Математическая биология и биоинформатика
2017. Т. 12. № 1. С. 83–101. doi: 10.17537/2017.12.83

========== **БИОИНФОРМАТИКА** ==========

УДК: 519

Исследование межспиральных углов в структурных мотивах, образованных двумя спиралями

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 анализ распределения межспиральных углов в парах связанных между собой перетяжками спиралей в пространственных структурах белковых молекул. Были разработаны правила отбора спиральных пар в структурах белковых молекул из Protein Data Bank. Полученное множество спиральных пар было проанализировано с целью его классификации и установления закономерностей структурной организации. По критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей, полученное множество было разбито на три подмножества. Показано, что распределение всех типов спиральных пар, не имеющих пересечения проекций спиралей, охватывает практически весь диапазон углов с одним максимумом в области прямого угла. Большинство пар этого множества составляют спиральные пары, состоящие из α- и 3_{10}-спиралей, а множества с пересечением только проекций спиралей – спиральные пары, образованные двумя α-спиралями. Также показано, что образованные двумя α-спиралями спиральные пары составляют абсолютное большинство пар множества с пересечением проекций и осей спиралей. При этом значительная часть указанных пар имеет острый угол $(20° \le \varphi \le 60°)$ между осями спиралей. Проведен анализ распределения всех типов спиральных пар, принадлежащих различным множествам, в зависимости от длины перетяжки. Показано, что во всех множествах больше всего исследуемых структур с короткой перетяжкой.

Ключевые слова: *структурные мотивы белков, точечная модель, спиральные пары в белковых молекулах, угол между осями спиралей.*

ВВЕДЕНИЕ

Ранее были разработаны простые правила укладки белковой цепи, приведшие к построению структурных деревьев, которые описывают многообразие белковых структур [1, 2]. Это поставило задачу автоматического распознавания и предсказания различных структурных мотивов белков.

В настоящее время объектом пристального внимания исследователей являются структурные мотивы, имеющие уникальные укладки полипептидной цепи в пространстве [1]. Такой интерес к структурным мотивам обусловлен уникальностью структур и их возможностью быть зародышами в процессе сворачивания белков, к которым остальные участки цепи могут пристраиваться в соответствии с простыми правилами. Однако, независимо от того, по какому механизму происходит

*dmitry.tikhonov@gmail.com
**likulikova@mail.ru

сворачивание белка, структурные мотивы могут быть использованы в качестве стартовых структур для поиска возможных укладок полипептидной цепи при моделировании структуры белков [1, 3].

Простейшими являются структурные мотивы, состоящие из двух элементов вторичной структуры, имеющих уникальные укладки полипептидной цепи в пространстве [1–4]. В предыдущих работах [5, 6] решалась задача распознавания структурных мотивов белков типа α-α-уголков с помощью аналитического описания основной цепи белковой глобулы и спектрального метода распознавания повторов [7–9]. Объектом исследований был указанный тип структурных мотивов с короткой перетяжкой. Известно, что α-α-уголки являются одним из часто встречаемых структурных мотивов в белках [3, 10–15]. 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. Была проведена верификация найденных структур по конформационному шаблону, охарактеризованному предельными значениями углов на карте Рамачандрана. Методом молекулярной динамики нами было показано, что α-α-уголки с короткой перетяжкой как автономные структуры устойчивы в водной среде [5, 6]. Ранее устойчивость α-α-уголков была косвенно доказана экспериментально (*in vitro*) [16]. Независимо выдвинутая нами гипотеза об автономной устойчивости структурных мотивов была проверена *in silico* в вычислительных экспериментах молекулярной динамики.

Из литературы известно, что α-спирали в белках упакованы плотно. Структурные мотивы, образованные двумя α-спиралями, расположенными в полипептидной цепи друг за другом и соединенных между собой перетяжками описаны в работах [2–4]. Они представляют собой компактные пространственные структуры. Также 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 [2]. Исходя из вышесказанного, кажется весьма актуальным создание множества, включающего в себя все возможные указанные структурные мотивы и их дальнейшее исследование.

В опубликованной ранее работе [17] изучались структуры, в образовании которых участвуют две спирали любого типа: α-спирали, 3_{10}-спирали и π-спирали. Было показано, что 72.16 % от общего числа всех спиралей приходится на α-спирали; 27.73 % составляют 3_{10}-спирали; оставшиеся 0.1 % – π-спирали. В работе были отобраны из Protein Data Bank [18] все структурные мотивы, образованные двумя спиралями любого типа, расположенными в полипептидной цепи друг за другом и соединенными между собой перетяжками различной длины и конформации. Был проведен статистический анализ распределения межспиральных расстояний в спиральных парах, входящих в эту базу данных.

Мы продолжаем изучение найденных в предыдущей работе структур с точки зрения распределения углов между осями спиралей во всем отобранном множестве спиральных структур. Принципиальным отличием нашей работы от ранее известных публикаций по исследованию межспиральных углов в белковых структурах [19–23] является то, что мы изучаем углы между осями соединенных (связанных) между собой перетяжкой спиралей. Перетяжка представляет собой участок аминокислотной последовательности длиной в одну или несколько аминокислот, вторичная структура которой отличается от спиральной.

Нужно отметить, что, вероятно, эти углы являются малоинформативными для биологов и биохимиков, поскольку их определение не позволяет различать между собой известные супервторичные структуры, например, L-структуры и α-α-уголки и др.

Исследование межспиральных углов носит формальный характер, мы анализируем все характеристики спиральных пар (расстояния между спиралями, длина перетяжки, межспиральные углы, площадь и периметр полигона пересечения проекций спиралей и др.). Таким образом, подчеркиваем, вычисляемые нами углы φ между осями спиралей не являются межспиральными торсионными углами Ω. Анализу торсионных углов Ω в спиральных парах будет посвящена следующая работа.

Также необходимо отметить, что данная работа проводилась со спиральными парами, которые отбирались из Банка белковых данных без учета классификации белков. Данная классификация ведется исходя из особенностей вторичной и третичной структуры. Известно, что все множество белковых структур подразделяется по наличию и сочетанию α-спиралей и β-структур на несколько классов [24]: α-, β-, (α/β)-, (α + β)-белки и другие. Так, к α-белкам относятся белки, имеющие в своем составе только α-спирали, яркими представителями этого класса белков являются гемоглобин, миоглобин. Белки, имеющие в своем составе только β-структуры, относятся к β-белкам, примером таких белков является иммуноглобулин. В составе (α + β)-белков присутствуют – те и другие структуры в пределах одной полипептидной цепи (молекулы лизоцима), а класс α/β-белков содержит множество α- и β-структур, чередующихся вдоль полипептидной цепи (лактатдегидрогеназа). В дальнейшем мы планируем провести исследование распределения межспиральных углов и других характеристик спиральных пар в каждом отдельно взятом классе белков.

МЕТОДЫ

Для исследования структурных мотивов белковых молекул были разработаны специальные правила для распознавания и отбора структур, основными из которых являются нижеследующие.

1. Анализ вторичной структуры аминокислотной цепи осуществлялся с помощью метода, разработанного авторами Dictionary for Secondary Structure of Protein [25]. Находим участки спиралей. Спирали могут быть различных типов. Первый тип – это α-спираль или H в обозначениях DSSP. Второй тип (G-спираль) включает в себя 3_{10}-спирали. Третий тип (I) – это π-спираль.

2. Кандидатом на искомую структуру считается участок белка, который содержит две спирали и участок белковой нити между спиралями, который называется перетяжкой. Все подобные структуры можно также описать согласно типам входящих в них спиралей.

