

P10-2012-138

И. В. Бедняков¹, П. В. Зрелов¹, В. В. Иванов¹,
Р. В. Полозов^{1,2}, В. С. Сивожелезов³,
В. А. Степаненко¹, Ю. Н. Чиргадзе⁴

**КАРТОГРАФИРОВАНИЕ СТРУКТУР БЕЛКОВ
И НУКЛЕИНОВЫХ КИСЛОТ**

Направлено в журнал «Письма в ЭЧАЯ»

¹ Объединенный институт ядерных исследований, Дубна

² Институт теоретической и экспериментальной биофизики РАН,
Пушино

³ Институт биофизики клетки РАН, Пушино

⁴ Институт белка РАН, Пушино

Бедняков И. В. и др.

P10-2012-138

Картографирование структур белков и нуклеиновых кислот

Развиты методы и алгоритмы для исследования поверхностей глобулярных и фибриллярных белков, ДНК, РНК. Данные методы построения карт фрагментов этих объектов в разработанной нами оригинальной цилиндрической проекции существенно расширяют возможности изучения распределения зарядов и рельефа поверхности биологических структур. Этот подход в значительной мере дополняет качественные характеристики методов визуализации структур биополимеров.

Работа выполнена в Лаборатории информационных технологий ОИЯИ.

Препринт Объединенного института ядерных исследований. Дубна, 2012

Bednyakov I. V. et al.

P10-2012-138

Mapping Structures of Proteins and Nucleic Acids

Methods and algorithms for analysis of surfaces of globular and fibrillar proteins, DNA, and RNA have been developed. These methods of constructing cards of fragments of these objects, in the cylindrical projection original developed herein, essentially expand possibilities for studying distribution of charges and a relief of a surface of biological structures. This approach essentially supplements qualitative characteristics of methods of visualization of biopolymer structures.

The investigation has been performed at the Laboratory of Information Technologies, JINR.

Preprint of the Joint Institute for Nuclear Research. Dubna, 2012

ВВЕДЕНИЕ

В настоящее время исследования структуры белков и нуклеиновых кислот тесно связаны с вопросами расширения возможностей современной медицины и генетики. Любые исследования такого рода основаны на результатах рентгеноструктурного анализа кристаллов биополимеров на уровне отдельных атомов, и они требуют применения самой современной аппаратуры и методов [1]. Быстрорастущие объемы компьютерных банков данных [2] структур биомолекул привели к созданию новой области науки — биоинформатики. В настоящее время данные о расшифрованных белках, а также ДНК и РНК хранятся в специальном PDB-формате [3], где для каждого атома указаны параметры (номер цепочки, код нуклеотида, код аминокислоты и пространственные координаты центра атома). В банке данных белков (Protein Data Bank) и нуклеиновых кислот (Nucleic Acid Database) в настоящее время имеется около 100000 структур, и эти банки постоянно пополняются. Написан ряд программ, которые используются в повседневной практике при определении структуры биологических молекул и их визуализации. Мы разработали методы и алгоритмы для построения различных карт, отображающих рельеф поверхности белков и нуклеиновых кислот в распределенной вычислительной среде.

МЕТОДЫ КАРТОГРАФИРОВАНИЯ БЕЛКОВЫХ СТРУКТУР

Кратко изложим стандартную процедуру определения ориентации биополимера или его фрагмента, так как при картографировании требуется задание поворота относительно некоторого стандартного положения. PDB-файл фрагмента белковой структуры представляет собой упорядоченное множество N центров атомов, заданных в прямоугольной декартовой системе координат XYZ с метками, характеризующими свойства каждого атома фрагмента. Для построения карты рельефа поверхности биополимера относительно некоторой плоскости или центра масс требуется повернуть биополимер или его фрагмент в начальное стандартное положение и произвести масштабирование изображения на карте с учетом вандерваальсовых радиусов атомов [4, 5].

