

УДК: 519.2

Статистический анализ внутренних расстояний спиральных пар в белковых молекулах

Тихонов Д.А.^{*1}, Куликова Л.И.^{**1}, Ефимов А.В.²

¹*Институт математических проблем биологии РАН – филиал ИПМ им. М.В. Келдыша
РАН, Пущино, Московская область, Россия*

²*Институт белка РАН, Пущино, Московская область, Россия*

Аннотация. В данной работе проведен статистический анализ распределения межспиральных расстояний в парах связанных между собой перетяжками спиралей в пространственных структурах белковых молекул. Полученное по определенным правилам множество спиральных пар всех белковых молекул, включенных в Protein Data Bank, разбито на три подмножества по критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей. Показано, что распределения расстояний для спиральных пар, проекции которых не имеют пересечений, имеют более дальнедействующий характер, чем те, проекции которых пересекаются. При помощи регрессионного анализа исследован характер распределений, в частности, показано, что в подмножестве без пересечений распределения различных расстояний между осями спиралей относятся к гамма-распределениям. Показано, что подмножества пар с пересечением имеют с большой вероятностью малое отношение минимального расстояния к межплоскостному, в отличие от подмножества пар без пересечения, где наблюдается противоположная картина. Обосновывается вывод о том, что спиральные пары, проекции спиралей которых пересекаются, дополнительно стабилизируются за счет внутренних взаимодействий.

Ключевые слова: *структурные мотивы белков, спиральные пары в белковых молекулах, точечная модель, статистический анализ, межплоскостное расстояние, минимальное расстояние между осями спиралей.*

ВВЕДЕНИЕ

В настоящее время объектом пристального внимания исследователей являются структурные мотивы, имеющие уникальные укладки полипептидной цепи в пространстве. Такой интерес к структурным мотивам обусловлен уникальностью структур и их возможностью быть зародышами, к которым остальные участки цепи в процессе сворачивания белков могут пристраиваться в соответствии с простыми принципами [1]. Однако, независимо от того, по какому механизму происходит сворачивание белка, структурные мотивы могут быть использованы для создания системы структурной классификации белков [2], а также в качестве стартовых структур для поиска возможных укладок полипептидной цепи при моделировании структуры белков [1, 3]. Обычно (однако, не всегда) структурный мотив является комбинацией нескольких элементов вторичной структуры. Простейшими являются структурные

*dmitry.tikhonov@gmail.com

**likulikova@mail.ru

мотивы, состоящие из двух элементов вторичной структуры, имеющих уникальные укладки полипептидной цепи в пространстве. Объектом наших интересов являются структурные мотивы, образованные двумя спиралями, расположенными в полипептидной цепи друг за другом и соединенных между собой перетяжками [4, 5, 6].

Известно, что α -спираль в белках упакована оптимально плотно [6]. Именно она из всех видов спиралей преобладает в белках. Вероятно, это связано с тем, что является более стабильным элементом вторичной структуры. Довольно мало в белках π -спиралей. В виде коротких фрагментов в белках встречаются 3_{10} -спирали, в основном, правые [7]. Хорошо изученной вторичной структурой является α -спираль [8–12] и, наряду с β -структурой, – важнейшим элементом белковых молекул, во многом определяющим их общую архитектуру. Стабилизирующими элементами для α -спирали являются аланин, лейцин и метионин (Ala, Leu и Met), в то время как глицин (Gly) разрушает спираль и способствует образованию нерегулярных участков. Пролин (Pro) также не встречается в α -спирали (кроме ее N-концевого витка). Влияние аминокислотного состава на вторичную структуру можно не только оценить статистически [13] или экспериментально [14], но и *a priori* оценить теоретически [7]. Кроме того, стабильность α -спирали достигается определенным порядком чередования гидрофобных групп в цепи. Этот порядок приводит к образованию на спирали сплошной гидрофобной поверхности. Важность правильного чередования боковых групп для формирования вторичной структуры была показана в работе [15, 16].

Такие же описания можно составить и для более сложных структур, состоящих из двух или трех спиралей или комбинации нескольких элементов вторичной структуры [17].

Также из литературы известно, что наиболее плотная упаковка двух α -спиралей достигается при антипараллельной, перпендикулярной и так называемой скошенной ориентацией между спиралями, примерами такой упаковки являются рассмотренные нами супервторичные структуры: α - α -уголки, α - α -шпильки, L-образные и V-образные структуры [3].

Одним из часто встречаемых структурных мотивов в белках являются α - α -уголки [6]. Эта супервторичная структура образована двумя соседними по полипептидной цепи α -спиралями, связанными между собой перетяжками и упакованными ортогонально (крестообразно). В белках α - α -уголки встречаются в форме левой суперспирали. Их последовательности имеют определенное расположение в цепи гидрофобных, гидрофильных и глициновых остатков.

Супервторичной структурой из двух соседних по цепи α -спиралей, которые связаны перетяжкой и упакованы антипараллельно, являются α - α -шпильки. Данная супервторичная структура может быть левой или правой в зависимости от того, как расположена вторая α -спираль относительно первой – справа или слева. Длина перетяжек между спиралями тоже может быть разной, при этом каждая стандартная α - α -шпилька должна иметь строго определенное и свойственное только ей расположение в цепи гидрофобных, гидрофильных и глициновых остатков [3].

L-образные структуры также образованы двумя спиралями. Особую роль в образовании таких структур играет пролин (Pro), именно он способствует излому между двумя α -спиралями. Указанные структуры могут быть правыми или левыми [5].

В образовании V-образных структур тоже участвуют две α -спирали. Они очень похожи на α - α -шпильки, у которых сильно удалены друг от друга несвязанные концы α -спиралей, также похожи и на L-образные структуры. В V-образных структурах α -спирали имеют длину, как правило, не превышающую трех–четырёх витков [3].

В предыдущих работах [18, 19] нами решались задачи распознавания [20, 21], анализа устойчивости и конформационного анализа структурных мотивов белков типа α - α -уголков в вычислительном эксперименте молекулярной динамики. Объектом наших исследований был указанный тип структурных мотивов с короткой перетяжкой.

Ранее (в 1993 г.) устойчивость α - α -углов была косвенно доказана *in vitro* [22]. Независимо выдвинутая нами гипотеза об автономной устойчивости структурных мотивов была проверена *in silico* в вычислительных экспериментах молекулярной динамики [18].

В настоящей работе мы поставили две основные задачи:

- отобрать из банка белковых данных PDB все структурные мотивы, образованные двумя любого типа спиралями, расположенными в полипептидной цепи друг за другом и соединенных между собой перетяжками различной длины и конформации, и создать соответствующую базу данных для дальнейших исследований таких двухспиральных структур;

- провести на первом этапе статистический анализ распределения межспиральных расстояний в спиральных парах, входящих в эту базу данных.

ТОЧЕЧНАЯ МОДЕЛЬ СПИРАЛЬНОЙ ПАРЫ

Известно, что спирали в белках являются очень часто встречаемой вторичной структурой [6]. Есть белки, которые состоят только из спиралей и нерегулярных участков различной конформации (см. гистограмму на рисунке 1). Из гистограммы видно, что в PDB имеется много белков, которые на 60–70 % состоят из спиралей. При этом гистограмма демонстрирует, что распределение числа белков в зависимости от отношения количества аминокислот в спиральных к общему количеству аминокислот в структурах, имеет максимум при 40 %.

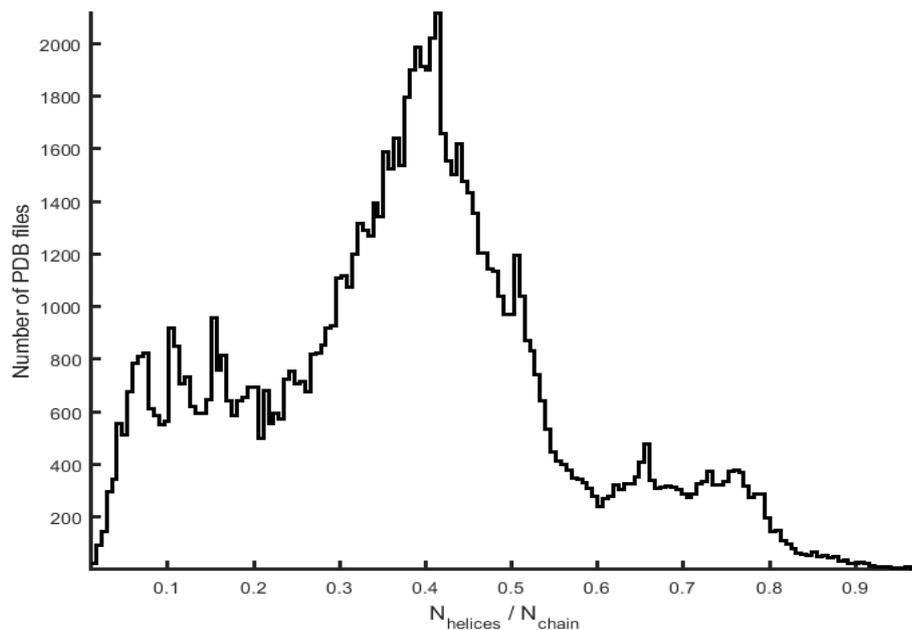


Рис. 1. Гистограмма распределения числа белков в PDB в зависимости от отношения количества аминокислот в спиральных к общему количеству аминокислот в белке.

