

ДЕСКРИПТИВНАЯ СТАТИСТИКА «ЗАПРЕЩЕННЫХ» ОБЛАСТЕЙ И РАЗЛИЧНЫХ ТИПОВ ВТОРИЧНОЙ СТРУКТУРЫ БЕЛКА В КОНТЕКСТЕ ИССЛЕДОВАНИЯ «ПЕРЕКРУЧЕННЫХ» β -ШПИЛЕК

© 2016 г. И.Ю. Торшин, Л.А. Урошлев*, Н.Г. Есипова*, В.Г. Туманян*

Химический факультет Московского государственного университета им. М.В. Ломоносова, 119991, Москва, Ленинские горы, 1/3;

**Институт молекулярной биологии им. В.А. Энгельгардта РАН, 119991, Москва, ул. Вавилова 32*

E-mail: vladimir.tumanyan@gmail.com

Поступила в редакцию 10.11.15 г.

Проведен детальный анализ конформаций остова полипептидной цепи белков на участках, примыкающих к областям β -изгибов, включая позоции «запрещенных» конформаций. Проведено перекрестное сравнение конформаций, соответствующих «запрещенным» областям карты Рамачандрана, и конформаций основных типов β -изгибов и примыкающих вторичных структур. Полученные результаты позволяют сделать вывод, что запрещенная область типа 2(II, II') на карте Рамачандрана соответствует преимущественно β -шпилькам и, более точно, именно «перекрученным» β -шпилькам. В составе последних остатки с углами ϕ_i, ψ_i в области 2(II, II') встречаются в 140 раз чаще, чем в «обычных» β -шпильках.

Ключевые слова: запрещенные конформации, β -шпилька, β -изгиб, конформационный анализ, статистический анализ.

Работа посвящена проблеме использования так называемых «запрещенных» конформаций в конформационном пространстве глобулярных белков.

Вопрос ставится так: каковы физические причины сохранения таких конформаций, которые, как представляется, уводят объект от структуры, соответствующей минимуму энергии (по крайней мере, в локальной области)?

В настоящее время благодаря развитию традиционных методов определения структуры и появлению новых, существование так называемых «запрещенных» конформаций полипептидной цепи в белках можно считать окончательно доказанным. Как правило, речь идет о появлении в областях карты конформаций Рамачандрана для дипептидной единицы допустимых по стереохимическим критериям только для глицина других типов аминокислотных остатков. Феномен «запрещенных» конформаций исследуется как теоретическими [1–3], так и экспериментальными [4,5] методами. Авторы работы [1] насчитывают шесть областей «запрещенных» конформаций.

В работе [6] на примере β -изгиба β -шпильки в SH3-доме α -спектрина показано, что «запрещенная» конформация не имеет альтернативы из-за так называемого топологического

замка. Дело в том, что фиксация аминокислотных остатков β -шпильки, непосредственно примыкающих к β -изгибу, соединенных водородной связью, приводит к формированию псевдоцикла, и уменьшение числа независимых переменных из-за замыкания цикла сильно ограничивает число возможных решений.

В работе [7] были установлены основные свойства «перекрученных» β -шпилек, схожих с дистальной шпилькой 42–55 в SH3-доме α -спектрина. «Перекрученные» β -шпильки характеризуются увеличенным значением «индекса хиральности» [8]; измененной сеткой водородных связей типа «остов–остов» по сравнению с обычными β -шпильками, такими как, например, остатки 17–32 бычьего панкреатического ингибитора трипсина; повышенной стабильностью (согласно данным денатурирующей молекулярной динамики), несмотря на стерические препятствия в β -изгибе.