Так как общее число атомов N может быть достаточно большим, для вычисления координат центра тяжести обычно используют некоторое подмножество одинаковых атомов n , характеризующих конфигурацию биополимера или его фрагмента с достаточной точностью. Формулы для расчета координат центра масс C_x, C_y, C_z фрагмента имеют следующий вид:

$$C_x = \frac{1}{n} \sum_{i=1}^n x_i, \quad C_y = \frac{1}{n} \sum_{i=1}^n y_i, \quad C_z = \frac{1}{n} \sum_{i=1}^n z_i,$$

где n — число атомов подмножества; x, y, z — координаты атомов в начальной системе отсчета.

Для начальной ориентации фрагмента в пространстве требуется вычислить моменты инерции относительно каждой из осей новых декартовых прямоугольных координат, проходящих через центр масс. По определению, моментом инерции J_u биополимера относительно оси u называют выражение

$$J_u = \sum_{i=1}^n m_i r_i^2,$$

где m_i — масса атома фрагмента биомолекулы, r_i — расстояние этой точки от оси u .

Момент инерции характеризует распределение массы биополимера относительно оси и требует довольно больших вычислений. В новой системе координат выражение момента инерции представляется в виде

$$J_u = J_x \cos^2 \alpha + J_y \cos^2 \beta + J_z \cos^2 \gamma - 2J_{xy} \cos \alpha \cos \beta - 2J_{xz} \cos \alpha \cos \gamma - 2J_{yz} \cos \beta \cos \gamma, \quad (1)$$

где

$$\begin{aligned} J_x &= \sum m_i (y_i^2 + z_i^2), & J_{xy} &= \sum m_i x_i y_i, \\ J_y &= \sum m_i (x_i^2 + z_i^2), & J_{xz} &= \sum m_i x_i z_i, \\ J_z &= \sum m_i (x_i^2 + y_i^2), & J_{yz} &= \sum m_i y_i z_i \end{aligned}$$

— осевые и центробежные моменты инерции тела. В матричной форме записи:

$$J_u = \mathbf{e}^T \mathbf{J} \mathbf{e},$$

где \mathbf{e} — орт оси u , заданный в указанной системе координат своими направляющими косинусами $(\cos \alpha, \cos \beta, \cos \gamma)$, а \mathbf{e}^T — транспонированный вектор этих значений.

Матрица

$$J = \begin{vmatrix} J_x & -J_{xy} & -J_{xz} \\ -J_{xy} & J_y & -J_{yz} \\ -J_{xz} & -J_{yz} & J_z \end{vmatrix}$$

определяет тензор второго ранга — тензор инерции тела.

Изменяя направление оси u и, следовательно, изменяя направляющие косинусы в выбранной системе координат, по известным компонентам тензора инерции можно найти момент инерции для любой оси, проходящей через центр масс атомов фрагмента O . Понятие момента инерции допускает геометрическую интерпретацию распределения моментов инерции относительно осей пучка направлений с центром в этой точке.

Возьмем на оси u точку $M(x, y, z)$ на расстоянии, равном $K/\sqrt{J_u}$ от точки O (K — размерный коэффициент). Тогда координаты этой точки можно вычислить следующим образом:

$$x = \left(K/\sqrt{J_u}\right) \cos \alpha, \quad y = \left(K/\sqrt{J_u}\right) \cos \beta, \quad z = \left(K/\sqrt{J_u}\right) \cos \gamma.$$

Подставляя полученные из этих равенств значения $\cos \alpha$, $\cos \beta$, $\cos \gamma$ в выражение (1), получим уравнение поверхности 2-го порядка

$$J_x x^2 + J_y y^2 + J_z z^2 - 2J_{xy}xy - 2J_{xz}xz - 2J_{yz}yz = K^2. \quad (2)$$

Эта поверхность является геометрическим местом точек M при всевозможных направлениях оси u . Поскольку момент инерции относительно произвольной оси величина отличная от нуля, то поверхность не имеет точек на бесконечности и, значит, является эллипсоидом. Этот эллипсоид называют эллипсоидом инерции. Каждой точке тела соответствует свой эллипсоид инерции.