Спиральная пара – фрагмент белковой цепи, состоящий из двух рядом расположенных по цепи спиралей, между которыми находится одна или более аминокислот, вторичная структура которых не относится по типу к спиральям. Этот элемент является, по-видимому, самым простым элементом супервторичной структуры. С точки зрения анализа такой объект является чрезвычайно удобным, поскольку может быть описан всего четырьмя точками в пространстве. Действительно, если мы аппроксимируем обе спирали цилиндрами, на которые намотаны спирали, образованные нитью, проходящие через C_{α} -атомы, то начала и концы осей цилиндров дадут нам те четыре точки, которые полностью опишут данную супервторичную

структуру. На рисунке 2 изображен пример спиральной пары. Это фрагмент цепи белка из Банка белковых данных (PDB ID 3A0B, C_{α} : 1000–1037). Показаны цилиндры, аппроксимирующие спираль, а также плоскости, проходящие через оси цилиндров. Кривая аппроксимирована по положениям C_{α} -атомов белковой цепи, атомы на кривой обозначены точками.

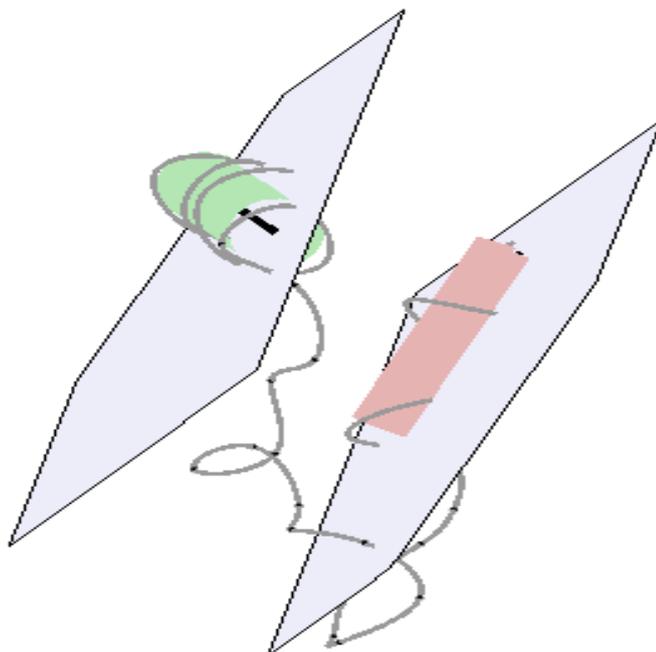


Рис. 2. Пример спиральной пары. Фрагмент цепи белка из банка белковых данных длиной 38 аминокислот (PDB ID 3A0B, координаты участка: 1000–1037). Показаны аппроксимирующие спираль цилиндры, а также плоскости, проходящие через оси цилиндров. Кривая аппроксимирована по положениям C_{α} -атомов белковой цепи, атомы на кривой обозначены точками.

С точки зрения описания взаимного расположения спиралей совершенно естественно возникают три расстояния. Во-первых, это межплоскостное расстояние. Как известно, через две непараллельные прямые в пространстве можно единственным образом построить две параллельные плоскости, кратчайшее расстояние между которыми является также кратчайшим расстоянием между прямыми. Если считать, что оси цилиндров принадлежат прямым, то кратчайшее расстояние между этими прямыми мы будем называть межплоскостным расстоянием спиральной пары и обозначать его d . Второе расстояние спиральной пары – кратчайшее расстояние между отрезками, которые являются осями цилиндров. Кратчайшее расстояние мы будем обозначать r . Очевидно, что минимальное расстояние спиральной пары r всегда будет больше или равно межплоскостному расстоянию d , поэтому, даже формально, можно ввести величину катета l , определив его как $l = \sqrt{r^2 - d^2}$. Катет l будет тем третьим расстоянием, которое описывает относительное расположение спиралей в спиральной паре. Точечная модель спиральной пары приведена на рисунке 3. Показаны только оси спиральной пары. Отрезок $[A_1, A_2]$ – ось цилиндра первой спирали, $[B_1, B_2]$ – ось цилиндра второй спирали. На этом рисунке изображены также и все возможные расстояния d , r и l между спиральными парами спиральной пары.

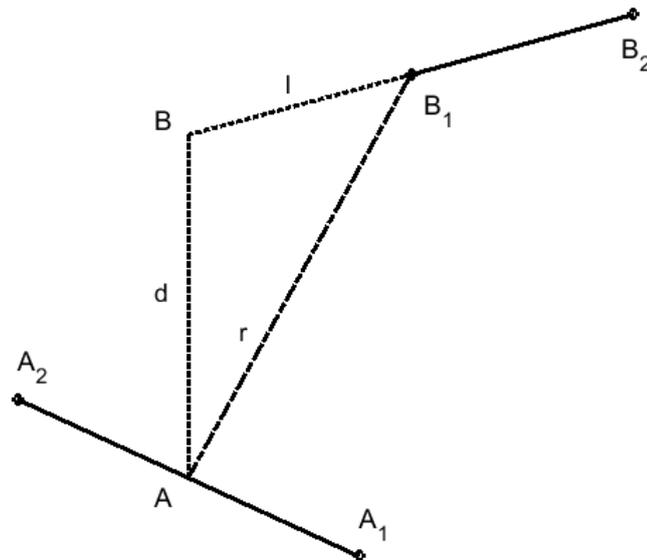


Рис. 3. Геометрия спиральной пары – четыре точки, образующие два отрезка в пространстве.

На рисунке 3 изображен частный случай взаимного расположения осей, когда одна из точек, откуда выходит перпендикуляр, соединяющий плоскости, в данном случае это точка A , является внутренней точкой оси спирали $[A_1, A_2]$. Вторая точка B – не принадлежит оси спирали $[B_1, B_2]$. В таком случае катет l является продолжением оси спирали до точки B , куда входит перпендикуляр, соединяющий плоскости. Другим частным случаем может быть случай, когда обе точки A и B не принадлежат своим осям. Тогда минимальное расстояние между отрезками будет равно расстоянию между ближайшими к точкам входа и выхода перпендикуляра концами осей спирали. Тогда катет l не будет иметь простого геометрического изображения, но всегда может быть вычислен формально.

Особый случай расположения спиралей в паре, когда обе точки A и B являются внутренними точками осей $[A_1, A_2]$ и $[B_1, B_2]$ соответственно. Тогда межплоскостное расстояние d совпадает с минимальным расстоянием r , и катет l , соответственно, равен нулю. Это случай пересечения проекций осей цилиндров. Мы выделим его отдельно, и подробно рассмотрим связанное с ним распределение межплоскостного расстояния d . Этот случай можно считать сингулярным, если относить в общий случай распределения соотношения межплоскостного к минимальному расстоянию r/d , вычисленного для каждой спиральной пары. В случае пересечения осей соотношение r/d равно 1. В данной работе мы исследуем общий вид статистического распределения для соотношения r/d . Так же точно будут исследованы статистические распределения и для всех расстояний в спиральной паре.

КРИТЕРИЙ РАЗБИЕНИЯ ВСЕХ СПИРАЛЬНЫХ ПАР НА МНОЖЕСТВА

В данной работе мы будем использовать площадь полигона пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей в спиральной паре. Все множество спиральных пар мы разобьём на три подмножества согласно следующим правилам:

1. Множество $\{A\}$ образуют все спиральные пары, у которых площади проекций спиралей не имеют пересечений;
2. Множество $\{B\}$ образуют все спиральные пары с ненулевым пересечением проекций, за исключением тех спиральных пар, где полигон пересечения включает в себя точку пересечения проекций осей спиралей;

3. Множество $\{C\}$ образуют все спиральные пары, где полигон пересечения включает в себя точку пересечения проекций осей спиралей.

Площадь полигона пересечения проекций спиралей, помимо геометрии, зависит еще от диаметра цилиндра спирали. Причем диаметр цилиндра определяется не только типом спирали, но и величиной среднего размера боковой цепи DR . Общепринятое значение $DR = 3.6 \text{ \AA}$. На рисунке 4 изображен полигон пересечения проекций спирали для спиральной пары, изображенной на рисунке 2. Ширины прямоугольников, которые являются проекциями соответствующих спиралей, равны диаметрам спиралей, которые определяются их типом, плюс удвоенное значение DR . В данном случае обе спирали являются α -спиралями с диаметром 4.6 \AA . Таким образом, ширина прямоугольников составляет 11.8 \AA . Полигон пересечения выделен цветом, указана его площадь S и периметр P . Отмечена точка проекций пересечения осей спиралей. На рисунке указана также и величина межплоскостного расстояния d .

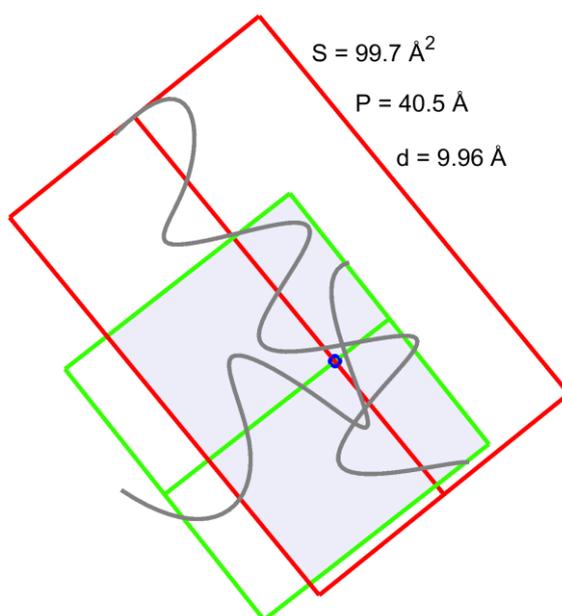


Рис. 4. Пересечение проекций цилиндров спиралей спиральной пары. Полигон пересечения проекций спирали для спиральной пары, изображенной на рисунке 2 (PDB ID 3A0B, координаты участка:1000–1037).

