

# ГЕОМЕТРИЯ ЛОМАННЫХ И ПОЛИПЕПТИДОВ

А.О. Иванов, А.С. Мищенко, А.А. Тужилин

*Московский государственный университет имени М.В. Ломоносова,  
Московский государственный технический университет им. Н.Э.Баумана*

aoiva@mech.math.msu.su, asmish@mech.math.msu.su, tuz@mech.math.msu.su

Поступила 29.04.2014

В работе показано, как конформацию полипептидов можно изучать по двум функциям дискретным аналогам кривизны и кручения пространственных кривых. Переход к таким функциям не только дает наглядную кодировку пространственной структуры, но и позволяет автоматизировать выделение различных структурных элементов конформации, таких, например, как спирали.

УДК 514.8, 51-76, 57.087

## Введение

Цель данной работы — продемонстрировать возможности приложений дифференциальной геометрии и ее дискретных аналогов к изучению полипептидов. Хорошо известно, что химические свойства полипептидов зависят от их пространственной конфигурации (конформации), причем нарушение последней может существенно повлиять на эти свойства. Изменение конформации белков клеток живого организма иногда является причиной тяжелых болезней (например, коровьего бешенства).

Одной из наиболее важных задач молекулярной биологии клетки является изучение связи между первичной структурой белка (последовательностью составляю-

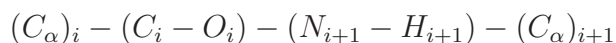
щих белок аминокислот) и его третичной структурой (пространственным расположением). Кроме того, интересна также и задача описания различных геометрических структур, встречающихся в тех или иных местах пептидного остова (вторичная структура). Традиционно выделяют разного рода спирали и слои.

Изучению пространственного строения биологических полимеров посвящено огромное количество работ. Вот некоторые из них [1]–[9], имеющие то или иное отношение к нашему исследованию.

## 1. Построение пространственной ломаной по полипептиду

В качестве первого шага, молекуле полипептида сопоставляется некоторая пространственная ломаная. Обычно для этой цели используется так называемый пептидный остов (backbone). Напомним, что полипептид — это цепочка последовательно соединенных между собой пептидов (аминокислот). Каждый пептид состоит из так называемого альфа-углерода  $C_\alpha$ , к четырем валентным связям которого крепятся аминогруппа  $NH_3$  (“азотистый конец”), карбоксильная группа  $COOH$ , содержащая еще один атом углерода  $C$  (“углеродный конец”), водород  $H$  и так называемый аминокислотный остаток (или радикал)  $R$ . Разные аминокислоты различаются именно радикалами. Два пептида соединяются друг с другом в бипептид так: аминогруппа одного взаимодействует с карбоксильной группой другого, выделяется вода и возникает так называемая пептидная связь между углеродом и азотом. У полученного бипептида снова есть “азотистый конец” и “углеродный конец”, поэтому процесс может продолжаться. Цепочка атомов вида  $NC_\alpha C - NC_\alpha C - \dots - NC_\alpha C$ , где черточки соответствуют пептидным связям, и образует пептидный остов белка. В качестве вершин ломаной  $L$  берутся или последовательные атомы пептидного остова, или только атомы  $C_\alpha$  альфа-углеродов последовательных пептидов.

**Замечание.** И тот и другой выбор имеет свои преимущества. Химические связи вдоль пептидного остова считаются наиболее сильными и устойчивыми, расстояния между соседними атомами — почти постоянными. Кроме того, считается установленным, что шестерки атомов



(здесь индексы  $i$  и  $i + 1$  обозначают номер пептида в цепочке, а атомы кислорода и водорода крепятся к стоящим с ними в одних скобках атомам углерода и азота из пептидного остова соответственно), расположены в одной двумерной плоскости. Эта шестерка атомов формирует так называемую пептидную группу  $P_{ii+1}$ , которая образуется за счет возникновения пептидной связи между  $C_i$  и  $N_{i+1}$ .