3. Для каждой спирали структуры находим ось цилиндра, на который эта спираль намотана. Это несложно выполнить, зная параметры идеальных спиралей по методу наименьших квадратов. Ось цилиндра будет найдена тем точнее, чем ближе спираль к идеальной. Качество оценки осей характеризуется величиной среднеквадратичного отклонения. Мы отбираем те спирали (и соответственно структуры), для которых точность оценки оси удовлетворяет наперед заданному критерию.

4. Две оси спиралей полностью определяют пространственную структуру. Известно, что через две не пересекающиеся прямые в пространстве можно провести две параллельные плоскости так, что первая ось принадлежит первой плоскости, а вторая – второй. Ось, лежащую в одной плоскости, можно спроецировать на другую плоскость. Таким образом, пространственная структура будет полностью описана расстоянием между параллельными плоскостями и проекциями осей спиралей на плоскость.

5. Все найденные согласно разработанным правилам спиральные пары разбиваем на три множества. Эта разбиение проводится по критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей:

– множество $\{A\}$ образуют все спиральные пары, у которых проекции спиралей не имеют пересечений;

– 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;

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

6. По известным координатам точек A_1, A_2, B_1 и B_2, являющимися начальными и конечными точками осей двух спиралей [12], вычисляем координаты векторов $\overrightarrow{A_2A_1}$, $\overrightarrow{B_1B_2}$ и находим косинус угла между этими векторами $\cos(\overrightarrow{A_2A_1}, \overrightarrow{B_1B_2})$. Точечная модель спиральной пары приведена на рисунке 1. Показаны 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 спиралями спиральной пары. С точки зрения описания взаимного расположения спиралей совершенно естественно возникают три расстояния. Во-первых, это межплоскостное расстояние. Как известно, через две непересекающихся 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 d. Второе расстояние спиральной пары – кратчайшее расстояние между отрезками, которые являются осями цилиндров. Кратчайшее расстояние обозначим r. Очевидно, что минимальное расстояние спиральной пары r всегда будет больше или равно межплоскостному расстоянию d, поэтому можно ввести величину катета l, определив его как $l = \sqrt{r^2 - d^2}$. Катет l является третьим расстоянием, которое описывает относительное расположение спиралей в спиральной паре.

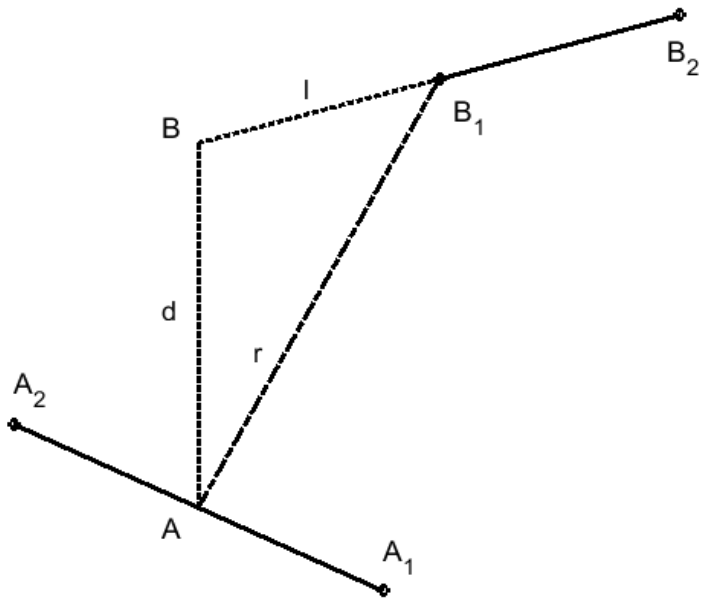

Рис. 1. Геометрия спиральной пары – четыре точки, образующие два отрезка в пространстве.

По разработанному алгоритму были созданы база структурных мотивов и комплекс программ обработки структур, позволяющий получать данные для каждой структуры, в частности, межспиральные углы φ отобранных спиральных пар.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Ранее опубликованная нами работа [12] была посвящена анализу распределения межспиральных расстояний в парах связанных между собой перетяжками спиралей в пространственных структурах белковых молекул. Мы продолжаем исследовать найденные в предыдущей работе структуры, но уже с точки зрения распределения углов между осями спиралей. Сначала дадим краткое обсуждение проведенной работы.

Для решения поставленной задачи предложена точечная модель описания искомых структур, состоящих из двух спиралей с различной длиной перетяжки между ними. Просматривая непрерывные цепочки всех белков, отбираем структуры, состоящие из двух спиралей и нерегулярных участков различной длины между ними.

Каждая спираль представляется в виде цилиндра, по методу наименьших квадратов находится ось цилиндра. Точность определения оси цилиндра оценивается величиной среднеквадратичного отклонения. Таким образом, в дальнейшем рассматриваются только те структуры, в состав которых входят спирали максимально приближенные к идеальной, и точность оценки оси спирали соответствует наперёд заданному критерию [26, 27]. На рисунке 2 дано геометрическое представление структуры, образованной двумя спиралями и перетяжкой между ними. Рисунок демонстрирует пример спиральной пары, фрагмент аминокислотной цепи белка из Банка белковых данных (PDB ID 3VKH, chain B, участок цепи: 2846–2891). Показаны аппроксимирующие спираль цилиндры, а также плоскости, проходящие через оси цилиндров. Кривая аппроксимирована по положениям C_α-атомов белковой цепи. Данная спиральная пара состоит из двух α-спиралей (двух спиралей типа H), длина перетяжки равна 12. На рисунке обозначены отрезки (красная и синяя), это наименьшие расстояния (6.67 Å и 7.43 Å, соответственно) между спиралями. Об этом измерении в исследуемых структурах будет детально описано в следующей работе.

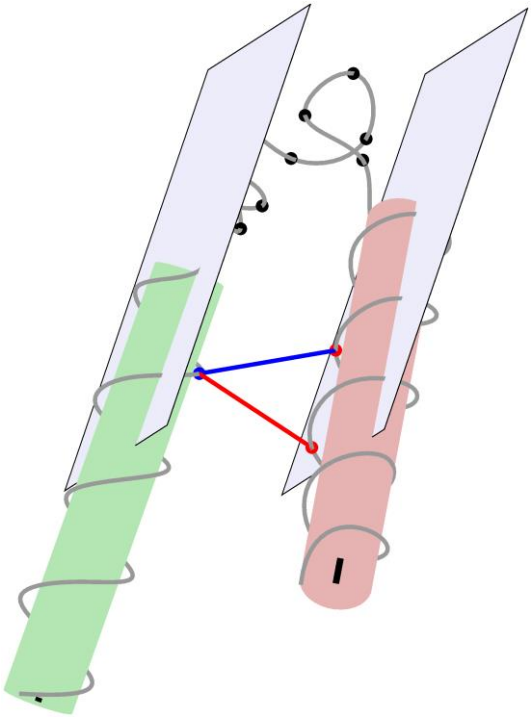

Рис. 2. Пример спиральной пары. Фрагмент цепи белка из Банка белковых данных (PDB ID 3VKH, chain B, участок цепи: 2846–2891). Показаны аппроксимирующие спираль цилиндры, оси цилиндров, а также плоскости, проходящие через оси цилиндров. Кривая аппроксимирована по положениям C_α-атомов белковой цепи, атомы на кривой обозначены точками. Красный и синий отрезки – наименьшие расстояния (6.67 Å и 7.43 Å, соответственно) между спиралями.

Данные о количестве обработанных нами белковых молекул, зарегистрированных в PDB, а также количестве найденных спиралей различных типов представлены 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

Из таблицы видно, что абсолютное большинство составляет количество встречаемых в белках спиралей типа H.