В ходе проведенных исследований возникли вопросы о других отличиях «обычных» β -шпилек от «перекрученных» и о встречаемости остатков с «запрещенной» конформацией из «запрещенной» области II (также обозначается как II') карты Рамачандрана в различных контекстах вторичной структуры (в частности, в β -

Таблица 1. Статистика двугранных углов в «перекрученных» и обычных β -шпильках

Позиция	Угол	«Перекрученная» шпилька		«Нормальная» шпилька		<i>P</i> (тест Стьюдента)
		Среднее	Ст. откл.	Среднее	Ст. откл.	
+1	φ	-115,182	36,24931	-101,603	50,04559	$1,15 \cdot 10^{-16}$
	ψ	128,1733	52,24468	121,0908	61,36564	0,0004
	ω	0,681919	3,029602	0,574871	3,049357	Н.Д.
+2	φ	-120,722	36,87637	-106,748	37,28244	$1,02 \cdot 10^{-24}$
	ψ	127,9713	55,20566	127,2371	51,81036	Н.Д.
	ω	0,841387	2,989053	0,539521	3,057723	0,0032
+3	φ	-108,993	27,66664	-109,681	32,69486	Н.Д.
	ψ	121,0149	32,58941	128,4939	49,83235	$1,44 \cdot 10^{-6}$
	ω	0,534339	3,0559	0,69802	3,024323	0,0703
+4	φ	-114,479	40,83464	-103,642	34,26247	$9,57 \cdot 10^{-16}$
	ψ	59,13389	105,0258	127,285	50,03335	$9,24 \cdot 10^{-117}$
	ω	-0,74255	3,020817	0,924678	2,964402	$7,52 \cdot 10^{-51}$
+5	φ	2,727375	66,2179	-89,5261	31,76319	$< 10^{-220}$
	ψ	-20,9553	68,9428	120,8922	63,77129	$< 10^{-220}$
	ω	0,165955	3,098988	-0,76388	3,013367	$5,71 \cdot 10^{-17}$
+6	φ	-38,7967	90,03045	-58,5171	27,16606	$1,31 \cdot 10^{-18}$
	ψ	-0,88734	28,24469	-23,5312	49,59893	$1,57 \cdot 10^{-46}$
	ω	-0,38251	3,088481	-0,45612	3,069764	Н.Д.
+7	φ	-66,9524	96,10579	-58,4209	38,03471	0,0003
	ψ	96,64765	94,51683	-25,679	28,69098	$< 10^{-220}$
	ω	0,907627	2,967471	0,347801	3,094437	$2,59 \cdot 10^{-7}$
+8	φ	-110,137	30,0589	-75,7804	48,98203	$6,22 \cdot 10^{-98}$
	ψ	131,2041	38,28264	-12,688	42,07971	$< 10^{-220}$
	ω	0,70811	3,023489	-0,16054	3,10683	$7,49 \cdot 10^{-15}$
+9	φ	-118,175	35,5861	-54,5145	57,44962	$2,56 \cdot 10^{-207}$
	ψ	129,6163	48,58421	19,14602	78,37219	$7,07 \cdot 10^{-297}$
	ω	0,875874	2,976993	0,140344	3,107098	$3,16 \cdot 10^{-11}$

Примечание. Позиция +1 соответствует остатку № 41 в SH3-домене α -спектрина, позиция +2 – остатку № 42 и т.д. Н.Д. – отличие статистически недостоверно.

шпильках). В настоящей работе данные вопросы исследуются методами дескриптивной статистики на основе представительной выборки белков из PDB.

МАТЕРИАЛЫ И МЕТОДЫ

Проведен анализ базы данных PDB последних выпусков (2015 год). Из базы данных были удалены файлы, соответствующие идентичным последовательностям, и файлы структур, определенных с разрешением хуже, чем 2,0 Å. В результате была сформирована выборка из 60135 PDB-файлов, в целом включающая струк-

туры 101348 полипептидных цепей белков и 48750000 аминокислотных остатков с известными координатами всех не водородных атомов. Для каждой из полипептидных цепей были вычислены для каждого остатка значения двугранных углов φ , ψ , ω и проведено отнесение значений углов к шести «запрещенным» областям на карте Рамачандрана в соответствии с работой [1]. Вторичная структура исследованных белковых цепей была размечена с помощью программы STRIDE со стандартными значениями параметров [9].

