

Распознавание и исследование спиральных пар белковых молекул

Тихонов Д.А.^{1,2}, Куликова Л.И.¹, Руднев В.Р.², Ефимов А.В.³

¹Институт математических проблем биологии РАН – филиал ИПМ
им. М.В. Келдыша РАН, Пуццино, Россия

²Институт теоретической и экспериментальной биофизики РАН, Пуццино, Россия

³Институт белка РАН, Пуццино, Россия

likulikova@mail.ru

Исследование проводилось на множестве спиральных пар, отобранных из структур белковых молекул, представленных в Банке белковых структур (Protein Data Bank). Были разработаны правила отбора спиральных пар в структурах белковых молекул из PDB. Полученное множество спиральных пар было проанализировано с целью его классификации и установления закономерностей структурной организации и разбито на три подмножества по критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей. Спиральные пары были разделены также по типу спиралей, образующих исследуемые структуры. Проведен всесторонний анализ геометрических характеристик отобранных структур: межспиральных расстояний и углов, торсионных углов, длин перетяжек. Разработанная база данных выложена в сети интернет. Ее интерфейс позволяет выполнять операции поиска, сортировки, фильтрации структур по всем параметрам, просматривать гистограммы распределения отобранных спиральных пар в зависимости от выбранных параметров, визуализировать 3D модели исследуемых структурных мотивов.

Ключевые слова: структурные мотивы, спиральные пары в белковых молекулах, точечная модель, торсионный угол между осями спиралей, межплоскостное расстояние, минимальное расстояние между осями спиралей, межспиральный угол.

Recognition and Study of Helical Pairs in Protein Molecules

Tikhonov D.A.^{1,2}, Kulikova L.I.¹, Rudnev V.R.², Efimov A.V.³

¹Institute of Mathematical Problems of Biology RAS – the Branch of Keldysh Institute of Applied Mathematics, RAS,
Pushchino, Russia

²Institute of Theoretical and Experimental Biophysics, RAS, Pushchino, Russia

³Institute of Protein Research, RAS, Pushchino, Russia

The study was conducted on the database for helical pairs compiled using the Protein Data Bank taking into account the developed rules. The database was analyzed in order to elaborate its classification and find out novel structural features in helix packing. The database was subdivided into three subsets according to criterion of crossing helix projections on the parallel planes passing through the axes of the helices. All selected helical pairs were also divided according to the type of the helices forming the structures under study. A comprehensive analysis of the geometric characteristics of all the pairs of the obtained database has been carried out: the interplanar distance, the minimum distance between the axes of the helices, inter-helical angle, the torsion angle between axes of the helices. Constructed database was posted on the Internet at the network. Its interface allows performing searches, sorting, filtering of structures for all parameters, viewing histograms of the distribution of selected helical pairs depending on the chosen parameters, visualizing the 3D models of the investigated structural motifs.

Key words: structural motifs of proteins, pairs of helices, the point model, the torsion angle between axes of the helices, the interplanar distance, the minimum distance between the axes of the helices, inter-helical angle.

1. Введение

В данной работе решаются три основные задачи:

— распознавание и отбор в структурах белковых молекул, зарегистрированных в банке белковых структур PDB [1] все структурные мотивы,

образованные двумя спиралями любого типа, расположенными в полипептидной цепи друг за другом и соединенными между собой перетяжками различной длины, имеющими разные конформации; — всестороннее исследование геометрии отобранных спиральных пар; — разработка базы данных найденных структурных мотивов и размещение ее в сети интернет.

Актуальность создания базы данных указанных структур и их дальнейшее изучение следует из интереса исследователей к структурным мотивам, имеющим уникальные укладки полипептидной цепи в пространстве [2]. Структурные мотивы, образованные двумя α -спиралями, расположенными в полипептидной цепи друг за другом и соединенными между собой перетяжками описаны в работах [2, 3]. 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 спиралями. Примерами таких упаковок являются супервторичные структуры: α - α -уголки, α - α -шпильки, L-образные и V-образные структуры [3], однако всеобъемлющий анализ этих структурных мотивов во всех белках, представленных в PDB, до сих пор не проводился.