Итак, последовательные аминокислоты стыкуются друг с другом по пептидным связям, образуя на каждом таком стыке (почти плоскую) пептидную группу. Последовательные пептидные группы стыкуются по общим альфа-углеродам. При этом, плоскость пептидной группы  $P_{i+1i+2}$  можно получить из плоскости группы  $P_{ii+1}$  двумя последовательными поворотами вокруг оси связей  $N_{i+1} - (C_\alpha)_{i+1}$  и  $(C_\alpha)_{i+1} - C_{i+1}$ . Углы этих поворотов называются торсионными углами. Принято считать, что именно за счет изменения торсионных углов меняется геометрия белковой молекулы. Таким образом, локальное устройство ломаной, составленной из звеньев пептидного остова, известно лучше и представляется более жестким. С другой стороны, эта ломаная имеет в 4 раза больше вершин и несет повторяющуюся информацию об более-менее

одинаковых плоских пептидных группах, что может усложнить анализ геометрии ломаной в целом.

**Замечание.** Если внимательно посмотреть на приведенное выше описание аминокислоты, то становится ясным, что ее пространственное строение не определено однозначно: направление обхода  $HNC$  вокруг связи  $RC_\alpha$  может происходить как в одну, так и в другую сторону. Действительно, в природе встречаются и те, и другие (так называемые левые и правые) разновидности одних и тех же аминокислот. Любопытный (и, насколько мы знаем, до сих пор никак не объясненный) факт состоит в том, что в полипептидах один из двух типов аминокислот (левые) встречается гораздо чаще, чем другой.

## 2. Геометрия пространственных ломаных и автоматическое выделение спиралей

Рассмотрим произвольную *геометрическую реализацию* полипептида, т.е. некоторое конкретное расположение его в пространстве. Построим ломаную  $L$ , выбрав в качестве ее вершин последовательные альфа-углероды полипептида. Что нужно знать про ломаную, чтобы по имеющимся данным она восстанавливалась бы однозначно с точностью до сохраняющего ориентацию движения пространства? Ответ более менее очевиден: нужно знать длины ребер, углы поворота между парами последовательных ребер, а также углы поворота между последовательными парами плоскостей, первая из которых натянута на  $(i-1)$ -ое и  $i$ -ое ребра, а вторая — на  $i$ -ое и  $(i+1)$ -ое.

Определим описанными только что углы более строго. Пусть пространственная ломаная  $L = A_0A_1 \cdots A_n$  задана последовательностью своих *вершин*  $A_i$ . Ребро  $e_i$  ломаной  $L$  — это пара  $A_{i-1}A_i$ . Нам будет удобно представлять ребра  $e_i$  как векторы  $A_i - A_{i-1}$ . Мы будем всегда предполагать, что последовательные векторы  $e_i, e_{i+1}$  неколлинеарны. Пусть  $\alpha_i$  — величина угла между векторами  $e_i, e_{i+1}$ , т.е.

$$\alpha_i = \arccos \frac{\langle e_i, e_{i+1} \rangle}{\|e_i\| \cdot \|e_{i+1}\|},$$

где  $\langle v, w \rangle$  — стандартное скалярное произведение векторов  $v$  и  $w$ , а  $\|v\|$  — соответствующая норма вектора, т.е.  $\|v\| = \sqrt{\langle v, v \rangle}$ . Из сделанного предположения вытекает, что  $0 < \alpha_i < \pi$ .

Рассмотрим теперь произвольное *внутреннее* (т.е. неконцевое) ребро  $e_i$  ломаной  $L$ , тогда имеются соседние с ним ребра  $e_{i-1}$  и  $e_{i+1}$ . Мы хотим определить величину  $\beta_i$  угла между плоскостями, первая из которых натянута на ребра  $e_{i-1}, e_i$ , а вторая — на  $e_i, e_{i+1}$ . Мы поступим несколько хитрее, а именно, будем определять углы между этими плоскостями, но естественным образом *ориентированными*.

