

Развитие базы данных двухспиральных мотивов белковых молекул и вычислительные сервисы для их анализа

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

¹Институт теоретической и экспериментальной биофизики РАН, Пуццино, Россия

²Институт математических проблем биологии РАН – филиал ИПМ РАН им. М.В. Келдыша, Пуццино, Россия

³Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича, Москва, Россия

volodyrv@mail.ru

Описываемая в настоящей работе база данных «Structural Elements Database» объединяет сервисы хранения данных и вычислительные алгоритмы для их анализа. Structural Elements DB на сегодня содержит свыше 45000 аннотированных белковых мотивов – спиральных пар. Спиральная пара представляет собой супервторичную структуру и состоит из двух соседних по цепи спиралей и неструктурированный участок (перетяжку) между ними. Пространственная ориентация двух спиралей определяет тип спиральной пары: α - α -уголок, V-структура, L-структура, α - α -шпилька и др. Был создан новый интерфейс базы данных «Structural Elements DB», который дополнен возможностями графического представления данных отдельной спиральной пары, а также статистической обработки выборки отсортированных спиральных пар. Предусмотрена возможность загрузки результатов пространственного и математического анализа отдельных спиральных пар, а также выборки спиральных пар.

Ключевые слова: база данных, спиральные пары, геометрические характеристики, структурный анализ.

Development of a Database of Two-Helical Motifs of Protein Molecules and Computational Services for Their Analysis

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

¹Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, Russia

²Institute of Mathematical Problems of Biology RAS – the Branch of Keldysh Institute of Applied Mathematics RAS, Pushchino, Russia

³V.N. Orekhovich Institute of Biomedical Chemistry, Moscow, Russia

The Structural Elements Database combines data storage services and computational algorithms for their analysis. Structural Elements DB currently contains over 45000 annotated protein motifs – helical pairs. The helical pair is a super-secondary structure and consists of two α -helices and an unstructured coil. The spatial orientation of the two α -helices determines the type of helical pair: α - α -corner, V-structure, L-structure, α - α -hairpin et al. A new interface for the "Structural Elements DB" database was created, which was supplemented with the capabilities of graphical presentation of data for an individual spiral pair, as well as statistical processing of a sample of sorted helical pairs. It is possible to download the results of spatial and mathematical analysis of individual helical pairs, as well as a sample of helical pairs.

Key words: database, helical pairs, geometric characteristics, structural analysis.

1.База данных структурных элементов

Двумя отечественными коллективами – Институтом теоретической и экспериментальной биофизики РАН (Пуццино) и Институтом математических проблем биологии РАН (Пуццино) и при экспертной поддержке Института биомедицинской химии им. А.Н. Ореховича

(Москва) были продолжены работы по улучшению графических и математических сервисов для Structural Elements DB [1]. База данных предназначена для исследователей в области биомедицины и фармацевтики, позволяет анализировать геометрические особенности супервторичных структур в белковой глобуле, проводить анализ изменений, обусловленных

вероятными аминокислотными заменами и посттрансляционными модификациями. Актуальность создания базы данных структурных мотивов, имеющих уникальные укладки полипептидной цепи в пространстве и их дальнейшее изучение, следует из интереса исследователей к указанным структурам [2–10].

1.1. Типы спиральных пар

Компактные в пространстве и стабильные мотивы супервторичной структуры белка состоят из двух спиралей, соединенных перетяжкой, так называемые спиральные пары. Супервторичные структуры включают α - α -уголки, α - α -шпильки, L-структуры и V-структуры [10–13]. Структуры спиральных пар компактны в пространстве и содержат гидрофобное ядро и полярную оболочку. Каждая α -спираль имеет по крайней мере один гидрофобный остаток на спираль; в этом случае гидрофобные остатки располагаются по одну сторону от α -спирали и образуют сплошной гидрофобный кластер [11].

Для изучения структурных мотивов в белковых молекулах мы сформулировали правила распознавания и выбора спиральных пар.

1.2. Правила распознавания и выбора спиральной пары для пептидного фрагмента

Правила распознавания и отбора спиральных пар описаны в ранее опубликованных нами работах [1,14–15]. Здесь для удобства представим их в краткой форме.

