

База данных структурных мотивов белковых молекул и вычислительные сервисы для их анализа

Руднев В.Р.¹, Тихонов Д.А.^{2,1}, Куликова Л.И.², Губин М.Ю.³, Ефимов А.В.⁴

¹Институт теоретической и экспериментальной биофизики РАН, Пуццино, Россия

²ИМПБ РАН – филиал ИПМ им. М.В. Келдыша РАН, Пуццино, Россия

³Национальный исследовательский Томский политехнический университет

⁴Институт белка РАН, Пуццино, Россия

volodyrv@mai.ru

В данной работе представлена база данных структурных мотивов белковых молекул, состоящих из двух элементов вторичной структуры, имеющих уникальные укладки полипептидной цепи в пространстве. Исследуемые мотивы представляют собой пары любого типа спиралей, соединёнными между собой различной ненулевой длины и различной конформации перетяжки. Для каждой структуры рассчитаны геометрические параметры. Интерфейс базы данных позволяет выполнять операции поиска, сортировки, фильтрации по всем параметрам, что дает возможность получать выборки структур с заданными геометрическими характеристиками. Также интерфейс системы позволяет проводить статистический анализ и строить гистограммы распределения различных характеристик структур в выборке. Таким образом, имеется возможность исследовать корреляцию геометрии пространственных структур исследуемых мотивов с аминокислотной последовательностью. Система позволяет выгружать полученные выборки в файловую систему в формате CSV. Формат выгрузки и набор выгружаемых данных настраивается в интерфейсе. Выгрузка может содержать как все геометрические характеристики интересующих структур, так и соответствующие аминокислотные последовательности. Интерфейс базы данных позволяет также просматривать 3D модели интересующих структурных мотивов.

Ключевые слова: структурные мотивы белковых молекул, база данных, спиральные пары, межспиральные расстояния, углы между осями спиралей, площадь пересечения проекций спиралей.

Database of structural motifs of protein molecules and computer services for their analysis

Rudnev V.R.¹, Tikhonov D.A.^{2,1}, Kulikova L.I.², Gubin M.Yu.³, Efimov A.V.⁴

¹IMPB RAS – Branch of KIAM RAS, Pushchino, Russia

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

³National Research Tomsk Polytechnic University, Tomsk, Russia

⁴Institute of Protein Research, RAS, Pushchino, Russia

This paper presents a database of structural motifs of protein molecules consisting of two elements of a secondary structure that have unique spatial stacking of a polypeptide chain. The motives investigated are pairs of any type of helix, connected by a different non-zero length and different conformation of the connection. For each structure, geometric parameters are calculated. The database interface allows you to perform searches, sorting, filtering on all parameters, which makes it possible to obtain samples of structures with given geometric characteristics. Also, the system interface allows for statistical analysis and the construction of histograms for the distribution of various characteristics of structures in the sample. Thus, it is possible to investigate the correlation of the geometry of the spatial structures of the investigated motifs with the amino acid sequences. The system allows uploading of the received samples to the file system in CSV format. The upload format and set of uploaded data is configured in the interface. Downloading can contain both all the geometric characteristics of the structures of interest, and the corresponding amino acid sequences. The database interface also allows you to view 3D models of structural motifs.

Key words: protein structural motifs, database, helix pairs, inter-helix distances, helix axis angles, area of helix projection intersection.

1. Введение

Актуальность создания базы данных структурных мотивов, имеющих уникальные укладки полипептидной цепи в пространстве и их дальнейшее изучение, следует из интереса исследователей к указанным структурам [1–8]. Структурные мотивы, образованные двумя α -спиралями, расположенными в полипептидной цепи друг за другом и соединенными между собой перетяжками описаны в работах [5, 6]. 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 спиралью. Примерами таких упаковок являются супервторичные структуры: α - α -уголки, α - α -шпильки, L-образные и V-образные структуры [6]. Таким образом, создание базы данных всех структурных мотивов белковых молекул, зарегистрированных в банке белковых структур PDB [9], позволяющий проводить всеобъемлющий анализ структур, является крайне важной и актуальной задачей [10–16]. Данная работа посвящена созданию базы данных спиральных пар, образованных двумя спиралью любого типа, расположенными в полипептидной цепи друг за другом и соединенными между собой перетяжками различной длины, имеющими разные конформации. В опубликованных работах [17–19] был описан разработанный нами подход отбора спиральных пар в структурах белковых молекул, представленных в PDB. 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).