Более подробно, рассмотрим векторные произведения  $\xi_i = [e_{i-1}, e_i]$  и  $\xi_{i+1} = [e_i, e_{i+1}]$ . Первый из этих векторов перпендикулярен первой плоскости, а второй — второй. Рассмотрим плоскость  $\Pi$ , перпендикулярную  $e_i$ . Тогда оба  $\xi_i$  и  $\xi_{i+1}$  ей параллельны. Введем в плоскости  $\Pi$  декартовы координаты. Для этого в качестве начала координат выберем точку пересечения плоскости  $\Pi$  и прямой, проходящей через ребро  $e_i$ ; ось  $x$  пустим в направлении  $e_x$  вектора  $\xi_i$ ; ось  $y$  выпустим в таком направлении  $e_y$ , чтобы тройка  $\xi_i, e_y$  и  $e_i$  была положительно ориентирована. Иными словами,  $e_y = [e_i/\|e_i\|, \xi_i/\|\xi_i\|]$ . Теперь рассмотрим направление  $e = \xi_{i+1}/\|\xi_{i+1}\|$  вектора  $\xi_{i+1}$ . Тогда в базисе  $e_x, e_y$  вектор  $e$  имеет координаты  $(\cos \beta_i, \sin \beta_i)$  для некоторого однозначно определенного угла  $\beta_i \in [-\pi, \pi)$  (именно его мы будем использовать в качестве угла между рассматриваемыми плоскостями).

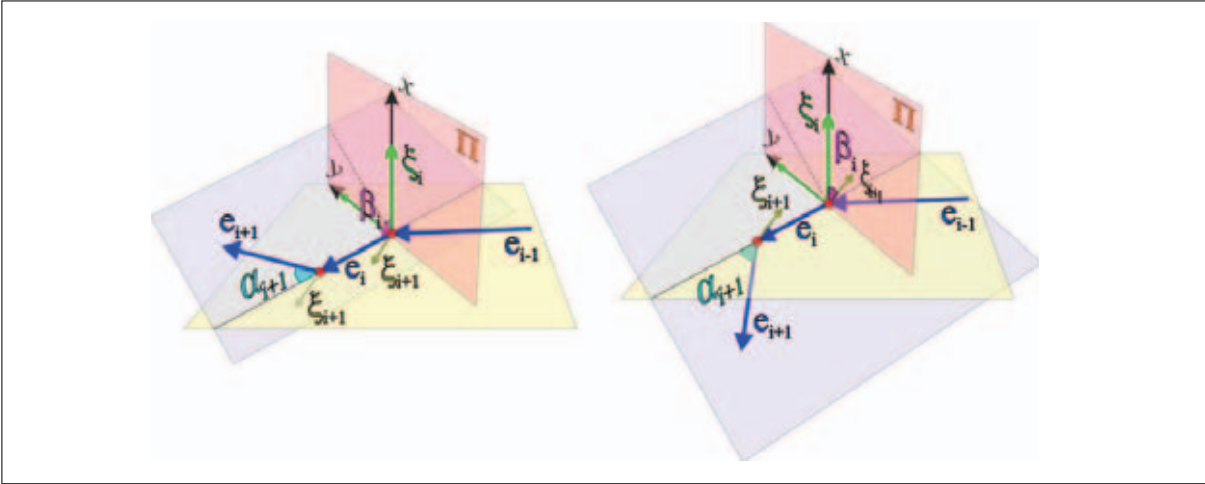


Рис. 1. Угловые характеристики ломаной

Найдем угол  $\beta_i$  явно. Заметим, что  $\cos \beta_i = \langle e, e_x \rangle$ ,  $\sin \beta_i = \langle e, e_y \rangle$ , поэтому

$$\begin{aligned} \beta_i &= \text{sign}(\sin \beta_i) \arccos(\cos \beta_i) = \text{sign}(\langle e, e_y \rangle) \arccos(\langle e, e_x \rangle) = \\ &= -\text{sign}[\det(\xi_i, e_i, \xi_{i+1})] \arccos \frac{\langle \xi_i, \xi_{i+1} \rangle}{\|\xi_i\| \|\xi_{i+1}\|}. \end{aligned}$$

Покажем, что длины  $s_i$  ребер ломаной  $L$  ее углы  $\alpha_i, \beta_j$  определяют ломаную  $L$  однозначно с точностью до движения.