Выбор осей симметрии эллипсоида инерции (рис. 1) в качестве координатных осей позволяет привести уравнение эллипсоида инерции (2) к канонической форме. Оси ξ , η , ζ симметрии эллипсоида инерции, построенного в некоторой точке, — это главные оси инерции. В системе координат ξ, η, ζ для этой точки выражение момента инерции J_u приобретает вид

$$J_u = J_\xi \cos^2 \alpha + J_\eta \cos^2 \beta + J_\zeta \cos^2 \gamma.$$

Таким образом, распределение масс тела характеризуется связанными между собой понятиями тензора инерции тела и

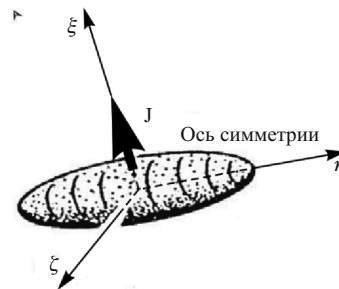


Рис. 1. Эллипсоид инерции

эллипсоида инерции. В главных осях тензор инерции имеет матрицу диагонального вида. Так как направлениям главных осей инерции соответствуют оси симметрии эллипсоида инерции, а значит и стационарные значения моментов инерции, то задача нахождения направлений главных осей сводится к рассмотрению экстремума выражения (2) при условии

$$f(\cos \alpha, \cos \beta, \cos \gamma) = \cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma - 1 = 0, \quad (3)$$

или, что то же, к рассмотрению необходимого условия экстремума функции

$$\phi(\cos \alpha, \cos \beta, \cos \gamma) = J_u(\cos \alpha, \cos \beta, \cos \gamma) - \lambda f(\cos \alpha, \cos \beta, \cos \gamma),$$

где λ — неопределенный множитель (метод Лагранжа). Условие стационарности функции $\phi(\cos \alpha, \cos \beta, \cos \gamma)$ дает

$$\frac{\partial \phi}{\partial X_i} = \frac{\partial J_u}{\partial X_i} - \lambda \frac{\partial f}{\partial X_i} = 0 \quad (i = \overline{1, 3}), \quad (4)$$

где $X_1 = \cos \alpha$, $X_2 = \cos \beta$, $X_3 = \cos \gamma$.

Если корни уравнения (4) различны, то, подставляя значения λ в систему (3), найдем для каждого λ_i свой набор $\cos \alpha$, $\cos \beta$, $\cos \gamma$, который определяет соответствующее направление главной оси в выбранной системе координат XYZ . Известно, что если уравнение (4) имеет два равных корня, то два главных момента инерции равны между собой, а эллипсоид инерции является эллипсоидом вращения. В случае равенства всех корней уравнения (4) эллипсоид инерции превращается в сферу.

Если биополимер или его фрагмент имеет ось материальной симметрии, то эта ось является главной центральной осью, и в таком случае при наличии плоскости материальной симметрии для любой точки этой плоскости одна из главных осей совпадает с перпендикуляром, восстановленным к плоскости симметрии в данной точке.

Для вычисления направляющих применен прямой метод Якоби, который позволяет одновременно определять все собственные значения и собственные векторы. Исходя из этого, можно сказать, что данный метод представляет собой численное преобразование к главным осям потому, что собственные векторы X_j являются непосредственно вектор-столбцами матрицы преобразования X , а собственные значения λ_j — это диагональные элементы диагональной матрицы D в преобразовании $X^T A X = D$.