Была предложена точечная модель спиральной пары [17], описывающая структуры четырьмя точками в пространстве. Действительно, если мы аппроксимируем обе спирали цилиндрами, на которые намотаны спирали, образованные нитью, проходящие через C_α -атомы, то начала и концы осей цилиндров дадут нам те четыре точки, которые полностью опишут данную супервторичную структуру. На рисунке 1 дано геометрическое представление супервторичной структуры, образованной двумя спиралью и перетяжкой между ними. Это фрагмент цепи белка из Банка белковых данных (PDB ID 3A0B, C_α : 1000–1037). Каждая спираль представлена в виде цилиндра, определены оси цилиндров. Показаны также плоскости, проходящие через оси цилиндров.

Важными характеристиками спиральных пар являются межспиральные расстояния (межплоскостное расстояние, минимальное расстояние и расстояние, которое описывает относительное расположение спиралей в

спиральной паре), углы между осями спиралей и площадь пересечения проекций спиралей [20].

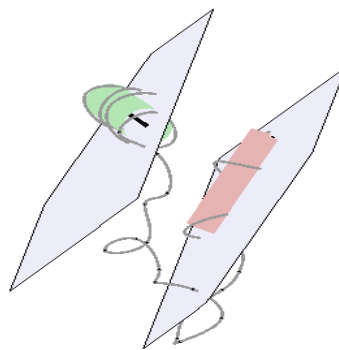


Рис. 1. Геометрическое представление спиральной пары.

По критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей, все множество спиральных пар разбито на три подмножества, включающие в себя различные типы супервторичных структур. Площадь полигона пересечения проекций спиралей, помимо геометрии, зависит еще от диаметров цилиндров спиралей [21]. Диаметр цилиндров определяется типом спирали и величиной среднего размера боковой цепи. На рисунке 2 изображен полигон пересечения проекций спирали для спиральной пары, представленной на рисунке 1. Полигон пересечения выделен цветом, указаны его площадь S , периметр P и величина межплоскостного расстояния d . Отмечена точка пересечения проекций осей спиралей.

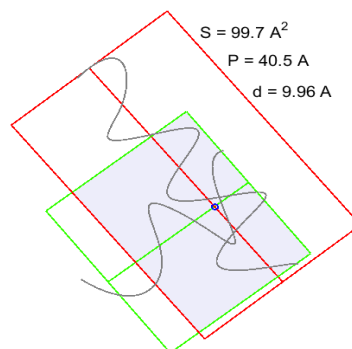


Рис. 2. Пересечение проекций цилиндров спиралей спиральной пары (PDB ID 3A0B, координаты участка: 1000–1037).

Основные характеристики исследуемых структур и их описания представлены ниже:

- код PDB – уникальный код белковой молекулы в Protein Data Bank, в которой обнаружена структура;
- начало структуры – номер аминокислотного остатка, соответствующего началу первой спирали в белковой молекуле;
- начало перетяжки – номер аминокислотного остатка, соответствующего началу перетяжки;

- конец перетяжки – номер аминокислотного остатка, соответствующего концу перетяжки;
- конец структуры – номер аминокислотного остатка, соответствующего концу второй спирали;
- угол между осями спиралей – межспиральный угол ϕ [18];
- расстояние между плоскостями – межплоскостное расстояние d [17];
- площадь пересечения проекций цилиндров – площадь полигона пересечения проекций цилиндров;
- вторичная структура DSSP – результат обработки структуры программой DSSP [22];
- первичная структура Helix A – аминокислотная последовательность, соответствующая первой спирали структуры;
- первичная структура перетяжки – аминокислотная последовательность, соответствующая перетяжки;
- первичная структура Helix B – аминокислотная последовательность, соответствующая второй спирали;
- длина перетяжки – количество аминокислотных остатков в перетяжке;
- длина структуры Helix A – количество аминокислотных остатков в первой спирали;
- длина структуры Helix B – количество аминокислотных остатков во второй спирали.