(1) Выберем произвольное расположение начальной вершины  $A_0$  ломаной  $L$ .

(2) Разместим вторую вершину  $A_1$  ломаной  $L$  произвольным образом на расстоянии  $s_1$ .

(3) Выпустим из вершины  $A_1$  луч  $r_1$  под углом  $\alpha_1$  к лучу  $r_0 = A_0A_1$  (это можно сделать многими способами) и отложим на  $r_1$  отрезок  $A_1A_2$  длины  $s_2$ .

Если у каждого из предыдущих построений была некоторая свобода выбора, то все оставшиеся шаги будут определены однозначно. Мы покажем это для шага, на котором строится вершина  $A_{i+1}$ . Итак, пусть все вершины  $A_0, \dots, A_i$  уже построены. В частности, у ломаной  $L$  уже реализованы ребра  $e_{i-1}$  и  $e_i$ .

(4) Вычисляем вектор  $\xi_i = [e_{i-1}, e_i]$  и полагаем  $e_x = \xi_i / \|\xi_i\|$ .

(5) Вычисляем вектор  $e_y = [e_i, e_x] / \|e_i\|$ .

(6) Тогда, по определению, имеем  $\xi_{i+1} / \|\xi_{i+1}\| = \cos \beta_i e_x + \sin \beta_i e_y$ .

(7) Чтобы задать положение точки  $A_{i+1}$ , мы должны в плоскости, проходящей через ребро  $e_i$  и перпендикулярной  $\xi_{i+1}$ , отложить от  $A_i$  вектор  $e_{i+1}$  длины  $s_{i+1}$  в направлении, составляющем с вектором  $e_i$  угол  $\alpha_{i+1}$ . Это можно сделать двумя способами. Однако, по определению вектора  $\xi_{i+1}$ , он должен быть равен  $[e_i, e_{i+1}]$ , так что подходит лишь один из этих двух способов. А именно,

$$e_{i+1} = s_{i+1} \left( \cos \alpha_{i+1} \frac{e_i}{\|e_i\|} + \sin \alpha_{i+1} \left[ \cos \beta_i e_x + \sin \beta_i e_y, \frac{e_i}{\|e_i\|} \right] \right).$$

Пусть теперь все ребра ломаной  $L$  имеют одинаковую длину. Тогда постоянство отличных от нуля и  $\pi$  углов  $\alpha_i$  и  $\beta_i$  равносильно тому, что вершины ломаной лежат на некоторой спирали. Это соображение позволяет отлавливать спирали в полипептидах. Напомним, что в оставшейся после очистки базе данных во всех полипептидах

