Несомненно, такой подход является весьма полезным и во многих ситуациях незаменимым. Но он годится лишь для вывода на экран. Кроме того, исследователю приходится держать в зрительной памяти весь фильм. Поставим задачу получить информацию о динамике контактов вдоль траектории.

Предложим следующий алгоритм построения результирующего рисунка, в дальнейшем называемого «лентами контактов».

Одна лента – это информация о контакте двух структур. Лента представляется цветной полоской, максимальная яркость которой определяется побитовым сложением предопределенных цветов двух структур. На этот раз разбивки на каналы не будет.

Лента контакта подобна киноленте. Пусть она будет иметь фиксированную ширину кадра и высоту. Кадр представляет собой узкий залитый прямоугольник. Яркость цвета пропорциональна сумме элементов матрицы контактов внутри прямоугольника, отвечающего за контакт между требуемыми элементами структуры.

Количество лент зависит от квадрата количества структур. Ленты контактов вида ij и ji являются полностью совпадающими. Нет необходимости строить две одинаковые ленты.

5. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

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

Изложенный теоретический материал о контактных картах реализован в программно-алгоритмическом комплексе «Анализатор траекторий молекулярной динамики» (TAMD) [1]. 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);
- построение ленты контактов.

[†] Программный комплекс ПУМА для моделирования молекулярной динамики полимеров и биополимеров. – ИМПБ РАН, техническая документация.

6. ПРИМЕР ИСПОЛЬЗОВАНИЯ

Методику исследования динамической белковой структуры при помощи динамических расширенных контактных карт и лент контактов рассмотрим на примере V1 домена иммуноглобулина связывающего белка L (в дальнейшем 2PTL) [2]. 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 двух бета-шпилек, одной альфа-спирали и хвоста (рис. 2).

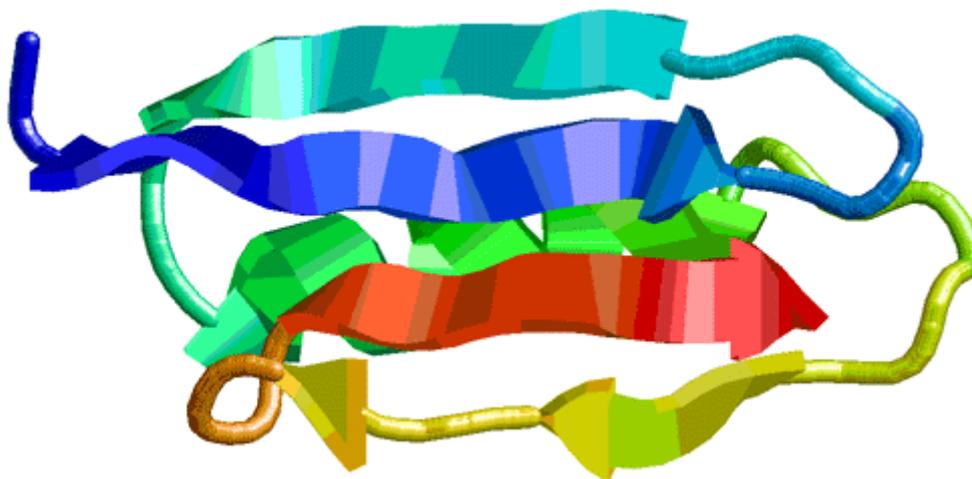


Рис. 2. Изображение структуры белка 2ptl.

6.1. Методика проведения исследования. Интерпретация результатов

Сначала построим монохромную контактную карту белковой глобулы до начала растяжения.

Для построения карты мы использовали систему координат, в которой элемент матрицы $a_{1,1}$ расположен в левом нижнем углу окна (рис. 3).