Интерфейс базы данных позволяет выполнять операции поиска, сортировки, фильтрации по всем параметрам, что дает возможность получать выборки структур с заданными геометрическими характеристиками. Также интерфейс системы позволяет строить гистограммы распределений выбранных параметров в выборке.

Разработанный программный комплекс позволяет хранить и обрабатывать информацию, описывающую структурные мотивы белковых молекул. Данные могут быть сформированы либо как описание фрагмента (начальные, конечные точки и другие характеристики фрагмента), либо как детализированная последовательность атомов. При этом один способ хранения данных не исключает другого. Система также предоставляет возможность поиска, фильтрации, агрегации структур по набору их характеристик, а также обработки (статистический анализ данных, выявление закономерностей и др.).

Реализованный базовый функционал просмотра структурных элементов с ограниченным набором аналитических инструментов подразумевает дальнейшее расширение возможностей. Также планируется реализация функционала импорта структур, созданных посредством приложения Matlab.

2. Используемые технологии

Общая схема решения представлена на рисунке 3.

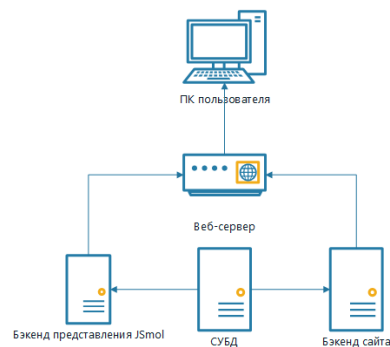


Рис. 3. Общая схема компонентов решения.

2.1. Серверная часть

Серверная часть портала написана на языке программирования Python. Этот язык был выбран по нескольким причинам:

1. Python является одним из наиболее часто используемых языков для разработки приложений, связанных с Data science, и обладающим широким спектром инструментария для обработки данных;
2. для Python есть развитый инструментарий работы с биологическими данными, например, BioPython [23];
3. существование готовых модулей для импорта файлов Matlab в Python [24].

Также важную роль сыграли простота и удобство разработки интернет-приложений на Python, особенно в связке с существующими фреймворками. В данном проекте использовался фреймворк Django в качестве основы для сайта. Django предоставляет следующую функциональность, использованную в проекте:

Для хранения данных на данном этапе используется реляционная СУБД (Система Управления Базами Данных) MySQL [25]. Выбор СУБД был определен, в первую очередь, простотой установки и интеграции с приложением. В дальнейшем для поддержки более широких и гибких возможностей планируется переход на нереляционную (реляционные СУБД работают с данными в виде таблиц и связями между ними; к нереляционным относятся любые другие СУБД) СУБД Elasticsearch, которая обладает следующими возможностями и преимуществами:

1. высокая производительность при работе с большими объемами данных;
2. удобный и функциональный интерфейс для интеграции с программными средствами;
3. возможность хранения и индексирования произвольных структур данных, в том числе вложенных;
4. широкие возможности для поиска данных.

2.2. Клиентская часть

Клиентская часть разделяется на три части:

1. интерфейс просмотра табличных данных;
2. интерфейс просмотра распределений;
3. интерфейс визуализации молекул и функциональных элементов.

Интерфейс просмотра табличных данных написан на JavaScript + HTML + CSS для соблюдения основных стандартов, применяемых к технологиям для разработки сайтов в сети Интернет и отсутствия необходимости в сложной логике на стороне клиента, которая могла бы потребовать использования препроцессоров CSS, визуальных редакторов HTML или фреймворков JavaScript.

Интерфейс просмотра гистограмм распределений создан с использованием библиотеки Google Charts [26], обеспечивающей необходимую функциональность при простоте интеграции и достаточно высокой скорости работы.

Интерфейс визуализации использует систему отображения JSmol [27], позволяющую в веб-браузере просматривать визуализацию молекул и фрагментов молекул. Данное решение является безальтернативным для такой функциональности. При этом отображаемые фрагменты молекул размещаются в файловой системе как файлы в формате PDB. Для работы JSmol также используется его собственный серверный скрипт на php, который в перспективе планируется заменить на собственный бэкенд, позволяющий использовать данные из базы данных, а не напрямую в виде файлов.

3. Инструменты пользователя

3.1. Описание страницы просмотра табличных данных

Страница просмотра табличных данных приведена на рисунке 4.