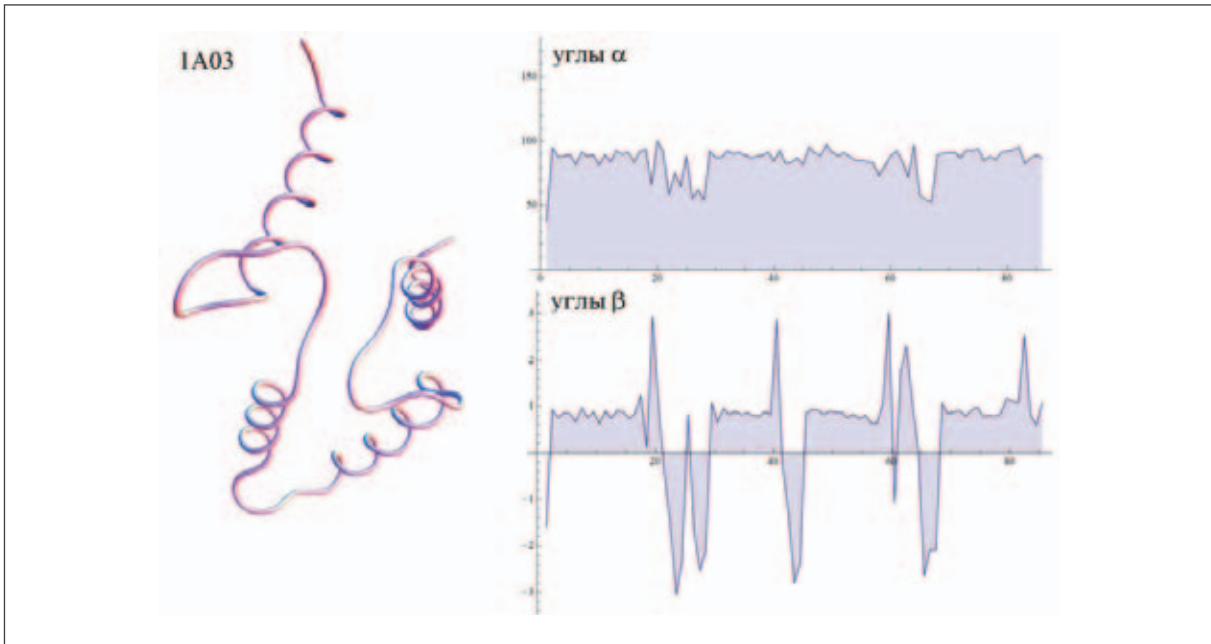


Рис. 2. Аппроксимация пептидного остова полипептида 1A03 кривой (слева) и графики последовательных углов  $\alpha_i$  и  $\beta_i$  (справа).

расстояния между последовательными альфа-углеродами отклоняются от среднего значения не более, чем на 7.5%, поэтому эти расстояния можно считать почти постоянными. Таким образом, если для ломаной  $L$ , проведенной через альфа-углероды, на каком-то ее фрагменте углы  $\alpha_i$  и  $\beta_i$  окажутся почти постоянными (и отличными от 0 и  $\pi$ ), то этот фрагмент будет являться спиралью. Продемонстрируем, как это работает на примере.

Рассмотрим полипептид 1A03 из базы данных Protein Data Bank. Для наглядности мы провели кривую через альфа-углероды этого полипептида, см. рис. 2. На рис. 2 (слева) хорошо видно, что этот полипептид имеет четыре спирали.

Справа на этом рисунке приведены последовательности  $\alpha$ - и  $\beta$ -углов ломаной, вершины которой — последовательные альфа-углероды полипептида 1A03. На этих графиках хорошо просматриваются участки локального постоянства. Таких участков также имеется четыре штуки, и они соответствуют спиральям полипептида. Несложно написать программу, которая будет автоматически определять участки постоянства последовательностей углов, тем самым, выделять спирали не прибегая к изображению.

Для большей наглядности можно изобразить оба графика одновременно (растянув для наглядности один из них), рис. 3.

**Замечание.** Было бы интересно найти критерий того, что произвольная ломаная (не обязательно с равными, но близкими по длине звеньями) вписана в спираль. Это условие позволило бы проводить аналогичное автоматическое определение спиралей и для полипептидов с большими отклонениями расстояний между последовательными альфа-углеродами.

В чем преимущество описанного только что подхода? Для коротких полипептидов можно, конечно же, определять фрагменты структуры и на глаз, но если полипептид достаточно длинный, то непосредственное изучение его трехмерного изображения является достаточно тяжелой работой. Вместо этого мы предлагаем смотреть на две