После того как все спирали в аминокислотной цепочке были найдены, спиральные пары составлялись из соседних спиралей. Правила отбора спиральных структур детально описаны в работе [12]. Все подобные структуры можно также описать согласно типам входящих в них спиралей. Всего таких типов структур будет шесть: HH, GG, II, HG, HI, GI. Если название типа состоит из двух одинаковых букв, то спиральная пара данного типа образована двумя спиралями одинакового типа. Например, спиральная пара HH-типа 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 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.

Решив задачу вычисления полигона пересечений проекций спиралей [28] на параллельные плоскости, проходящие через оси спиралей, все найденные спиральные пары мы разбиваем на три множества. Эта разбивка проводится по критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей. В результате получаем три множества: $\{A\}$, $\{B\}$ и $\{C\}$ [12]:

– множество $\{A\}$ образуют все спиральные пары, у которых площади проекций спиралей не имеют пересечений;

– 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;

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

В таблице 2 мы привели результат такой разбивки всех типов спиральных пар по множествам спиральных пар. Проведем краткий анализ данных в таблице. 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.

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 %.

Таблица 2. 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**

На рисунке 3 изображено пересечение проекций спиралей для спиральной пары, изображенной на рисунке 2 (PDB ID 3VKH**,** цепь B, координаты участка: 2846–2891). Прямоугольниками разных цветов представлены проекции спиралей данной спиральной пары и оси соответствующих спиралей. Из рисунка видно, что проекции спиралей данной пары имеют пересечение, а оси спиралей не пересекаются. На рисунке полигон пересечения проекций спиралей выделен цветом, указаны величины площади S и периметра P полигона пересечения. В данном случае обе спирали являются α-спиралями. Таким образом, данная спиральная пара типа HH является представителем множества $\{B\}$. На рисунке указана также и величина межплоскостного расстояния d.

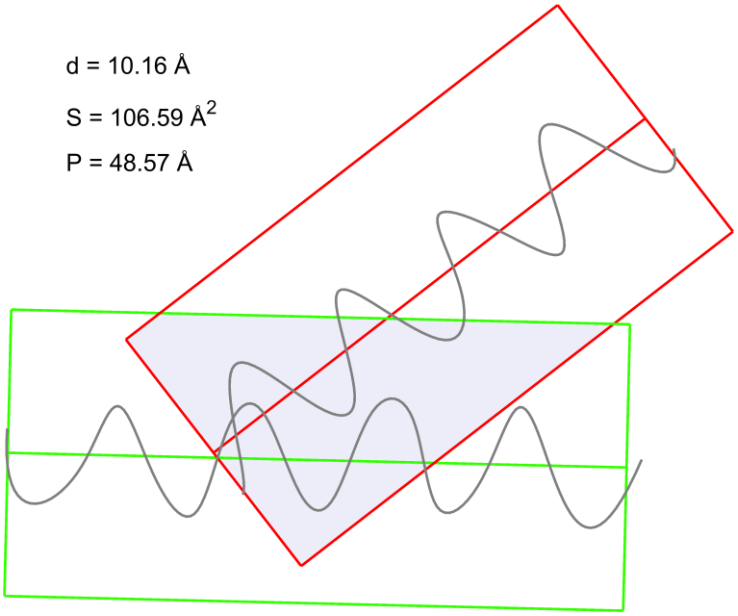

Рис. 3. Пересечение проекций цилиндров спиралей спиральной пары. Полигон пересечения проекций спиралей для спиральной пары, изображенной на рис. 2 (PDB ID 3VKH**,** цепь B, координаты участка: 2846–2891). Красным и зеленым цветом выделены прямоугольники, являющиеся проекциями спиралей, и оси спиралей. Цветом выделен полигон пересечения проекций цилиндров спиралей, указаны величины площади S, периметра P полигона пересечения и межплоскостного расстояния d.

Общее число найденных спиральных пар всех типов равно 2206605. Значительный объём данных, полученный при обработке PDB-структур, обеспечивает достаточную достоверность результатов.

Итак, 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 спиралям – и называется перетяжкой. Мы исследуем структуры, образованные двумя различного типа спиралями (в том числе и α-спиралями), соединёнными между собой различной ненулевой длины N_P и различной конформации перетяжкой. Представим результаты расчетов распределения всех типов спиральных пар, принадлежащих различным множествам, в зависимости от длины перетяжки.

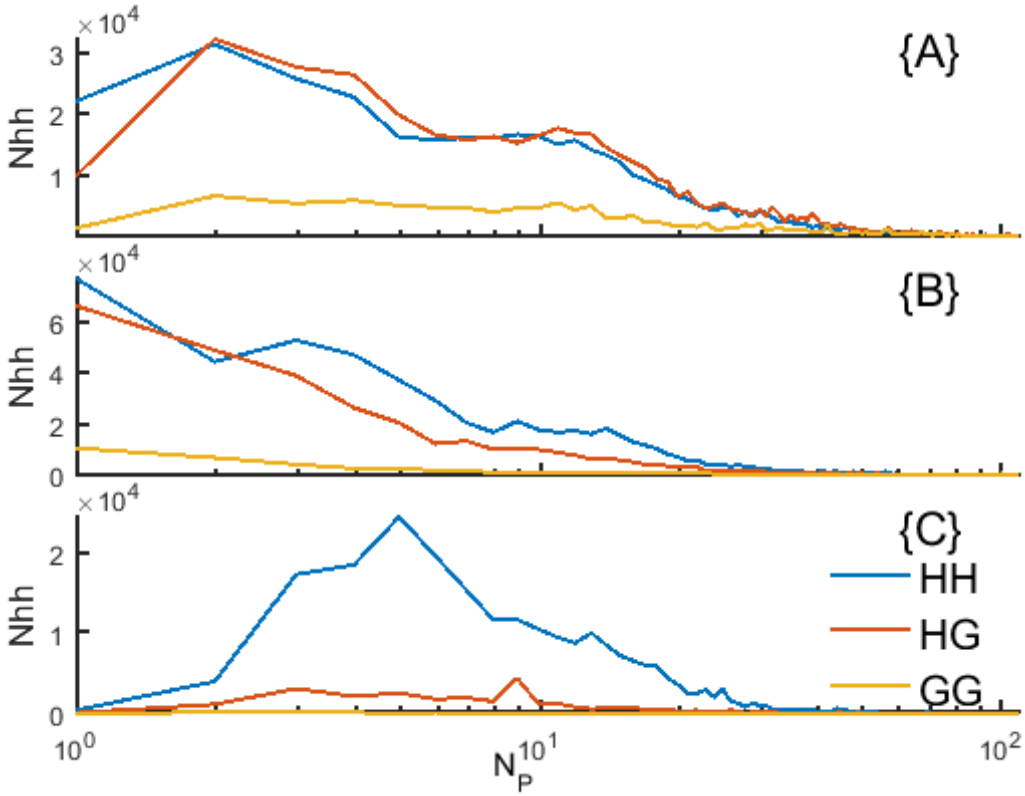

Рис. 4. Распределение числа отобранных структур, принадлежащих различным типам и различным множествам, в зависимости от числа аминокислот в перетяжке. (Ось X – длина перетяжек; ось Y – фактическое число отобранных структур с данной длиной перетяжки; Синяя линия HH-структурные мотивы белков, образованные двумя α-спиралями; красная линия HG – мотивы, образованные α-спиралями и G-спиралями; желтая линия GG – мотивы, образованные G-спиралями. $\{A\}$, $\{B\}$ и $\{C\}$ – множества спиральных пар, полученные по критерию пересечения проекций спиралей).