Граница между множествами $\{A\}$ и $\{B\}$ достаточно условна и определена с точностью до выбора величины DR . При меньшей величине DR элементы из множества $\{B\}$ будут переходить в множество $\{A\}$. Справедливо и обратное утверждение. Величина DR носит усредненный по всем аминокислотным остаткам приблизительный характер. На рисунке 4 приведено значение межплоскостного расстояния $d = 9.96 \text{ \AA}$, оно существенно меньше 11.8 \AA , которое должно было быть для двух α -спиралей в контакте с учетом среднего размера боковой цепи $DR = 3.6 \text{ \AA}$. Однако, для определенности мы выполнили все наши вычисления именно с такой общепринятой величиной DR .

ДЕТАЛИ ВЫЧИСЛЕНИЙ

Для работы нам необходимы точечные модели спиральных пар, а также площади проекций спиральных пар на параллельные плоскости, проходящие через оси спиралей. В данной части работы мы опишем способ их получения.

Источником молекулярных структур являлся банк белковых данных PDB. Мы провели исчерпывающий анализ всех учтенных в нем белковых молекул. Анализ вторичной структуры аминокислотной цепи осуществлялся с помощью метода,

разработанного авторами Dictionary for Secondary Structure of Protein [23]. Метод является наиболее широко используемым, кроме того, в публичном доступе находится программа, реализующая метод, которая обрабатывает непосредственно PDB-файлы, возвращая файл с информацией о вторичной структуре. Прежде всего, нас интересовало наличие таких элементов вторичной структуры, как спирали. Всего в программе DSSP различаются три типа спиралей. Первый тип, кодируемый буквой H, – это α -спираль, второй тип, кодируемый буквой G, – это 3_{10} -спираль. Третий тип, кодируемый буквой I, – это π -спираль. Обработывались выходные файлы программы DSSP с целью найти начала и концы непрерывных участков, кодируемых как спирали. Если PDB-структура содержала белок, в котором имелось не больше одной спирали, она отбрасывалась. Отбрасывалась структура и в том случае, если рядом находились две спирали различных типов, связанных перетяжкой нулевой длины. Таким образом, у нас остались только те структуры, которые содержат спиральные пары. Данные о количестве таких структур приведены в таблице 1.

Таблица 1. Общие данные – число всех обработанных белковых молекул и число найденных спиралей различных типов

Количество обработанных белковых структур PDB	Количество обработанных аминокислотных остатков	Количество обработанных аминокислотных цепочек	Кол-во спиралей типа H	Кол-во спиралей типа G	Кол-во спиралей типа I
100397	66546491	384666	1952658	750605	2908

Отметим, что в процентном отношении количество спиралей типа H составляет абсолютное большинство – 72.16 % к общему числу всех найденных спиралей. Спирали типа G составляют 27.73 %. Все остальные (0.1 %) приходятся на спирали типа I.

После того как все спирали в аминокислотной цепочке были найдены, спиральные пары составлялись из соседних спиралей. Например, если аминокислотная цепочка содержит N спиралей, то спиральных пар она содержит $N - 1$. Если обозначить h_i – спираль любого типа внутри цепочки, то спиральные пары можно описать как (h_i, h_{i+1}) , где i меняется от 1 до $N - 1$.

После того, как выделены спиральные пары, нужно получить оси спиралей. Для этого мы воспользуемся методом, широко применяемым в молекулярных вычислениях. Мы знаем тип спирали, число аминокислот спирали n_h , следовательно, можем построить координаты идеальной спирали в удобной системе координат. Идеальная спираль определяется всего тремя параметрами, угловым шагом φ_h , линейным шагом z_h и радиусом r_h . Если выбрать ось идеальной спирали вдоль оси z , то координаты C_α -атомов будут выражены через параметры идеальной спирали следующим образом:

$$\begin{aligned}x_i &= r_h \cos(\varphi_h(i-1)), \\y_i &= r_h \sin(\varphi_h(i-1)), \\z_i &= z_h(i-1), \quad i = 1, \dots, n_h.\end{aligned}$$

Нам также известны реальные координаты C_α -атомов спиралей спиральной пары. Для совмещения реальных и идеальных координат воспользуемся алгоритмом, предложенным в работах [24, 25]. Алгоритм основан на сингулярном разложении ковариационной матрицы идеальных и реальных координат. Он позволяет вычислить матрицу вращения и вектор трансляции идеальных координат так, чтобы они наилучшим образом совпали с реальными координатами. Качество совпадения

оценивается при помощи величины RMSD, которая является квадратным корнем от среднего квадрата отклонений координат. Матрица вращений и вектор трансляций выбраны из условия минимума величины RMSD. В таблице 2 приведены значения параметров идеальных спиралей для различных их типов, а также средние оценки величины RMSD, которые были получены нами при аппроксимации реальных спиралей идеальными.

Таблица 2. Параметры идеальных спиралей различных типов, а также статистические оценки распределения величины RMSD, полученные при аппроксимации реальных спиралей идеальными (φ_h – конформационные углы идеальной спирали, z_h – шаг идеальной спирали, r_h – радиус идеальной спирали)

параметры	Н-спираль	G-спираль	I-спираль
φ_h , град.	100.00	120.00	81.82
z_h , Å	5.40	6.00	4.80
r_h , Å	2.30	1.90	2.80
мода RMSD	0.20	0.05	1.98
медиана RMSD	0.31	0.13	2.00
среднее RMSD	0.40	0.20	2.10

Из таблицы видно, что радиус идеальной α -спирали r_{hH} равен 2.3, радиус идеальной G-спирали r_{hG} равен 1.9. Анализируя величины в таблице 2, легко заметить, что аппроксимация идеальными спиральями для типов Н и G весьма удовлетворительна, среднее RMSD меньше, чем современное разрешение в кристаллографии. Однако качество аппроксимации спиралей типа I – невысокое.

Теперь мы переходим к задаче вычисления полигона пересечений проекций спиралей на параллельные плоскости, проходящие через оси спиралей. Оси спиралей мы находим при помощи матрицы вращений и вектора трансляций, воздействуя ими на идеальную ось, которая имеет начало в нуле координат и конец на оси z в точке $n_h z_h$. После того как обе оси найдены, выполняем проекцию второй оси на плоскость, в которой лежит первая ось. Первая ось вместе с проекцией второй оси однозначно определяет первую плоскость. Строим нормаль к этой плоскости и вычисляем матрицу поворота, совмещающую первую плоскость с плоскостью xy . После поворота обе оси спирали – это отрезки на плоскости. Выясняем, пересекаются ли эти отрезки. Затем строим проекции спиралей на плоскости, которые являются прямоугольниками, высоты которых равны длинам осей, и ширины равны удвоенному значению $r_h + DR$. Выясняем, есть ли пересечение этих прямоугольников. Если есть, то далее стоит задача нахождения полигона пересечения этих прямоугольников. Эта задача решается с использованием пакета MatGeom [26], как и другие задачи, о которых говорилось выше. Как только полигон пересечения найден, мы вычисляем его площадь и периметр. Таким образом, для каждой спиральной пары мы выясняем две вещи: первое – существует ли точка пересечения проекции осей спиралей; второе – пересекаются ли проекции спиралей. Дополнительно вычисляется площадь и периметр полигона пересечения проекций спиралей, если таковой существует. В соответствии с этой информацией мы относим спиральную пару к одному из трех множеств {A}, {B} или {C} по тем правилам, которые приведены выше. В таблице 3 мы привели число спиральных пар, попавших в разные множества с учетом их типа. Всего имеем 6 типов спиральных пар: HH, GG, II, HG, HI, GI. Если название типа состоит из двух одинаковых букв, то спиральная пара данного типа образована двумя спиральями одинакового типа. Например, спиральная пара типа HH состоит из двух H-спиралей.

Спиральная пара, состоящая из спиралей различных типов, в названии типа имеет название типов спиралей, из которых она состоит. Например, HG – спиральная пара, в которой одна спираль типа H, другая – G. Данные в таблице симметризованы, это означает, что если спиральная пара состоит из спиралей различных типов, мы не различаем порядок следования спиралей. Например, спиральная пара HG и GH будет отнесена к типу HG.

Таблица 3. Число различных типов спиральных пар в множествах спиральных пар

Множества спиральных пар	Типы спиральных пар						Число элементов множеств
	HH	HG	GG	HI	GI	II	
{A}	402912	441055	125766	1588	643	0	971964
{B}	570830	349024	45513	1677	244	1	967289
{C}	234000	31719	1598	26	9	0	267352
Общее число спиральных пар по типам	1207742	821798	172877	3291	896	1	2206605

Анализируя данные в таблице, отметим некоторые особенности. Общее число спиральных пар в множестве {A} равно числу элементов в множестве {B}, каждое из них дает вклад по 44 % к общему числу спиральных пар, остальные 12 % приходятся на элементы из {C}. Это означает, что большая часть спиральных пар в аминокислотных цепочках (56 %) имеют пересечения проекций спиралей.

Что касается разбиений по типам спиралей, отметим, что, несмотря на то, что H-спирали составляют подавляющее большинство (72 %), спиральные пары типа HH составляют только 54.7 % к общему числу спиральных пар. Спиральные пары с участием спиралей G типа – 45 %, меньше 0.2 % приходится на пары с участием спиралей I типа.

Можно отметить еще одну закономерность: в множестве {A}, где спиральные пары не имеют пересечений проекций, большинство пар 45 % относятся к типу HG. В множестве {B} большинство (59 %) составляют пары HH типа. А в том случае, когда проекции осей спиралей пересекаются (множество {C}), спиральные пары HH типа составляют подавляющее большинство, 87.5 %. Можно сказать, что спирали HH типа более склонны к внутреннему взаимодействию, чем остальные. Однако, чтобы связать пересечение проекций спиралей со склонностью к внутреннему взаимодействию, необходимо проанализировать распределения расстояний между спиралями в паре. Этому будет посвящена следующая часть работы.