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

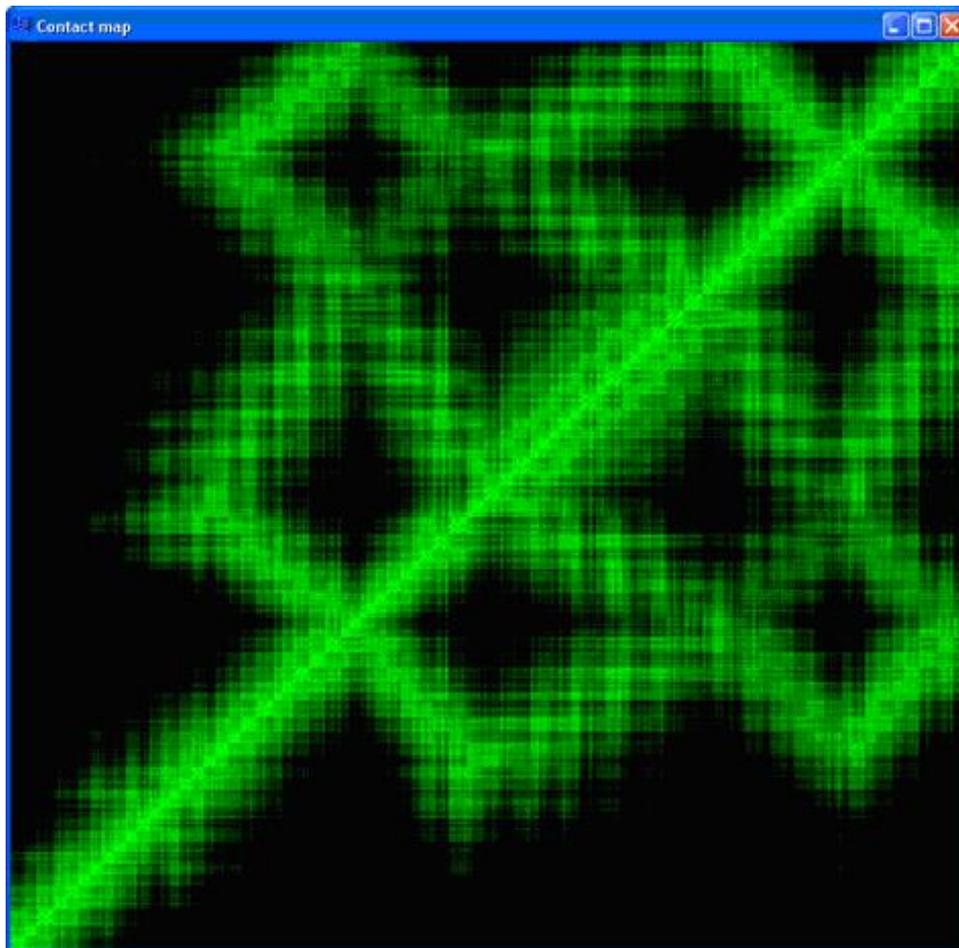


Рис. 3. Контактная карта белка. Начало эксперимента.

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.

6.2. Нанесение информации о структуре молекулы на карту

Известно, что белковая глобула состоит из двух бета-спилек, одной альфа-спирали и хвоста. Занесем известную нам информацию в программу. Типичный вид окна программы с занесенной структурой приведен на рис. 4.