Рисунок 4 демонстрирует гистограммы распределений различных типов спиральных пар, принадлежащих различным множествам ($\{A\}$, $\{B\}$ и $\{C\}$), в зависимости от длины нерегулярного участка между спиралями. Максимумы распределений спиральных пар, принадлежащих к множествам $\{A\}$ и $\{B\}$, приходятся на малые величины перетяжки (от одной до трех аминокислот). Распределение спиральных пар, принадлежащих множеству $\{A\}$, имеет существенно более дальнодействующий характер, чем распределение спиральных пар, принадлежащих множеству $\{B\}$. Из рисунка 4 видно, что распределение всех рассмотренных типов спиральных пар, принадлежащих множеству $\{A\}$, имеют максимум при длине перетяжки равной два. В то время как вычисленное по этим распределениям среднее значение N_P сильно отличается. Так для пар из множества $\{A\}$ среднее значение

перетяжки составляет 15–20 аминокислот, а для исследуемых структур, принадлежащих множеству $\{B\}$, это значение равно 9–11 аминокислотам (см. табл. 3). Очень большое количество спиральных пар, вошедших в множество $\{B\}$, относится к типам HH и HG и имеет длину перетяжки равной один. Примером таких спиральных пар могут быть шпильки, а эта супервторичная структура в белках действительно встречается очень часто. Спиральных пар типа GG в этом множестве очень мало, при этом и они имеют максимум при длине перетяжки равной один. Абсолютное большинство всех пар множества $\{C\}$ составляют спиральные пары HH. Показано, что среди всех отобранных по сформулированным критериям структур максимальное число структур с длиной перетяжки, равной пять. При этом среднее значение длины перетяжки спиральных пар из данного множества примерно равно среднему значению N_P спиральных пар из множества $\{B\}$. Спиральных пар GG в этом множестве почти нет. Небольшое количество пар HG, принадлежащих множеству $\{C\}$, имеют максимум при $N_P = 9$.

Нужно отметить, что среди всех типов найденных структур, принадлежащих множеству с пересечением осей спиралей (множество $\{C\}$), спиральных пар с длиной перетяжки в одну аминокислоту ($N_P = 1$) не обнаружено.

Также показано, что во всех множествах больше всего структурных мотивов с короткой перетяжкой.

Если учесть, что среднее значение, в основном, зависит от дальнодействующего характера распределения, то можно заключить, что дальнодействие распределения пар для множеств $\{B\}$ и $\{C\}$ примерно равное (см. табл. 3). В таблице 3 представлены статистические оценки длин перетяжек различных типов спиральных пар, принадлежащих различным множествам.

Таблица 3. Статистические оценки распределений длин перетяжек N_P для спиральных пар различных типов, принадлежащих различным множествам

Статистические оценки	длина перетяжки N_P различных типов спиральных пар		
	HH	HG	GG
мода $\{A\}$	2	2	2
медиана $\{A\}$	11	12	14
среднее $\{A\}$	15.34	17.67	24.26
ср.-кв.откл $\{A\}$	17.14	20.52	28.93
мода $\{B\}$	1	1	1
медиана $\{B\}$	6	4	4
среднее $\{B\}$	10.59	9.01	11.09
ср.-кв.откл $\{B\}$	11.78	13.41	18.72
мода $\{C\}$	5	9	3
медиана $\{C\}$	9	9	9
среднее $\{C\}$	11.97	12.65	18.34
ср.-кв.откл $\{C\}$	10.36	13.22	24.25

Таким образом, мы провели детальное исследование распределения спиральных пар в зависимости от длины перетяжки, потому что длина нерегулярного участка между спиралями, несомненно, играет определенную роль в пространственной структуре спиральных пар.

А теперь перейдем к исследованию найденных структур с точки зрения распределения углов между осями спиралей, в том числе и в зависимости от длины перетяжки и расстояния между осями спиралей. Межспиральным углом φ или углом между осями спиралей будем считать угол между векторами, лежащими на осях спиралей, где начало первого вектора есть конец первой спирали, конец первого вектора – начало этой спирали, а второй вектор начинается в начале второй спирали, и конец второй спирали есть конец второго вектора.

Рисунок 5 демонстрирует гистограммы распределения спиральных пар различных типов, принадлежащих различным множествам, в зависимости от угла между осями спиралей. В таблице 4 представлены статистические оценки распределений межспиральных углов φ в спиральных парах различных типов, принадлежащих различным множествам.

Из рисунка 5 хорошо видно существенное различие в распределении структур в зависимости от угла между осями спиралей, принадлежащих различным множествам. Верхний рисунок – распределение пар, принадлежащих множеству $\{A\}$, в зависимости от угла. Распределение достаточно широкое, охватывающее практически весь диапазон углов, с одним максимумом, который немного отличается для различных типов спиральных пар. Для спиральных пар типа HH он немного смещен в сторону тупых углов, что подтверждается средним значением угла 100.17° (см. табл. 4). Для спиральных пар других типов можно отметить скорее асимметрию распределения, чем смещение. Максимум приходится примерно на прямой угол, средние значения это подтверждают (93.99° и 91.45° для типов HG и GG, соответственно).

Таблица 4. Статистические оценки распределений межспиральных углов φ для спиральных пар различных типов, принадлежащих различным множествам

Статистические оценки	межспиральные углы φ различных типов спиральных пар		
	HH	HG	GG
мода $\{A\}$	82.78	104.97	67.54
медиана $\{A\}$	100.54	94.03	89.01
среднее $\{A\}$	100.17	93.99	91.45
ср.-кв.откл $\{A\}$	42.4	39.5	39.81
мода $\{B\}$	58.83	66.72	58.64
медиана $\{B\}$	71.65	72.41	75.69
среднее $\{B\}$	80.14	77.04	79.46
ср.-кв.откл $\{B\}$	47.62	37.05	38.43
мода $\{C\}$	44.29	96.17	109.48
медиана $\{C\}$	60.62	86.08	71.98
среднее $\{C\}$	73.56	84.49	75.56
ср.-кв.откл $\{C\}$	43.5	31.98	39.08

Средний и нижний рисунки иллюстрируют распределения всех типов спиральных пар, принадлежащих множествам $\{B\}$ и $\{C\}$. Распределения спиральных пар, принадлежащих множествам $\{B\}$ и $\{C\}$, отличаются от распределения структур, принадлежащих множеству $\{A\}$. Для этих спиральных пар характерны распределения с двумя максимумами, лежащими один – в области острых, другой – в области тупых углов.