Используется метод вращений Якоби для нахождения собственных значений матрицы:

$$T_{pq} = \begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 1 & 0 & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \cos \varphi & \cdot & \cdot & \cdot & \sin \varphi & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 & 1 & \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 & \cdot & \cdot & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & -\sin \varphi & \cdot & \cdot & \cdot & \cos \varphi & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & 1 \end{pmatrix}$$

$\begin{array}{cc} | & | \\ p\text{-й} & q\text{-й} \\ \text{столбец} & \text{столбец} \end{array}$

$\begin{array}{l} -p\text{-я} \text{ ячейка} \\ -q\text{-я} \text{ ячейка} \end{array}$

Для вычисления матрицы преобразования X мы использовали метод итераций:

- 1) выбор элемента $a_{pq}^{(k)}$ из матрицы $A^{(k)}$;
- 2) вычисление $A^{(k+1)} = T_{pq}^T A^{(k)} T_{pq}$, $k = 0, 1, \dots$;
- 3) конец вычислений при $\sum_{1, i \neq j}^n |a_{ij}^{(k+1)}|^2 < \varepsilon$, в противном случае увеличение k на 1 и переход к п.1.

В результате вычислений получаем матрицу, в которой с точностью $\varepsilon > 0$ диагональные элементы являются приближениями собственных значений, а столбцы произведения всех переменных T_{pq} — приближениями собственных векторов.

В найденной системе пространственных прямоугольных декартовых координат мы можем выбрать ось Z совпадающей с максимальным значением момента инерции (для горизонтального расположения молекулы) или совпадающей с минимальным значением момента инерции (для вертикального расположения молекулы).

Второй вариант больше подходит для исследования протяженных спиралевидных структур, и при этом оси Z и Y как бы меняются местами, а ось X соответствует среднему значению момента инерции эллипсоида. Таким образом, мы получаем правую декартову систему координат; ось направлена к наблюдателю, ось Y — влево, а ось Z — вверх. Масштабирование осуществляется по оси Z таким образом, чтобы биополимер или его фрагмент были видны наилучшим образом.

Вокруг центра тяжести фрагмента строится сфера с радиусом $R = 1$. Поверхность сферы покрывается равномерной сеткой с заданным числом узлов K в зодиакальном направлении и $2K$ в азимутальном. Узлы сетки на поверхности сферы соответствуют азимутальному и зодиакальному углам сферических координат для заданного атома следующим образом:

- φ — угол между радиусом-вектором атома и осью X в плоскости XY ;
- ψ — угол между радиусом-вектором атома и полярной осью Z .

Этот подход позволяет сравнительно просто получать карту рельефа для глобулярных белков посредством построения проекции из центра координат вдоль радиуса-вектора в заданный узел координатной сетки на сфере. Если при этом в узел попадают точки поверхностей нескольких атомов, то выбирается наиболее удаленная из них. Таким образом, мы получаем таблицу значений для каждого узла сетки, которая в дальнейшем используется для построения изолиний рельефа и границ атомов на карте. Сформированное на сфере отображение поверхности биополимера или его фрагмента является промежуточным, базовым. Для получения карт молекулярных поверхностей можно использовать либо проекцию на плоскость, либо цилиндрическую развертку, как будет описано дальше.

Для глобулярных белков первоначально была использована проекция поверхности сферы на плоскость по методу Аитова–Хаммера [8], который часто используется в астрономии. Координаты проекции связаны с углами φ и ψ по формулам

$$x = 2m \sin \psi \sin \frac{\varphi}{2}, \quad y = m \cos \frac{\psi}{2},$$

где

$$m = 1 / \sqrt{1 + \sin \psi \cos \frac{\varphi}{2}}, \quad -\pi \leq \varphi \leq \pi, \quad -\frac{\pi}{2} \leq \psi \leq \frac{\pi}{2}.$$

При этом предполагается, что наблюдатель находится в центре сферы и в центральной части карты, ограниченной координатными метками Z и Y , видит изображение перед собой, а справа и слева от этой области на карте, соответственно, проецируется изображение задней полусферы. В результате получается равновеликая эллиптическая проекция, характеризующаяся значительными нелинейными искажениями (рис. 2).

Искажения по краям центральной полосы карты можно частично компенсировать вращением карты вокруг оси Z , но вблизи полюсов значительные нелинейные искажения остаются в любом случае. Кроме того, атомы, находящиеся вблизи центра тяжести биополимера или его фрагмента, «загораживаются» более удаленными атомами молекулы (рис. 3).

