



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

**Тихонов Д.А., Куликова Л.И.,
Ефимов А.В.**

Анализ площадей и
периметров полигонов
пересечения проекций
спиралей в спиральных
парах белковых молекул

Рекомендуемая форма библиографической ссылки: Тихонов Д.А., Куликова Л.И., Ефимов А.В. Анализ площадей и периметров полигонов пересечения проекций спиралей в спиральных парах белковых молекул // Препринты ИПМ им. М.В.Келдыша. 2018. № 59. 24 с. doi:[10.20948/prepr-2018-59](https://doi.org/10.20948/prepr-2018-59)
URL: <http://library.keldysh.ru/preprint.asp?id=2018-59>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

Д.А. Тихонов, Л.И. Куликова, А.В. Ефимов

**Анализ площадей и периметров полигонов
пересечения проекций спиралей
в спиральных парах белковых молекул**

Москва — 2018

Тихонов Д.А., Куликова Л.И., Ефимов А.В.

Анализ площадей и периметров полигонов пересечения проекций спиралей в спиральных парах белковых молекул

В данной работе проведен анализ распределения площадей и периметров полигонов пересечения проекций спиралей в парах связанных между собой перетяжками спиралей в пространственных структурах белковых молекул. Исследование проводилось на множестве спиральных пар, отобранных нами согласно разработанным правилам из структур белковых молекул, представленных в PDB. Показано, что спиральные пары всех типов, принадлежащие подмножеству пар, имеющих пересечение проекций спиралей, но не имеющих пересечения осей спиралей, имеют площади и периметры полигонов пересечения, значения которых малы или близки к нулю. Среди всех типов спиральных пар, принадлежащих подмножеству пар, имеющих пересечение проекций и осей спиралей, нет пар, имеющих малые или близкие к нулю значения площадей и периметров полигонов пересечения проекций спиралей. Максимум распределения пар, образованных двумя α -спиралями и принадлежащих данному подмножеству, отмечен при значениях площадей и периметров полигонов пересечения проекций спиралей, лежащих в области от 70 Å до 150 Å и в области от 30 Å до 45 Å соответственно.

Ключевые слова: структурные мотивы, точечная модель, спиральные пары в белковых молекулах, площадь и периметр полигона пересечения проекций спиралей.

Dmitry A. Tikhonov, Liudmila I. Kulikova, Alexander V. Efimov

Analysis of the areas and perimeters of polygons of the helices projections intersection in helical pairs of protein molecules

In this study, an analysis of distribution of the areas and perimeters of polygons of the helices projections intersection in pairs of connected helices found in known proteins has been performed. The database for helical pairs was compiled using the Protein Data Bank taking into account the definite rules suggested earlier. It was shown that helical pairs of all types that belong to a subset of pairs that have helices projections intersection and not having their axes intersection have polygons intersection areas and perimeters whose values are small or close to zero. Among all types of helical pairs belonging to a subset of pairs with helices projections intersections and axes intersections, there were no pairs that have small or close to zero values of areas and perimeters of helices projections intersection polygons. The maximum of the distribution of pairs formed by two α -helices and belonging to a given subset was found at the values of polygons of the helices projections intersection areas and perimeters in the region (70 – 150) Å and (30 – 45) Å, respectively.

Key words: structural motifs of proteins, the point model, pairs of helices, area and perimeter of polygon of the helices projections intersection.

Введение

Всестороннему исследованию всевозможных структурных мотивов, состоящих из двух рядом расположенных по цепи спиралей, соединенных одним или более аминокислотными остатками, конформация которых не является спиральной, распознанных и отобранных нами ранее [1] в структурах белковых молекул, зарегистрированных в банке белковых структур PDB [2], посвящена данная и опубликованные ранее [1, 3–4] работы.

В опубликованных работах [1, 3–4] был описан разработанный нами подход отбора спиральных пар в структурах белковых молекул. В указанных работах изучались структуры, в образовании которых участвуют две спирали любого типа: α -спирали, спирали 3_{10} и π -спирали. Было показано, что 72.16 % от общего числа всех спиралей приходится на α -спирали; 27.73 % составляют спирали 3_{10} ; оставшиеся 0.1 % – π -спирали. Несмотря на то, что α -спирали составляют подавляющее большинство (72.16 %), спиральные пары, состоящие из двух α -спиралей, составляют только 54.7 % к общему числу спиральных пар. Спиральные пары с участием спиралей 3_{10} составляют 45 %, меньше 0.2 % приходится на пары с участием π -спиралей. Также было показано, что большинство спиральных пар (45 %), не имеющих пересечений проекций спиралей, образованы α -спиралью и спиралью 3_{10} . В множестве спиральных пар, проекции спиралей которых имеют пересечение, большинство (59 %) составляют пары, образованные двумя α -спиралями. В множестве пар, включающем структуры, имеющие пересечение проекций и осей спиралей, спиральные пары, образованные двумя α -спиралями, составляют подавляющее большинство (87.5 %). При этом общее число спиральных пар, не имеющих пересечений проекций спиралей, равно числу спиральных пар, проекции спиралей которых имеют пересечение, каждое из них дает вклад по 44 % к общему числу спиральных пар, остальные 12 % приходятся на структуры, имеющие пересечение проекций спиралей и осей спиралей. Это означает, что большая часть спиральных пар в аминокислотных цепочках (56 %) имеют пересечения проекций спиралей.

Указанные работы были посвящены исследованию межспиральных расстояний [1], углов между осями спиралей и количества аминокислот между спиралями [3], торсионных углов [4] в спиральных парах, отобранных согласно разработанным правилам отбора спиральных пар в структурах белковых молекул из Банка белковых структур. Было показано, что распределение всех типов спиральных пар, не имеющих пересечения проекций спиралей, охватывает практически весь диапазон углов с одним максимумом в области прямого угла. 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 ($20^\circ \leq \varphi \leq 50^\circ$) между осями спиралей. Межплоскостное расстояние для всех этих

структур равно 10 Å. 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 (от одной до семи аминокислот) перетяжкой, а также что среди всех отобранных по сформулированным критериям структур, состоящих из двух α -спиралей и имеющих пересечение проекций спиралей и осей спиралей, максимальное число структур с длиной перетяжки, равной пяти аминокислотам. Проведен анализ распределения торсионных углов между осями спиралей в парах связанных между собой перетяжками спиралей в пространственных структурах белковых молекул. Показано, что распределение всех типов спиральных пар, не имеющих пересечений проекций спиралей, в зависимости от торсионного угла охватывает практически весь диапазон углов с двумя пиками в области 0° и 180° . Распределение спиральных пар, образованных двумя α -спиралями и представляющих значительную часть подмножества с пересечением проекций спиралей, имеет максимум в области значений торсионного угла между осями спиралей в области (20° – 25°). Для образованных двумя α -спиралями спиральных пар, составляющих абсолютное большинство пар подмножества с пересечением проекций и осей спиралей, характерно распределение с тремя максимумами, лежащими в области острых углов: один – в области отрицательных значений (-50° ÷ -25°), другой – в области положительных значений углов (20° ÷ 25°) и еще один – в области прямого угла (70° ÷ 110°).

Данная работа посвящена исследованию площади и периметра полигона пересечения проекций спиралей в спиральных парах белковых молекул, при этом анализируются межспиральные взаимодействия только тех спиралей, которые расположены в полипептидной цепи друг за другом и соединены между собой перетяжкой, которая может иметь различную длину и конформацию. В работах других исследователей [5–9] изучаются взаимодействия спиралей в белковых молекулах без учета того, на сколько аминокислот отстоят друг от друга спирали и какую вторичную структуру они образуют. Данное исследование проводилось на множестве спиральных пар из структур белковых молекул, зарегистрированных в PDB, без учета классификации белков. Как правило, такие классификации основаны на особенностях вторичной и третичной структуры белков [10]. В дальнейшем планируется исследование распределения характеристик спиральных пар в каждом отдельно взятом классе белков.

Актуальность создания базы данных указанных структур и их дальнейшего изучения следует из интереса исследователей [11–15] к структурным мотивам, имеющим уникальные укладки полипептидной цепи в пространстве [16]. Структурные мотивы, образованные двумя α -спиралями, расположенными в полипептидной цепи друг за другом и соединенными между собой перетяжками, описаны в работах [17–19]. Они представляют собой компактные пространственные структуры. Также из литературы известно, что наиболее

плотная упаковка двух α -спиралей достигается при антипараллельной, перпендикулярной и так называемой скошенной ориентации между спиральями. Примерами таких упаковок являются супервторичные структуры: α - α -уголки, α - α -шпильки, L-образные и V-образные структуры [19], однако всеобъемлющий анализ этих структурных мотивов во всех белках, представленных в PDB, до сих пор не проводился.

Объект исследования

Из банка белковых данных PDB были отобраны все структурные мотивы, образованные двумя спиральями любого типа, расположенными в полипептидной цепи друг за другом и соединенными между собой перетяжками различной длины, имеющими разные конформации. Общее число найденных спиральных пар всех типов равно 2206605 [1, 3–4]. При этом среди всех отобранных спиральных пар обнаружено 1207742 структур, образованных двумя α -спиральями, 821798 структур, состоящих из α -спирали и 3_{10} -спирали, 172877 – образованных двумя 3_{10} -спиральями структур, 3291 структур, состоящих из α -спирали и π -спирали, 896 структур, образованных 3_{10} - и π -спиралью, и одна спиральная пара, образованная двумя π -спиральями. Всего было обработано 100397 белковых структур, количество обработанных аминокислотных остатков составило 66546491, количество обработанных аминокислотных цепочек – 384666. Значительный объем данных, полученный при обработке PDB-структур, обеспечивает достаточную достоверность результатов.