Рис. 4. Форма просмотра табличных данных.

Страница состоит из шапки, тела (таблицы с параметрами элементов молекул) и панели навигации.

Шапка состоит из: заголовка, главного меню, кнопки отображения панели фильтров, кнопки экспорта в Excel (CSV) (по нажатию позволяет скачать файл CSV с полным содержанием таблицы, с учетом фильтров и сортировки), указания количества результатов поиска.

Таблица содержит записи, найденные с помощью фильтра на текущей странице. Набор отображаемых столбцов настраивается вместе с фильтром по полям. С помощью треугольников в

заголовках столбцов таблицы можно задавать сортировку и направление сортировки:

Чтобы отсортировать таблицу по какому-либо столбцу, нужно нажать на треугольник в этом столбце. Сортировка выполнится сразу же, без запроса подтверждения. Треугольник станет закрашенным, а значения в таблице отсортируются. Для изменения направления сортировки в том же столбце нужно нажать на уже закрашенный треугольник ещё раз (см. рис. 5).

Код d PDB	Начало структуры	Начало перетяжки	Конец перетяжки	Конец структуры	Первичная структура
В▲	уры△	жкж△	жкж△	ры△	Helix A△
9XIM	13	16	34	44	LWTV
9XIM	405	408	426	436	LWTV
9XIM	796	799	817	827	LWTV

Рис. 5. Пример сортировки таблицы по полю «Код PDB» по возрастанию.

Панель навигации обеспечивает возможность перехода к первой (“<<<”), предыдущей (“<”), следующей (“>”), последней (“>>>”) страницам таблицы, перехода к произвольной странице. Также позволяет быстро перейти к ближайшим 10-ти страницам кликом по номеру страницы.

3.2. Форма фильтров и настройки отображения

Форма фильтров и настройки отображения открывается на той же странице по нажатию кнопки “Настроить фильтрацию и отображение полей” и может быть скрыта повторным нажатием на ту же кнопку. Внешний вид формы приведен на рисунке 6.

Форма представляет собой таблицу, состоящую из фиксированного набора столбцов, каждая строка столбца:

- «Поле» есть название параметра элемента структуры, к которому относятся настройки справа;
- «отображать» содержит флаг, управляющий отображением столбца параметра в таблице с результатами. Если флаг установлен, то столбец параметра будет отображаться в таблице. Важно: фильтры по скрытым столбцам всё равно применяются;
- «Фильтр» содержит поля фильтров. Вид фильтра зависит от типа данных, к которому относится параметр. Таким образом, есть два разных варианта фильтра: для чисел и для строк. Фильтр для чисел состоит из двух полей: “от”, “до”. Значениями по умолчанию для обоих являются, соответственно, наибольшее и наименьшее значения этого поля, найденные в базе данных. Фильтр для строк состоит из поля “Подстрока” и флага “regex”, по умолчанию значение пустое, флаг “regex” не проставлен. Пустое значение означает, что фильтр не будет применяться. Чтобы отфильтровать значения, содержащие некоторый набор символов нужно ввести этот набор символов в строку. Поиск является регистронезависимым, т.е. по запросу “XYZ” будут найдены те же записи, что и по “xyz”.

– «Распределение» содержит ссылку «Построить распределение», по нажатию на которую откроется в новой вкладке график распределения структур в отфильтрованной выборке в зависимости от выбранного параметра.

Поле	Отображать	Фильтр	Распределение
Уникальный идентификатор (id)	<input type="checkbox"/>	от 0 до 1627948	Построить распределение
Код в PDB (pdb)	<input checked="" type="checkbox"/>	Поиск: <input type="text"/>	Построить распределение
Начало структуры (A)	<input checked="" type="checkbox"/>	от 2 до 11829	Построить распределение
Начало перетяжки (A)	<input checked="" type="checkbox"/>	от 5 до 11846	Построить распределение
Конец перетяжки (B)	<input checked="" type="checkbox"/>	от 9 до 11872	Построить распределение
Конец структуры (B)	<input checked="" type="checkbox"/>	от 14 до 11879	Построить распределение
Угол между осями спиралей (field_A)	<input type="checkbox"/>	от 2,8 до 176	Построить распределение
Расстояние между плоскостями (field_L)	<input type="checkbox"/>	от 4,35 до 28,06	Построить распределение
Площадь пересечения плоскостей (field_S)	<input type="checkbox"/>	от 8,3 до 298	Построить распределение
Вторичная структура DSSP (ssr)	<input type="checkbox"/>	Поиск: <input type="text"/>	Построить распределение
Первичная структура Helix A (aaA)	<input checked="" type="checkbox"/>	Поиск: <input type="text"/>	Построить распределение
Первичная структура перетяжки (ssr)	<input checked="" type="checkbox"/>	Поиск: <input type="text"/>	Построить распределение
Первичная структура Helix B (aaB)	<input checked="" type="checkbox"/>	Поиск: <input type="text"/>	Построить распределение