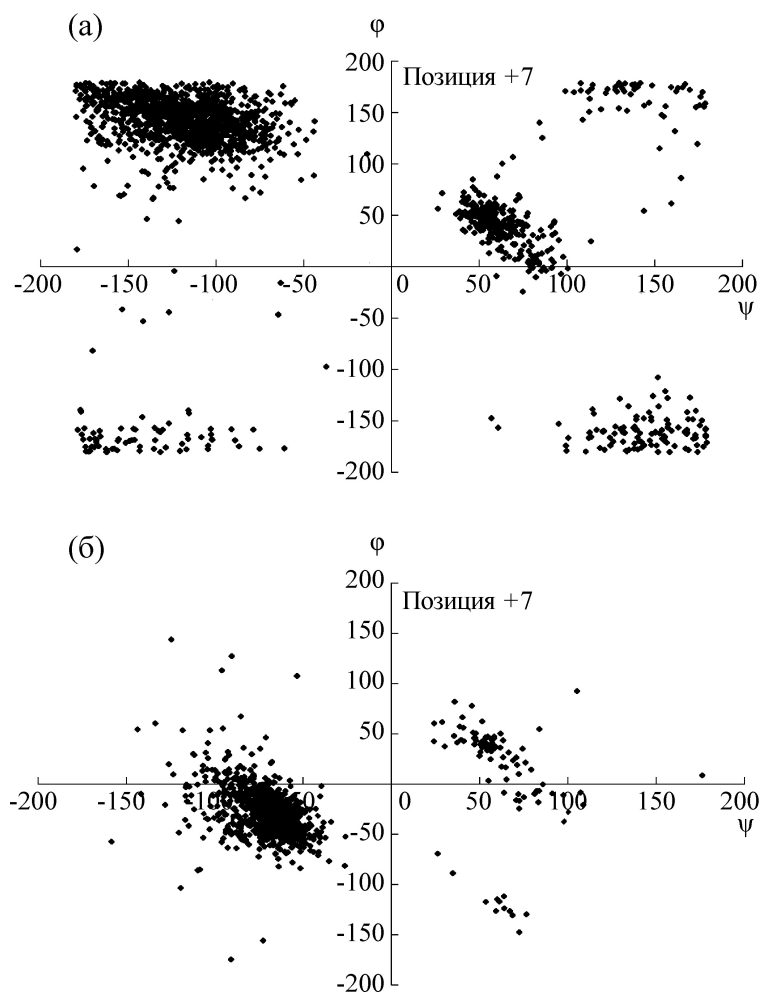


Рис. 1. Пример распределений значений углов ϕ , ψ в позиции +7 для «перекрученных» и обычных β -шпилек. (а) – «Перекрученные» β -шпильки (например, 42–52 в SH3-домене), (б) – «обычные» β -шпильки (например, 17–32 в ингибиторе трипсина).

РЕЗУЛЬТАТЫ

В работе представлены результаты сравнения конформаций «обычных» и «перекрученных» β -шпилек на основе двугранных углов главной цепи (т.е. ϕ , ψ , ω). Затем приведены результаты анализа распространенности различных «запрещенных» конформаций в контексте β -шпилек и других типов вторичных структур белка. В настоящей статье «запрещенные» области I–VI [1] обозначаются арабскими цифрами: I – 1, II (или II') – 2 и т. д. во избежание путаницы, связанной с тем, что римскими цифрами обозначаются и запрещенные области и β -повороты.

В работе [7] были использованы достаточно жесткие критерии для нахождения в PDB β -шпилек, структурно подобных «перекрученной» β -шпильке в SH3 домене α -спектрина: а) сред-

неквадратичное отклонение (RMSD) по всем C_{α} -атомам $< 0,3 \text{ \AA}$; (б) RMSD каждого атома C_{α} – не более $1,5 \text{ \AA}$. С использованием этих критериев в 25%-безизбыточной выборке PDB было найдено более 1800 β -шпилек, схожих с «перекрученной» шпилькой SH3-домена.

Анализ значений двугранных углов в «перекрученных» и в обычных β -шпильках позволил получить указание на статистически достоверные различия в средних значениях углов практически в каждой позиции «перекрученной» β -шпилек (табл. 1). Существование столь выраженных отличий указывает на существенные отличия в локальных конформациях главной цепи между «перекрученными» и «обычными» шпильками (см., например, отличия в позиции +7, рис. 1).

