

# Распознавание, отбор структурных мотивов, образованных двумя спиральями в белковых молекулах, и исследование межспиральных углов в спиральных парах

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

ИМПБ РАН - филиал ИПМ им. М.В. Келдыша РАН  
ИТЭБ РАН  
Институт белка РАН

Конференция: Математические методы распознавания  
образов ММРО-19

Москва, 26-29 ноября 2019

# Задача

- Возможно существуют стабильные элементы супервторичной структуры, аналогично тому как существуют элементы вторичной структуры повторяющиеся во всех белках.
- Очевидным претендентом является спиральная пара: элемент образованный двумя спиралями с неспиральным промежутком (перетяжкой) между ними.
- Мы собираемся выделить и проанализировать все спиральные пары белковых молекул в надежде найти повторяющиеся в плане пространственной структуры элементы.

4K9Z:2H:2D:9YAL: 037A:7D:9H:1A:5I: 1:K4D: 9:39



3B84:9H:0T:9YAL: 0K:0J:8: 0I: 0K2:9K:9: 1:K2L: 9:20



3HQ2:9D:2H:9YAL: 0B:0B:8: 0D: 0D6:0D4: 1:K2L: 9:20

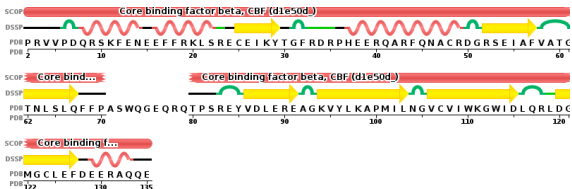


3QV:2M:37H:9YAL: 037A:7D: 03D:1A:2: 1:K4K: 9:39



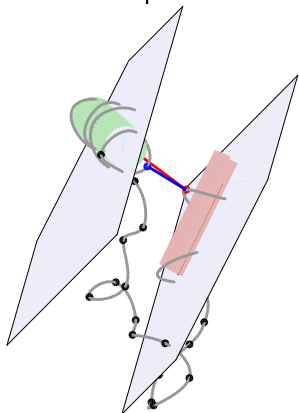
# Алгоритм выделения спиральных пар из PDB

1. Для каждой белковой цепи находим вторичную структуру по методу DSSP <https://swift.cmbi.umcn.nl/gv/dssp/>.
2. Выделяем начало и конец фрагмента спираль — перетяжка — спираль (спирали любого типа:  $\alpha$ ,  $3_{10}$ ,  $\pi$ )
3. По координатам атомов спиралей по МНК находим параметры идеальных цилиндров на которые спирали намотаны *Kabsch W. A Acta Cryst A 1976;32:9223.*

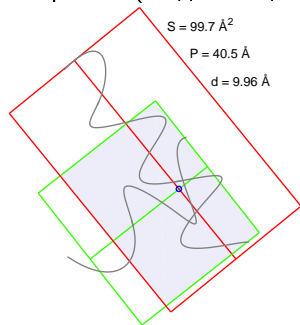


## Простое геометрическое описание спиральной пары

Пример спиральной пары: PDBid: 3A0B:1000-1037. Оси цилиндров лежат в параллельных плоскостях.



Пересечение проекций спиралей.  $S$  и  $P$  площадь и периметр полигона пересечения проекций спиралей (выделен цветом).

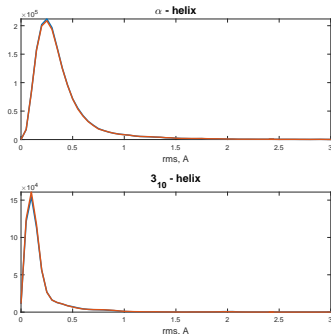


# Результаты работы алгоритма выделения спиральных пар по вторичной структуре с последующей аппроксимацией спиралей цилиндрами

Всего обработано 100397 записей PDB. Найдено 2209912 спиральных пар. Качество аппроксимации спиралей цилиндрами приведено на графиках.