ГИСТОГРАММЫ РАСПРЕДЕЛЕНИЯ ВСЕХ ТИПОВ СПИРАЛЬНЫХ ПАР В ЗАВИСИМОСТИ ОТ РАССТОЯНИЙ МЕЖДУ СПИРАЛЯМИ

В данной части работы мы представим результаты расчетов гистограмм распределения всех типов спиральных пар в зависимости от расстояний между спиралями, также статистические оценки различных расстояний. Значительный объём данных, полученный при обработке PDB-структур, обеспечивает достаточную достоверность результатов. Для спиральных пар, принадлежащих конкретному множеству и типу спиралей, мы получаем три набора расстояний: минимальное расстояние r , межплоскостное расстояние d и длина катета $l = \sqrt{r^2 - d^2}$. Для этих наборов мы строим гистограммы. Помимо указанных гистограмм, мы оцениваем соотношения расстояний r/d .

Все гистограммы, приведенные в данной части работы, вычислялись с постоянным шагом по величине расстояния, который варьировался, в зависимости от типа расстояния. Везде мы рассчитывали число спиральных пар N_{hh} в данный интервал расстояний.

Исключение составило распределение для соотношения расстояний r/d . Для этой величины шаг выбирался в логарифмической шкале, что связано со значительными особенностями данного распределения вблизи нуля. Кроме того, распределение r/d нормировалось, чтобы получить оценку плотности вероятности.

Мы начнем описание результатов с простейшего случая, когда спиральные пары имеют пересечение проекций осей спиралей на плоскость. Эти спиральные пары принадлежат множеству $\{C\}$ по нашей классификации. На рисунках 2 и 4 приведено изображение спиральной пары именно такого типа.

На рисунке 5 представлена гистограмма распределения спиральных пар, принадлежащих множеству $\{C\}$, в зависимости от межплоскостного расстояния d (в данном случае межплоскостное расстояние d равно минимальному r). Это единственная гистограмма, описывающая спиральные пары из данного множества, поскольку катет равен нулю, и, соответственно, соотношение r/d равно 1.

Из рисунка видно, что гистограмма имеет отчетливый максимум в области 10 Å и побочный максимум в области меньших расстояний для спиральных пар типа НН. В области до 7 Å мы не наблюдаем наличия спиральных пар, что связано со стерическими ограничениями для спиральных пар из этого множества. Для всех типов спиралей гистограммы имеют ненулевые значения вплоть до величин 30–35 Å. Спиральные пары типов НГ и GG не имеют выраженного пика в области 10 Å. Они являются многомодовыми распределениями, причем трудно сопоставить ординаты их локальных пиков с радиусами спиралей. Действительно, с учетом суммы радиусов НГ спиралей равной 4.2 Å, следовало ожидать, что максимум гистограммы для спиральных пар типа НГ сместится в сторону меньших расстояний, в сравнении с парами типа НН, где сумма радиусов больше на 0.4 Å. Однако мы этого не наблюдаем. Для спиралей GG типа картина осложняется недостаточной статистикой, поскольку, как раньше отмечалось, спирали данного типа составляют всего около 0.6 % от общего числа спиральных пар, принадлежащих множеству $\{C\}$.

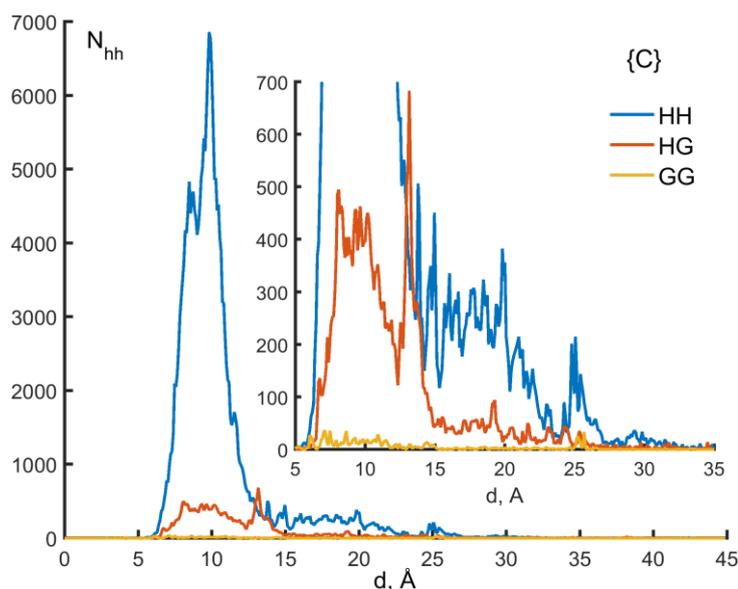


Рис. 5. Гистограмма распределения спиральных пар, принадлежащих множеству $\{C\}$, в зависимости от межплоскостного расстояния d (в данном случае межплоскостное расстояние d равно минимальному r).

В таблице 4 приведены статистические оценки распределений d и r для спиральных пар различных типов, принадлежащих различным множествам. Анализируя данные, относящиеся к множеству $\{C\}$, легко видеть, что моды (максимумы) распределений и средние значения расстояний не следуют качественному поведению, которое можно было бы ожидать, исходя из суммы радиусов спиралей разных типов. Однако следует отметить, что для всех трех типов спиральных пар распределения достаточно хорошо локализованы, поскольку среднеквадратичное отклонение от средних расстояний более чем в два раза меньше самих средних расстояний.

Таблица 4. Статистические оценки распределений межплоскостного d и минимального r расстояний для спиральных пар различных типов, принадлежащих различным множествам.

Статистические оценки	$d_{\text{НН}}$	$d_{\text{НГ}}$	$d_{\text{ГГ}}$	$r_{\text{НН}}$	$r_{\text{НГ}}$	$r_{\text{ГГ}}$
мода $\{A\}$	0.88	0.41	0.41	11.88	10.56	8.86
медиана $\{A\}$	5.55	6.71	8.14	14.78	15.72	19.20
среднее $\{A\}$	7.68	8.74	10.42	15.79	17.46	20.39
ср.-кв.откл $\{A\}$	7.32	7.83	8.53	8.35	9.15	9.80
мода $\{B\}$	8.10	0.23	4.40	9.20	8.47	5.11
медиана $\{B\}$	7.76	6.64	5.63	9.20	8.66	8.23
среднее $\{B\}$	8.02	7.57	7.98	10.07	9.80	10.17
ср.-кв.откл $\{B\}$	5.71	6.13	7.22	4.77	5.12	6.26
мода $\{C\}$	9.82	13.20	25.68	9.82	13.20	25.68
медиана $\{C\}$	9.78	10.83	10.34	9.78	10.83	10.34
среднее $\{C\}$	10.68	12.11	12.49	10.68	12.11	12.49
ср.-кв.откл $\{C\}$	3.66	4.84	6.59	3.66	4.84	6.59

Пользуясь данными таблицы 4, можно попробовать пересчитать общепринятую величину среднего размера боковых групп DR . Величина DR может быть связана со средним расстоянием простыми формулами:

$$\begin{aligned} DR_{\text{НН}} &= d_{\text{НН}} / 2 - r_{\text{НН}}, \\ DR_{\text{НГ}} &= d_{\text{НГ}} / 2 - r_{\text{НН}} / 2 - r_{\text{НГ}} / 2, \\ DR_{\text{ГГ}} &= d_{\text{ГГ}} / 2 - r_{\text{НГ}}. \end{aligned}$$

При подстановке в формулы значений из таблиц 2 и 4, получаем три значения DR для спиральных пар разных типов: $DR_{\text{НН}} = 3.04 \text{ \AA}$, $DR_{\text{НГ}} = 3.95 \text{ \AA}$ и $DR_{\text{ГГ}} = 4.35 \text{ \AA}$. Общепринятое значение $DR_{\text{НН}} = 3.6 \text{ \AA}$ оказывается переоцененным для НН пар и недооцененным для НГ и ГГ спиральных пар.

На рисунке 6 представлены гистограммы распределения всех типов спиральных пар, принадлежащих множествам $\{A\}$ и $\{B\}$ в зависимости от межплоскостного (первая колонка) и минимального расстояний (вторая колонка). Из рисунка видно, что кривая распределения спиральных пар из множества $\{A\}$ в зависимости от межплоскостного расстояния является монотонной функцией, напоминающей экспоненциальную. Это справедливо для всех типов пар спиралей. Для спиральных пар из множества $\{B\}$ – функция имеет локальный максимум в районе 10 \AA . Причем, для пар типа НН этот

максимум является максимумом всего распределения, для других типов – максимум остается локальным. Так, для пар типа GH локальный максимум в районе 10 Å сохраняется, для пар типа GG – нет. Что касается дальнего действия распределения межплоскостного расстояния, то оно примерно одинаково для пар из {A} и {B}, распределение в обоих случаях простирается до величины порядка 35–40 Å.

Правая колонка графика – распределение спиральных пар, принадлежащих множествам {A} и {B}, в зависимости от минимального расстояния между осями спиралей. Здесь мы так же отмечаем существенные различия в поведении распределений. В отличие от левой колонки, кривая распределения спиральных пар в зависимости от минимального расстояния начинается не от нуля, а с некоторой величины, которая обусловлена минимальным радиусом, на который могут сблизиться оси двух спиралей. Он составляет около 2 Å в обоих случаях и связан с радиусом Ван-дер-Ваальсового взаимодействия атомов спиралей. Распределение спиральных пар из множества {A} в зависимости от минимального расстояния представляет собой довольно широкое ассиметричное распределение, подобное гамма-распределению. Дальнее действие довольно существенно, распределение простирается дальше 45–50 Å.