Методы

Для исследования структурных мотивов белковых молекул были разработаны специальные правила распознавания и отбора спиральных пар. Они были описаны в опубликованных нами работах [1, 3–4], здесь приводятся для полноты изложения.

Анализ вторичной структуры аминокислотной цепи осуществлялся с помощью метода, разработанного авторами Dictionary for Secondary Structure of Protein [20]. 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-спираль) – это π -спираль.

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

Для каждой спирали структуры находится ось цилиндра, на который эта спираль намотана. Ось цилиндра определяется по методу наименьших квадратов, при котором минимизируется отклонение параметров спирали от параметров идеальной спирали [21, 22]. Качество оценки осей характеризуется

величиной среднеквадратичного отклонения. Отбираются те спирали (и, соответственно, структуры), для которых точность оценки оси удовлетворяет заданному критерию.

Две оси спиралей полностью определяют пространственную структуру. Известно, что через две непараллельные 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.

Все найденные спиральные пары разбиваются на три множества 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 [23] 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.

Для распознавания спиральной пары предложена точечная модель. По известным координатам точек A_1 , A_2 , B_1 и B_2 , являющимися начальными и конечными точками осей двух спиралей, вычисляются координаты векторов $\overline{A_2A_1}$, $\overline{B_1B_2}$. Точечная модель спиральной пары приведена на рис. 1.

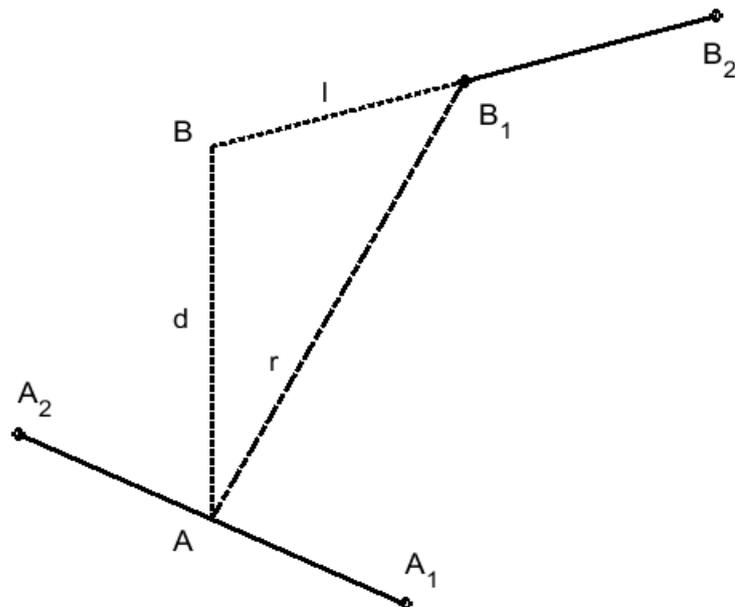


Рис. 1. Точечная модель спиральной пары. Показаны оси спиралей в спиральной паре. Отрезок $[A_1, A_2]$ – ось цилиндра первой спирали, $[B_1, B_2]$ – ось цилиндра второй спирали.

С точки зрения описания взаимного расположения спиралей совершенно естественно возникают три расстояния. Первое – это межплоскостное расстояние d . Как известно, через две непересекающиеся прямые в пространстве можно единственным образом построить две параллельные плоскости, кратчайшее расстояние между которыми является расстоянием между прямыми. Расстояние между прямыми, на которых лежат оси цилиндров, будем называть межплоскостным расстоянием d спиральной пары. Второй характеристикой спиральной пары является кратчайшее расстояние r между осями цилиндров. Очевидно, что минимальное значение r всегда будет не меньше межплоскостного расстояния d , оно равно межплоскостному расстоянию d в случае, когда проекции осей спиральных пар пересекаются. Катет l ($l = \sqrt{r^2 - d^2}$) является третьим расстоянием, которое описывает относительное расположение спиралей в спиральной паре.

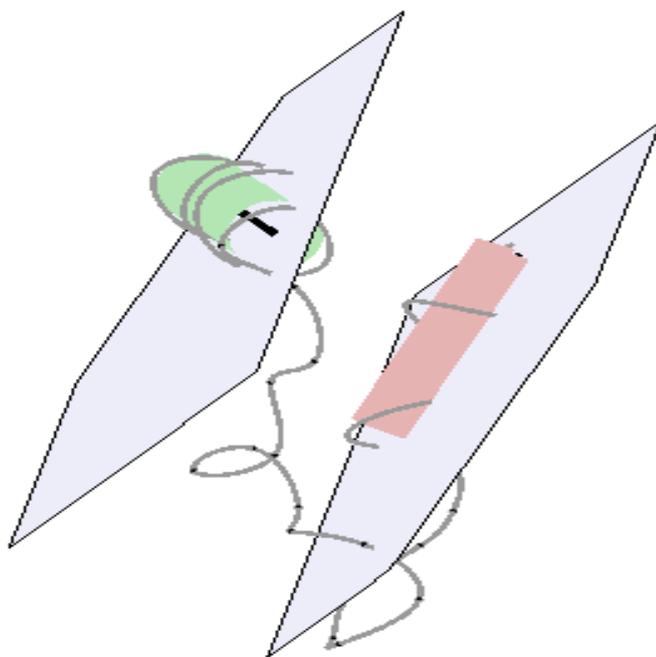


Рис. 2. Пример спиральной пары. Фрагмент цепи белка из банка белковых данных длиной 38 аминокислот (PDB ID 3A0B, координаты участка: 1000–1037). Показаны аппроксимирующие спираль цилиндры, а также плоскости, проходящие через оси цилиндров. Кривая аппроксимирована по положениям C_{α} -атомов белковой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

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, общепринятое его значение – 3.6 Å. На рис. 3 изображен полигон пересечения проекций спирали для спиральной пары, изображенной на рис. 2.

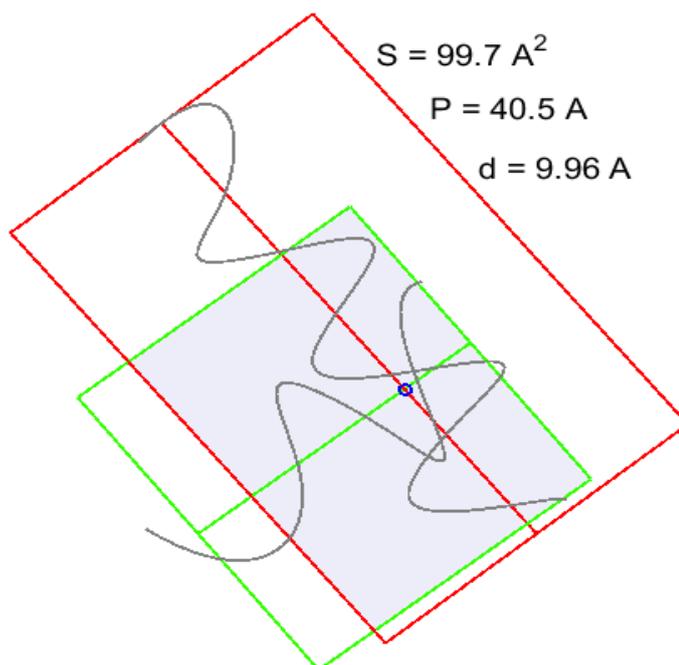


Рис. 3. Пересечение проекций цилиндров спиралей спиральной пары. Полигон пересечения проекций спирали для спиральной пары, изображенной на рисунке 2 (PDB ID 3A0B, координаты участка:1000–1037).

Ширины прямоугольников, которые являются проекциями соответствующих спиралей, равны диаметрам спиралей, которые определяются их типом, плюс удвоенное значение размера боковой цепи. В данном случае обе спирали являются α -спиралями с диаметром 4.6 Å. Таким образом, ширина прямоугольников составляет 11.8 Å. 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 .

Результаты и обсуждение

Все множество распознанных и отобранных нами спиральных пар разбито на три подмножества по критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей [1]. Работа посвящена исследованию площадей и периметров полигонов пересечения проекций спиралей в спиральных парах белковых молекул. А это значит, что из рассмотрения выпадает подмножество $\{A\}$, состоящее из спиральных пар, у которых проекции спиралей не имеют пересечений. В данной работе рассматриваются подмножества $\{B\}$ и $\{C\}$, включающие в себя спиральные пары с ненулевым пересечением проекций. При этом все отобранные спиральные пары разбиты по типу спиралей, образующих исследуемые структуры. Например, спиральная пара типа НН состоит из двух спиралей типа Н (α -спиралей), НG-мотивы образованы двумя спиральями, одна из которых –

спираль типа Н (α -спираль), другая – типа G (спираль 3_{10}); спиральная пара типа GG – это мотивы, образованные двумя спиральями типа G.