В силу указанных причин в такой проекции практически невозможно использовать шкалу расстояний. По меткам можно измерить расстояние только в пространстве между выбранными центрами видимой пары атомов биополимера или его фрагмента. Для диалоговой обработки изображения на карте

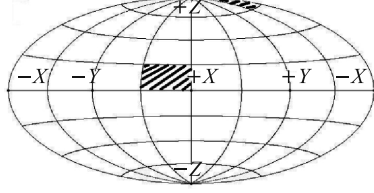


Рис. 2. Общий вид равновеликой эллиптической проекции Аитова–Хаммера (заштрихованы участки карты равной площади на поверхности сферы)

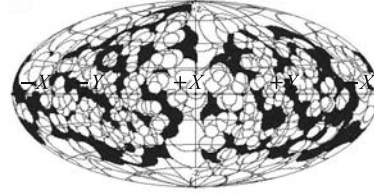


Рис. 3. Пример карты атомов глобулярного белка с функциональной раскраской атомов азота в проекции Аитова–Хаммера

(нанесения текстовых меток, выделения или функциональной раскраски атомов) переход в сферическую систему координат производится по формулам в предположении, что радиус сферы $R = 1$:

$$R = \sqrt{x^2 + y^2 + z^2}, \quad \varphi = \arctan(y/x), \quad \phi = \arctan(\sqrt{x^2 + y^2}/z).$$

Проецирование точек поверхности атомов на сетку сферы единичного радиуса существенно упрощает вычислительные формулы при изменении вида карты.

Для исследования вытянутых фибриллярных структур проекция Аитова–Хаммера мало подходит. По этой причине нами был разработан метод построения карты в виде ортогональных сечений вдоль оси Z с проецированием на сферу точек поверхности атомов в радиальной плоскости. Посредством предложенного псевдоконформного преобразования SC (Sphere–Cylinder) удалось «развернуть» поверхность сферы на прямоугольную поверхность цилиндра по формулам

$$x = x_0 + x_0 \sin \frac{\varphi}{4} \operatorname{sign}(\sin \psi) \sqrt{2}, \quad y = y_0 - y_0 \cos \psi / (1 + \frac{|\sin \psi|}{2}), \quad (5)$$

где x_0, y_0 — декартовы координаты центра карты. Вводя поправку на расстояние до оси Z , мы получаем карту атомов белка с существенно меньшими искажениями и со шкалой по оси Z в ангстремах. Соответственно, была введена поправка, зависящая от расстояния до оси цилиндра:

$$\cos \psi = Z_i \frac{S}{R_S}, \quad \sin \varphi = \frac{y}{r}.$$

где i — номер соответствующего узла на сфере в азимутальном направлении, S — коэффициент масштабирования, R_S — радиус сферы, r — радиальное расстояние атома до оси Z , Z_i — координата заданного сечения.

Вандерваальсовы радиусы атомов R_a были изменены следующим образом:

$$R_a = R_a \cdot r \frac{S}{R_S}.$$

Обратный перевод из системы декартовых координат карты в полярную систему координат сферы производится по формулам

$$\varphi = 4 \arctan \left(\frac{\bar{x}}{4} (\sqrt{8 - \bar{x}^2}) \right), \quad \phi = \arcsin (\bar{y} \sqrt{2 - \bar{y}^2}) + \pi/2, \quad (6)$$

где \bar{x} и \bar{y} соответственно равны

$$\bar{x} = 2 \frac{x - x_0}{x_0}, \quad \bar{y} = \frac{y_0 - y}{y_0}.$$

Если в результате этих вычислений угол $\varphi \leq 0$, то увеличиваем его на 2π .

Методика картографирования глобулярных белков. Глобулярные белки представляют собой сплетение цепочек атомов в виде клубка. На первом этапе по координатам С-альфа атомов углерода находится положение центра масс и вычисляются главные моменты инерции. Координаты всех атомов фрагмента белка пересчитываются в собственную систему координат, оси которой совпадают с главными моментами инерции. Обычно в качестве входных параметров задают имя PDB-файла и число разбиений области картографической проекции, которое определяет разрешение карты.