Установленные выше отличия в средних значениях двугранных углов ϕ , ψ , ω между «пе-

Таблица 2. Интервалы значений двугранных углов φ и ψ (град), соответствующие «запрещенным» областям на карте Рамачандрана, определенные по данным работы [1]

Запрещенные конформации	1 (I)	2 (II)	3 (III)	4 (IV)	5 (V)	6 (VI)
φ	-175 ÷ -40	35 ÷ 65	65 ÷ 100	70 ÷ 160	70 ÷ 160	45 ÷ -75
ψ	-145 ÷ -70	-145 ÷ -65	-75 ÷ -20	80 ÷ 160	-45 ÷ 15	-160 ÷ 160

Таблица 3. Частота встречаемости различных «запрещенных» конформаций в обычных и в «перекрученных» β -шпильках

Запрещенные конформации	1 (I)	2 (II)	3 (III)	4 (IV)	5 (V)	6 (VI)	Всего
ν_1	0,00658	0,01268	0,00561	0,00207	0,01804	0,00067	0,045
ν_2	0,00429	0,00210	0,00166	0,00228	0,00728	0,00105	0,018
ν_1/ν_2	1,532	6,025	3,365	0,908	2,479	0,637	2,44
$P(\chi^2)$	0,08	0,0002	0,01	0,21	0,04	0,12	0,04

Примечание. ν_1 – Частота встречаемости в «перекрученных» шпильках, ν_2 – частота встречаемости в обычных шпильках.

Таблица 4. Статистически достоверные отличия в частотах встречаемости различных «запрещенных» конформаций по позициям «перекрученных» и обычных β -шпилек по позициям β -шпильки

Позиция	Запрещенные конформации	ν_1	ν_2	ν_1/ν_2	$P(\chi^2)$
+1	4	0,000556	0,006324	0,0878	0,008
	5	0,001111	0,004743	0,2342	0,031
+2	4	0,000556	0,005534	0,1003	0,012
	5	0,010556	0,003162	3,3381	0,042
+6	2	0,111667	0,000791	141,25	$7 \cdot 10^{-6}$
	3	0,035006	0,001581	22,130	0,001
	5	0,130556	0,002372	55,050	$4 \cdot 10^{-4}$
+7	3	0,000556	0,002372	0,2342	0,047
	4	0,008889	0,000791	11,244	0,009
+9	1	0,001111	0,005534	0,2007	0,040
	5	0,001111	0,015020	0,0739	0,016

Примечание. Позиция «+1» соответствует остатку 41 в SH3-домене α -спектрина и т.д.; ν_1 – частота встречаемости в «перекрученных» шпильках; ν_2 – частота встречаемости в обычных β -шпильках. Затененными полосами выделены наиболее характерные отличия.

рекрученными» и обычными β -шпильками позволяют предположить, что и встречаемость различных типов «запрещенных» конформаций в них также может различаться. В соответствии с определениями «запрещенных» областей, предложенными в работе [1] (табл. 2), были рассчитаны частоты встречаемости различных «запрещенных» конформаций в обычных и в «перекрученных» β -шпильках.

В целом в «перекрученных» β -шпильках запрещенные конформации встречались достоверно в 2,44 раза чаще. В частности, в «перекрученных» β -шпильках была установлена досто-

верно более высокая частота встречаемости запрещенных конформаций особенно для конформации типа 2. Тип 2 встречается в шесть раз чаще, тип 3 – в 3,4 раза чаще и тип 5 – в 2,5 раза чаще (см. табл. 3).

Анализ частоты встречаемости «запрещенных» конформаций на различных позициях в обычных и «перекрученных» β -шпильках указал на достоверно более высокую частоту встречаемости «запрещенных» конформаций из областей 1 и 2 в позиции +5 β -шпилек, «запрещенных» конформаций 3 и 5 – в позиции +6 и «запрещенной» конформации 4 – в позиции +7 (табл. 4).