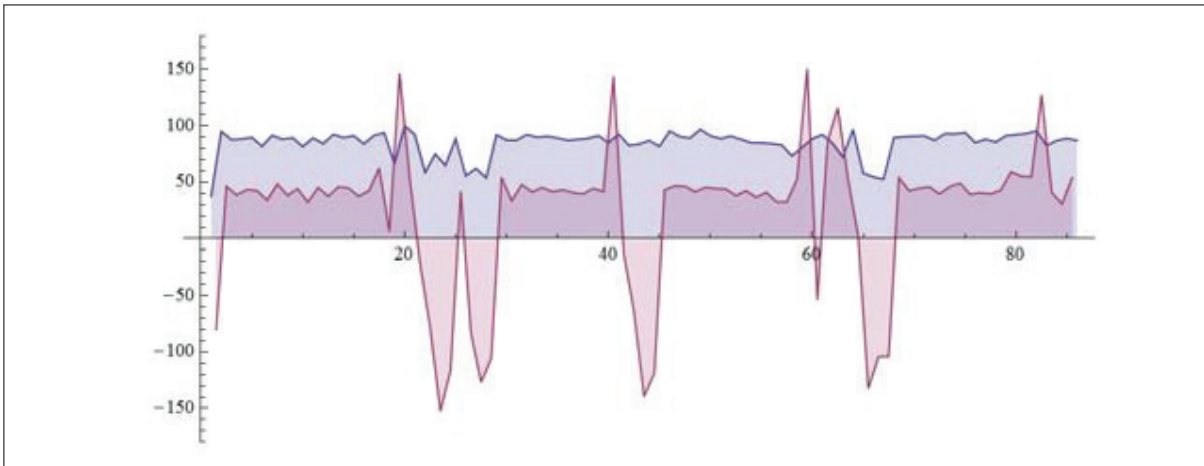


Рис. 3. Одновременное изображение обоих графиков  $\alpha$ - и  $\beta$ -углов.

функции и исследовать их (или на три, если учитывать расстояния). Кроме того, различные свойства участков функций, скажем, монотонность или выпуклость, могут привести к целому ряду новых типов фрагментов полипептидов и, как следствие, к более глубокому пониманию геометрической структуры.

Изложенные здесь соображения характеристики спиралей в терминах углов являются дискретными аналогами хорошо известных в классической дифференциальной геометрии результатов, описывающих форму пространственных кривых. В следующем разделе мы напомним основные определения теории кривых, чтобы предложить читателю еще ряд тем для поиска аналогий.

### 3. Кривизна и кручение пространственных кривых

Рассмотрим гладкую кривую  $\gamma(t) = (x_1(t), x_2(t), x_3(t))$  в пространстве  $\mathbf{R}^3$ . Здесь гладкость означает, что координатные функции  $x_i(t)$  непрерывно-дифференцируемы столько раз, сколько нам потребуется. *Скоростью кривой  $\gamma$  в точке  $\gamma(t)$*  называется вектор  $\dot{\gamma}(t) = (\dot{x}_1(t), \dot{x}_2(t), \dot{x}_3(t))$ , составленный из первых производных координатных функций. Если вместо первых производных рассмотреть вторые производные, то получим *вектор ускорения*  $\ddot{\gamma}(t) = (\ddot{x}_1(t), \ddot{x}_2(t), \ddot{x}_3(t))$ . Кривая, у которой скорость всюду отлична от нуля, называется *регулярной*, а у которой скорость и ускорение в каждой точке линейно независимы — *бирегулярной*.

Если длина вектора скорости кривой  $\gamma(t)$  всюду равна 1, то параметр  $t$  называется *натуральным*. Для натурально параметризованной кривой скорость и ускорение взаимно перпендикулярны в каждой точке кривой. Для такой кривой длина ускорения используется для характеристики искривленности: чем больше ускорение, тем сильнее кривая поворачивает. Длина вектора ускорения натурально параметризованной кривой называется *кривизной* и обозначается через  $k$ . Если кривизна отлична от нуля, т.е. ускорение ненулевое и, значит, кривая бирегулярна, то направление  $\dot{\gamma}/\|\dot{\gamma}\|$  вектора ускорения  $\ddot{\gamma}$  называется *главной нормалью кривой  $\gamma$*  и обозначается через  $\nu$ .

Итак, пусть кривая  $\gamma(t)$  натурально параметризована. Положим  $\tau = \dot{\gamma}$ , тогда  $\tau$  и  $\nu$  — единичные взаимно перпендикулярные векторы. Их векторное произведение  $\beta = [\tau, \nu]$  называется *бинормалью*. Тройка  $(\tau, \nu, \beta)$  единичных взаимно перпендикулярных векторов называется *репером Френе*.