В опубликованных работах [4–6] 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 (спираль типа H), спирали 3_{10} (G-спираль) и π -спирали (I-спираль). Было показано, что 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, включающего структуры, имеющие пересечение проекций и осей спиралей, спиральные пары, образованные двумя α -спиралями, составляют подавляющее большинство (87.5%). Большая часть спиральных пар в аминокислотных цепочках (56%) имеют пересечения проекций спиралей.

2. Методы

2.1. Правила отбора спиральных пар

Разработанные правила для распознавания, отбора спиральных пар и для дальнейшего

исследования структурных мотивов описаны в опубликованных нами ранее работах [4–6]. Для полноты изложения приведем их здесь.

Анализ вторичной структуры аминокислотной цепи осуществлялся с помощью метода, разработанного авторами Dictionary for Secondary Structure of Protein [7]. Находим участки спиралей.

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

Для каждой спирали структуры по методу наименьших квадратов определяется ось цилиндра, на который эта спираль намотана, при котором минимизируется отклонение параметров спирали от параметров идеальной спирали [8, 9].

На рисунке 1 представлено геометрическое представление спиральной пары.

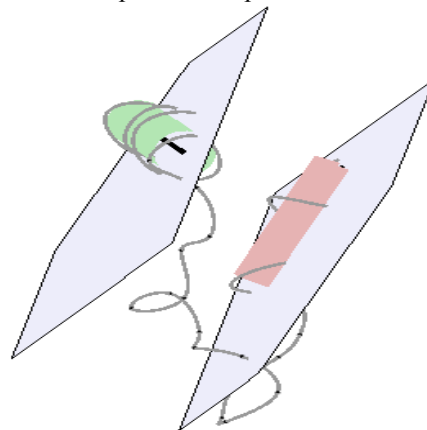


Рис. 1. Пример спиральной пары. Фрагмент цепи белка из банка белковых данных длиной 38 аминокислот (PDB ID 3A0B, координаты участка: 1000–1037). Показаны аппроксимирующие спираль цилиндры, а также плоскости, проходящие через оси цилиндров.

Две оси спиралей, расстояние между параллельными плоскостями, проведенными через оси, и проекции осей спиралей на плоскость полностью описывают пространственную структуру исследуемых мотивов.

Все множество спиральных пар разбивается на три подмножества по критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей:

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

— подмножество $\{B\}$ образуют все спиральные пары с ненулевым пересечением проекций, за исключением тех спиральных пар, где полигон пересечения [10] включает в себя точку пересечения проекций осей спиралей;

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

2.2. 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 A_2A_1, B_1B_2 . Точечная модель спиральной пары приведена на рисунке 2. Показаны оси спиральной пары. 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 (см. рис. 2).

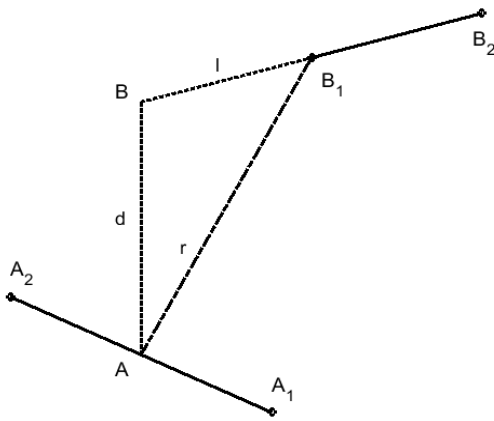


Рис. 2. Точечная модель спиральной пары.

2.3. Определения межспиральных расстояний, площади и периметра полигонов пересечения проекций спиралей и торсионного угла между осями в спиральных парах

Второй характеристикой спиральной пары является кратчайшее расстояние r между осями цилиндров (см. рис. 2). 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.

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.