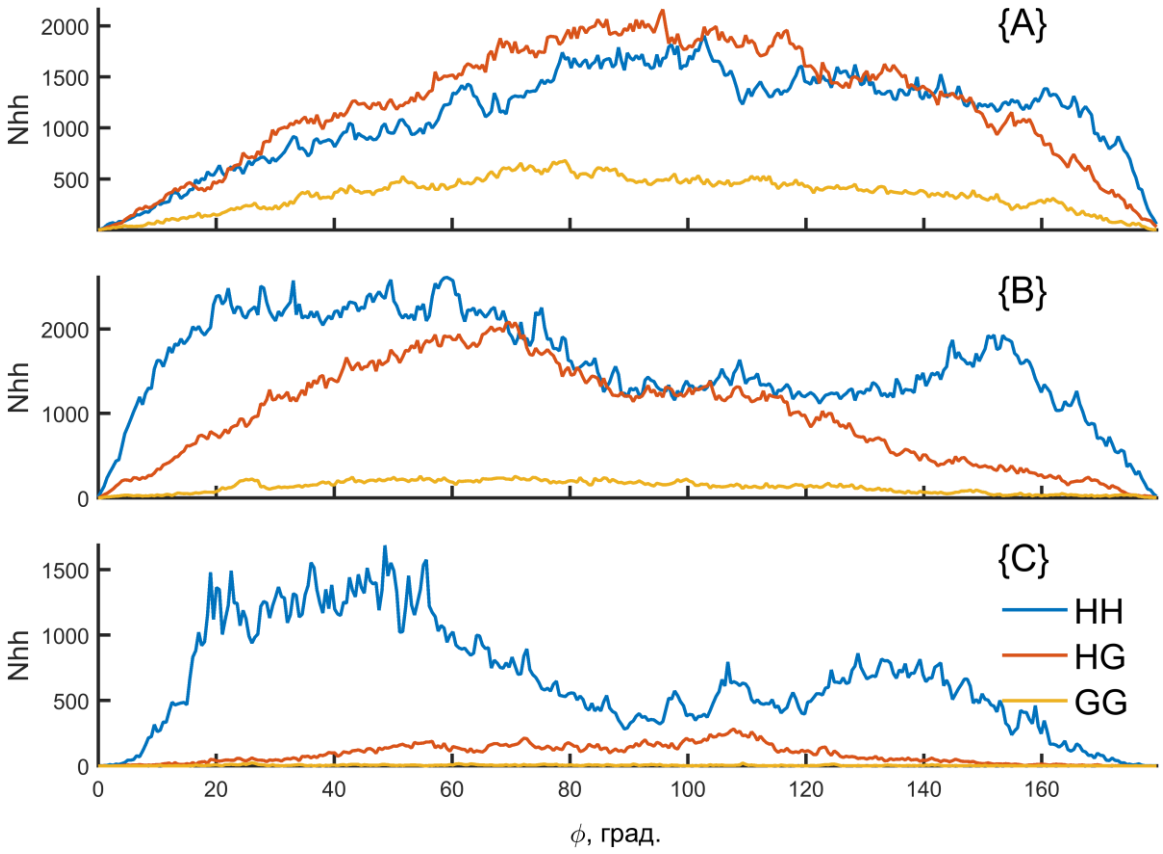

Рис. 5. Распределение числа спиральных пар различных типов, принадлежащих различным множествам, в зависимости от угла φ между осями спиралей.

Таблица 5. Статистические оценки распределений межспиральных углов φ в области острых и тупых углов для спиральных пар различных типов, принадлежащих различным множествам

Статистические оценки	Межспиральные углы φ в области острых углов для различных типов спиральных пар			Межспиральные углы φ в области тупых углов для различных типов спиральных пар		
	HH	HG	GG	HH	HG	GG
мода $\{B\}$	58.83	66.72	58.64	151.72	93.29	132.26
медиана $\{B\}$	46.77	56.46	55.81	133.12	114.84	117.47
среднее $\{B\}$	**46.68**	**54.49**	**54.73**	**131.77**	**118.83**	**120.51**
ср.-кв.откл. $\{B\}$	23	20.9	21.3	23.6	20.46	21.75
мода $\{C\}$	44.29	78.16	22.46	133.69	96.17	109.48
медиана $\{C\}$	45.04	60.56	52.79	129.12	109.31	117.15
среднее $\{C\}$	**45.92**	**59.35**	**51.8**	**127.82**	**112.76**	**120.22**
ср.-кв.откл. $\{C\}$	19.84	19.02	21.66	19.72	15.99	21.43

При этом максимум в области острых углов является значительно более выраженным. Провал распределения приходится на область прямого угла между осями спиральной пары. Особенно это заметно для спиральных пар, принадлежащих множеству $\{C\}$, где провал распределения достигает примерно одной четвертой от

максимума. Рисунок демонстрирует наличие двух пиков для структур, образованных двумя α-спиралями: один пик приходится на острый угол (20°–60°) и тупой угол (130°–150°). Средние значения углов: в области острых углов – 45.92° и тупых углов – 127.82°, что подтверждает таблица 5.

Для структур, в образовании которых участвуют одна α-спираль и еще одна спираль другого типа, наблюдается разброс от 50° до 120°. Для спиральных пар из множества $\{B\}$ провал хоть и заметен, но не такой существенный. Пик распределения смещен в область тупых углов, и мы не наблюдаем провала в области прямого угла.

Рисунок 6 демонстрирует гистограммы распределения спиральных пар различных типов, принадлежащих различным множествам, в зависимости от угла между осями спиралей и с учетом длины перетяжки. Рассмотрим рисунок 6,а, первая строка которого иллюстрирует распределение спиральных пар различных типов, принадлежащих множеству $\{A\}$, в зависимости от угла между осями спиралей при длине перетяжки $N_P < 4$ и $N_P > 3$.

Мы видим, что спиральных пар, не имеющих пересечение проекций спиралей, гораздо больше с длиной перетяжки $N_P > 3$, чем подобных структур с короткой перетяжкой (одна, две или три аминокислоты). Причём пар типа HG больше, чем спиральных пар, образованных двумя α-спиралями. Все распределения имеют одинаковый характер, но максимум распределения пар типа HH и HG отмечается при угле $\varphi \approx 90°$, а максимум для пар типа GG чуть смещен в сторону острых углов ($\varphi \approx 75°$). Распределения спиральных пар с длиной перетяжки $N_P < 4$ совершенно отлично от распределения пар с перетяжкой $N_P > 3$. Пар с короткой перетяжкой и острым углом между осями спиралей $(0° \le \varphi \le 60°)$ почти не наблюдается, структуры с такой перетяжкой появляются при $\varphi \approx 60°$ и количество их растет с увеличением угла между спиралями. Максимум распределения для пар типа HG отмечается при значениях угла $120° \le \varphi \le 130°$, для пар типа HH при $\varphi \approx 170°$. Можно с уверенностью сказать, что этот максимум распределения спиральных пар типа HH с короткой перетяжкой при $\varphi \approx 170°$ дают L-структуры.

Распределение пар типа GG имеет равномерный характер без каких-либо выбросов (пиков) при всех значениях углов $60° \le \varphi \le 180°$.

Вторая строка рисунка 6,а демонстрирует распределение спиральных пар различных типов, принадлежащих множеству $\{B\}$, в зависимости от угла между осями спиралей при длине перетяжки $N_P > 6$ и $N_P < 7$. Мы видим, что спиральных пар, принадлежащих множеству $\{B\}$ (пары имеют пересечение проекций спиралей, но не имеют пересечение осей спиралей) очень много, особенно много пар типа HH. При длине перетяжки $N_P > 6$ распределение пар различных типов имеет разный характер. Для спиральных пар типа HH характерно распределение с двумя максимумами: один – в области острых $(30° \le \varphi \le 50°)$, другой – в области тупых углов $(140° \le \varphi \le 150°)$. При этом максимум в области тупых углов является значительно более выраженным. Провал распределения приходится на область прямого угла между осями спиральной пары. Распределение исследуемых структур типа HG достаточно широкое, оно охватывает практически весь диапазон углов, с одним максимумом в области прямого угла ($\varphi \approx 90°$). Максимум распределения спиральных пар типа HG приходится на область, где отмечается провал распределения спиральных пар типа HH. А пар типа GG, принадлежащих множеству $\{B\}$, с отмеченной длиной перетяжки очень мало, они равномерно распределены по всему диапазону углов.