*Формулы Френе* показывают, с какими скоростями меняются векторы репера Френе. Известно, что скорость  $\beta$  изменения бинормали  $\beta$  коллинеарна вектору  $\nu$ .

Коэффициент пропорциональности между  $\beta$  и  $\nu$ , взятый со знаком минус, называется *кручением* и обозначается через  $\kappa$ .

Приведем некоторые примеры, показывающие, как свойства кривизны и кручения выделяют те или иные естественные классы кривых.

**Пример.** Множество всех прямых — это в точности множество всех кривых нулевой кривизны (отметим, что для прямых кручение не определено).

**Пример.** Кривизна окружности радиуса  $r$  равна  $1/r$ . Лежащая в плоскости кривая является окружностью, если и только если ее кривизна постоянна и отлична от нуля.

**Пример.** У *винтовой линии* или *спирали*, которая в некоторой системе координат имеет вид  $(a \cos t, a \sin t, bt)$ , где  $a$  и  $b$  — ненулевые постоянные, кривизна и кручение постоянны и не равны нулю. Обратно, каждая бигулярная кривая, у которой кривизна и кручение постоянны и отличны от нуля, является винтовой линией.

**Пример.** Бигулярная кривая лежит в некоторой плоскости, если и только если ее кручение равно нулю.

Хорошо известно, что функции  $k(t)$  и  $\kappa(t)$  полностью определяют форму натурально параметризованной бигулярной кривой  $\gamma(t)$ : если у двух кривых эти функции одинаковы, то кривые можно совместить. Более того, каждые две функции  $f(t)$  и  $g(t)$ , где  $f(t)$  везде положительна, являются кривизной и кручением некоторой натурально параметризованной бигулярной кривой. Процедура построения пространственной кривой с заданными кривизной и кручением сводится к решению некоторой системы обыкновенных дифференциальных уравнений.

В действительности, характеристика ломаных в терминах углов — это “дискретный аналог” характеристики кривых их кривизнами и кручениями. При этом условие на равенство длин ребер ломаной аналогично натуральности параметра.

Если на кривой  $\gamma(t)$  параметр  $t$  не натуральный, то кривизну и кручение можно вычислить по следующим формулам:

$$k(t) = \frac{\|[\dot{\gamma}, \ddot{\gamma}]\|}{\|\dot{\gamma}\|^3}, \quad \kappa(t) = \frac{\det(\dot{\gamma}, \ddot{\gamma}, \ddot{\gamma}')}{\|[\dot{\gamma}, \ddot{\gamma}]\|^2}.$$

Для этих формул можно найти соответствующие дискретные аналоги. Кроме того, имеется много различных формул для “кривизны” и “кручения” ломаной, которые выдерживают предельный переход: если в бигулярную кривую начать вписывать ломаные, все более измельчая длины их ребер, и для таких ломаных вычислять по этим формулам кривизну и кручение, то в пределе получатся кривизна и кручение самой кривой. В свое время авторы опробовали этот подход, протестировав целый ряд функций, однако соответствующие графики качественно не отличались от приведенных нами выше графиков для  $\alpha$ - и  $\beta$ -углов (мы рассматривали ломаные с ребрами почти постоянной длины).

**Замечание.** Многие авторы, см. например [4], [3], рассматривают дискретные аналоги реперов Френе для ломаных и определяют кривизну и кручение ломаной, исходя из анализа поведения ее репера Френе. В обозначениях предыдущего пункта, в качестве репера Френе в точке  $A_i$ ,  $i = 1, \dots, n - 1$ , можно взять три вектора  $(\tau_i, \nu_i, \beta_i)$ , определенные так:

$$\tau_i = \frac{e_i}{\|e_i\|}, \quad \beta_i = \frac{\xi_{i+1}}{\|\xi_{i+1}\|}, \quad \nu_i = [\beta_i, \tau_i].$$

Аналогом формул Френе в этом случае будет связь между последовательными реперами  $R_i = (\tau_i, \nu_i, \beta_i)$  и  $R_{i+1} = (\tau_{i+1}, \nu_{i+1}, \beta_{i+1})$ , которые отличаются на ортогональное преобразование:  $R_{i+1} = O_i R_i$ . Если, например, представить преобразование  $R_i$  как композицию поворота в плоскости  $e_i, e_{i+1}$  и поворота вокруг оси  $e_{i+1}$ , то углы этих поворотов можно взять в качестве аналогов кривизны и кручения в точке  $A_i$ .