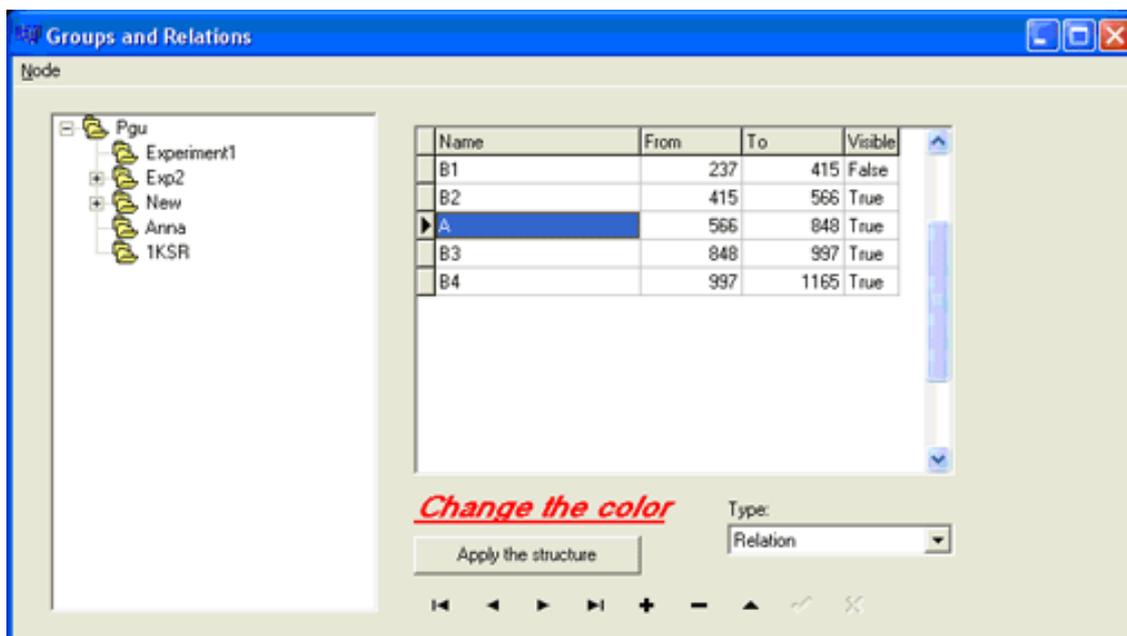


Рис. 4. Окно создания и выбора структуры программы TAMD.

После применения структуры получим расширенную контактную карту, представленную на рис. 5.

В левом верхнем углу прямоугольных областей обозначаются анализируемые структуры.

Наличие четкой линии в закрашенном фрагменте говорит о наличии контакта между структурами макромолекулы. Если закрашенная область пустая, то это означает отсутствие контактов по заданному критерию. Выбор критерия, что считать контактом, может быть индивидуальным для каждой задачи.

7. ДИНАМИКА

Как следует 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 (рис. 6).

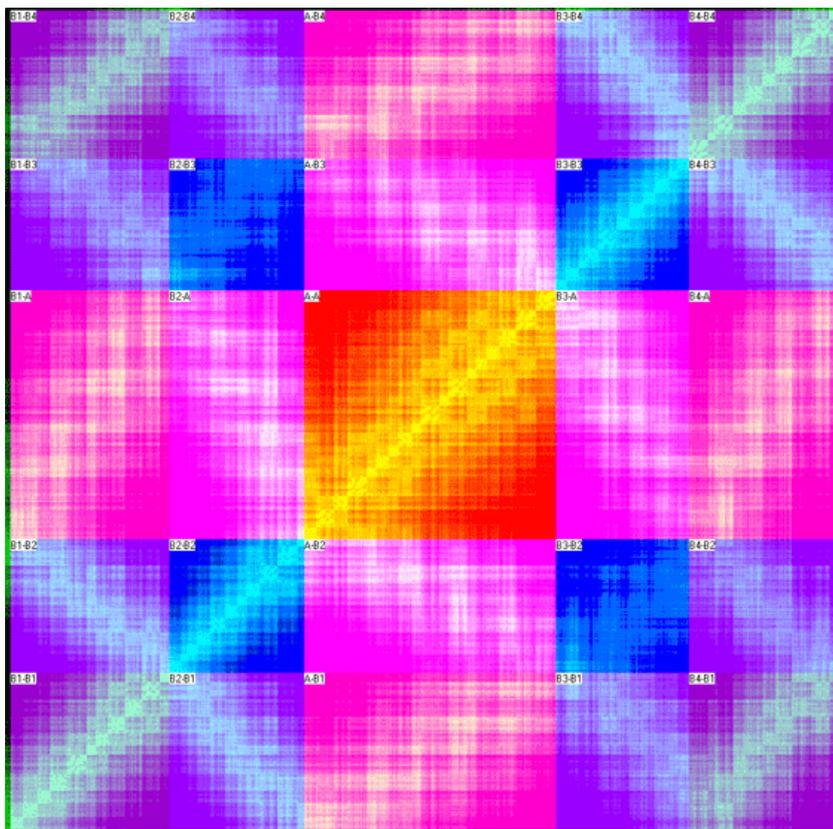


Рис. 5. Фрагмент контактной карты с обозначением структуры.