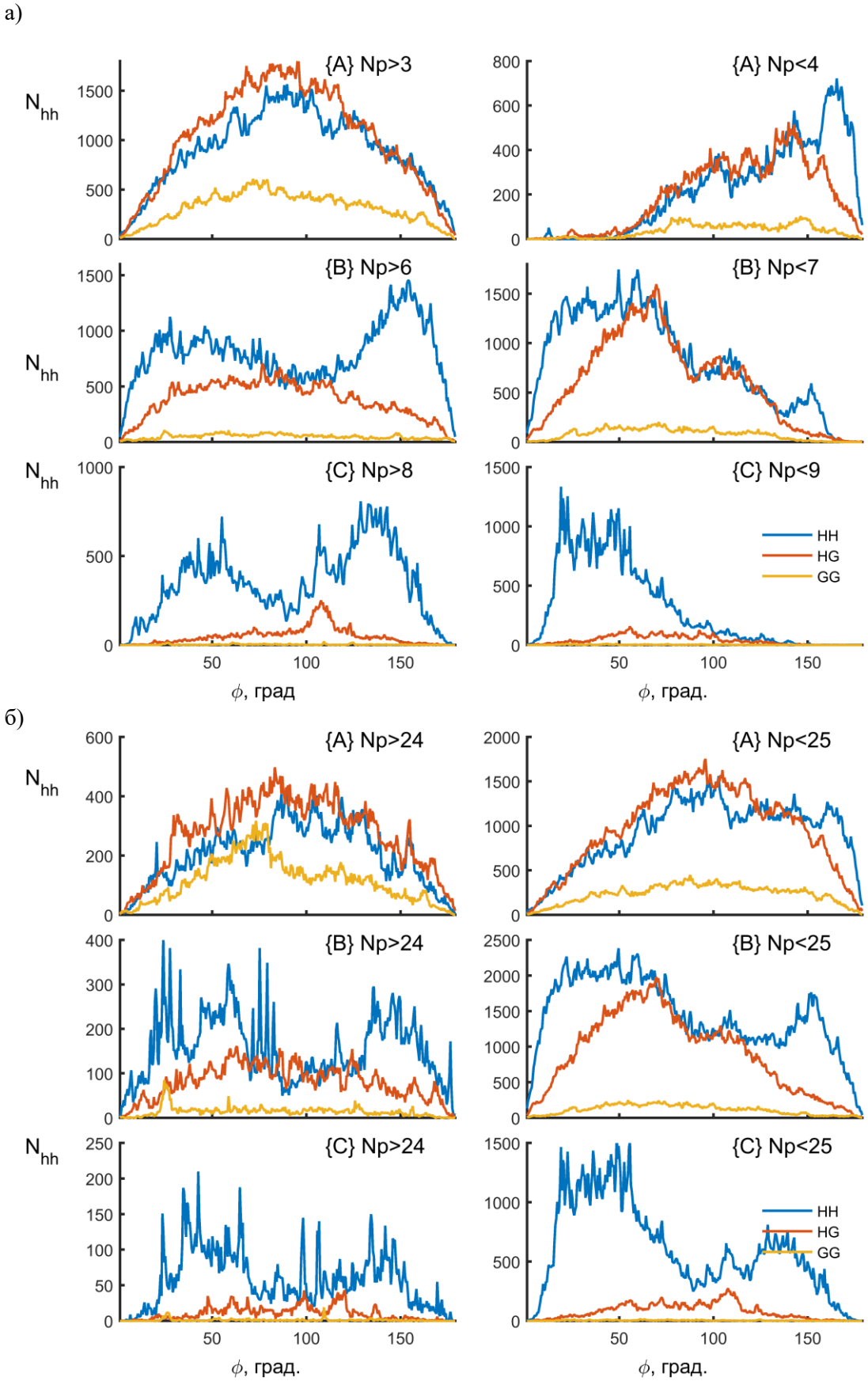

Рис. 6. Распределение спиральных пар различных типов, принадлежащих различным множествам, в зависимости от угла φ между осями спиралей и длины перетяжки.

Распределение спиральных пар различных типов, являющихся представителями множества $\{B\}$, и имеющих длину перетяжки $N_P < 7$, носит другой характер. Из рисунка видно, что большинство пар с указанными свойствами являются парами типа HH. Причем основная часть HH пар локализована в области острых углов $(10° < \varphi < 90°)$, количество их примерно одинаково на всем указанном диапазоне углов, при этом максимум распределения приходится на углы $(50° \le \varphi \le 60°)$. В области тупых углов наблюдаются еще два локальных максимума: один – в области $(100° < \varphi < 120°)$, другой – при $\varphi \approx 160°$. Большинство спиральных пар типа HG с указанной длиной перетяжки имеют острый угол между осями спиралей. Причем таких пар при очень малых углах крайне мало, количество их растет с увеличением угла между спиралями в области острых углов. В области тупых углов $(100° < \varphi < 120°)$ наблюдается также пик распределения, но он носит менее выраженный характер. И этот пик совпадает по значению с пиком распределения спиральных пар типа HH.

Третья строка рисунка 6,а иллюстрирует распределение спиральных пар различных типов, принадлежащих множеству $\{C\}$, в зависимости от угла между осями спиралей при длине перетяжки $N_P > 8$ (левый рисунок) и $N_P < 9$ (правый рисунок). Видно, что спиральных пар указанного множества $\{C\}$ с длиной перетяжки от одной до восьми аминокислот значительно больше, чем спиральных пар с более длинной перетяжкой. Причем почти все такие пары (см. правый рисунок) распределены в области острых углов $(10° < \varphi < 90°)$, очень небольшое количество пар наблюдается в области 100°. Спиральных пар множества $\{C\}$ с такой длиной перетяжки и тупым углом между осями спиралей $(\varphi > 100°)$ почти нет. Максимум распределения приходится на диапазон $(25° < \varphi < 60°)$. Очень мало спиральных пар типа HG, в основном они локализованы в области $(40° < \varphi < 100°)$. Пар GG не наблюдается вовсе. Нужно отметить, что правые рисунки для множеств $\{B\}$ и $\{C\}$, иллюстрирующие распределение спиральных пар с короткой перетяжкой, являются демонстрацией распределения спиральных пар α-белков.

Распределение спиральных пар с пересечением проекций спиралей и пересечением осей спиралей с длиной перетяжки $N_P > 8$ (левый рисунок) напоминает распределение спиральных пар множества $\{C\}$ в общем случае, без учета длины перетяжки (см. рис. 5). Но максимум в данном случае приходится на область тупых углов $(130° < \varphi < 145°)$. Отмечается некий всплеск при угле 110°, при этом же значении угла наблюдается максимум распределения структур типа HG.

На рисунке 6,б представлены гистограммы распределения различных типов спиральных пар, принадлежащих множествам $\{A\}$, $\{B\}$ и $\{C\}$, с длиной перетяжки $N_P > 24$ и $N_P < 25$. Верхняя строка - распределение спиральных пар, принадлежащих множеству $\{A\}$, с указанной длиной перетяжки. Из рисунка видно, что распределения всех типов спиральных пар (HH, HG и GG), принадлежащих множеству $\{A\}$, с длиной перетяжки $N_P < 25$ напоминают распределения пар из этого множества в общем случае, без учета длины перетяжки (см. рис. 5). Можно также отметить, что подавляющее большинство спиральных пар данного множества имеют перетяжку длиной до 25 аминокислот. Исследуемых структур с более длинной перетяжкой значительно меньше, распределение таких пар имеет несколько иной характер. Если пары типа HG также имеют широкое распределение на всем диапазоне углов и сохраняют максимум при угле $\varphi \approx 90°$, в распределении спиральных пар типа HH наблюдается некий провал в области $\varphi \approx 75°$, максимум распределения смещен в зону тупых углов $(90° < \varphi < 125°)$. Просматривается локальный максимум в области острых углов $\varphi \approx 60°$. Распределение спиральных пар типа GG также меняет характер, в

Математическая биология и биоинформатика. 2017. Т. 12. № 1. doi: 10.17537/2017.12.83

области $\varphi \approx 75°$, где фиксируется провал для HH пар, наблюдается максимум распределения, провал в области 100° и локальный максимум при $\varphi \approx 120°$.