На втором этапе в этой системе координат для каждой точки карты вычисляется расстояние вдоль радиуса-вектора из центра инерции для построения карты рельефа остатков или атомов. Если радиус-вектор пересекает поверхности сфер нескольких атомов, то в этом случае берется максимально удаленная точка.

Для устранения неопределенностей, связанных с разрывом поверхности, состоящей из атомов конечного размера, при расчете вводится добавка R_{add} к радиусам всех атомов модели. Этот прием приводит к сглаживанию деталей поверхности модели белка.

Рельеф поверхности описывается изолиниями, которые соответствуют точкам поверхности, находящимся на равном удалении от центра масс молекулы. Для более гладкого изображения этих линий рекомендуется задавать более частое покрытие сферы узлами координатной сетки, так как изолинии строятся по принципу поиска ближайшей равноудаленной от оси Z точки в двух ортогональных направлениях. Для быстрой оценки точности этой аппроксимации используется интегрирование методом Симпсона с автоматическим выбором шага. Данный алгоритм хорошо описывает рельеф поверхности молекулы, форма которой может быть задана эллипсоидом с небольшим соотношением осей. Для белков сложной формы рекомендуется исследовать поверхность каждого домена по отдельности. Заметные искажения

поверхности на проекции наблюдаются вблизи оси Z при значениях широты от 70° до 90° (рис. 4).

Могут быть построены три основных типа карт: карта рельефа, карта остатков, карта атомов. На картах второго и третьего типа выводятся границы остатков или атомов, находящиеся на поверхности молекулы. Соответственно, эти карты являются картами функциональной раскраски по желанию пользователя. Карты указанных типов можно наложить друг на друга.

Недостатком данного подхода является зависимость размера атомов от расстояния до центра тяжести фрагмента белка.

Методика картографирования фибриллярных белков. При исследовании структуры фибриллярных белков проекция Аитова–Хаммера малопримодна, да и сам фрагмент удобно представить вдоль оси Z . Для построения цилиндрической проекции был разработан алгоритм построения радиальных сечений, ортогональных оси Z (рис. 5).

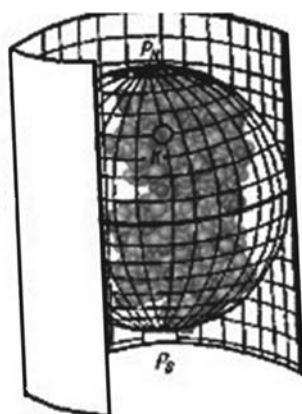


Рис. 5. Схема построения цилиндрической проекции

раках вокруг центров атомов, максимально удаленных от оси Z , для каждого полярного вектора ρ . В случае отсутствия атомов в заданном направлении расстояние $L = -99$ (фон). Для остальных точек на азимутальной оси с постоянным шагом Δz в цикле проводятся аналогичные вычисления.

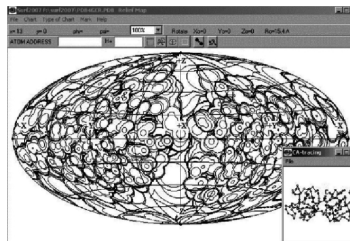


Рис. 4. Пример наложения карты схемы С-альфа цепочки атомов на карту рельефа поверхности того же глобулярного белка с двумя ядрами в проекции Аитова–Хаммера

Данный алгоритм хорошо описывает рельеф поверхности вытянутого фрагмента молекулы. Как уже отмечалось, SC-преобразование имеет незначительные искажения по краям карты (рис. 6) в сравнении с известными в картографии цилиндрическими проекциями (например, для проекции Меркатора) [8]. При этом на карте по оси Z мы имеем шкалу расстояний, а по оси Y — углы направления φ в плоскости сечения.