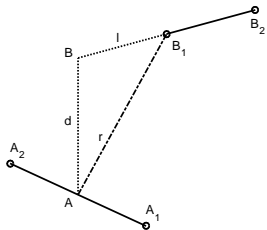
**Таблица.** Число найденных спиральных пар по типам их образующих спиралей.

	$\alpha$	$3_{10}$
$\alpha$	1210276	418644
$3_{10}$	403926	172877

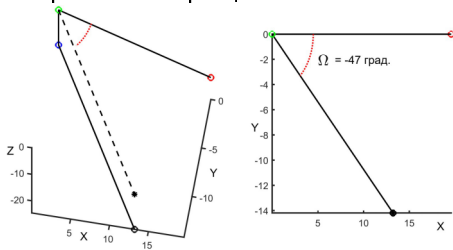


## 3D - структура спиральной пары $\mathbf{V} = [d, r, \alpha, \Omega, S]$

$A_1A_2B_1B_2$  - координаты осей цилиндров;  $d$  и  $r$  - межплоскостное и минимальное расстояние. Угол  $\alpha$  спиральной пары - угол между векторами  $\mathbf{A}_{12}$   $\mathbf{B}_{21}$ .



Торзионный угол  $\Omega$  - угол между проекциями осей цилиндров на ортогональную вектору  $A_2B_1$  плоскость. Знак определяет направление вращения от А к В.



# База данных структурных мотивов

<http://protdb.org/static/bd/clusters.htm>

protdb.org/static/bd/clusters.htm

ВСЕ   короткий   короткий   длинный   короткий   длинный   длинный   длинный   другие

#	IDBId	D	R	a	e	area	parim	pA	pB	IA	IP	IN
#1	IFXS:94:138:A	5.8	9.8	30.7	-22.8	198.4	61.3	0.36	0.31	1.54	2.53	1.45
#2	IG08:494:538:C	9.8	9.8	30.1	-22.5	199.3	61.5	0.35	0.27	1.54	2.55	1.45
#3	IG08:94:138:C	9.7	9.7	30.3	-22.5	199.8	61.6	0.34	0.31	1.54	2.55	1.45
#4	IIDA:94:138:A	9.7	9.7	29.6	-22.0	199.6	61.0	0.40	0.22	1.56	2.54	1.45
#5	IIDA:94:138:C	9.7	9.7	29.6	-22.0	199.6	61.7	0.37	0.25	1.55	2.51	1.45
#6	IQSP:94:138:A	9.7	9.7	29.8	-22.1	200.2	61.8	0.30	0.31	1.52	2.51	1.48
	mean value	9.7	9.7	30.0	-22.3	199.5	61.6	0.35	0.28	1.54	2.53	1.45


1 N:17(1)6(1)18(1) G:-22(1)30(1)9.7(0.13)

#	o	a	d	n	d1	d2	d3	d4		
#1	-23	30	5.7	179	-dpv	nkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	-5.18	5.71	5.77	5.89
#2	-23	30	5.6	131	-dpv	nkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	-5.45	5.66	6.01	6.04
#3	-22	30	5.6	18	vdpv	nkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	-5.76	5.94	7.45	7.52
#4	-22	30	5.7	14	-dpv	nkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	-5.28	5.80	5.81	5.89
#5	-23	30	5.8	13	--vn	nkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	-5.28	5.48	5.80	5.85
#6	-22	30	5.6	12	vdpv	nkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	-5.32	5.65	5.75	5.85
#7	-23	31	5.7	12	-dpv	nkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	-5.45	5.79	6.01	6.48
#8	-23	31	5.7	11	--pv	nkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	-5.24	5.93	5.96	5.98
#9	-23	31	5.8	11	--vn	nkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	-5.28	5.80	5.88	6.07

cluster #1394 Ncls = 14  
DPVNFKLLSHSLLVTLAHLSDPFPVAVHAALDKFLAVSVTLTS

ID	n	D	R	a	e	Sequence
#1	6	10	10	30	-22	dpvfkllshcllvtlaahLSDPFPavhaaldkflavsvtlts
#2	4	10	10	30	-22	dpvfkllshcllvtlaahLSDPFPavhaaldkflavsvtlts
#3	10	10	31	-23	dpvfkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	
#4	10	10	30	-22	dpvfkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	
#5	10	10	30	-22	dpvfkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	
#6	10	10	30	-22	dpvfkllshcllvtlaahLSDPFPavhaaldkflavsvtlts	