Пусть есть четыре точки A_1, A_2, B_1, B_2 , являющиеся начальными и конечными точками осей двух спиралей, и три вектора A_2A_1, A_2B_1 и B_1B_2 . Если поместить точки A_1, A_2 и B_1 на координатную плоскость XU таким образом, что точка A_2 попадает в начало координат, а точка A_1 – на положительное направление оси $OX+$, то получается угол между осью $OX+$ и проекцией на плоскость XU вектора B_1B_2 , который называется торсионным углом Ω между осями спиралей. Положительным углом ($0^\circ \leq \Omega \leq 180^\circ$) считается направление против часовой стрелки и, соответственно, отрицательным ($-180^\circ \leq \Omega \leq 0^\circ$) – направление по часовой стрелке.

3. Результаты и обсуждение

Каждое исследование, посвященное изучению межспиральных взаимодействий в спиральных парах, расширяет понимание и дополняет знания о структурных мотивах, имеющих уникальные укладки полипептидной цепи в пространстве. Нами были исследованы межспиральные расстояния, характерное количество аминокислот между спиральями, торсионные углы Ω , межспиральные углы φ в отобранных структурах. Известно много работ [11–15], в которых изучаются взаимодействия спиралей в белковых молекулах. Отличие нашей работы от указанных состоит в том, что нами анализировались 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. Авторы других работ изучали взаимодействия спиралей в белковых молекулах без учета того факта, на сколько аминокислот отстоят друг от друга спиралей, и какую вторичную структуру они образуют. Также нужно отметить, что исследование нами проводилось на множестве спиральных пар из структур белковых молекул, зарегистрированных в PDB, без учета классификации белков, основанной на особенностях вторичной и третичной структуры белков [16].

Изучая все типы спиральных пар, принадлежащих различным подмножествам, было обнаружено, что во всех подмножествах больше всего исследуемых структур с короткой (от одной до семи аминокислот) 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, 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 ($N_p = 5$).

3.1. Анализ межспиральных расстояний

В работе проведен анализ межспиральных расстояний (межплоскостного, минимального и расстояния, описывающего относительное расположение спиралей) в спиральных парах всех типов и подмножеств. В ходе исследований нами показано, что распределения расстояний для спиральных пар, проекции которых не имеют пересечений, имеют более дальнедействующий характер, чем те, проекции которых пересекаются.

Распределение минимального расстояния r для всех спиральных пар, принадлежащих подмножеству $\{A\}$ и $\{B\}$, начинается не от нуля, а с некоторой величины, которая обусловлена минимальным радиусом, на который могут сблизиться оси двух спиралей. Он составляет около 2 \AA в обоих случаях и связан с радиусом Ван-дер-Ваальсового взаимодействия атомов спиралей.

Межплоскостное расстояние d для всех спиральных пар, имеющих пересечение осей спиралей (в данном случае межплоскостное расстояние d равно минимальному r), находится в диапазоне от 7 \AA до 12 \AA , максимум приходится на $d \approx 10 \text{ \AA}$. Абсолютное большинство структур этого множества являются спиральными парами НН типа. Среди структур указанного типа нет ни одной пары, имеющей межплоскостное расстояние d меньше 7 \AA , что связано со стерическими ограничениями для спиральных пар из этого множества [17–21]. Распределения межплоскостного расстояния d для пар типов НG и GG не имеют выраженного пика в области 10 \AA , носят многомодовый характер.

3.2. Анализ межспиральных углов в структурных мотивах, образованных двумя спиральями

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 -спиральями. Также показано, что значительная часть спиральных пар, образованных двумя α -спиральями и составляющих абсолютное большинство пар множества с пересечением проекций и осей спиралей, имеет острый угол ($-180^\circ \leq \Omega \leq 0^\circ$) между осями спиралей.

3.3. Анализ торсионных углов между осями спиралей в спиральных парах белковых молекул

В результате работы показано, что распределение пар, принадлежащих подмножеству $\{A\}$, в зависимости от угла Ω достаточно широкое, охватывающее практически весь диапазон углов. Для распределения спиральных пар НН и НG типа характерно наличие двух пиков в области 0° и 180° . Распределение пар типа GG имеет равномерный характер на всем диапазоне углов.