- Анализ вторичной структуры аминокислотной цепи осуществляется с помощью Dictionary for Secondary Structure of Protein [16]. Поиск участков α -, 3_{10} -, π -спирали.

- Спиральной парой считается участок структуры белка, которая образована двумя соседними по полипептидной цепи спиралью, соединенными неструктурированным фрагментом (перетяжка) произвольной длины и конфигурации.

- Спиральные пары в белках встречаются следующих типов: α - α -уголок, α - α -шпилька, V- и L-структуры.

- Каждая спираль пары может быть вписана в цилиндр. Ось цилиндра определяется по методу наименьших квадратов с минимальным отклонением параметров спирали от идеальной спирали.

- Две оси спиралей определяют пространственную структуру спиральной пары. Пространственная структура спиральной пары полностью описывается межплоскостным расстоянием (d) и проекциями осей спиралей на плоскость.

- Площадь полигона (S) пересечения проекций спиралей, зависит от взаимного расположения спиралей (геометрии) и от диаметра цилиндров спиралей. Диаметр цилиндра определяется типом

спирали и величиной среднего размера боковой цепи.

- Спиральные пары могут быть сгруппированы в три множества по критерию пересечения проекций спиралей на параллельные плоскости, проходящие через ось спиралей:

- множество {A} образуют спиральные пары, у которых проекции спиралей не имеют пересечений;

- множество {B} образуют 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 спиральные пары, где полигон пересечения включает в себя точку пересечения проекций осей спиралей.

1.3. Отличительные свойства типов спиральных пар

1. α - α уголок

- плотная ортогональная упаковка спиралей;
- проекции спиралей и оси спиралей пересекаются;

- значения межплоскостного расстояния и минимального расстояния равны ($d = r$);

- расстояние между осями спиралей $d \approx (9-11) \text{ \AA}$;

- площадь и периметр пересечения проекций спиралей больше, чем у других двухспиральных мотивов;

- межспиральные углы торсионный θ и плоский ϕ имеют значения $70^\circ-110^\circ$;

- длина перетяжки (N_p) обычно составляет 4–5 аминокислот.

2. α - α -шпилька

- плотная антипараллельная упаковка спиралей;
- проекции спиралей пересекаются, но оси спиралей не пересекаются;

- диапазон межплоскостного расстояния d шире, чем у α - α -уголка;

- углы между спиралью $0^\circ-30^\circ$;

- длина перетяжки (N_p) может варьироваться;

- обычно состоит из двух длинных спиралей.

3. L-образная структура

- плотная скошенная упаковка спиралей;

- проекции спиралей и оси спиралей не пересекаются ($S \approx 0 \text{ \AA}^2$, $P \approx 0 \text{ \AA}$);

- угол между спиралью имеет значение $\phi \approx 90^\circ$;

- длина перетяжки (N_p) может отличаться.

4. V-образная структура

- плотная скошенная упаковка спиралей;

- проекции спиралей (чаще всего) и оси спиралей не пересекаются;

- площадь (S) и периметр (P) имеют значения, близкие к нулю;

- угол между спиралью $\phi \approx 50^\circ$;

- перетяжка (N_p), как правило, состоит из одной аминокислоты;

- длина спиралей, как правило, не превышает трех-четырёх витков.

2. База данных структурных элементов

Structural Elements Database дополнена информацией о количестве найденных по запросу пользователя релевантных статей. Форма фильтров и настройки отображения дополнена параметрами сравнения. Внешне форма выглядит следующим образом (рис. 1).