IFXS 94 138 A VAL: (9.78,9.78) (30.7,-22.8) (0.36, 0.31)



## Разбиение спиральных пар на три множества

Разбиение 2209912 спиральных пар по геометрии:  $[d, r, S]$

- $\{A\}$ :  $S = 0$
- $\{B\}$ :  $S > 0$ ,  $r > d$
- $\{C\}$ :  $S > 0$ ,  $r = d$

Splitting by geometry of helical pairs

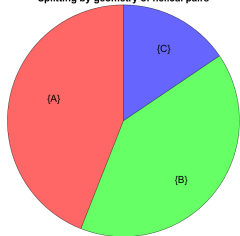
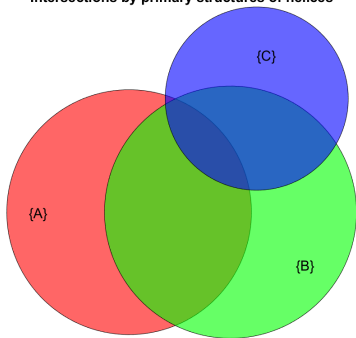


Диаграмма Венна для 421087 уникальных ПП спиралей входящих в спиральные пары

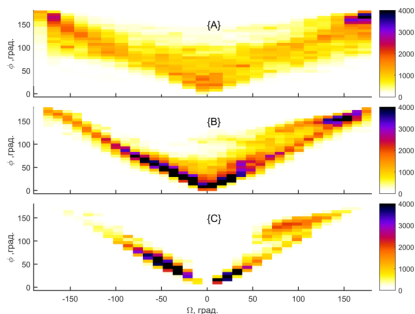
Intersections by primary structures of helices



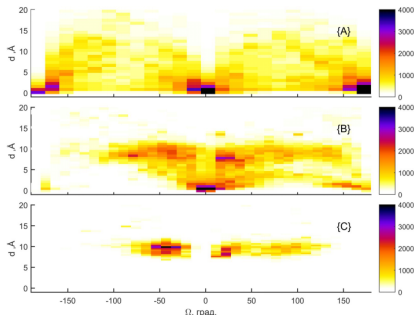


# Распределения геометрических величин $[d, \alpha, \Omega]$ в трех множествах

## Гистограмма $\alpha - \Omega$



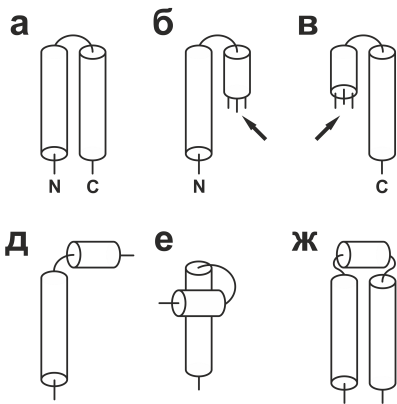
## Гистограмма $d - \Omega$



Анализ гистограмм показывает, что из множества  $\{C\}$  выделяется подмножество спиральных структур с квазидвумерной геометрией, где  $|\Omega| = \alpha$ .

## Двумерные схемы взаимодействия в спиральных структурах (в представлении биохимиков)

Схемы структур



$\alpha$ -спирали в белках упакованы так, что их гидрофобные поверхности погружены в гидрофобное ядро, а гидрофильные боковые цепи и полярные группы основной цепи взаимодействуют между собой или с молекулами воды. Три NH-группы на N-конце и три CO-группы на C-конце  $\alpha$ -спирали свободны и должны быть доступны молекулам воды или иметь других партнеров для образования водородных связей.