Среди структур, принадлежащих подмножеству $\{B\}$, очень мало пар типа GG, и они распределены на всем диапазоне углов. Распределение спиральных пар НG типа имеет максимум в области ($20^\circ \leq \Omega \leq 25^\circ$). В этой же области ($20^\circ \leq \Omega \leq 25^\circ$) отмечен максимум распределения спиральных пар, состоящих из двух α -спиралей, кроме того, имеется еще три локальных максимума: один – в области ($-15^\circ \leq \Omega \leq -10^\circ$), другой – в области $\Omega \approx -30^\circ$ и третий в области ($140^\circ \leq \Omega \leq 150^\circ$).

Торсионные углы большинства структур, имеющих пересечение проекций и осей спиралей и образованных двумя α -спиральями, имеют три локализации: ($-60^\circ \leq \Omega \leq -30^\circ$), в области 20° и 90° .

3.4. База данных двухспиральных мотивов белковых молекул

База данных двухспиральных мотивов белковых молекул выложена в сети интернет URL: <http://protdb.org>. Интерфейс базы данных позволяет:

- выполнять операции поиска, сортировки, фильтрации по всем параметрам;
- проводить статистический анализ и исследовать корреляцию геометрии пространственных структур исследуемых мотивов с аминокислотной последовательностью;
- просматривать 3D модели структурных мотивов.

4. Благодарности

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).

5. Заключение

Исследование проводилось на множестве спиральных пар, отобранных из структур белковых молекул, представленных в Protein Data Bank. Были разработаны правила отбора спиральных пар в структурах белковых молекул из PDB. Общее число найденных спиральных пар всех типов равно 220660. 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.

Спиральные пары были разделены по типу спиралей, образующих исследуемые структуры. Полученное множество спиральных пар также было разбито на три подмножества по критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей. Проведен всесторонний анализ геометрических характеристик отобранных структур, принадлежащих различным подмножествам: межспиральных расстояний и углов, торсионных углов, длин перетяжек. Показаны распределения характеристик различных типов структур внутри каждого подмножества. Разработанная база данных выложена в сети интернет.

6. Список литературы

1. 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. Т. 28. С. 235–242.
2. Efimov A.V. Standard structures in proteins. *Prog. Biophys. Molec. Biol.* 1993. V. 60. P. 201–239.
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. 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).
5. 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).
6. Тихонов Д.А., Куликова Л.И., Ефимов А.В. Исследование межспиральных углов в структурных мотивах, образованных двумя спиральями препринт *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).
7. 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.
8. Kabsch W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica*. 1976. V. 32. P. 922–923.
9. 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.
10. Legland D. *MatGeom: Matlab geometry toolbox for 2D/3D geometric computing*. URL: <http://github.com/dlegland/matGeom> (дата обращения: 12.02.2018).
11. Crick F.H.C. The Packing of α -helices: simple coiled-coils. *Acta Crystallographica*. 1953. V. 6. P. 689–697.
12. 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).
13. Walther D., Eisenhaber F., Argos P. Principles of helix-helix packing in proteins: the helical lattice superposition model. *Molecular Biology*. 1996. V. 255. P. 536–553.
14. 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.
15. Chothia C., Levitt M., Richardson D. Helix to helix packing in proteins. *Molecular Biology*. 1981. V. 145. P. 215–250.
16. Levitt M., Chothia C. Structural patterns in globular proteins. *Nature*. 1976. V. 261. P. 552–558.
17. 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. № 5. P. 1101–1115.
18. Calhoun J.R., Nastri F., Maglio O., Pavone V., Lombardi A., DeGrado W.F. Artificial diiron proteins: From structure to function. *Peptide Science*. 2005. V. 80. № 2–3. P. 264–278.
19. Chino M., Maglio O., Nastri 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)
20. Chino M., Leone L., Maglio O., Lombardi A. Designing Covalently Linked Heterodimeric Four-Helix Bundles. *Methods in enzymology*. 2016. V. 580. № 21. P. 471–499. doi: [10.1016/bs.mie.2016.05.03](https://doi.org/10.1016/bs.mie.2016.05.03).
21. 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.