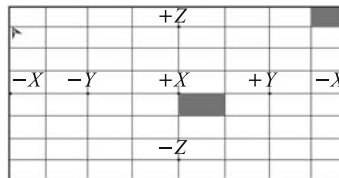


Рис. 6. Результат SC-преобразования системы координат сферы на боковую поверхность цилиндра (затемнены равновеликие на поверхности сферы области)

Методика картографирования ДНК. При разработке алгоритма для построения карт ДНК требовалось иначе выбирать остов фрагмента ДНК, изменить функциональную раскраску нуклеотидов, изменить общий вид карты. Это представляет интерес для детального изучения распределения зарядов атомов на поверхности большой бороздки ДНК [10], к которым обычно присоединяются белки. Для построения карт в SC-цилиндрической проекции использовали уравнения (5) и (6).

Так как в ДНК и РНК имеются совсем небольшие различия в длине нуклеотидов и вместо тимина в ДНК мы имеем урацил в РНК, то для картографирования спиральных РНК не потребовалось разрабатывать отдельной программы. В качестве эталона брали одновитковый фрагмент В-ДНК, в котором на уровне PDB-файла оставляли одну или две цепочки. Важно отметить возможность включения в фрагмент макромолекулы одновременно белка и ДНК, РНК (рис. 7). Рельеф будет строиться для всех выбранных атомов, а функциональную раскраску атомов белка в этом случае необходимо задавать вручную.

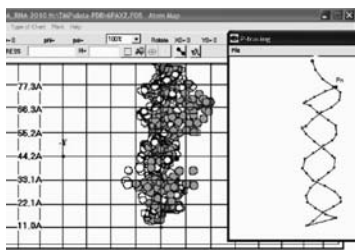
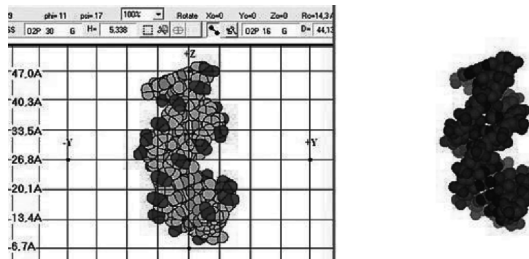


Рис. 7. Пример картографирования фрагмента ДНК с белком

Попытки одновременного проецирования атомов из выбранной точки на оси Z в прямом (от оси Z) и обратном направлении (извне) для каждого радиального среза молекулы не увенчались успехом. Во-первых, алгоритм упорядочивания атомов белка на карте в проекции Аитова–Хаммера ранее был запрограммирован на работу только с положительными расстояниями до оси Z , а во-вторых, при малых расстояниях атомов относительно оси симметрии цилиндрической проекции наблюдалось удвоение контуров отдельных атомов по причине потери точности вычислений, а эту область и требовалось исследовать. Использование аппарата двойной точности было неприемлемо, так как это еще в 2–3 раза сокращало допустимое разрешение карт. Наложение атомов из задней полусферы

Рис. 8. Сравнение результатов картографирования фрагмента ДНК с результатами 3D-моделирования с помощью программы PyMol [11]



на изображения атомов в центральной области карты после дополнительного масштабирования и ортогонального проецирования на плоскость ZOY позволило ввести равные шкалы расстояний в плоскости карты. При этом рельеф поверхности ДНК определялся на основании вычисления расстояний точек атомов до плоскости ZOY , а не по радиусу-вектору до центра масс атомов, как это было в случае белков.

Насколько удачно это было реализовано, можно судить по рис. 8, где для сравнения выведены изображения независимого 3D-моделирования (справа) и остов фрагмента этой молекулы (слева).

На этом рисунке хорошо видны большая и малая бороздки ДНК. При этом практически полностью сохранились все функциональные возможности картографирования и можно проводить следующие типы анализа фрагмента:

- 1) анализ особенностей рельефа и определение формирующих его атомных групп;
- 2) поиск кластеров, содержащих атомные группы определенного типа, например, заряженные;
- 3) анализ областей связывания спиральных молекул ДНК и РНК с лигандами и белками;
- 4) сравнительный анализ структуры поверхности мутантных по нуклеотидному составу спиральных молекул ДНК и РНК;
- 5) анализ поверхности, например, локализацию функционально важных участков последовательности, поиск возможных мест связывания данной молекулы с другими молекулами и многое другое.