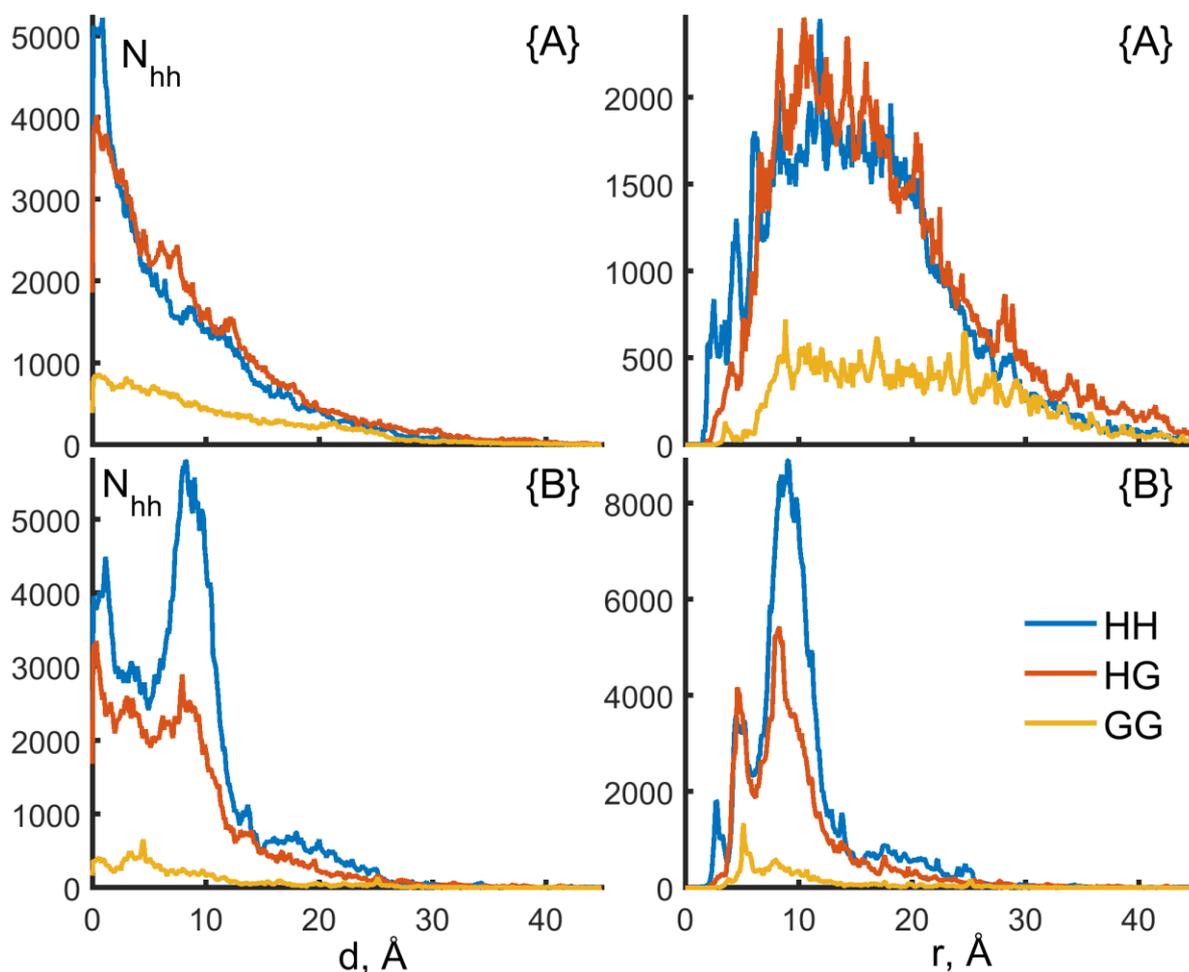


Рис. 6. Гистограммы распределения всех типов спиральных пар, принадлежащих множествам {A} и {B} в зависимости от межплоскостного (первая колонка) и минимального расстояний (вторая колонка).

Основные параметры гистограмм, приведенных на рисунке 6, помещены в таблице 4. Данные таблицы в целом подтверждают выводы, сделанные при сравнении графиков распределений. Можно отметить только, что оценки средних значений межплоскостных расстояний близки по величине с величиной дисперсии для

спиральных пар из $\{A\}$, в то же время для минимальных расстояний дисперсия в среднем в два раза ниже среднего. Для пар из множества $\{B\}$ дисперсия как минимального, так и межплоскостного расстояний всегда меньше, чем средние значения расстояний. Что касается сравнения средних минимальных расстояний, то для пар из $\{A\}$ они в полтора–два раза больше, чем для спиральных пар, принадлежащих множеству $\{B\}$.

Возвращаясь к сравнению графиков на рисунке 6, можно сделать один важный вывод: качественно отличаются гистограммы для межплоскостного расстояния, поскольку для спиральных пар без пересечений проекций (множество $\{A\}$) распределение носит монотонный характер, в отличие от спиральных пар с пересечением проекций (множество $\{B\}$). Что касается минимальных расстояний, то мы не видим качественных отличий в гистограммах, разве что количественные отличия, связанные с дальнодействием.

Перейдем к рассмотрению производных величин от минимального и межплоскостного расстояний: величины катета, который характеризует расстояние между проекциями спиралей на плоскости, а также логарифм плотности вероятности соотношения минимального расстояния к межплоскостному.

Таблица 5. Статистические оценки распределений величин (отношения r/d и катета $l = \sqrt{r^2 - d^2}$) для различных спиральных пар, принадлежащих множествам $\{A\}$ и $\{B\}$.

Статистические оценки	$r/d_{\text{НН}}$	$r/d_{\text{НГ}}$	$r/d_{\text{ГГ}}$	$l_{\text{НН}}$	$l_{\text{НГ}}$	$l_{\text{ГГ}}$
мода $\{A\}$	1.17	1.20	1.17	7.94	8.28	8.27
медиана $\{A\}$	2.36	2.19	2.04	11.43	11.92	14.41
среднее $\{A\}$	15.32	22.38	32.36	12.68	13.88	16.06
ср.кв.откл. $\{A\}$	519.44	3869.94	7241.39	6.76	7.63	8.52
мода $\{B\}$	1.00	1.00	1.00	4.26	4.22	3.84
медиана $\{B\}$	1.16	1.22	1.23	4.36	4.51	4.51
среднее $\{B\}$	6.72	11.02	5.75	4.51	4.64	4.67
ср.кв.откл. $\{B\}$	718.61	1360.74	131.78	2.62	2.39	2.24

На рисунке 7 слева приведены гистограммы распределения спиральных пар различных типов, принадлежащих множествам $\{A\}$ (верхний рисунок) и $\{B\}$ (нижний), в зависимости от катета. Сравнивая эти два рисунка, отмечаем несколько принципиальных отличий: первое отличие состоит в том, что катет для пар из $\{B\}$ имеет множество значений близких к нулю, в отличие от пар из множества $\{A\}$, гистограмма которых в нуле имеет нулевое значение. Второе отличие – гистограмма для пар из $\{A\}$ – это широкое асимметричное распределение, подобное распределению в зависимости от минимального расстояния для того же множества (см. рис. 6). Напротив, гистограмма распределения пар из $\{B\}$ в зависимости от катета определена в ограниченной области величин катета, максимальные значения которого для разных типов спиралей отличаются. Эти максимальные значения равны 11.77 Å, 11.37 Å и 10.95 Å для спиральных пар НН, НГ и ГГ типа, соответственно. Эти величины достаточно близки к теоретическим максимальным величинам 11.8 Å, 11.4 Å и 11 Å, полученным в предположении, что максимальное значение катета равно тому минимальному расстоянию, на которое могут сблизиться две спирали, если оси их параллельны. Если значение катета оказывается большим, то спиральная пара уже не имеет пересечений проекций, соответственно не относится к множеству $\{B\}$. В

таблице 5 приведены основные параметры гистограмм, изображенных на рисунке 7. Сравнивая средние значения и дисперсии для величины катета спиральных пар из $\{A\}$ со средними значениями минимального расстояния для того же множества, можно сделать вывод о меньшем дальнодействии распределения катета. Качественно данное распределение очень похоже на распределение пар в зависимости от минимального расстояния, оба распределения носят характер гамма-распределения. Далее мы проверим это предположение.