На рис. 4 представлено распределение всех типов спиральных пар, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от площади полигона пересечения проекций спиралей. Из рисунка хорошо видно существенное различие в распределении структур, принадлежащих различным подмножествам. Верхний график демонстрирует распределение спиральных пар, принадлежащих подмножеству $\{B\}$, то есть пар с ненулевым пересечением проекций, но не имеющих пересечения осей спиралей. Для распределения спиральных пар всех типов этого подмножества характерно наличие одного максимума в области 0Å (или близкого к нулю значения), т.е. площадь полигона пересечения проекций спиралей всех типов пар этого подмножества мала, она стремится к нулю. Что касается различий между типами спиральных пар, то самый большой диапазон изменений площади S приходится на спиральные пары типа НН. Распределение исследуемых структур, состоящих из двух α -спиралей, достаточно широко, в отличие от распределения пар типа НG и GG, которое более локализовано.

На втором графике показано распределение пар, имеющих пересечение проекций и осей спиралей (множество $\{C\}$) в зависимости от площади полигона пересечения проекций спиралей. Для распределения всех типов пар, принадлежащих множеству $\{C\}$, характерно полное отсутствие структур с малой, близкой к нулю, площадью пересечения проекций спиралей. Распределения всех типов пар этого подмножества начинаются не с нулевого значения. Отличаются эти распределения шириной: распределение исследуемых структур, состоящих из двух α -спиралей, достаточно широко (от 20Å до 400Å), в отличие от распределения пар типа НG и GG, которое более локализовано. Максимум распределения пар типа НН приходится на $(80\text{--}150)\text{Å}$, есть еще локальный пик при значениях площади полигона пересечения проекций спиралей $(310\text{--}330)\text{Å}$.

В таблице 1 представлены средние значения площадей S полигонов пересечения спиралей для различных типов спиральных пар. Из таблицы видим, что для обоих подмножеств средние значения площадей S полигонов пересечения спиралей для мотивов, состоящих из двух α -спиралей, больше средних значений площадей S полигонов пересечения спиралей для пар других типов. Также нужно отметить, что средние значения площадей полигонов пересечения спиралей спиральных пар, принадлежащих подмножеству $\{C\}$, больше средних значений площадей полигонов пересечения спиралей спиральных пар, принадлежащих подмножеству $\{B\}$, примерно в 3–4 раза.

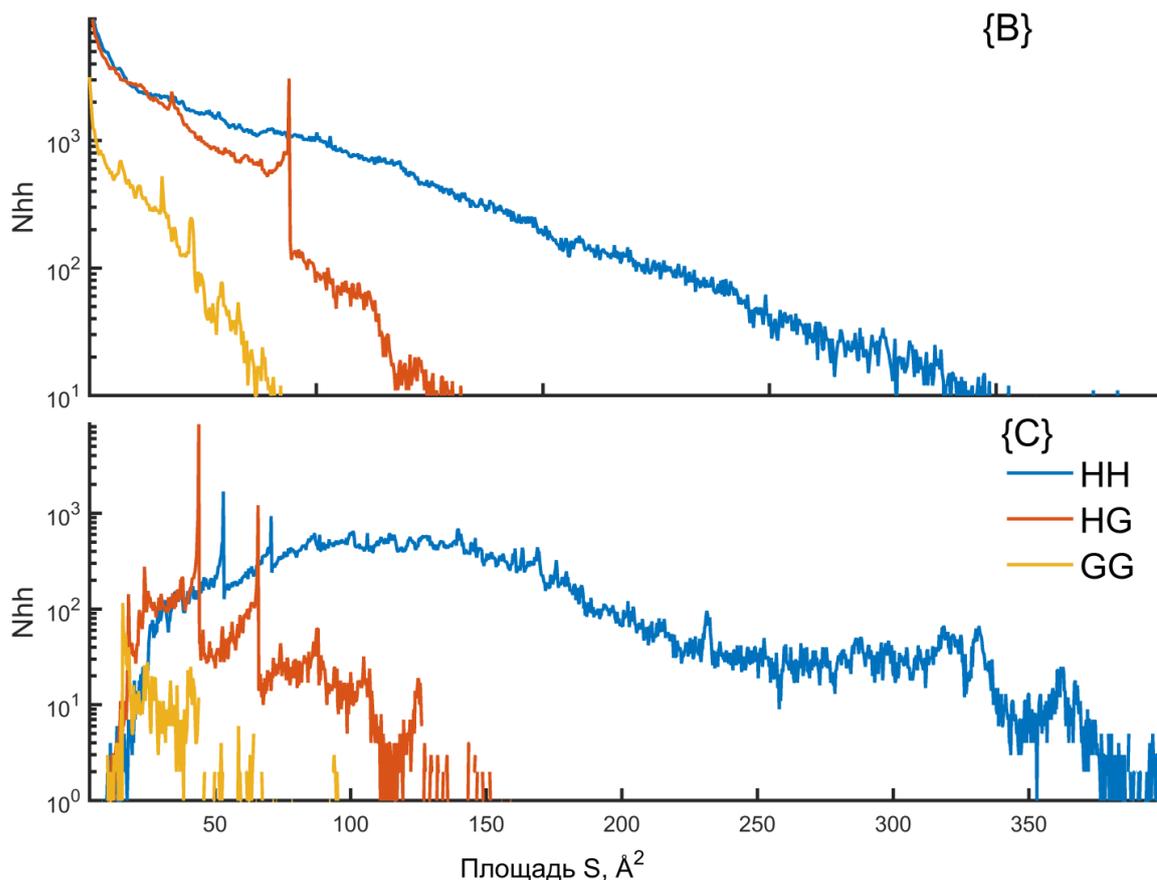


Рис. 4. Распределение всех типов спиральных пар, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от площади S полигона пересечения проекций спиралей. (По оси абсцисс – площадь S полигона пересечения проекций спиралей; по оси ординат – фактическое число отобранных структур. Синяя линия НН – структурные мотивы белков, образованные двумя α -спиралями; красная линия НГ – мотивы, образованные α -спиралями и спиральями 3_{10} ; желтая линия ГГ – мотивы, образованные 3_{10} -спиралями. $\{B\}$ и $\{C\}$ – множества спиральных пар, полученные по критерию пересечения проекций спиралей.)

Таблица 1.

Множества спиральных пар	Средние значения площадей S полигонов пересечения спиралей для различных типов спиральных пар		
	НН	НГ	ГГ
$\{B\}$	30.63	13.72	9.00
$\{C\}$	126.87	48.43	28.72

На рис. 5 представлено распределение всех типов спиральных пар, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от площади полигона пересечения проекций спиралей и с учетом количества аминокислот

между спиралями. Первая строка рис. 5 иллюстрирует распределение спиральных пар различных типов, принадлежащих множеству $\{B\}$, в зависимости от площади S полигона пересечения проекций спиралей при длине перетяжки $N_p > 3$ и $N_p < 4$. Видим, что распределения всех типов исследуемых структур при длине перетяжки $N_p > 3$ и $N_p < 4$ носит одинаковый характер.

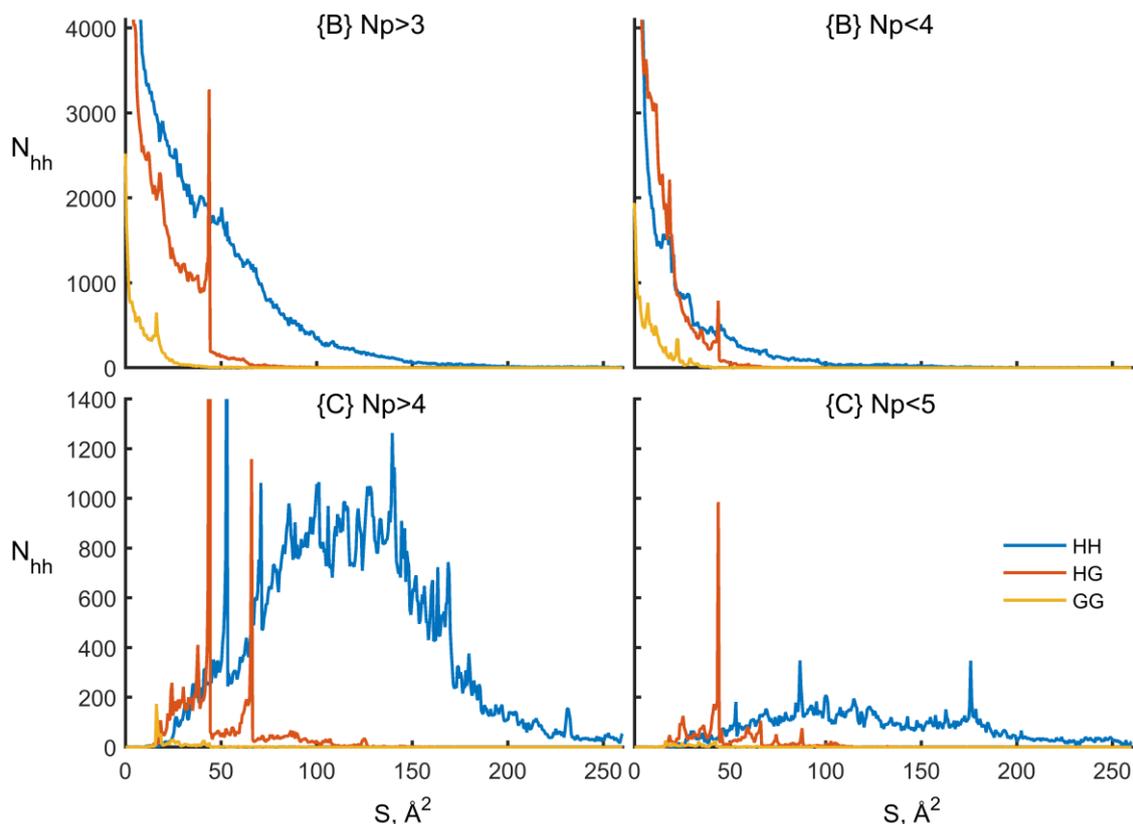


Рис. 5. Распределение спиральных пар различных типов, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от площади S полигона пересечения проекций спиралей и длины перетяжки N_p . (Ось X – площадь S полигона пересечения проекций спиралей, ось Y – фактическое число отобранных структур. Синяя линия HH – структурные мотивы белков, образованные двумя α -спиралями; красная линия HG – мотивы, образованные α -спиралями и спиралями 3_{10} ; желтая линия GG – мотивы, образованные 3_{10} -спиралями. $\{B\}$ и $\{C\}$ – подмножества спиральных пар, полученные по критерию пересечения проекций спиралей.)