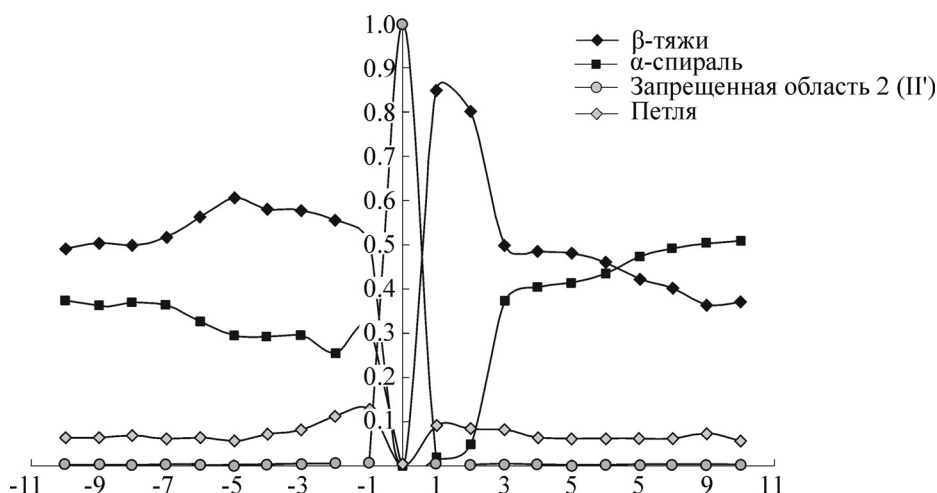


Рис. 2. Контекст вторичной структуры остатков из запрещенной области 2 (II').

Интересно отметить, что «запрещенная» конформация типа 2 остатка Asn47 в SH3-доме не является типичной именно для позиции +7 «перекрученных» β -шпилек – в настоящем исследовании не было установлено соответствующих статистических достоверных различий. В то же время «запрещенная» конформация 2(II, II') действительно характерна для «перекрученных» β -шпилек по сравнению с обычными, но не для позиции +7, а для позиции +5. Действительно, в позиции +5 «запрещенная» конформация из области 2 встречается в 141 раз чаще, чем в «обычных» β -шпильках (табл. 4, $P = 7 \cdot 10^{-6}$).

Таким образом, можно предположить, что существование «запрещенной» конформации типа 2 для остатка Asn47 в SH3-доме (т.е. в позиции, обозначаемой как +7) является индивидуальной особенностью «перекрученной» шпильки SH3-домена α -спектрина. Для «перекрученных» шпилек на 2 порядка чаще встречается «запрещенная» конформация типа 2 именно в позиции +5.

Возникает вопрос, насколько область запрещенных конформаций 2 (II, II') характерна для β -шпилек вообще, будь-то «перекрученные» или обычные β -шпильки. Контекст вторичной структуры каждого из остатков из запрещенных областей (в том числе области 2) также был проанализирован посредством дескриптивной статистики. Были подсчитаны частоты встречаемости различных типов вторичной структуры в позициях слева (–10, –9, ...) и справа (...+9, +10) относительно каждого остатка из запрещенной области 2 (позиция 0, рис. 2).

Проведенный анализ показал: запрещенные конформации из области 2 были ассоциированы преимущественно с β -структурой: в позициях –7..+3 β -структуры они встречались достоверно более часто, чем любой другой тип вторичной структуры ($P < 0,01$, тест χ^2). Отметим, что пик в позициях +1, +2 (рис. 2) соответствует β -изгибу, т.е. более 80% запрещенных конформаций из «запрещенной» области 2 на карте Рамачандрана реализуются в β -изгибах, преимущественно встречающихся в контексте β -структур (т.е. опять же в β -шпильках). Тем не менее 25–30% запрещенных конформаций из области 2 также встречаются и в контексте α -спиралей (см., например, позиции –2, –1, рис. 2).

В отличие от запрещенной области 2(II'), не было установлено выраженных предпочтений вторичной структуры для других запрещенных областей (например, для области 1, см. рис. 3).

Таким образом, полученные результаты позволяют сделать вывод, что «запрещенная» область типа 2 (II, II') на карте Рамачандрана соответствует преимущественно β -шпилькам и, более того, именно «перекрученным» β -шпилькам (в составе которых остатки с углами ϕ_i , ψ_i в области 2 (II, II') встречаются в 140 раз чаще, чем в обычных β -шпильках).