#### 4. Выводы

Таким образом, в данной работе демонстрируется, как простые геометрические соображения могут помочь в изучении пространственной структуры полипептидов, в частности, дают возможность написать программу автоматического выделения спиралей. Кроме того, этот подход показывает, что “кодировка” пространственной структуры существенно более простыми объектами (в нашем случае — двумя функциями) может оказаться полезной, причем не только для решения старых задач (нахождения спиралей и слоев), но и открывает широкие возможности для поиска новых структурных составляющих. Возможно, поиск других “удачных кодировок” является перспективным методом изучения конформаций.

#### Литература

1. *Glaser J.A. and Deutscher M.P. (ed.)* Introduction to Biophysical Methods for Protein and Nucleic Acid Research // Academic Press, London, N.-Y., 1995.
2. *Pauling L., Corey R.B., and Branson H.R.* The structure of proteins; two hydrogenbonded helical configurations of the polypeptide chain // Proc. Natl. Acad. Sci., USA, 1951, **37**, 205–211.
3. *Hu S., Lundgren M., and Niemi A.J.* The Discrete Frenet Frame, Infection Point Solitons And Curve Visualization with Applications to Folded Proteins // arXiv:1102.5658v1, 2011.
4. *Goriely A., Hausrath A., and Neukirch S.* The differential geometry of proteins and its applications to structure determination // Biophysical Reviews and Letters, 2008, **3** (1–2), 77–101.
5. *Боженко В.К., Иванов А.О., Мищенко А.С., Тужилин А.А., Шишкин А.М.* Геометрические модели в биологии: как и что можно моделировать // Электронный журнал “Научная визуализация”, 2009, **1** (1), 66–99.
6. *Балабаев Н.К., Шайтан К.В.* Компьютерное моделирование молекулярной динамики Компьютерное моделирование полимерных и биополимерных систем // Под ред. Иванова В.А., Рабиновича А.Л., Хохлова А.Р., М.: Книжный дом “ЛИБРОКОМ”, 2009, 35-62.
7. *Шайтан К.В., Турлей Е.В., Голик Д.Н., Терёшкина К.Б., Левцова О.В., Федик И.В., Шайтан А.К., Ли А., Кирпичников М.П.* Динамический молекулярный дизайн био- и наноструктур // Журнал российского химического общества, 2006, **50** (2), 53-65.
8. *Шайтан К.В.* Энергетическая поверхность и конформационная динамика молекул // Электрохимия, 2003, **39** (2), 212-219.
9. *Шайтан К.В., Беляков А.А., Леонтьев К.М., Сарайкин С.С., Михайлюк М.Г., Егорова К.Б., Орлов М.В.* Геометрия энергетической поверхности и конформационная динамика: от углеводородов к белкам и пептидам // Хим. физ., 2003, **22** (2), 57-68.



# GEOMETRY OF POLYGONAL LINES AND POLYPEPTIDES

A.O. Ivanov, A.S. Mishchenko and A.A. Tuzhilin

*Lomonosov Moscow State University,  
Bauman Moscow State Technical University*

aoiva@mech.math.msu.su, asmish@mech.math.msu.su, tuz@mech.math.msu.su

Received 29.04.2014

In the paper we present a method to investigate conformations of polypeptides in terms of two functions, discrete analogues of the curvature and the torsion of space curves. This change-over not only gives a visual encoding of the space structure of polypeptides, but also permits to mark out different structural elements of conformations, for instance, the spirals.