Важно, что при этом была обеспечена возможность задания начальных углов поворота фрагмента макромолекулы относительно каждой из осей декартовых координат в пространстве.

ЗАКЛЮЧЕНИЕ

Развитые в работе методы картографирования расширяют возможности для исследования и визуализации [6, 12].

В процессе разработки рассмотренных методов нами были решены следующие важные задачи:

- Развита новая методика построения цилиндрической системы координат для визуализации атомов белковых структур с минимальными искажениями при сохранении интерфейса программ картографирования, основанного на использовании промежуточной сферической системы координат.

- На основе этого подхода разработана методика картографирования фибриллярных белков, ДНК и РНК с введением шкалы измерений в ангстремах; устранена зависимость визуализации атомов от расстояния до центра масс.

- Для исследования белков и нуклеиновых кислот создано некоторое подобие «супермикроскопа» с возможностью построения рельефа на фоне функциональной раскраски атомов с равномерной шкалой в горизонтальном и вертикальном направлениях карты (в ангстремах).

Разработанные алгоритмы картографирования и соответствующее программное обеспечение могут использоваться для детального исследования новых белков и их комплексов, при разработке нового класса лекарств и решении некоторых задач генной инженерии.

Работа поддержана грантом РФФИ 11-07-00374.

ЛИТЕРАТУРА

1. Альбертс Б. и др. Молекулярная биология клетки. В 3-х т. М.: Мир, 1994.
2. <http://www.rcsb.org/pdb/statistics/holdings.do>
3. Westbrook J., Fitzgerald P. M. The PDB format, mmCIF formats and other data formats. Structural Bioinformatics. P. E. Bourne and H. Weissig. Hoboken, New York: John Wiley & Sons, Inc., 2003. P. 161–179.
4. Barlow D. J., Thornton J. M. Interactive Map Projection Algorithm for Illustrating Protein Surfaces // J. Molec. Graphics. 1986. V. 4. P. 97–100.
5. Bondi A. Van der Waals Volumes and Radii // J. Phys. Chem. 1964. V. 68. P. 441–451.
6. Афанасьев О. А. и др. Комплекс программ для картографирования и исследования белков и нуклеиновых кислот. Сообщение ОИЯИ Р10-2011-108. Дубна, 2011.
7. Лаврентьев М. А., Шабат Б. В. Методы теории функций комплексного переменного. 4-е изд., доп и испр. М.: Наука, 1973.
8. Porter W. Introduction to Map Projections. New York: Marcel Dekker, 1979.
9. Мецержков Г. А. Теоретические основы математической картографии. М.: Недра, 1968.
10. Чиргадзе Ю. Н., Ларионова Е. А. Определяющая роль кластеров полярных остатков в структурах белковых факторов при узнавании большой бороздки двухспиральной В-ДНК // Молекулярная биология. 2003. Т. 37, вып. 2. С. 266–276.
11. <http://www.pymol.org/>
12. Afanasiev O. A. et al. Software Complex for Computing Surface Maps of Helical Biopolymer Molecule Proteins and Nucleic Acids // Int. Conf. «Mathematical Modeling and Computational Physics», Dubna, July 7–11, 2009, Book of Abstracts. P. 171.

Получено 18 декабря 2012 г.

Редактор *М. И. Зарубина*

Подписано в печать 03.04.2013.

Формат 60 × 90/16. Бумага офсетная. Печать офсетная.

Усл. печ. л. 0,94. Уч.-изд. л. 1,11. Тираж 225 экз. Заказ № 57953.

Издательский отдел Объединенного института ядерных исследований
141980, г. Дубна, Московская обл., ул. Жолио-Кюри, 6.

E-mail: publish@jinr.ru

www.jinr.ru/publish/