Все распределения имеют максимумы в области 0, с той лишь разницей, что распределения пар с короткой перетяжкой (одна, две или три аминокислоты) более локализованы, а распределения всех типов пар при длине перетяжки $N_p > 3$ более широки, и в первую очередь это касается структур, образованных двумя α -спиралями. Также можно отметить, что среди пар, принадлежащих множеству $\{B\}$, с короткой перетяжкой больше пар типа HG, а с длиной перетяжки $N_p > 3$ больше пар типа HH.

Вторая строка рис. 5 демонстрирует распределение спиральных пар различных типов, принадлежащих множеству структур, имеющих пересечение проекций и осей спиралей, в зависимости от площади S полигона пересечения проекций спиралей при длине перетяжки $N_p > 4$ и $N_p < 5$. Мы видим, что характер и широта распределения всех типов спиральных пар, принадлежащих множеству $\{C\}$, при различной длине перетяжки одинаков. Спиральных пар, принадлежащих множеству $\{C\}$, гораздо больше с длиной перетяжки $N_p > 4$, чем подобных структур с перетяжкой $N_p < 5$ (одна, две, три или четыре аминокислоты). Причем среди пар с длиной перетяжки $N_p > 4$ большинство структур, образованных двумя α -спиралями, а с короткой перетяжкой – больше структурных мотивов, образованных α -спиралью и спиралью 3_{10} . Максимум распределения спиральных пар типа НН при длине перетяжки $N_p > 4$ приходится на (100–150) Å, распределение спиральных пар типа НН с короткой перетяжкой носит более равномерный характер, хотя имеет тот же максимум.

На рис. 6 дано распределение спиральных пар, образованных двумя α -спиралями, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от площади S полигона пересечения проекций спиралей и длины перетяжки N_p . Из рисунка видно, что распределения пар, принадлежащих обоим подмножествам, сильно локализованы. Спиральные пары, принадлежащие подмножеству $\{B\}$ (см. верхнюю диаграмму), локализованы в области (0–100) Å и длине перетяжки от 1 до 30 аминокислот. Абсолютный максимум распределения отмечен при длине перетяжки от 1 до 5 аминокислот и площади S полигона пересечения проекций спиралей, лежащей в области значений от близкого к нулю до 25 Å. Почти все мотивы этого подмножества имеют малую площадь полигона пересечения проекций спиралей.

На второй диаграмме рис. 6 представлено распределение спиральных пар, образованных двумя α -спиралями и имеющих пересечение проекций и осей спиралей, в зависимости от площади S полигона пересечения проекций спиралей и количества аминокислот между спиральями. Среди указанных структур нет (или встречаются крайне редко) пар, площадь полигона пересечения проекций спиралей которых меньше 50 Å и больше 180 Å. Структуры локализованы в области (50–180) Å и при длине перетяжки от 3 до 13–15 аминокислот. Максимум распределения наблюдается при длине перетяжки 5 аминокислот и площади S полигона пересечения проекций спиралей, лежащей в области (160–170) Å.

На рис. 7 представлено распределение всех типов спиральных пар, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от периметра P полигона пересечения проекций спиралей. На верхнем графике показано распределение спиральных пар с ненулевым пересечением проекций, но не имеющих пересечения осей спиралей. Распределение всех типов мотивов этого подмножества носит одинаковый характер, большинство пар имеют периметр полигона пересечения проекций спиралей в области значений ($0 < P \leq 50$) Å.

Самый большой диапазон изменений периметра приходится на спиральные пары, образованные двумя α -спиралями. Распределение исследуемых структур типа НН достаточно широко, до 100 \AA , в отличие от распределения пар типа НГ и GG, которое очень локализовано.

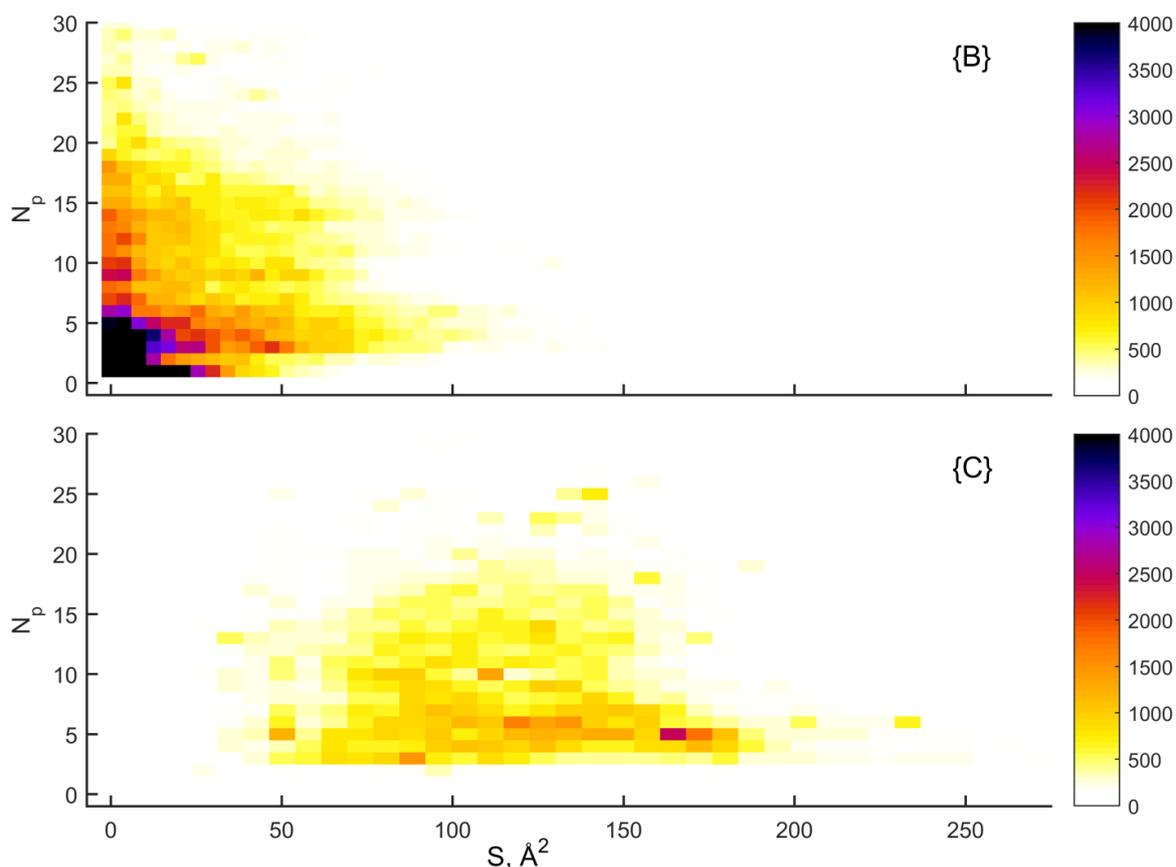


Рис. 6. Распределение спиральных пар типа НН, принадлежащих множествам $\{B\}$ и $\{C\}$, в зависимости от площади S полигона пересечения проекций спиралей и длины перетяжки N_p . Ось X – площадь полигона пересечения проекций спиралей, ось Y – количество аминокислот между спиральями. $\{B\}$ – подмножество спиральных пар, имеющих пересечение проекций спиралей. $\{C\}$ – подмножество спиральных пар, имеющих пересечение проекций и осей спиралей. Справа дана шкала соответствия цвета (от белого до черного) количеству спиральных пар.

На нижнем графике рис. 7 дано распределение пар, имеющих пересечение проекций и осей спиралей, в зависимости от периметра полигона пересечения проекций спиралей. Для распределения всех типов пар, принадлежащих множеству $\{C\}$, характерно отсутствие структур с близким к нулевому значению периметром полигона пересечения проекций спиралей. Отличаются эти распределения шириной: распределение структур типа GG находится в области значений периметра ($5 < P \leq 35$) \AA , структур типа НГ – ($10 < P \leq 45$) \AA , а распределение исследуемых структур, состоящих из двух α -спиралей,

достаточно широко и находится в области $(10 < P \leq 100)$ Å. Максимум распределения пар типа НН, а их большинство в этом подмножестве, находится при значениях периметра в области $(40 < P \leq 45)$ Å, имеется один локальный пик в области $(85 < P \leq 95)$ Å.