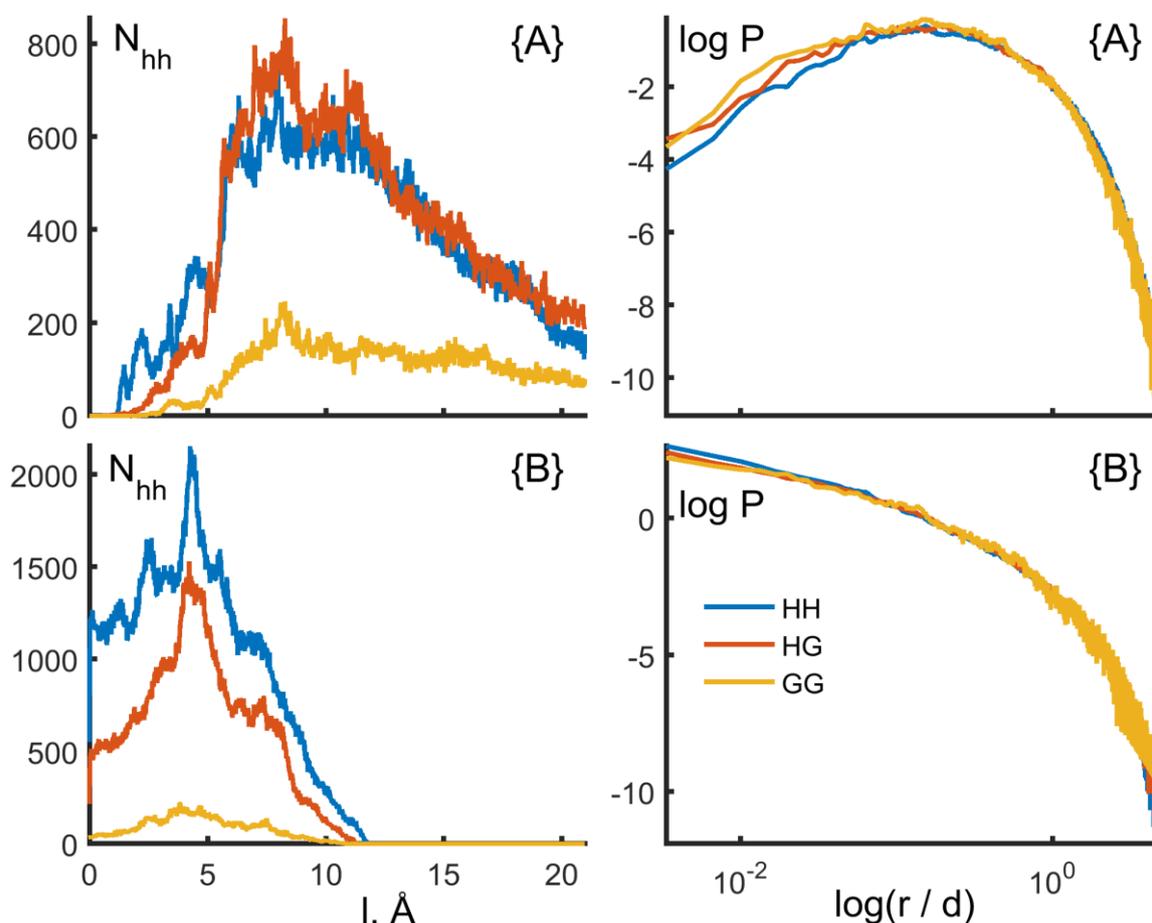


Рис. 7. Гистограмма распределения всех типов спиральных пар в зависимости от величины катета $l = \sqrt{r^2 - d^2}$ (слева) и логарифм вероятности отношения r/d , вычисленного для каждой спиральной пары, принадлежащей множествам $\{A\}$ и $\{B\}$ (справа).

Перейдем к анализу логарифма плотности вероятности отношения минимального к межплоскостному расстоянию. На рисунке 7 справа приведены эти логарифмы отношений для спиральных пар различных типов, принадлежащих разным множествам. Логарифмы плотности вероятности изображены как функции логарифма отношения r/d . Плотности вероятности для спиральных пар, принадлежащих различным множествам, сильно отличаются только вблизи нуля величины $\log(r/d)$, поэтому абсцисса графика выбрана в логарифмическом масштабе. Принципиальным отличием плотности вероятности для пар из множества $\{A\}$ является ее немонотонное поведение. Для пар из $\{B\}$ плотность вероятности – монотонно убывающая функция. В таблице 5 приведены основные характеристики распределений. Максимум плотности вероятности (значение моды) для пар из $\{A\}$ наблюдается при соотношении $r/d = 1.17$. Из-за того, что масштаб абсциссы выбран логарифмическим, мы ничего не можем сказать про асимптотику плотности вероятности. Далее мы покажем, что эта асимптотика пропорциональна $P(r/d) \rightarrow 1/(r/d)^2$ для всех спиральных пар,

независимо от принадлежности к множествам. Именно этот факт не позволяет нам вычислить среднее значение, поскольку его величина логарифмически расходится при увеличении выборки. В таблице 6 приведены оценки средних на конечной выборке, а также оценки дисперсии средних. Большие значения дисперсии в сравнении со средними указывают на факт расходимости.

В следующей части работы мы более детально проанализируем гистограммы распределения расстояний, в частности, проверим наше предположение о том, что гистограммы распределений величин минимального расстояния для спиральных пар могут быть описаны формулой гамма-распределения. Проверим, что этой же формулой описывается распределение величины катета для спиральных пар из множества $\{A\}$. Проведем анализ распределения межплоскостного расстояния и соотношения r/d . Мы предложим модель случайной геометрии спиральной пары, которая в целом описывает распределение r/d , в особенности асимптотическое поведение этого распределения. Проанализируем основные отличия распределения r/d для спиральных пар, найденных из эксперимента с моделью случайной геометрии.

РЕГРЕССИОННЫЙ АНАЛИЗ ГИСТОГРАММ РАСПРЕДЕЛЕНИЯ СПИРАЛЬНЫХ ПАР В ЗАВИСИМОСТИ ОТ РАССТОЯНИЙ МЕЖДУ СПИРАЛЯМИ

Перейдем от гистограмм, которые мы рассматривали раньше отдельно для каждого типа спиральных пар, к усредненной по всем спиральным парам плотности вероятности. Для этого достаточно суммировать все гистограммы, относящиеся к разным типам спиральных пар, и нормировать эту сумму. Всего мы рассматриваем четыре плотности вероятности, обозначая их: P_d , P_r , P_l и P . Нижний индекс обозначает расстояние: межплоскостное d , минимальное r и катет l . Плотность вероятности без индекса относится к соотношению r/d .

Нас будет интересовать асимптотическое поведение плотностей вероятности. Как уже говорилось, мы предполагаем, что все три плотности вероятности наилучшим образом могут быть описаны формулой гамма-распределения. По этой причине мы будем искать асимптотику плотностей вероятности как $x^n \exp(-sx)$, где n – целое число, а $s < 0$. Относительно выбора степени n у нас есть следующие физические соображения: минимальное расстояние связано с вектором в трехмерном пространстве, потому он предполагает при вычислении плотности вероятности изотропный множитель $4\pi r^2 dr$. А межплоскостное расстояние и катет связаны с векторами на плоскости. Для катета, очевидно, – это просто проекция минимального расстояния на плоскость пересечения проекций спиралей. А для межплоскостного – не так очевидно, но если ввести плоскость ортогональную плоскости пересечения проекций и катету, то тоже понятно. Изотропный множитель для вектора на плоскости это $2\pi r dr$. Потому правильно искать асимптотики для минимального расстояния как $x^2 \exp(-sx)$, а для катета и межплоскостного как $x \exp(-sx)$.

Начнем с анализа плотностей вероятностей расстояний для спиральных пар, проекции которых не имеют пересечений. Поскольку мы предполагаем, что плотности вероятностей экспоненциально спадают с расстоянием, проведем регрессионный анализ их логарифмов.

На рисунке 8 приведены: логарифм плотности вероятности межплоскостного расстояния, деленный на расстояние, P_d/d ; логарифм плотности вероятности минимального расстояния, деленный на квадрат расстояния, P_r/r^2 ; логарифм плотности вероятности катета, также деленный на расстояние, P_l/l . На тех же рисунках приведены прямые линейной регрессии, построенные по методу робастного оценивания параметров линейной регрессии [26]. Исходя из общего вида графиков, мы выбирали интервалы, на которых логарифм вероятности спадает линейно с расстоянием.

Все три линейные зависимости мы оценивали для данных в интервале 15–60 Å. Коэффициенты линейной регрессии приведены на рисунке. Отмечаем достаточно хорошее согласие логарифмов вероятностей с регрессионными кривыми, за исключением малой области до 10–12 Å. Кроме этого, следует отметить, что и сами коэффициенты s оказываются близкими по величине для всех расстояний. Последнее свидетельствует в пользу того, что наше предположение об изотропности векторов расстояний является хорошим приближением в случае спиральных пар из множества $\{A\}$.

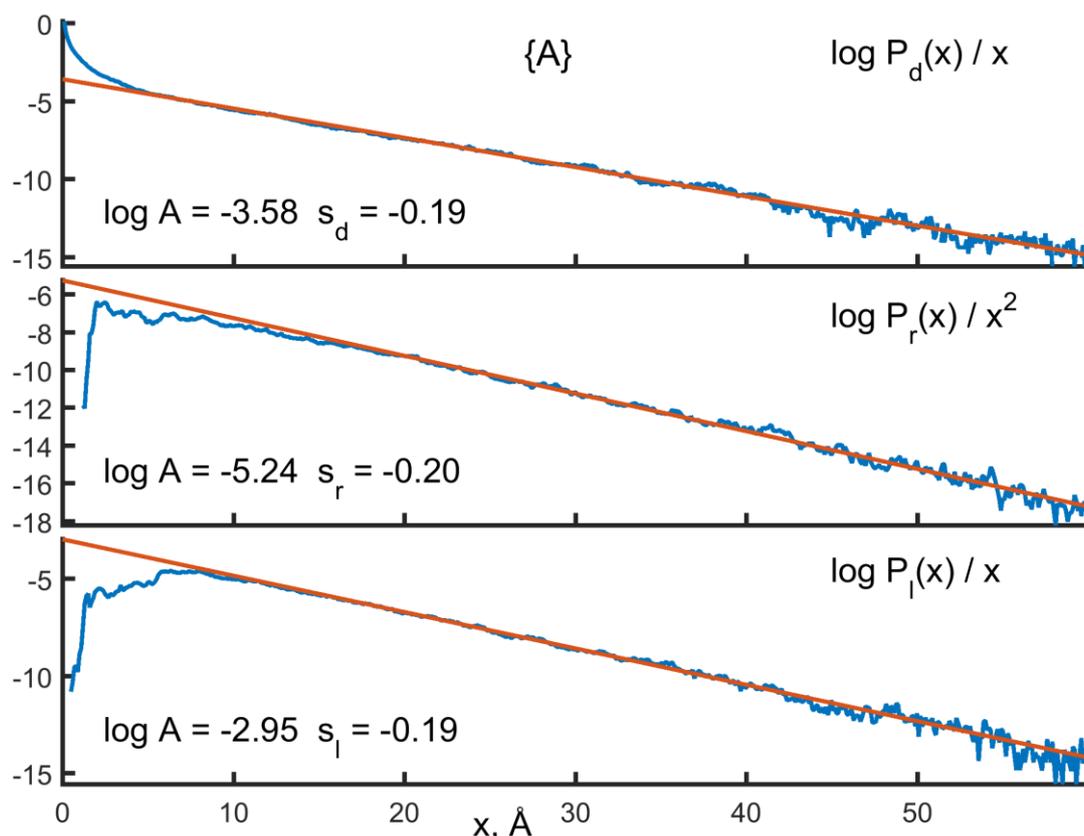


Рис. 8. Регрессионный анализ вероятности обнаружить данное межплоскостное, минимальное расстояние, а также величину катета для спиральных пар из $\{A\}$.