Рис. 6. Внешний вид формы фильтров.

3.3. Просмотр 3D-модели структурного мотива белковой молекулы

Форма просмотра 3D-модели элемента молекулы открывается по нажатию на ссылку в столбце «просмотр в 3D» основной таблицы. На рисунке 7 представлен пример отображения структурного мотива. Форма содержит апплет JSmol для отображения пространственной структуры исследуемого мотива.

Страница состоит из:

- апплета 3D-просмотра, позволяющий с помощью курсора мыши изменять угол поворота элемента молекулы относительно позиции взгляда пользователя и колесиком мыши менять масштаб;
- параметров спиральных пар: код PDB, начало и конец последовательности;
- консоли интерактивных скриптов Jmol [28];
- кнопки «Скачать», при нажатии которой скачивается на компьютер пользователя pdb-файл с отображаемым элементом белковой молекулы.

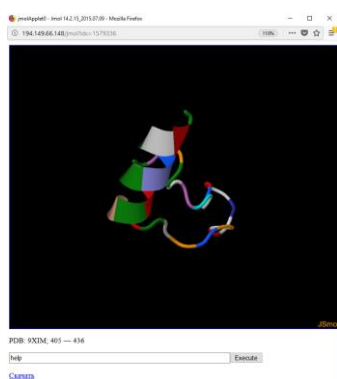


Рис. 7. Пример отображения двухспиральных структур белковых молекул.

3.4. Страница отображения распределения

Страница открывается по нажатию ссылки «Построить распределение» на форме настройки фильтров. На форме отображается график

распределения структур в зависимости от выбранного параметра (см. рис. 8). При построении распределения учитываются все настроенные фильтры.

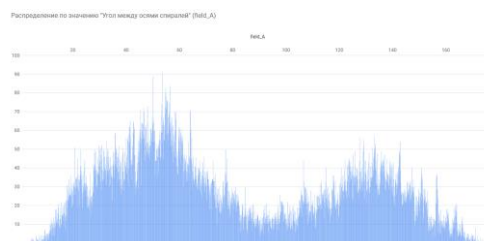


Рис. 8. Страница отображения распределения.

4. Заключение

В работе представлена база данных структурных мотивов белковых молекул, состоящих из двух элементов вторичной структуры, имеющих уникальные укладки полипептидной цепи в пространстве. Исследуемые мотивы представляют собой пары любого типа спиралей, соединёнными между собой различной ненулевой длины и различной конформации перетяжки. Для каждой структуры рассчитаны геометрические параметры. База данных двухспиральных мотивов белковых молекул выложена в сети интернет URL: <http://protdb.org>. Разработанная база данных позволяет строить выборки структурных мотивов по интересующим геометрическим параметрам, а также предоставляет возможность исследовать взаимосвязь геометрии пространственных структур с аминокислотной последовательностью с помощью разработанных инструментов.

Интерфейс базы данных позволяет выполнять операции поиска, сортировки, фильтрации по всем параметрам, что дает возможность получать выборки структур с заданными геометрическими характеристиками. Также интерфейс системы позволяет проводить статистический анализ и строить гистограммы распределения различных характеристик структур в выборке. Система позволяет выгружать полученные выборки в файловую систему в формате CSV. Выгрузка может содержать как все геометрические характеристики интересующих структур, так и соответствующие аминокислотные последовательности. Интерфейс базы данных позволяет также просматривать 3D модели структурных мотивов.

5. Благодарности

Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований, проекты № 18-07-01031-а и № 16-01-00692-а.