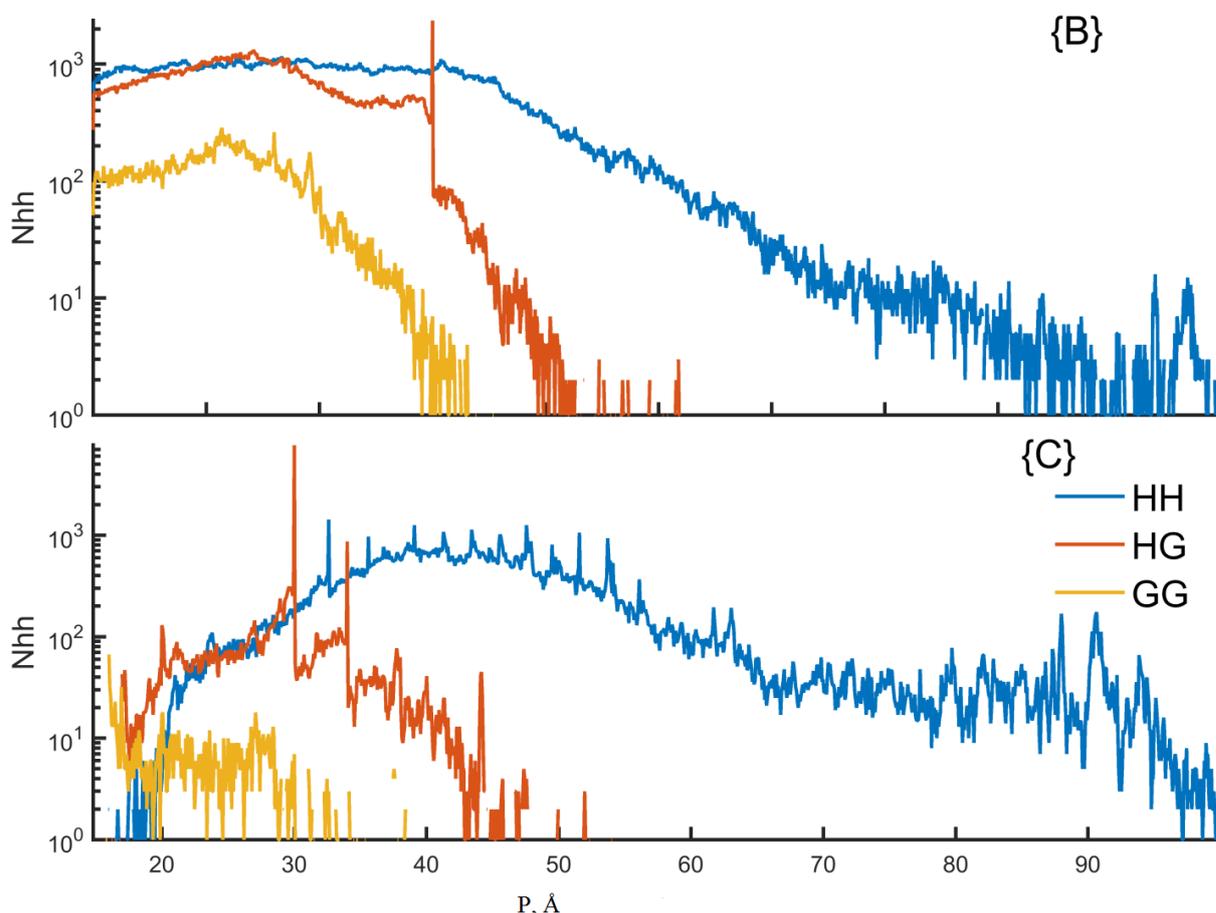


Рис. 7. Распределение всех типов спиральных пар, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от периметра P полигона пересечения проекций спиралей. (По оси абсцисс – периметр P полигона пересечения проекций спиралей; по оси ординат – фактическое число отобранных структур. Синяя линия НН – структурные мотивы белков, образованные двумя α -спиралями; красная линия НГ – мотивы, образованные α -спиралями и спиралью 3_{10} ; желтая линия ГГ – мотивы, образованные 3_{10} -спиралями. $\{B\}$ и $\{C\}$ – подмножества спиральных пар, полученные по критерию пересечения проекций спиралей.)

В таблице 2 представлены средние значения периметров P полигонов пересечения спиралей для различных типов спиральных пар. Из таблицы видим, что для обоих подмножеств средние значения периметров P полигонов пересечения спиралей для мотивов, состоящих из двух α -спиралей, больше средних значений периметров полигонов пересечения спиралей для пар других типов. Показано, что средние значения периметров P полигонов пересечения

спиралей для спиральных пар, принадлежащих множеству $\{C\}$, больше средних значений периметров P полигонов пересечения спиралей для спиральных пар, принадлежащих множеству $\{B\}$, примерно в 2 раза.

Таблица 2.

Множества спиральных пар	Средние значения периметров P полигонов пересечения для различных типов спиральных пар		
	НН	НГ	ГГ
$\{B\}$	21.03	14.18	11.29
$\{C\}$	45.32	29.32	22.57

Рис. 8 иллюстрирует распределение спиральных пар различных типов, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от периметра P полигона пересечения проекций спиралей при фиксированных значениях длины перетяжки N_p . Первая строка рис. 8 иллюстрирует распределение спиральных пар различных типов, принадлежащих множеству $\{B\}$, в зависимости от периметра P полигона пересечения проекций спиралей при длине перетяжки $N_p > 3$ и $N_p < 4$. Видим, что распределения всех типов исследуемых структур при длине перетяжки $N_p > 3$ и $N_p < 4$ носит примерно одинаковый характер. Разница в том, что среди пар с короткой перетяжкой $N_p < 4$ чаще встречаются пары, образованные α -спиралью и спиралью 3_{10} , распределение имеет максимум при значениях периметра полигона пересечения (16–18) Å. Распределение структур типа НН этого подмножества, имеющих короткую перетяжку, сдвинуто к меньшим значениям периметра и имеет максимум при значениях периметра полигона пересечения (5–10) Å. Среди спиральных пар с длиной перетяжки $N_p > 3$ больше пар, состоящих из двух α -спиралей, распределение которых имеет максимум при значениях периметра полигона пересечения (30–35) Å. Максимум распределения спиральных пар типа НГ при длине перетяжки $N_p > 3$ приходится на (10–18) Å.

Вторая строка рис. 8 демонстрирует распределение спиральных пар различных типов, принадлежащих подмножеству структур, принадлежащих множеству $\{C\}$, в зависимости от периметра P полигона пересечения проекций спиралей при длине перетяжки $N_p > 4$ и $N_p < 5$. Мы видим, что характер и широта распределения всех типов спиральных пар при различной длине перетяжки одинаков. Различие состоит лишь в количестве: спиральных пар, принадлежащих множеству $\{C\}$, гораздо больше с длиной перетяжки $N_p > 4$, чем подобных структур с перетяжкой $N_p < 5$ (одна, две, три или четыре аминокислоты). Максимум распределения спиральных пар типа НН при длине перетяжки $N_p > 4$ наблюдается при значениях около 40 Å, а спиральных пар,

образованных α -спиралью и спиралью Z_{10} , – при значениях периметра около 30 \AA . Распределение спиральных пар с короткой перетяжкой $N_p < 5$ носит более равномерный характер, хотя имеет тот же максимум. Среди указанных структур нет пар с периметром полигона пересечения проекций спиралей меньше 20 \AA .