Теперь интересно сравнить асимптотическое поведение плотностей вероятности для спиральных пар из множеств $\{B\}$ и $\{C\}$. Для пар, принадлежащих данным множествам, мы анализируем только межплоскостное и минимальное расстояния, поскольку катет в случае $\{B\}$ ограничен по величине максимальным значением, а в случае $\{C\}$ равен нулю.

На рисунке 9 приведены результаты регрессионного анализа для пар из множеств $\{B\}$ и $\{C\}$. Все расчеты проведены так же, как и для данных, приведенных на рисунке 7, с той поправкой, что линейные зависимости анализировались в интервале 15–45 Å для пар из $\{B\}$ и 25–35 Å для пар из $\{C\}$. Сравнивая коэффициенты регрессии s , отмечаем, что они различны для разных расстояний. Это говорит о том, что распределения расстояний в парах с пересечением проекций спиралей неизотропны. Кроме этого, величины коэффициентов регрессии больше по модулю, чем в случае пар из $\{A\}$. Это означает, что пары с пересечением проекций более компактны.

Можно так же сказать, что анализ расстояний спиральных пар без пересечений отражает только среднее распределение размеров свернутой аминокислотной цепи, в то время как для пар с пересечением проекций расстояния относятся именно к описанию

координаций спиралей друг относительно друга. Или другими словами, наличие пересечений проекций спиралей обуславливает их координационную близость. Факт *a priori* вовсе не очевидный.

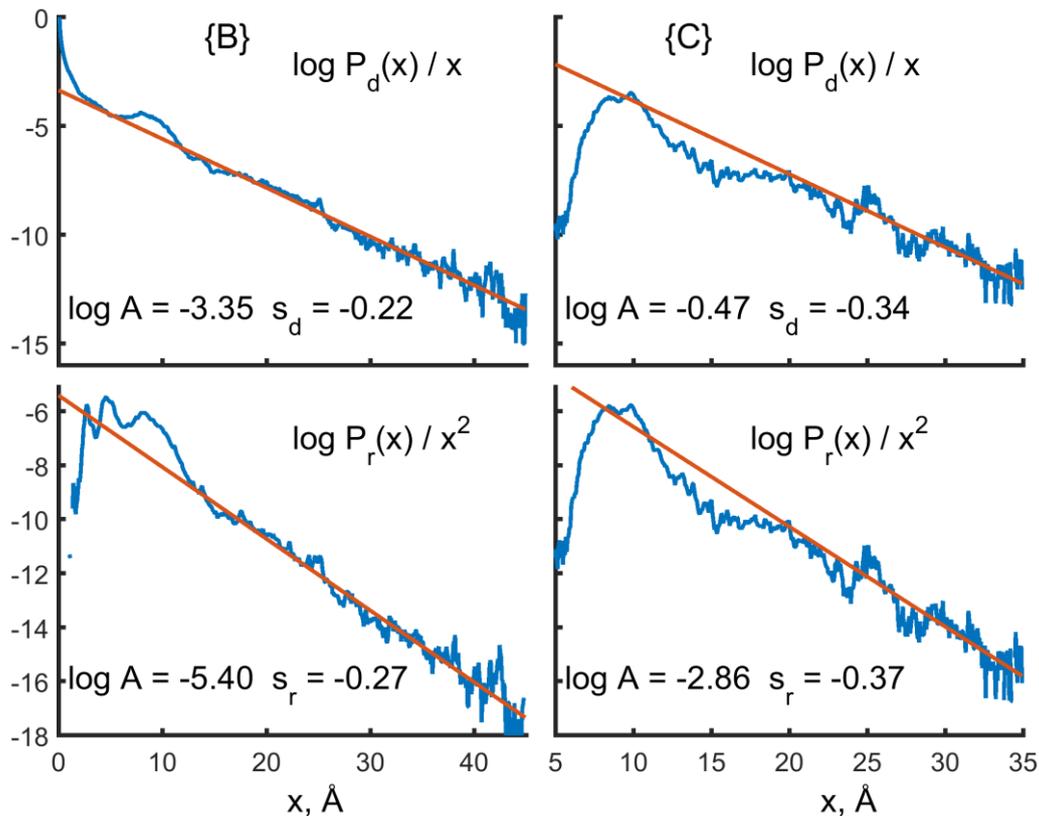


Рис. 9. Регрессионный анализ вероятности обнаружить данное межплоскостное, минимальное расстояние для спиральных пар из {B} и {C}.

Мы видим, что сам по себе критерий пересечения проекций спиралей позволяет разделить спирали на два множества. Первое ничем не выделяет спиральную пару от других фрагментов аминокислотной цепи в отношении расстояний между этими фрагментами. В другое множество входят компактно координированные спиральные пары. Для дальнейшего анализа статистических характеристик этих множеств, и, чтобы уйти от влияния конкретных размеров свернутой аминокислотной цепи на результаты, исследуем распределение соотношения минимального расстояния в спиральной паре к ее межплоскостному расстоянию. Общий вид таких распределений приведен на рисунке 7. Мы отмечали, что эти распределения имеют далекодействующую асимптотику, которая приводит к расходимости средних величин, а также то, что отличия в распределениях для пар с пересечением и без наблюдаются только в малой окрестности вблизи нуля для величины $\log(r/d)$. Близость распределений позволяет предположить, что их свойства лишь в малой степени связаны с тем, что данные для них брались из анализа геометрии белковых молекул. В качестве эталона сравнения предлагается модель случайной геометрии спиральной пары. Приведем детали алгоритма генерации случайных геометрий для четырехточечной модели спиральной пары.

Выберем три сочлененных отрезка $[A_1, A_2]$, $[A_2, B_1]$ и $[B_1, B_2]$ (см. рис. 3). Длины отрезков l_1 , l_2 , l_3 выбираются из случайного равномерного распределения от нуля до некоторого максимального значения L . Выберем также два случайных вектора ξ_1 и ξ_2 в трехмерном пространстве с единичными модулями. Расположим второй отрезок вдоль координатной оси x . Тогда координаты точечной модели получаются по следующим формулам:

$$B_2 = [0, 0, 0],$$

$$A_1 = [l_2, 0, 0],$$

$$B_1 = B_2 - l_1 \xi_1,$$

$$A_2 = A_1 - l_3 \xi_2.$$

Вычисляя минимальное и межплоскостное расстояние между первым $[A_1, A_2]$ и третьим $[B_1, B_2]$ отрезком, затем набирая статистику для соотношения этих расстояний, получим искомую гистограмму для модели случайной геометрии спиральных пар. Мы выбирали разные значения L . Расчеты показывают, что гистограмма не зависит от величины L , начиная с некоторого значения.

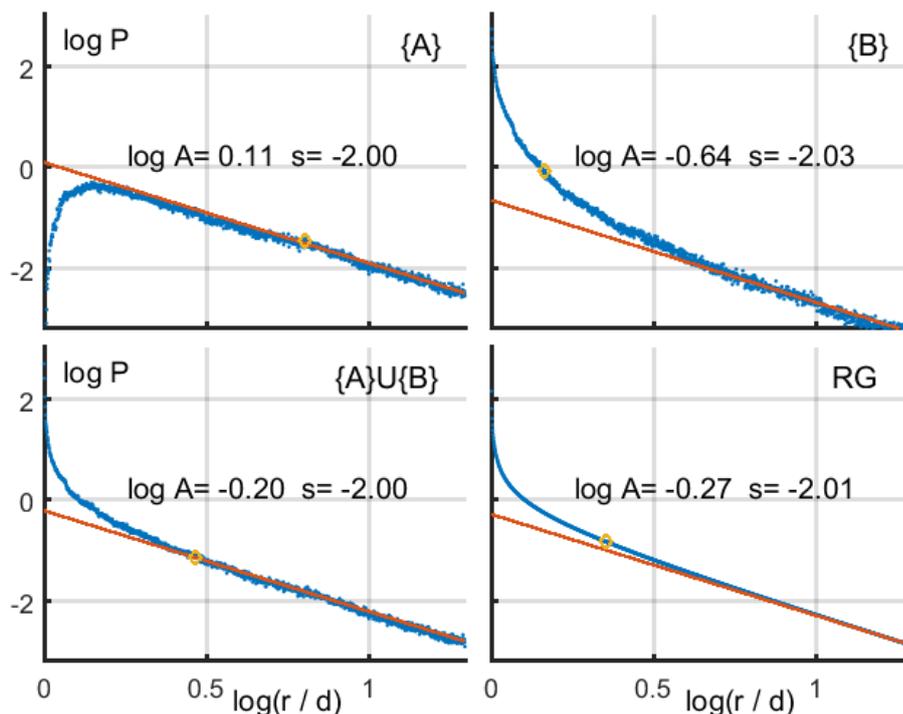


Рис. 10. Регрессионный анализ вероятности отношения r/d для спиральных пар из множеств $\{A\}$ и $\{B\}$ в сравнении с объединением этих множеств и случайным распределением.