Вторая строка рисунка 6,б иллюстрирует распределения различных типов спиральных пар, принадлежащих множествам $\{B\}$, с длиной перетяжки $N_P > 24$ и $N_P < 25$. Из рисунка видно, что спиральных пар данного множества с длиной перетяжки $N_P < 25$ подавляющее большинство. Распределения всех типов спиральных пар множества $\{B\}$ и $\{C\}$ с такой длиной перетяжки также сохраняют характер распределения пар этого множества в общем случае, без учета длины перетяжки (см. рис. 5). Количество спиральных пар с длинной перетяжкой значительно меньше, и распределение таких пар меняет свой характер. Распределение структур типа HG множества $\{B\}$ с длинной перетяжкой охватывает практически весь диапазон углов, с одним максимумом в области прямого угла. Спиральных пар типа GG в множествах $\{B\}$ и $\{C\}$ очень мало, они равномерно распределены на всем диапазоне углов с одним локальным пиком в области острых углов ($\varphi \approx 25°$). В распределении спиральных пар типа HH наблюдается множество пиков на всей области определения. Хотя можно отметить, что общая тенденция (провалы, два максимума в области острых и тупых углов) сохраняется.

На рисунке 7 представлено распределение спиральных пар, состоящих из двух α-спиралей (спиральная пара типа HH), принадлежащих различным множествам, в зависимости двух параметров: угла φ между осями спиралей и межплоскостного расстояния *d*. Верхний рисунок – распределение спиральных пар типа HH, принадлежащих множеству $\{A\}$. Из рисунка видно, что спиральные пары множества $\{A\}$ имеют межспиральные углы $(20° < \varphi < 180°)$, а межплоскостное расстояние от 0 до 15 ангстрем. Существует достаточно большое количество пар, спирали которых расположены под прямым углом друг к другу, а межплоскостное расстояние – в пределах 0–5 Å. Очень много спиральных пар, проекции спиралей которых не пересекаются, а имеют межплоскостное расстояние 2–3 Å и угол между осями спиралей лежит в области $(145° < \varphi < 170°)$. При этом максимальное количество пар при $\varphi \approx 165°$.

Второй рисунок – распределение спиральных пар типа HH, принадлежащих множеству $\{B\}$. Для исследуемых структур этого множества свойственно другое распределение. Большинство пар имеет межплоскостное расстояние 5–10 ангстрем и угол между осями спиралей в области острых углов, при этом наблюдаются два максимума при значениях угла φ: $(15° < \varphi < 25°)$, $(35° < \varphi < 50°)$ и межплоскостном расстоянии $d \approx 8$ Å. Нужно отметить, что имеются еще два локальных максимума в области тупых углов, $\varphi \approx 110°$ и $\varphi \approx 150°$, межплоскостное расстояние при этом примерно равно 2 Å.

Нижний рисунок - распределение спиральных пар, состоящих из двух α-спиралей, принадлежащих множеству $\{C\}$ в зависимости от угла φ между осями спиралей и межплоскостного расстояния *d*. Из рисунка видно, что почти все спиральные пары типа HH множества с пересечением проекций спиралей и осей спиралей локализованы в области острых $(10° < \varphi < 80°)$ и тупых $(110° < \varphi < 150°)$ углов и имеют межплоскостное расстояние 8–10 Å. Спиральных пар с острым углом φ между осями спиралей абсолютное большинство. При этом отмечены два пика: при значениях угла φ $(30° < \varphi < 35°)$ и $(40° < \varphi < 50°)$; межплоскостное расстояние в обоих случаях равно 10 Å.

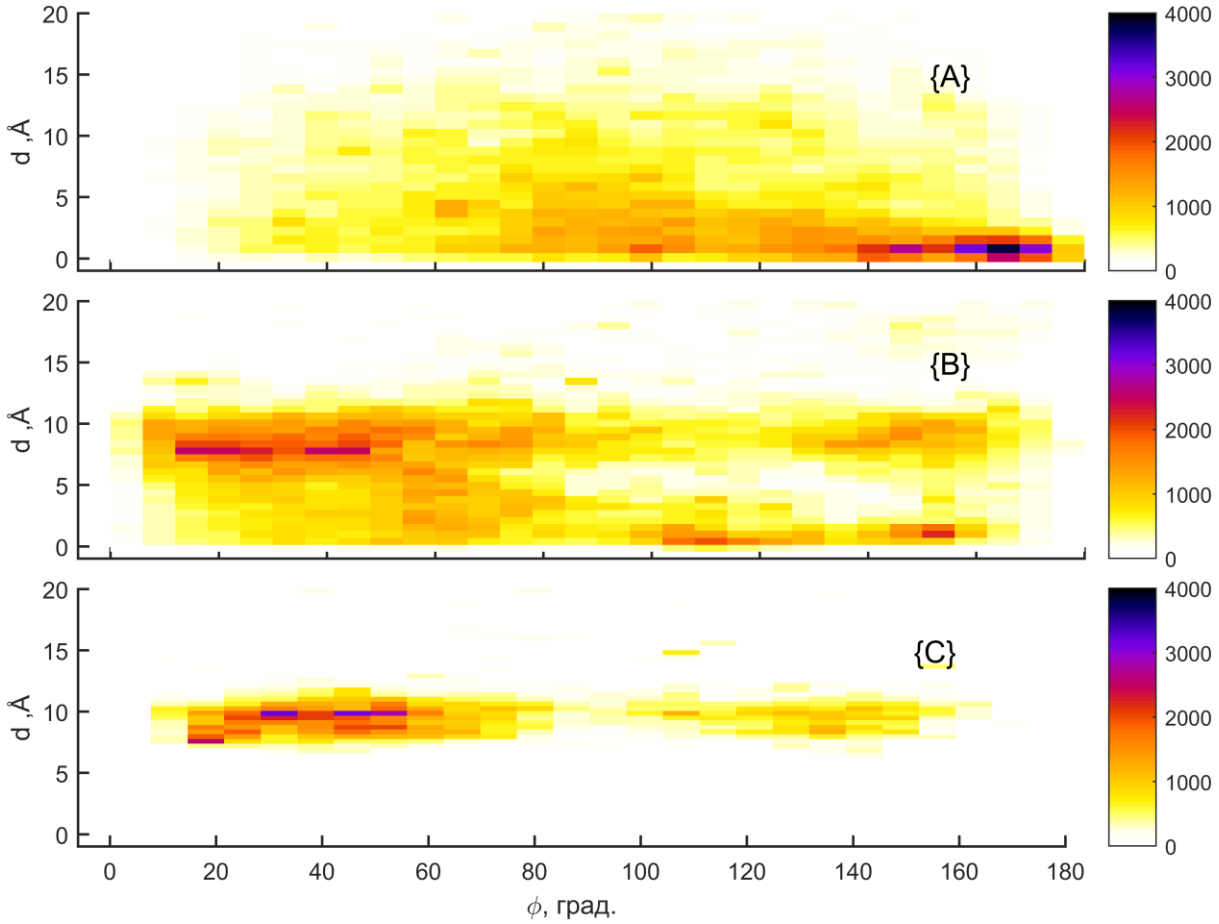

Рис. 7. Распределение спиральных пар типа НН, принадлежащих различным множествам, в зависимости от угла между осями спиралей и межплоскостного расстояния *d*. Ось *Ox* – угол между осями спиралей φ, ось *Oy* – межплоскостное расстояние *d*. {*A*}, {*B*} и {*C*}- множества спиральных пар, полученные по критерию пересечения проекций спиралей. Справа дана шкала соответствия цвета (от белого до черного) количеству спиральных пар.