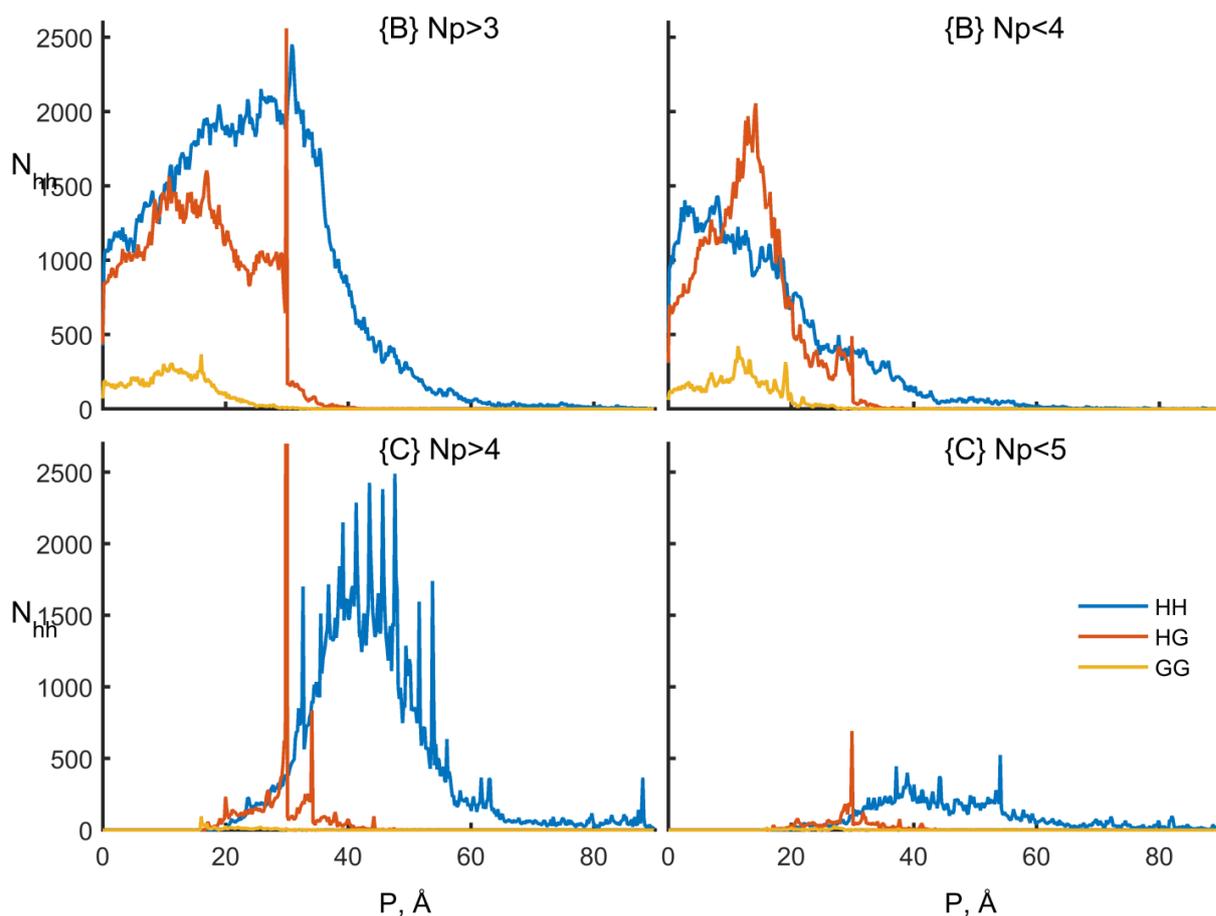


Рис. 8. Распределение спиральных пар различных типов, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от периметра P полигона пересечения проекций спиралей и длины перетяжки N_p . Ось X – периметр P полигона пересечения проекций спиралей, ось Y – фактическое число отобранных структур. Синяя линия NH – структурные мотивы белков, образованные двумя α -спиральями; красная линия HG – мотивы, образованные α -спиральями и спиральями Z_{10} ; желтая линия GG – мотивы, образованные Z_{10} -спиральями. $\{B\}$ и $\{C\}$ – подмножества спиральных пар, полученные по критерию пересечения проекций спиралей.

На рис. 9 представлено распределение образованных двумя α -спиральями спиральных пар, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от периметра P полигона пересечения проекций спиралей и длины перетяжки N_p . Видим, что распределения пар, принадлежащих обоим подмножествам, сильно локализованы. Спиральные пары, принадлежащие подмножеству $\{B\}$ (верхняя

диаграмма), локализованы в области $(0-40)$ Å и длине перетяжки от 1 до 15 аминокислот. Абсолютный максимум распределения отмечен при длине перетяжки от 1 до 4 аминокислот и периметра P полигона пересечения проекций спиралей, лежащей в области значений от близкого к нулю до 20 Å.

На второй диаграмме рис. 9 представлено распределение спиральных пар, образованных двумя α -спиральями и имеющих пересечение проекций и осей спиралей, в зависимости от периметра полигона пересечения проекций спиралей и количества аминокислот между спиральями. Среди указанных структур нет (или встречаются крайне редко) пар, площадь полигона пересечения проекций спиралей которых меньше 30 Å и больше 60 Å. Структуры локализованы в области $(30-60)$ Å при длине перетяжки от 3 до 15–17 аминокислот.

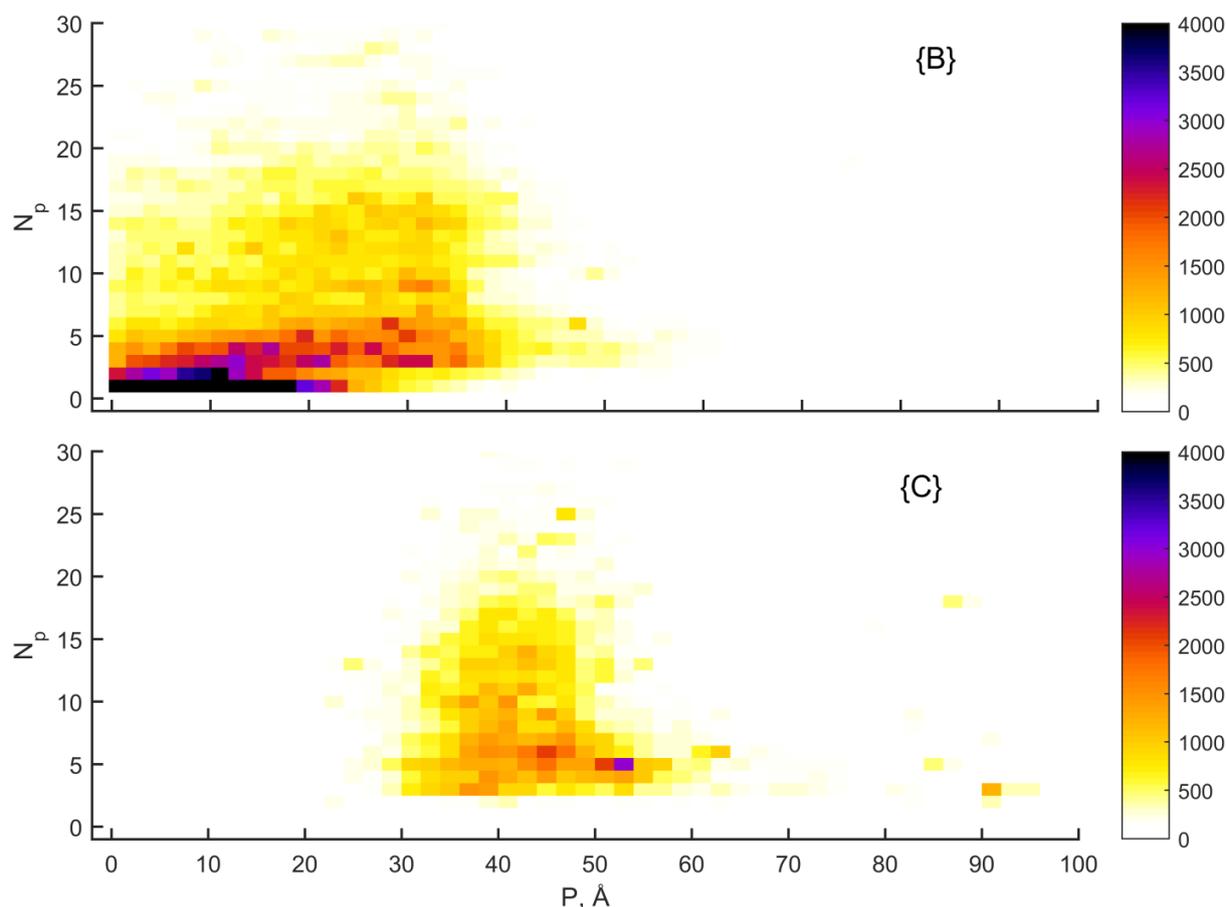


Рис. 9. Распределение спиральных пар типа НН, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от периметра P полигона пересечения проекций спиралей и длины перетяжки N_p . Ось X – площадь полигона пересечения проекций спиралей, ось Y – количество аминокислот между спиральями. $\{B\}$ – подмножество спиральных пар, имеющих пересечение проекций спиралей. $\{C\}$ – подмножество спиральных пар, имеющих пересечение проекций и осей спиралей. Справа дана шкала соответствия цвета (от белого до черного) количеству спиральных пар.

Максимум распределения наблюдается при длине перетяжки 5 аминокислот и периметра полигона пересечения проекций спиралей, лежащей в области (50–55) Å.

Рис. 10 демонстрирует распределение спиральных пар типа НН, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от площади S и периметра P полигона пересечения проекций спиралей. Из рисунка видно, что много исследуемых структур, принадлежащих подмножеству пар, проекции спиралей которых имеют пересечение (подмножество $\{B\}$), локализовано у близких к нулю значений периметра и площади. Нужно также отметить, что много пар этого подмножества имеют периметр полигона пересечения проекций спиралей в области значений от 10 Å до 20 Å, а площадь полигона пересечения проекций спиралей находится в области близких к нулю значений.