Помимо основных типов вторичной структуры (спирали, тяжи (стрэнды), петли), был проведен анализ совместной встречаемости запрещенных конформаций с носящими более локальный характер типами вторичной структуры, а именно β -изгибами (определения β -изги-

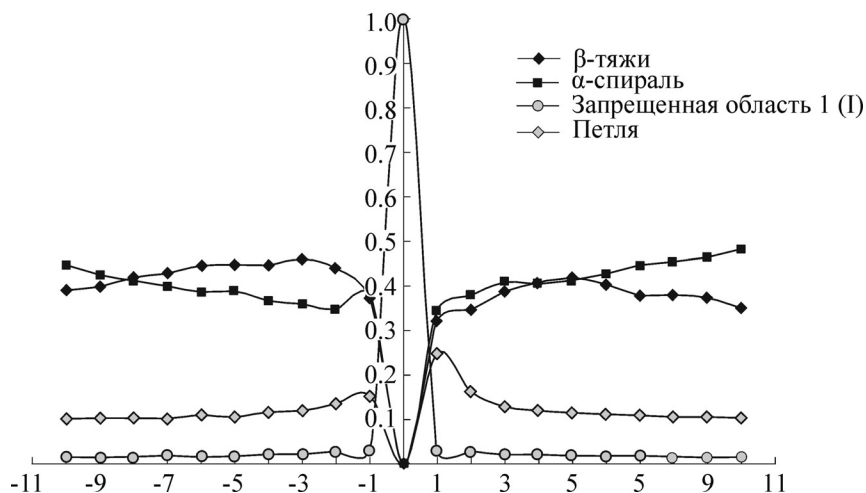


Рис. 3. Контекст вторичной структуры остатков из «запрещенной» области 1.

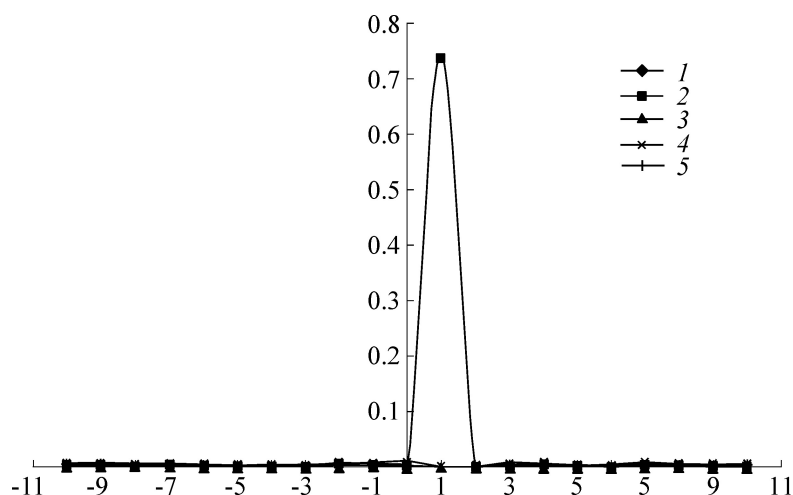


Рис. 4. β -изгибы типа II' в контексте остатков с «запрещенными» областями на карте Рамачандрана (определения из работы [1]). Арабскими цифрами обозначены типы запрещенных областей.

бов из работы [10]). Было установлено: с остатками с углами ϕ_i , ψ_i в «запрещенной» области 2 преимущественно ассоциированы β -изгибы типа II' (рис. 4), в то время как для других типов β -изгибов столь выраженных ассоциаций с запрещенной областью 2 не наблюдалось (см., например, данные для изгиба типа I, рис. 5).

ЗАКЛЮЧЕНИЕ

Анализ экспериментальных данных по структурам белков позволил выявить взаимозависимость между тремя видами объектов: «запрещенными» областями на карте Рамачандрана, основными типами вторичной структуры и наиболее распространенными типами β -изги-

бов. Найдены статистически достоверные различия в значениях двугранных углов в «перекрученных» и обычных β -шпильках. «Запрещенные» конформации остатков с высокой степенью достоверности чаще, чем в обычных β -шпильках, в «перекрученных» β -шпильках для большинства типов запрещенных областей, особенно для области 2 (II, II'). «Запрещенные» конформации в области 2, соответствующие β -изгибу, существуют преимущественно в контексте β -структуры, а также, хотя значительно реже, в контексте α -спирали. Для других «запрещенных» областей превалирования того или иного типа вторичной структуры не наблюдается. В итоге с «запрещенной» областью 2 (II, II') ассоциируется β -структура (причем в