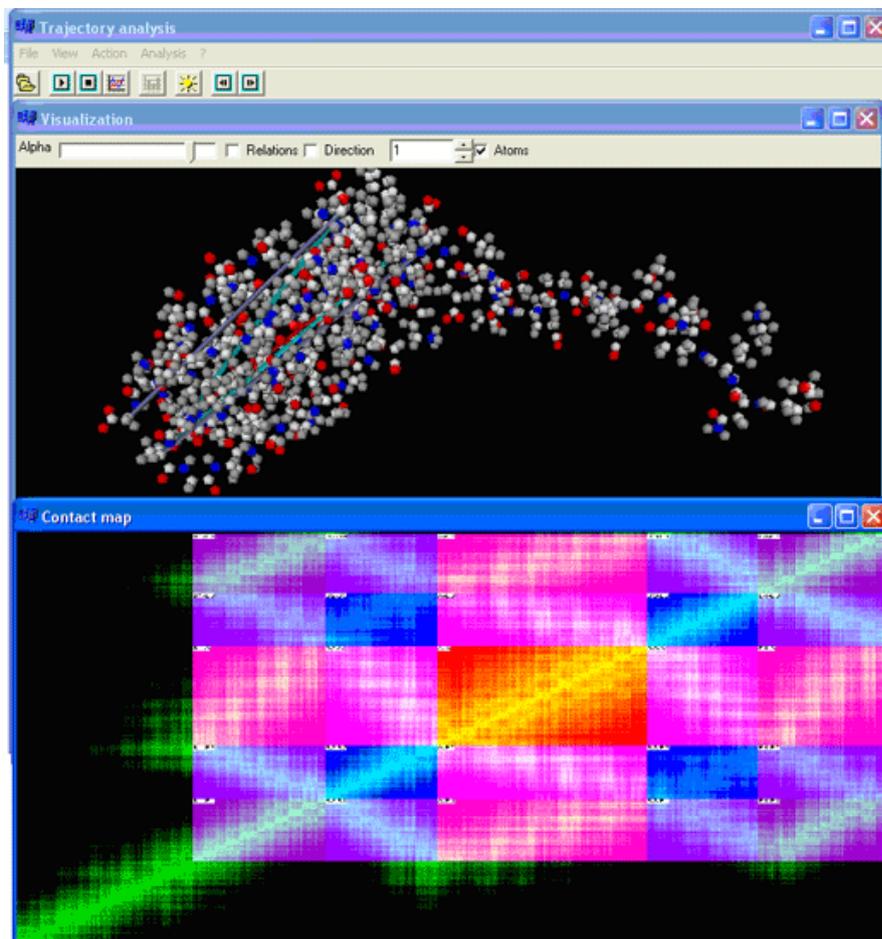


Рис. 6. Режим работы TAMM при одновременном построении перспективной проекции и контактной карты.

К сожалению, в статье нельзя показывать анимации. Ограничимся лишь главным окном программы с кадром трехмерной модели молекулы и контактной картой, соответствующей данному кадру.

Это изображение является динамически развивающимся. Приходится просматривать весь фильм с регулируемой скоростью, если речь идет о траекториях.

Перейдем к возможности построения ленты контактов. Траектория белка L представляется в виде ленты контактов (рис. 7).