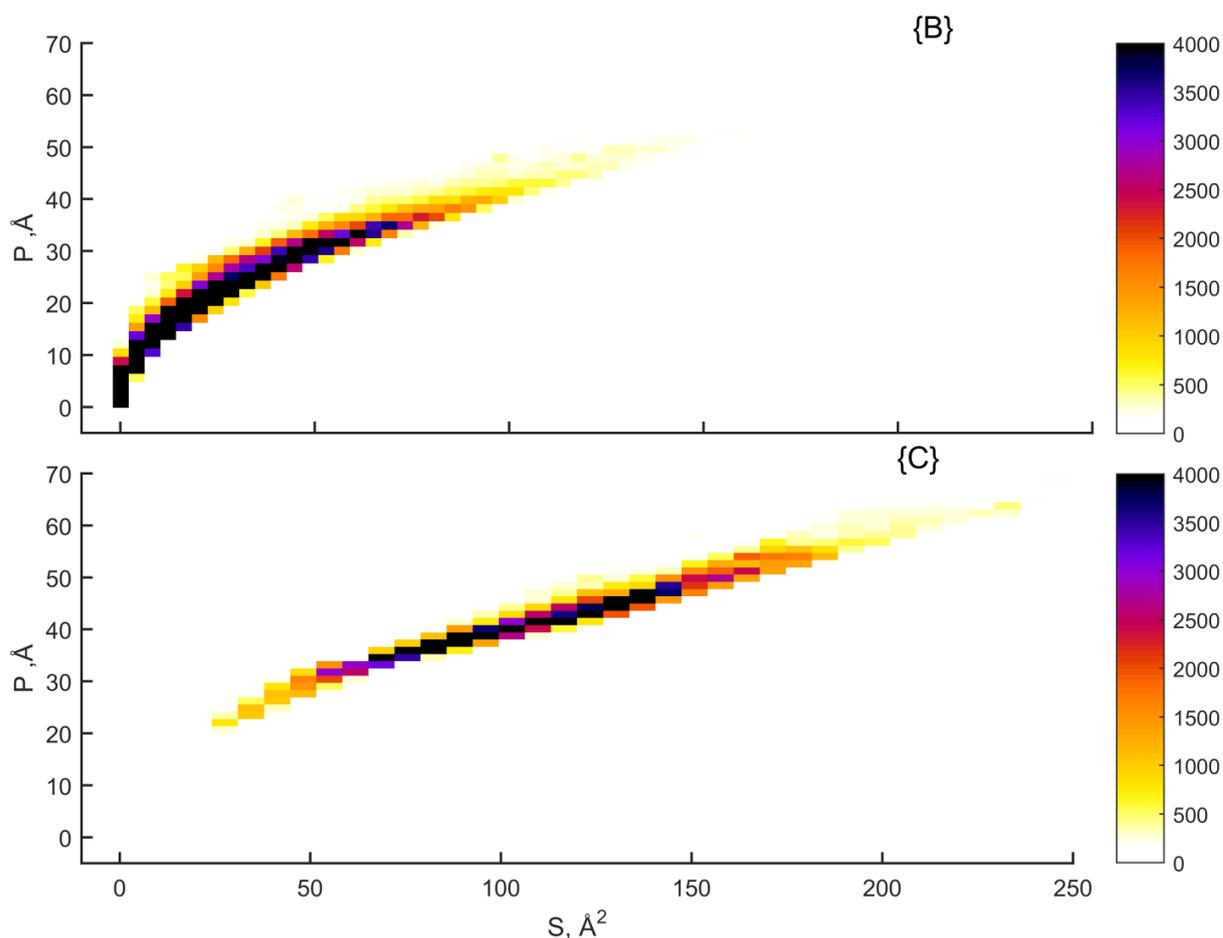


Рис. 10. Распределение спиральных пар типа НН, принадлежащих подмножествам $\{B\}$ и $\{C\}$, в зависимости от площади S и периметра P полигона пересечения проекций спиралей. Ось X – площадь S полигона пересечения проекций спиралей, ось Y – периметр P полигона пересечения проекций спиралей. $\{B\}$ – подмножество спиральных пар, имеющих пересечение проекций спиралей. $\{C\}$ – подмножество спиральных пар, имеющих пересечение проекций и осей спиралей. Справа дана шкала соответствия цвета (от белого до черного) количеству спиральных пар.

Для таких пар характерно «скользящее» пересечение спиралей, такие полигоны имеют при достаточно больших периметрах малые площади.

Большинство пар, принадлежащих подмножеству $\{B\}$, имеют периметр полигона пересечения проекций спиралей в области значений $(0 < P \leq 40) \text{ \AA}$ и площадь полигона пересечения проекций спиралей, лежащую в области значений от близкого к нулю до 80 \AA (см. верхнюю диаграмму).

На нижней диаграмме представлено распределение состоящих из двух α -спиралей пар, принадлежащих множеству спиральных пар, проекции спиралей и оси спиралей которых имеют пересечение, в зависимости от площади S и периметра P полигона пересечения проекций спиралей. Данное распределение отличается от распределения пар подмножества $\{B\}$ тем, что мотивов, имеющих близкие к нулю значения периметра и площади, не обнаружено. Распределение исследуемых структур находится в области значений периметра $(20 \leq P \leq 60) \text{ \AA}$ и площади S – от 30 \AA до 210 \AA . Максимум распределения отмечен при значениях площади S полигона пересечения проекций спиралей, лежащих в области значений от 70 \AA до 150 \AA , и значениях периметра P полигона пересечения проекций спиралей, лежащих в области от 30 \AA до 45 \AA .

Далее было интересно посмотреть, как в спиральных парах соотносятся значения площадей и периметров полигонов пересечения спиралей, т.е. изучить характер пересечения проекций спиралей в спиральных парах; при этом исследовать, как это связано с количеством аминокислот между спиральями.

На рис. 11 дано распределение спиральных пар типа НН, принадлежащих множествам $\{B\}$ и $\{C\}$, в зависимости от S/P полигона пересечения проекций спиралей и длины перетяжки N_p . На верхней диаграмме представлено распределение состоящих из двух α -спиралей пар, принадлежащих подмножеству $\{B\}$, в зависимости от S/P полигона пересечения проекций спиралей и количества аминокислот между спиральями. Видно, что значение длины перетяжки структур этого подмножества находится в диапазоне от 1 до 20 аминокислот, при этом значение S/P полигона пересечения проекций спиралей определено в области значений от близкого к нулю до 2. Наблюдаются два максимума распределения спиральных пар этого подмножества, оба отмечены при малых значениях S/P и длины перетяжки. Первый максимум – при S/P , находящемся в области значений от близкого к нулю до 0.5, и длине перетяжки N_p , принимающей значения от 1 до 2. Вторым максимумом определен при S/P , находящемся в области значений близком к 1, и длине перетяжки $N_p = 1$. Таким образом, видим, что большинство структур этого подмножества имеют площадь полигона пересечения проекций спиралей меньше периметра полигона пересечения или близкое к значению периметра полигона пересечения проекций спиралей. По всей видимости, представителями этого подмножества структур являются такие пары, как L-образные, V-образные структуры.

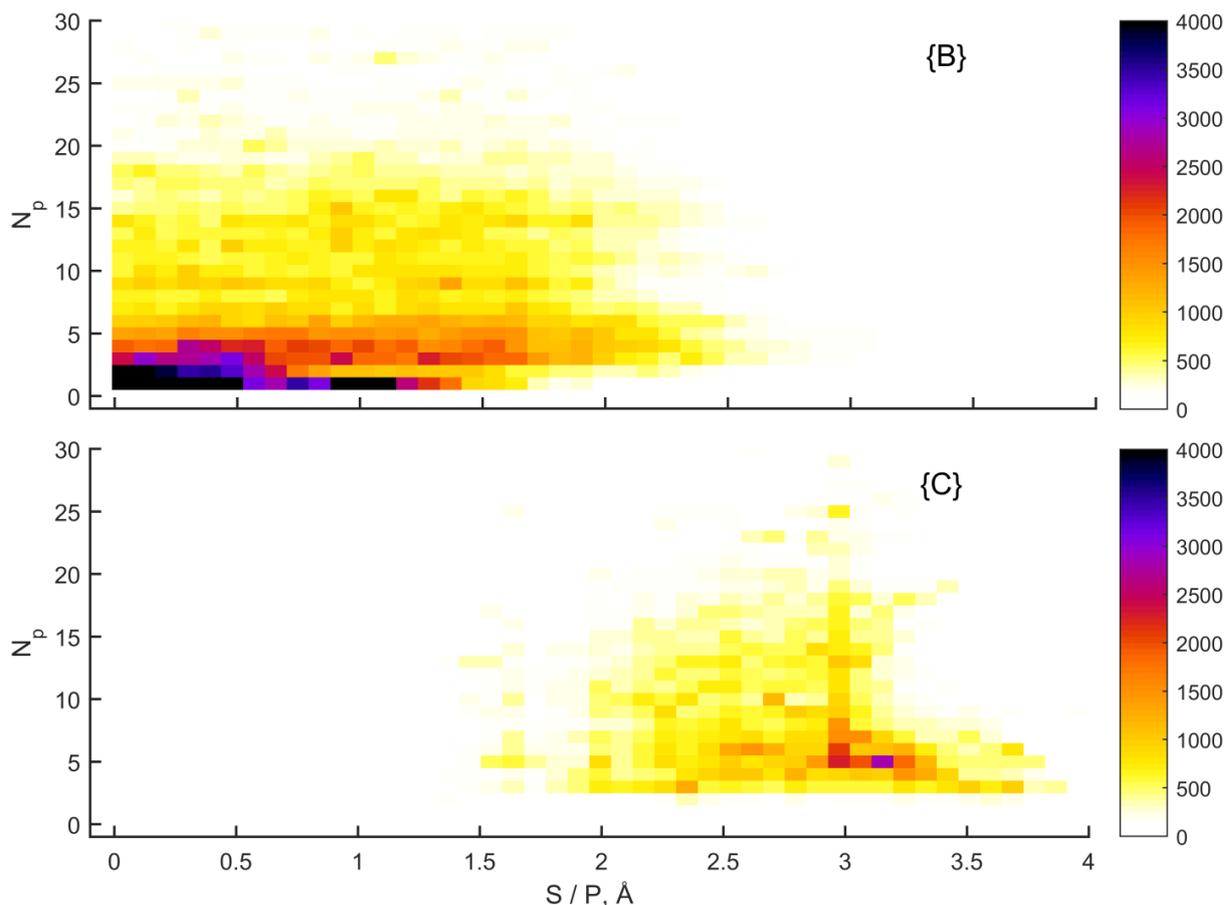


Рис. 11. Распределение спиральных пар типа НН, принадлежащих множествам $\{B\}$ и $\{C\}$, в зависимости от S/P полигона пересечения проекций спиралей и длины перетяжки N_p . Ось X – S/P полигона пересечения проекций спиралей, ось Y – количество аминокислот между спиральями. $\{B\}$ – подмножество спиральных пар, имеющих пересечение проекций спиралей. $\{C\}$ – подмножество спиральных пар, имеющих пересечение проекций и осей спиралей. Справа дана шкала соответствия цвета (от белого до черного) количеству спиральных пар.