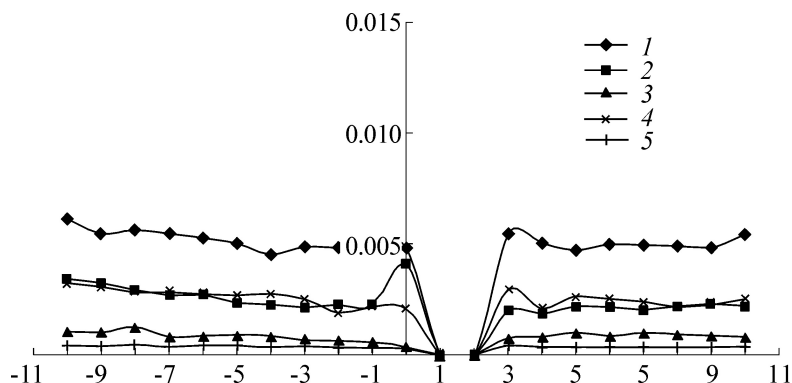


Рис. 5. β -изгибы типа I в контексте остатков с запрещенными областями на карте Рамачандрана (определения из работы [1]). Арабскими цифрами обозначены типы запрещенных областей.

форме «перекрученной» β -шпильки) и β -изгиб II'-типа.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 14-04-00639) и Программы Президиума РАН «Молекулярная и клеточная биология».

СПИСОК ЛИТЕРАТУРЫ

1. K. Gunasekaran, C. Ramakrishnan, and P. Balaram, *J. Mol. Biol.* **264**, 191 (1996).
2. D. Pal and P. Chakrabati, *Biopolymers* **63**, 195 (2002).
3. C. Ramakrishnan, B. Lakshmi, A. Kurien, et al., *Protein & Peptide Letters* **14**, 672 (2007).
4. M. C. Vega, J. C. Martinez, and L. Serrano, *Protein Sci.* **9**, 2322 (2000).
5. A. R. Viguera, M. A. Jimenez, M. Rico, and L. Serano, *J. Mol. Biol.* **255**, 507 (1996).
6. Л. А. Урошлев, И. Ю. Торшин, А. В. Батяновский и др., *Биофизика* **60** (1), 5 (2015).
7. I. Yu. Torshin, N. G. Esipova, and V. G. Tumanyan, *J. Biomol. Struct. Dynam.* **32** (2), 198 (2014).
8. A. Pietropaolo, L. Muccioli, R. Berardi, and C. Zannoni, *Proteins* **70**, 667 (2008).
9. M. Heinig and D. Frishman, *Nucl. Acids Res.* **32**, W500-2 (2004).
10. E. G. Hutchinson and J. M. Thornton, *Protein Sci.* **3**, 2207 (1994).

Descriptive Statistics of Disallowed Regions and Various Types of Protein Secondary Structure in the Context of Twisted β -Hairpin Study

I.Yu. Torshin*, L.A. Uroshlev**, N.G. Esipova**, and V.G. Tumanyan**

*Department of Chemistry, Lomonosov Moscow State University, Leninskie Gory 1/3, Moscow, 119991 Russia

**Engelhard Institute of Molecular Biology, Russian Academy of Sciences, ul. Vavilova 32, Moscow, 119991 Russia

A detailed analysis of polypeptide chain backbone conformations was carried out for segments of protein polypeptide chains joined to the regions of β -bends including sites of disallowed conformations. A cross comparison of conformations was performed for disallowed regions of the Ramachandran map and main types of β -bends and adjacent secondary structures. The results obtained permit us to draw a conclusion that disallowed region 2(II, II') in the Ramachandran plot coincides mainly with β -hairpins and moreover to the twisted β -hairpins. The latter hairpins include in its composition the residues with φ_i, ψ_i from 2(II, II') region, which is 140 times more frequent than common β -hairpins.

Key words: disallowed conformations, β -hairpin, β -turn, conformation analysis, statistical analysis