Поле	Отображать	Фильтр	
Уникальный идентификатор (idx)	<input type="checkbox"/>	От: 0 До: 1027510	Построить распределение
Код PDB (pdb)	<input checked="" type="checkbox"/>		Построить распределение
Начало структуры (iA)	<input checked="" type="checkbox"/>	От: 1 До: 11620	Построить распределение
Конец структуры (iB)	<input checked="" type="checkbox"/>	От: 1 До: 11640	Построить распределение
Начало петли (iA)	<input checked="" type="checkbox"/>	От: 1 До: 11670	Построить распределение
Конец петли (iB)	<input checked="" type="checkbox"/>	От: 1 До: 11670	Построить распределение
Угол между осями спиралей (field_A)	<input type="checkbox"/>	От: 2.0 До: 176	Построить распределение
Расстояние между плоскостями (field_L)	<input type="checkbox"/>	От: 0.0 До: 20.00	Построить распределение
Площадь пересечения проекций (field_S)	<input type="checkbox"/>	От: 0.0 До: 100	Построить распределение
Вращательная структура DSSP (ssr)	<input type="checkbox"/>		Построить распределение
Первичная структура Helix A (aaA)	<input checked="" type="checkbox"/>		Построить распределение
Первичная структура петли (asp)	<input checked="" type="checkbox"/>		Построить распределение
Первичная структура Helix B (aaB)	<input checked="" type="checkbox"/>		Построить распределение
Просмотр в 3D (3d)	<input checked="" type="checkbox"/>		Построить распределение
Имя субструктуры (Chain ID) которой принадлежит альфа структура (id)	<input type="checkbox"/>		Построить распределение
Разрешение для данной PDB, ангстрем (res)	<input type="checkbox"/>	От: 0 До: 16	Построить распределение
Среднеквадратичная погрешность аппроксимации альфа спирали молекулы идеальной спиралью для первой спирали A (rmsdA)	<input type="checkbox"/>	От: 0.0 До: 0.5	Построить распределение
Среднеквадратичная погрешность аппроксимации альфа спирали молекулы идеальной спиралью для второй спирали B (rmsdB)	<input type="checkbox"/>	От: 0.0 До: 0.5	Построить распределение
len (len)	<input type="checkbox"/>	От: 1 До: 1011	Построить распределение
len (aa)	<input type="checkbox"/>	От: 1 До: 1716	Построить распределение
len (ssr)	<input type="checkbox"/>	От: 1 До: 703	Построить распределение
len (asp)	<input type="checkbox"/>	От: 1 До: 1000	Построить распределение
len (aaA)	<input type="checkbox"/>	От: 1 До: 1006	Построить распределение
len (aaB)	<input type="checkbox"/>	От: 1 До: 1700	Построить распределение
len (aa)	<input type="checkbox"/>	От: 1 До: 1412	Построить распределение
Длина петли (lpa)	<input type="checkbox"/>		Построить распределение
Длина структуры Helix B (hlLength)	<input type="checkbox"/>		Построить распределение
Длина структуры Helix A (haLength)	<input type="checkbox"/>		Построить распределение

Рис. 1. Форма фильтров и настройки поиска структур белков.

Зарегистрированный пользователь может сохранить фильтр в файл для повторного использования в будущем или для передачи другому пользователю.

Форма просмотра 3D-модели элемента молекулы открывается по нажатию на ссылку в столбце “просмотр в 3D” основной таблицы. Форма содержит встраиваемое в веб-страницу обособленное приложение JSmol для отображения пространственной модели (рис. 2).

Страница отображения диаграмм открывается по нажатию ссылки “Построить распределение” на форме настройки фильтров. На форме отображаются настраиваемые диаграммы встречаемости и зависимости полей. Диаграмма встречаемости является столбчатой диаграммой и отображает частоту появления того или иного параметра в сформированной выборке (рис. 3).

График зависимости позволяет построить диаграмму по двум параметрам сравнения и тем самым выявить корреляции между ними. Параметры задаются в выпадающих списках над графиком.

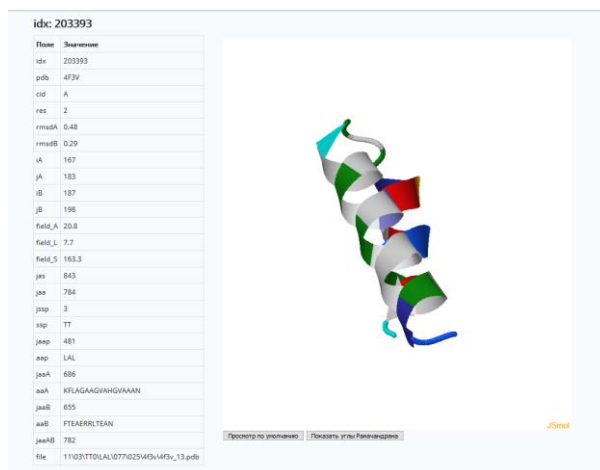


Рис. 2. Форма пространственной визуализации (JSmol) выбранной белковой структуры, которая содержит параметры элемента молекулы.