На нижней диаграмме представлено распределение состоящих из двух α -спиралей пар, принадлежащих подмножеству спиральных пар, проекции спиралей и оси спиралей которых имеют пересечение, в зависимости от S/P полигона пересечения проекций спиралей и количества аминокислот между спиральями. Данное распределение пар также локализовано и отличается от распределения структур подмножества $\{B\}$ тем, что среди спиральных пар этого подмножества нет мотивов, имеющих малое значение S/P . Значение S/P для этих структур находится в области от 2 до 3.7, а длина перетяжки – от 3 до 18 аминокислот. Максимум распределения спиральных пар, принадлежащих подмножеству $\{C\}$, отмечен при S/P , находящемся в области значений от 3 до 3.3, и длине перетяжки $N_p = 5$. Примером таких структурных мотивов, состоящих из двух α -спиралей и принадлежащих подмножеству спиральных

пар, проекции спиралей и оси спиралей которых имеют пересечение, являются, например, α - α -уголки.

Заключение

Используя точечную модель спиральных пар, мы распознали и отобрали согласно разработанным правилам множество всевозможных структурных мотивов, состоящих из двух рядом расположенных по цепи спиралей, соединенных одним или более аминокислотными остатками, конформация которых не является спиральной, из всех белковых молекул, зарегистрированных в PDB. Полученное множество спиральных пар было проанализировано с целью его классификации и установления закономерностей структурной организации. Множество спиральных пар разбито на три подмножества по критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей. Всестороннему исследованию отобранных структур посвящена данная и опубликованные нами ранее работы. В данной работе проведен анализ площадей и периметров полигонов пересечения проекций спиралей в отобранных спиральных парах белковых молекул. Показано, что спиральные пары всех типов (HH, HG и GG), принадлежащие подмножеству пар, имеющих пересечение проекций спиралей, но не имеющих пересечения осей спиралей, имеют площади и периметры полигонов пересечения, значения которых малы или близки к нулю. Большинство структурных мотивов данного подмножества имеют короткую длину перетяжки (одна, две аминокислоты). Среди всех типов спиральных пар, принадлежащих подмножеству пар, имеющих пересечение проекций и осей спиралей, нет пар, имеющих малые или близкие к нулю значения площадей и периметров полигонов пересечения проекций спиралей. Максимум распределения состоящих из двух α -спиралей пар, принадлежащих подмножеству спиральных пар, проекции и оси спиралей которых имеют пересечение, в зависимости от площади S и периметра P полигона пересечения проекций спиралей отмечен при значениях площади S полигона пересечения проекций спиралей, лежащих в области значений от 70 Å до 150 Å и значениях периметра P полигона пересечения проекций спиралей, лежащих в области от 30 Å до 45 Å. При этом длина перетяжки исследуемых структурных мотивов $N_p = 5$.

Также показано, что для обоих подмножеств средние значения периметров P и площадей S полигонов пересечения спиралей для мотивов, состоящих из двух α -спиралей, больше средних значений периметров и площадей полигонов пересечения спиралей других типов пар. Средние значения периметров P полигонов пересечения спиралей для спиральных пар, принадлежащих подмножеству пар, имеющих пересечение проекций и осей спиралей, больше средних значений периметров P полигонов пересечения спиралей для спиральных пар, принадлежащих подмножеству пар, имеющих пересечение проекций спиралей, но не имеющих пересечения осей спиралей, примерно в 2

раза. Средние значения площадей S полигонов пересечения спиралей для спиральных пар, принадлежащих подмножеству пар, имеющих пересечение проекций и осей спиралей, больше средних значений площадей S полигонов пересечения спиралей для спиральных пар, принадлежащих подмножеству пар, имеющих пересечение проекций спиралей, но не имеющих пересечение осей спиралей, примерно в 3–4 раза.

Список литературы

1. 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](https://doi.org/10.17537/2016.11.170)
2. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. The Protein Data Bank. *Nucleic Acids Research*. 2000. V. 28. P.235-242. doi: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235)
3. Тихонов Д.А., Куликова Л.И., Ефимов А.В. Исследование межспиральных углов в структурных мотивах, образованных двумя спиральями. *Математическая биология и биоинформатика*. 2017. Т. 12. № 1. С. 83–101. doi: [10.17537/2017.12.83](https://doi.org/10.17537/2017.12.83)
4. Тихонов Д.А., Куликова Л.И., Ефимов А.В. Анализ торсионных углов между осями спиралей в спиральных парах белковых молекул. *Математическая биология и биоинформатика*. 2017. Т. 12. № 2. С. 398–410. doi: [10.17537/2017.12.398](https://doi.org/10.17537/2017.12.398)
5. Crick F.H.C. The Packing of α -helices: simple coiled-coils. *Acta Crystallographica*. 1953. V. 6. P. 689–697. doi: [10.1107/S0365110X53001964](https://doi.org/10.1107/S0365110X53001964)
6. 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](https://doi.org/10.1007/s10930-007-9097-9)
7. 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. doi: [10.1006/jmbi.1996.0044](https://doi.org/10.1006/jmbi.1996.0044)
8. Chothia C., Levitt M., Richardson D. Structure of proteins: packing of α -helices and pleated sheets. *Proc. Natl. Acad. Sci.* 1977. V. 74. P. 4130–4134. doi: [10.1073/pnas.74.10.4130](https://doi.org/10.1073/pnas.74.10.4130)
9. Chothia C., Levitt M., Richardson D. Helix to helix packing in proteins. *Molecular Biology*. 1981. V. 145. P. 215–250. doi: [10.1016/0022-2836\(81\)90341-7](https://doi.org/10.1016/0022-2836(81)90341-7)

10. Levitt M., Chothia C. Structural patterns in globular proteins. *Nature*. 1976. V. 261. P. 552–558. doi: [10.1038/261552a0](https://doi.org/10.1038/261552a0)
11. Calhoun J.R., Kono H., Lahr S., Wang W., DeGrado W.F., Saven J.G. Computational design and characterization of a monomeric helical dinuclear metalloprotein. *Journal of Molecular Biology*. 2003. V. 334. No. 5. P. 1101–1115. doi: [10.1016/j.jmb.2003.10.004](https://doi.org/10.1016/j.jmb.2003.10.004)
12. Calhoun J.R., Natri F., Maglio O., Pavone V., Lombardi A., DeGrado W.F. Artificial diiron proteins: From structure to function. *Peptide Science*. 2005. V. 80. No. 2–3. P. 264–278. doi: [10.1002/bip.20230](https://doi.org/10.1002/bip.20230)
13. Chino M., Maglio O., Natri F., Pavone V., DeGrado W.F., Lombardi A. Artificial diiron enzymes with a de novo designed four-helix bundle structure. *European Journal of Inorganic Chemistry*. 2015. P. 3371–3390. doi: [10.1002/ejic.201500470](https://doi.org/10.1002/ejic.201500470)
14. Chino M., Leone L., Maglio O., Lombardi A. Designing Covalently Linked Heterodimeric Four-Helix Bundles. *Methods in enzymology*. 2016. 2016(21). V. 580. P. 471–499. doi: [10.1016/bs.mie.2016.05.036](https://doi.org/10.1016/bs.mie.2016.05.036)
15. Trovato A., Seno F. A new perspective on analysis of helix-helix packing preferences in globular proteins. *Proteins: structure, function, bioinformatics*. 2004. V. 55. P. 1014–1022. doi: [10.1002/prot.20083](https://doi.org/10.1002/prot.20083)
16. Efimov A.V. Standard structures in proteins. *Prog. Biophys. Molec. Biol.* 1993. V. 60. P. 201–239. doi: [10.1016/0079-6107\(93\)90015-C](https://doi.org/10.1016/0079-6107(93)90015-C)
17. 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. doi: [10.1016/j.bbrc.2010.05.136](https://doi.org/10.1016/j.bbrc.2010.05.136)
18. 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.
19. 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. doi: [10.1023/A:1004859003221](https://doi.org/10.1023/A:1004859003221)
20. 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. doi: [10.1002/bip.360221211](https://doi.org/10.1002/bip.360221211)
21. Kabsch W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica*. 1976. V. 32. P. 922–923. doi: [10.1107/S0567739476001873](https://doi.org/10.1107/S0567739476001873)

22. 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. doi: [10.1107/S0567739478001680](https://doi.org/10.1107/S0567739478001680)
23. Legland D. *MatGeom: Matlab geometry toolbox for 2D/3D geometric computing*. URL: <http://github.com/dlegland/matGeom> (дата обращения: 21.10.2017).

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проекты № 16-01-00692-а и № 18-07-01031-а, в рамках государственного задания (ГЗ № 0017-2018-0013).

Оглавление

Введение	3
Объект исследования	5
Методы	5
Результаты и обсуждение	8
Заключение.....	21
Список литературы.....	22