## Зависимость угла пересечения спиралей от их длины

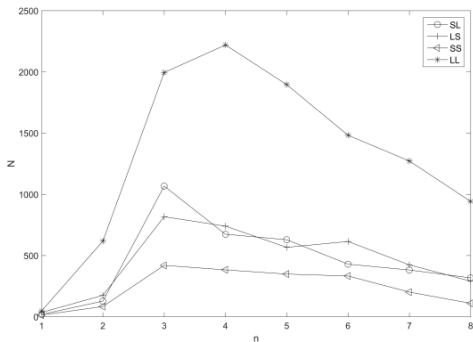
По материалам статьи : Тихонов Д.А., Куликова Л.И., Ефимов А.В. "Зависимость торсионных углов между осями  $\alpha$ -спиралей от их длины в структурных мотивах белковых молекул."  
**Молекулярная биология**, 2019 (в печати)

Во внутренних витках  $\alpha$ -спирали *NH*- и *CO*-группы основной цепи образуют насыщенную систему водородных связей, Их полное или частичное погружение в гидрофобное окружение без образования водородных связей энергетически крайне невыгодно. Из этого следует, что контакты концов спиралей с поверхностью спиралей энергетически невыгодны. Это значит углы между осями спиралей в случае (б,в) должны быть большими. Мы собираемся проверить данное утверждение пользуясь данными из нашей базы структурных мотивов.

## Отбор "двумерных" структур

Выбираем структуры где  $\max(\rho_A, \rho_B) < 0.5$  и  $|\alpha - |\Omega|| < 5$  и  $|d - r| < 0.1$  и  $N_P \leq 8$  четырех типов по правилам:

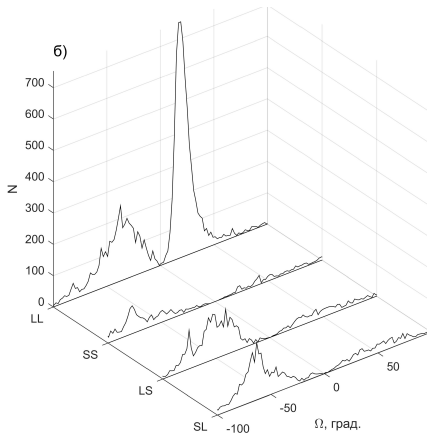
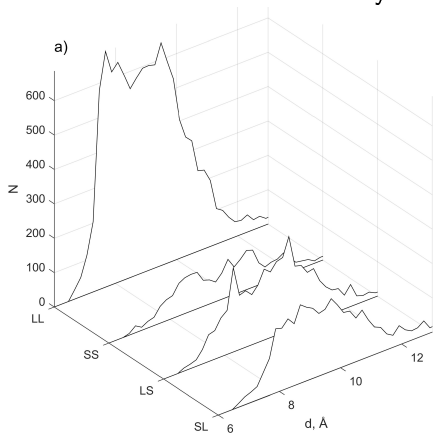
Число пар по типам



- SL (короткий-длинный) —  $N_B \geq 12$  и  $4 \leq N_A \leq 8$
- LS (длинный-короткий) —  $N_A \geq 12$  и  $4 \leq N_B \leq 8$
- SS (короткий-короткий) —  $4 \leq N_A \leq 8$  и  $4 \leq N_B \leq 8$
- LL (длинный-длинный) —  $N_A \geq 12$  и  $N_B \geq 12$

## Распределение $d$ и $\Omega$ в двумерных структурах по типам

Только LL структуры допускают параллельную ориентацию спиралей. Все остальные, как правило, ориентированы под углом.



## Выводы по работе

- Концепция структурных мотивов как симбиоза текстовых и геометрических свойств участка белковой цепи кажется весьма перспективным направлением дальнейшей работы.
- Сделана попытка выделить только одно слово в тексте описывающем белок на примере двухспирального мотива.
- Дальнейший прогресс должен состоять в существенной автоматизации процесса выделения мотивов.

Работа выполнена при финансовой поддержке РФФИ  
грант №18-07-01031