Проведем регрессионный анализ вероятности $\log(r/d)$ для спиральных пар в сравнении со случайной моделью. При вычислении вероятности мы не различаем типы спиральных пар. На рисунке 7 (колонка справа) мы уже приводили эти данные для пар из множеств $\{A\}$ и $\{B\}$. Из рисунка видно, что тип спиральной пары практически не влияет на характер распределения вероятности. Помимо суммирования по всем типам спиральных пар, мы расширим сравнение, включив в него вероятность $\log(r/d)$, вычисленную для объединения множеств $\{A\}$ и $\{B\}$. Асимптотическая часть логарифма вероятности как функция логарифма соотношения расстояний аппроксимировалась линейной зависимостью по методу [27]. Данные для регрессии выбирались в диапазоне $0.5 < \log(r/d) < 5$. Параметр L для случайной геометрии спиральной пары был равен 30 \AA . Мы также проверяли и другие значения параметра, распределение при этом не меняется.

На рисунке 10 приведен логарифм вероятности обнаружить отношение минимального расстояния между осями спиралей к межплоскостному расстоянию, вычисленного для каждой спиральной пары. Логарифм вероятности приведен в зависимости от логарифма этого отношения. В таком масштабировании асимптотика вероятности – линейная функция. Приведены сами вероятности и их линейные

аппроксимации для четырех случаев. Верхняя строка – для спиральных пар, принадлежащих множествам $\{A\}$ и $\{B\}$, нижняя строка – левый график – вероятность для всех спиральных пар без разбора типа множества. Правый график – вероятность, вычисленная для случайной геометрии. Маркерами на графиках отмечены значения медианы распределений. Также на рисунках приведены значения параметров линейной регрессии.

Видно, что нижний ряд зависимостей на рисунке 10 практически совпадает, что позволяет нам сделать вывод о том, что точечная геометрия спиральных пар в белках по этому критерию не отличается от случайной. Этот вывод подтверждает и близость значений медианы, которая равна 1.6 для всех спиральных пар $\{A\} \cup \{B\}$ и 1.43 для спиральных пар со случайной геометрией. В то же время, медианы данных распределений для спиральных пар из множеств $\{A\}$ и $\{B\}$ отличаются почти в два раза (2.4 для $\{A\}$ и 1.18 для $\{B\}$).

Асимптотическое поведение во всех случаях одинаково с поправкой на ошибку регрессии, вероятность соотношения r/d пропорциональна величине $(d/r)^2$ для $r/d > 1.6$.

ЗАКЛЮЧЕНИЕ

Используя точечную модель спиральных пар, мы отобрали множество исследуемых структур всех белковых молекул, зарегистрированных в PDB. Множество спиральных пар, полученное согласно разработанным правилам, разбито на три подмножества по критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей. Проведен анализ статистических свойств различных расстояний между рядом стоящими спиральями в белковых цепочках.

Показано, что распределения расстояний для спиральных пар, проекции которых не имеют пересечений, имеют более дальнедействующий характер, чем те, проекции которых пересекаются.

При помощи регрессионного анализа исследован характер распределений, в частности, показано, что в подмножестве без пересечений распределения различных расстояний между осями спиралей относятся к гамма-распределениям.

Из сравнений распределений для отношения минимального расстояния между осями спиралей к межплоскостному расстоянию с таким же для сгенерированных случайных пар показано, что в целом все спиральные пары мало отличаются от случайных, однако подмножества пар с пересечением проекций имеет распределение с большей вероятностью иметь малое отношение минимального к межплоскостному расстоянию. Напротив, распределение в подмножестве пар без пересечения проекций имеет существенный провал вероятности в области малых отношений минимального к межплоскостному расстоянию.

Обосновывается вывод о том, что спиральные пары, проекции спиралей которых пересекаются, дополнительно стабилизируются за счет внутреннего взаимодействия. Не только α -спирали образуют супервторичные структуры, спиральные пар HG-типа составляют такое же количество супервторичных структур. Критерием того, что исследуемый мотив есть супервторичная структура, является координация цилиндров спиралей в контакте. Оказалось, что такая координация обеспечивается условием пересечения проекций на плоскость осей цилиндров, что не очевидно. В процентном отношении такому условию удовлетворяет больше половины от общего числа найденных спиральных пар.

Средний размер боковой цепи DR для спиральных пар типа HG и GG, вычисленный по расположению цилиндров, оказывается больше общепризнанного значения (3.6 Å). Для спиральных пар, состоящих из двух α -спиралей, средний размер боковой цепи меньше общепризнанного значения.

Авторы выражают благодарность Соболеву О.В. за помощь в получении данных.

Работа выполнена при поддержке грантов РФФИ № 16-01-00692, № 14-07-00924 и № 15-29-07063.

СПИСОК ЛИТЕРАТУРЫ

1. Efimov A.V. Standard structures in proteins. *Prog. Biophys. Molec. Biol.* 1993. V. 60. P. 201–239.
2. Gordeev A.B., Kargatov A.M., Efimov A.V. PCBOST: Protein classification based on structural trees. *Biochemical and Biophysical Research Communications*. 2010. V. 397. P. 470–471.
3. Efimov A.V. Super-secondary structures and modeling of protein folds. In: *Methods in Molecular Biology*. Ed. Kister A.E. Clifton: Humana Press, 2013. V. 932. P. 177–189.
4. Brazhnikov E.V., Efimov A.V. Structure of α - α -hairpins with short connections in globular proteins. *Molecular Biology*. 2001. V. 35. № 1. P. 89–97.
5. Ефимов А.В. L-образная структура из двух α -спиралей с остатком пролина между ними. *Молекулярная биология*. 1992. Т. 26. С. 1370–1376.
6. Ефимов А.В. Новая супервторичная структура белков: α - α -уголок. *Молекулярная биология*. 1984. Т. 18. С. 1524–1537.
7. Финкельштейн А.В. *Физика белковых молекул*. М.–Ижевск: АНО «Ижевский институт компьютерных исследований», 2014. 424 с.
8. Птицын О.Б., Филькенштейн А.В. Проблемы предсказания структуры белка. В: *Итоги науки и техники*. Под ред. Волькенштейна М.Д. М: ВИНТИ, 1979. Т. 15. С. 6–41. (Серия «Молекулярная биология»).
9. Шульц Г.Е., Ширмер Р.Х. *Принципы структурной организации белков*. М.: Мир, 1982. 354 с.
10. Miller S., Janin J., Lesk A.M., Chothia C. Interior and surface of monomeric proteins. *J. Molecular Biology*. 1987. V. 196. P. 641–656.
11. Creighton T.E. *Proteins*. 2-nd edn. N.Y.: W.H. Freeman & Co, 1991
12. Степанов В.М. *Молекулярная биология. Структура и функции белков*. М.: Высшая школа, 1996. 336 с.
13. Финкельштейн А.В. Электростатические взаимодействия заряженных групп в водной среде и их влияние на образование вторичных структур полипептидной цепи. *Молекулярная биология*. 1970. Т. 11. С. 811–819.
14. Fersht A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. N.Y.: W.H. Freeman & Co, 1999.
15. Lim V.I. Structural principles of the globular organization of protein chains. A stereochemical theory of globular protein secondary structure. *J. Molecular Biology*. 1974. V. 88. P. 857–872.
16. Lim V.I. Algorithm for prediction of α -helices and β -structural regions in globular proteins. *J. Molecular Biology*. 1974. V. 88. P. 873–894.
17. Wierenga R.K., Terpstra P., Hol W.G.S. Prediction of the occurrence of the ADP-binding $\beta\alpha\beta$ -fold in proteins, using an amino acid sequence fingerprint. *J. Molecular Biology*. 1986. V. 187. P. 101–107.
18. Руднев В.Р., Панкратов А.Н., Куликова Л.И., Дедус Ф.Ф., Тихонов Д.А., Ефимов А.В. Распознавание и анализ устойчивости структурных мотивов типа α - α -уголок в глобулярных белках. *Математическая биология и биоинформатика*. 2013. Т. 8. № 2. С. 398–406. doi: [10.17537/2013.8.398](https://doi.org/10.17537/2013.8.398).
19. Руднев В.Р., Панкратов А.Н., Куликова Л.И., Дедус Ф.Ф., Тихонов Д.А., Ефимов А.В. Конформационный анализ структурных мотивов типа α - α -уголок в вычислительном эксперименте молекулярной динамики. *Математическая биология и биоинформатика*. 2014. Т. 9. № 2. С. 575–584. doi: [10.17537/2014.9.575](https://doi.org/10.17537/2014.9.575).
20. Дедус Ф.Ф., Куликова Л.И., Панкратов А.Н., Тетуев Р.К. *Классические ортогональные базисы в задачах аналитического описания и обработки*

информационных сигналов: Учебное пособие. М.: Издательский отдел ВМиК МГУ, 2004. 147 с.

21. Pankratov A.N., Gorchakov M.A., Dedus F.F., Dolotova N.S., Kulikova L.I., Makhortykh S.A., Nazipova N.N., Novikova D.A., Olshevets M.M., Pyatkov M.I., Rudnev V.R., Tetuev R.K., Filippov V.V. Spectral Analysis for identification and visualization of repeats in genetic sequences. *Pattern Recognition and Image Analysis*. 2009. V. 19. № 4. P. 687–692.
22. Tsai F.C., Sherman J.C. Circular dichroism analysis of a synthetic peptide corresponding to the α, α -corner motif of hemoglobin. *Biochemical and Biophysical Research Communications*. 1993. V. 196. № 1. P. 435–439.
23. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983. V. 22. № 12. P. 2577–2637.
24. Kabsch W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica*. 1976. V. 32. P. 922–923.
25. Kabsch W. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica*. 1978. V. 34. P. 827–828.
26. Legland D. *MatGeom: Matlab geometry toolbox for 2D/3D geometric computing*. URL: <http://github.com/dlegland/matGeom> (дата обращения: 11.03.2016).
27. Holland P.W., Welsch R.E. Robust regression using iteratively reweighted least-squares. *Communications in Statistic – Theory and Methods*. 1977. V. 6. № 9. P. 813–827.

Рукопись поступила в редакцию 27.04.2016.

Дата опубликования 04.08.2016.