6. Список литературы

1. Phillips D.C. The hen egg-white lysozyme molecule. *Proc. Natl. Acad. Sci. USA*. 1967. V. 57. P. 483–495. doi: [10.1073/pnas.57.3.483](https://doi.org/10.1073/pnas.57.3.483).

2. Птицын О.Б. Стадийный механизм самоорганизации белковых молекул. *Докл. АН СССР*. 1973. Т. 210. С. 1213–1215.
3. Wetlauffer D.B. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA*. 1973. V. 70. P. 697–701. doi: [10.1073/pnas.70.3.697](https://doi.org/10.1073/pnas.70.3.697).
4. Edwards M.S., Sternberg J.E., Thornton J.M. Structural and sequence patterns in the loops of $\beta\alpha\beta$ units. *Protein Engineering*. 1987. V. 1. P. 173–181. doi: [10.1093/protein/1.3.173](https://doi.org/10.1093/protein/1.3.173).
5. 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).
6. 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.
7. Lin S.I., Tsai J., Nussinov R. A study of 4-helix bundles—investigating protein folding via similar architectural motifs in protein cores and in subunit interfaces. *J. Mol. Biol.* 1995. V. 248. P. 151–161.
8. Chothia C., Levitt M., Richardson D. Structure of proteins: packing of α -helices and pleated sheets. *Proc. Natl. Acad. Sci.* 1977. V. 74. 4130–4134. doi: [10.1073/pnas.74.10.4130](https://doi.org/10.1073/pnas.74.10.4130).
9. 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. doi: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235).
10. 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.
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. № 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. № 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. V. 580. № 21. P. 471–499. doi: [10.1016/bs.mie.2016.05.03](https://doi.org/10.1016/bs.mie.2016.05.03).
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. 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).
17. Тихонов Д.А., Куликова Л.И., Ефимов А.В. Статистический анализ внутренних расстояний спиральных пар в белковых молекулах. *Мат. биол. и биоинф.* 2016. Т. 11. № 2. С. 170–190. doi: [10.17537/2016.11.170](https://doi.org/10.17537/2016.11.170).
18. Тихонов Д.А., Куликова Л.И., Ефимов А.В. Исследование межспиральных углов в структурных мотивах, образованных двумя спиральями препринт *Мат. биол. и биоинф.* 2017. Т. 12. № 1. С. 83–101. doi: [10.17537/2017.12.83](https://doi.org/10.17537/2017.12.83).
19. Тихонов Д.А., Куликова Л.И., Ефимов А.В. Анализ торсионных углов между осями спиралей в спиральных парах белковых молекул. *Мат. биол. и биоинф.* 2017. Т. 12. № 2. С. 398–410. doi: [10.17537/2017.12.398](https://doi.org/10.17537/2017.12.398).
20. Legland D. *MatGeom: Matlab geometry toolbox for 2D/3D geometric computing*. URL: <http://github.com/dlegland/matGeom> (дата обращения: 19.07.2018).
21. 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. Москва: ИПМ им. М.В.Келдыша, 2018. № 59. 24 с. doi: [10.20948/prepr-2018-59](https://doi.org/10.20948/prepr-2018-59).
22. Kabsch W., Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Bio polymers*. 1983. V. 22. № 12. P. 2577–2637. doi: [10.1002/bip.360221211](https://doi.org/10.1002/bip.360221211).
23. *Инструментарий ВуоPython*. URL: <https://biopython.org/> (дата обращения: 29.08.2018).
24. *Инструментарий для импорта Matlab-файлов в Python*. URL: <https://docs.scipy.org/doc/scipy/reference/tutorial/ipo.html> (дата обращения: 29.08.2018).
25. *Инструкция по возможностям регулярных выражений в MySQL*. URL: <http://www.mysql.ru/docs/man/Regexp.html> (дата обращения: 29.08.2018).
26. *Google Charts*. URL: <https://developers.google.com/chart/> (дата обращения: 29.08.2018).
27. *Jmol*. URL: <http://www.chemistry.wustl.edu/~edudev/LabTutorials/jmol.html> (дата обращения: 29.08.2018).
28. *Интерактивная консоль Jmol*. URL: <https://chemapps.stolaf.edu/jmol/docs/> (дата обращения: 29.08.2018).