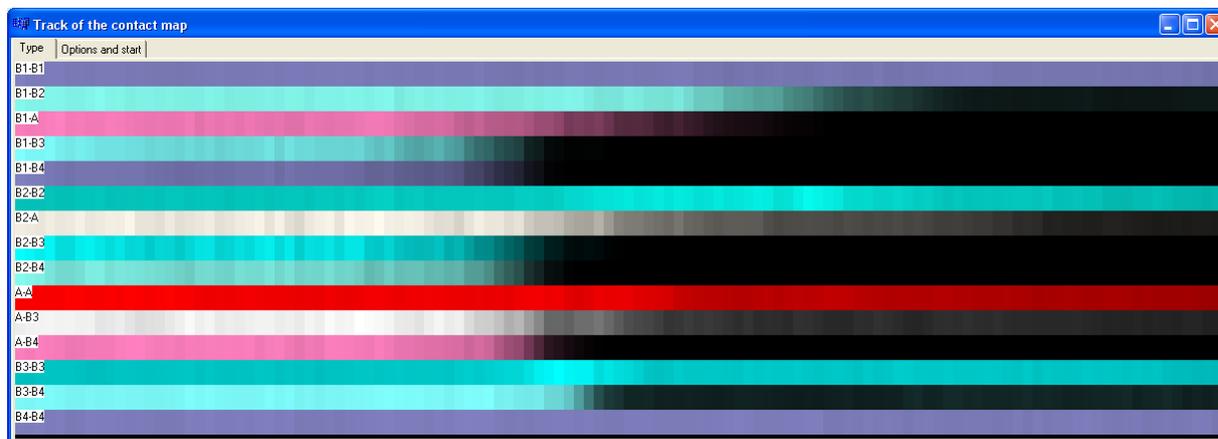


Рис. 7. Ленты контактов, построенные TAMD.

Возвращаясь к задаче о разворачивании белковой глобулы под действием постоянных сил, приложенных к ее концам, мы переходим к быстрой интерпретации данных, полученных на ленте контактов. По резкой границе смены интенсивности цвета на конкретной ленте мы мгновенно даем ответ на вопрос, когда и в какой последовательности распадаются белковые структуры. По лентам типа ii (отвечающих за одну структуру) мы видим, как быстро растягивается конкретная структура. Как и ожидалось, сами структуры начинают интенсивно растягиваться только после разрыва контактов с соседними структурами, бета-листами и альфа-спиралью.

ЗАКЛЮЧЕНИЕ

Контактные карты были созданы именно благодаря простоте построения и анализа. Но при бурном развитии вычислительных мощностей они не получали должного развития. Взамен пришли трехмерные динамические изображения молекулярных систем, так называемое молекулярное кино.

В данной работе авторы представляют алгоритм построения новых контактных карт, которые задействуют возможности современных устройств вывода графической информации, делая тем самым процесс анализа более наглядным и лучше воспринимаемым. Помимо того, это является мощным средством анализа статических и динамических структур.

Построение ленты контактов не требует больших вычислительных ресурсов. В то же время, информация, расположенная на лентах, говорит сразу обо всей траектории. Вместе с тем она достаточно информативна. Ленты четко показывают динамику развития контактов во времени между конкретными структурами молекулярной системы, представляя общую картину происходящих в системе процессов. Для получения более детальной информации имеется возможность построения дополнительных характеристик траектории, анализ которых заложен в разрабатываемый программный комплекс TAMD.

СПИСОК ЛИТЕРАТУРЫ

1. И.В. Лихачев, Н.К. Балабаев. Анализатор траекторий молекулярной динамики. *Математическая биология и биоинформатика*. 2007. Т. 2. № 1. С. 120–129. URL: [http://www.matbio.org/downloads/Likhachev2007\(2_120\).pdf](http://www.matbio.org/downloads/Likhachev2007(2_120).pdf)
2. Глякина А.В., Галзитская О.В., Балабаев Н.К. Молекулярная динамика разворачивания белковых глобул под действием внешних сил. *Вестник Тверского государственного университета*. 2008. Т. 68. № 8. С. 109–119.

Материал поступил в редакцию 06.02.2009, опубликован 11.03.2009