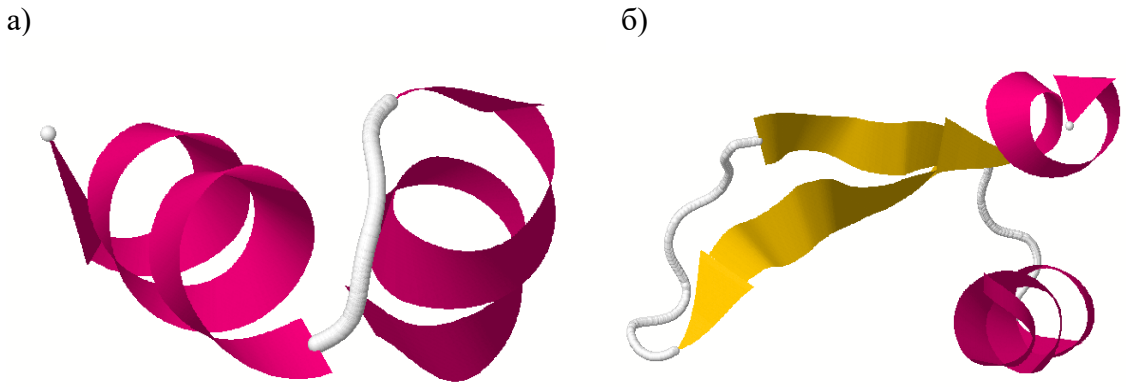

Рис. 8. Примеры компактных структур, образованных двумя спиралями, в данном случае двумя α-спиралями: а) структура с короткой перетяжкой, белок 4DL6 (координаты 275–291), б) структура с длинной перетяжкой, 4EAG (C_α: 211–249).

На рисунке 8 представлены примеры найденных по предложенным критериям компактных структур: структура с короткой перетяжкой (в данном случае это α-α-уголок) и структура с длинной перетяжкой, где между двумя α-спиралями находятся β-тяжи, а также нерегулярные участки различной длины.

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 распределения всех типов спиральных пар, принадлежащих к различным множествам, в зависимости от длины перетяжки. Показано, что во всех множествах больше всего структурных мотивов с короткой перетяжкой. Распределение всех рассмотренных типов спиральных пар, принадлежащих множеству $\{A\}$, имеет максимум при длине перетяжки равной два. Очень большое количество спиральных пар множества $\{B\}$ относится к типам HH и HG и имеет длину перетяжки равной одному.

Показано, что абсолютное большинство всех пар множества $\{C\}$ составляют спиральные пары HH. Среди всех типов найденных структур, принадлежащих множеству с пересечением осей спиралей (множество $\{C\}$), спиральных пар с длиной перетяжки в одну аминокислоту (N_P = 1) не обнаружено. Также показано, среди всех отобранных по сформулированным критериям структур максимальное число структур с длиной перетяжки, равной пяти аминокислотам (N_P = 5).

Проведен анализ распределения всех типов спиральных пар, принадлежащих различным множествам, в зависимости от углов между осями спиралей отобранных структур. В результате работы показано, что распределение всех типов спиральных пар, принадлежащих множеству $\{A\}$, охватывает практически весь диапазон углов с одним максимумом в области $(90° \le \varphi \le 100°)$. Спиральные пары типа HG составляют большинство всех пар, проекции спиралей которых не имеют пересечения. Для спиральных пар, принадлежащих множествам $\{B\}$ и $\{C\}$, характерны распределения с двумя максимумами, лежащими один – в области острых, другой – в области тупых углов. При этом максимум в области острых углов является значительно более выраженным. Показано, что значительная часть спиральных пар типа HH, составляющая абсолютное большинство пар множества $\{C\}$, имеет острый угол $(20° \le \varphi \le 50°)$ между осями спиралей. При этом межплоскостное расстояние для всех этих структур равно 10 Å.

Работа выполнена при частичной поддержке гранта РФФИ № 16-01-00692.

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. 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
6. 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

7. Дедус Ф.Ф., Махортых С.А., Устинин М.Н., Дедус А.Ф. *Обобщенный спектрально – аналитический метод обработки информационных массивов. Задачи анализа изображений и распознавания образов.* М.: «Машиностроение», 1999. 357 с.
8. Dedus F.F., Makhortykh S.A., and Ustinin M.N. Application of the Generalized Spectral-Analytic Method in Information Problems. *Pattern Recognition and Image Analysis.* V. 12. № 4. 2002. P. 429-437.
9. Панкратов А.Н., Тетуев Р.К., Пятков М.И., Тойгильдин В.П., Попова Н.Н. Спектрально-аналитический метод распознавания неточных повторов в символьных последовательностях. *Труды Института системного программирования РАН.* 2015. Т. 27. Вып. 6. С. 335–344.
10. Финкельштейн А.В. *Физика белковых молекул*. М.–Ижевск: АНО «Ижевский институт компьютерных исследований», 2014. 424 с.
11. Птицын О.Б., Филькенштейн А.В. Проблемы предсказания структуры белка. В: *Итоги науки и техники»*. Под ред. Волькенштейна М.Д. М: ВИНИТИ, 1979. Т. 15. С. 6–41. (Серия «Молекулярная биология»).
12. Шульц Г.Е., Ширмер Р.Х. *Принципы структурной организации белков*. М.: Мир, 1982. 354 с.
13. Miller S., Janin J., Lesk A.M., Chothia C. Interior and surfage of monomeric proteins. *MolecularBiology*. 1987. V. 196. P. 641–656.
14. Creighton T.E. *Proteins*. N.Y.: W.H. Freeman & Co, 1991.
15. Степанов В.М. *Молекулярная биология. Структура и функции белков*. М.: Высшая школа, 1996. 336 с.
16. 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.
17. Тихонов Д.А., Куликова Л.И., Ефимов А.В. Статистический анализ внутренних расстояний спиральных пар в белковых молекулах. *Математическая биология и биоинформатика.* 2016. Т. 11. № 2. С. 170–190. doi:10.17537/2016.11.170.
18. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. The Protein Data Bank. 2000. *Nucleic Acids Research*. T. 28. C. 235–242.
19. Crick F.H.C. The Packing of a-Helices: Simple Coiled-Coils. *Acta Crystallographica*. 1953. V. 6. P. 689–697.
20. Lee H.S., Choi J., Yoon S. QHELIX: A Computational Tool for the Improved Measurement of Inter-Helical Angles in Proteins. *Protein.* 2007. V. 26. P. 556–561. doi: 10.1007/s10930-007-9097-9
21. Walther D., Eisenhaber F. and Argos P. Principles of Helix-Helix Packing in Proteins: The Helical Lattice Superposition Model. *Molecular Biology.* 1996. V. 255. P. 536–553.
22. Chothia C., Levitt M. and Richardson D. Structure of proteins: Packing of α-helices and pleated sheets. *Proc. Natl. Acad. Sci.* 1977. V. 74. P. 4130–4134.
23. Chothia C., Levitt M. and Richardson D. Helix to Helix Packing in Proteins. *Molecular Biology*. 1981. V. 145. P. 215–250.
24. Levitt M., Chothia C. Structural patterns in globular proteins. *Nature*. 1976. V. 261. P. 552–558.
25. 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.

26. Kabsch W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica*. 1976. V. 32. P. 922–923.
27. 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.
28. Legland D. *MatGeom: Matlab geometry toolbox for 2D/3D geometric computing.* URL: http://github.com/dlegland/matGeom (дата обращения: 17.09.2015).

Рукопись поступила в редакцию 22.11.2016.
Дата опубликования 20.02.2017.