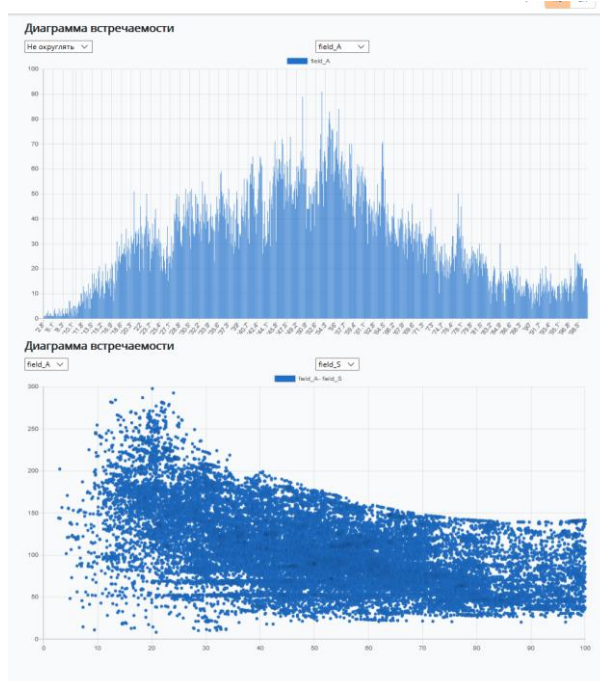


Рис. 3. Вид диаграммы встречаемости: столбчатый и точечный для выборки пространственных структур.

3. Благодарности

Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований, проект № 18-07-01031.

4. Список литературы

1. Rudnev V.R., Tikhonov D.A., Kulikova L.I., Gubin M.Yu., Efimov A.V. Database of two-helical motifs of protein molecules and computer services for their analysis. *Journal of Bioinformatics and Genomics*. 2019. 3(12). doi: <https://doi.org/10.18454/jbg.2019.3.12.1>

2. 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).
3. Adamian L.I., Liang J. Helix-helix packing and interfacial pairwise interactions of residues in membrane proteins. *J. Mol. Biol.* 2001. V. 311. P. 891–907.
4. 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).
5. 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).
6. 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).
7. Chino M., Leone L., Maglio O., Lombardi A. Designing Covalently Linked Heterodimeric Four-Helix Bundles. *Methods in enzymology.* 2016. V. 580. P. 471–499. doi: [10.1016/bs.mie.2016.05.036](https://doi.org/10.1016/bs.mie.2016.05.036).
8. 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).
9. 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. doi: [10.1006/jmbi.1996.0044](https://doi.org/10.1006/jmbi.1996.0044).
10. Efimov A.V. Standard Structures in Proteins. *Prog. Biophys. Mol. 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).
11. Brazhnikov E., Efimov A. Structure of α - α -Hairpins with Short Connections in Globular Proteins. *Mol. Biol.* 2001. V. 35. P. 89–97. doi: [10.1023/A:1004859003221](https://doi.org/10.1023/A:1004859003221).
12. Efimov A.V. L-shaped structure from two α -helices with a proline residue between them. *Mol. Biol. (Mosk.)* 1992. V. 26. P. 1370–1376.
13. 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.
14. Tikhonov D., Kulikova L., Kopylov A., Malsagova K., Stepanov A., Rudnev V., Kaysheva A. Super Secondary Structures of Proteins with Post-Translational Modifications in Colon Cancer. *Molecules.* 2020. V. 25. P. 3144. doi: [10.3390/molecules25143144](https://doi.org/10.3390/molecules25143144).
15. 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)
16. 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).