

УДК 577.3

## КОНФОРМАЦИОННЫЕ АСПЕКТЫ ФОРМИРОВАНИЯ СТРУКТУР ОСТОВА ПОЛИПЕПТИДНЫХ ЦЕПЕЙ В БЕЛКАХ. СВЯЗЬ КОНФОРМАЦИОННОЙ СТАБИЛЬНОСТИ/ЛАБИЛЬНОСТИ С ТИПОМ $\beta$ -ИЗГИБОВ

© 2023 г. И.Ю. Торшин<sup>\*,#</sup>, И.В. Филатов<sup>\*\*</sup>, А.В. Батяновский<sup>\*\*\*</sup>, К.В. Смирнов<sup>\*\*\*\*</sup>,  
А.А. Анашкина<sup>\*\*\*\*</sup>, Н.Г. Есипова<sup>\*\*\*\*</sup>, В.Г. Туманян<sup>\*\*\*\*</sup>

<sup>\*</sup>Вычислительный центр им. А.А. Дородницына Федерального исследовательского центра  
«Информатика и управление» РАН, ул. Вавилова, 40, Москва, 119333, Россия

<sup>\*\*</sup>Московский физико-технический институт, Институтский пер.,  
9, Долгопрудный Московской области, 141700, Россия

<sup>\*\*\*</sup>Институт биофизики клетки и клеточной инженерии НАН Беларуси,  
ул. Академическая, 27, Минск, 220072, Беларусь

<sup>\*\*\*\*</sup>Институт молекулярной биологии им. В.А. Энгельгардта РАН, ул. Вавилова, 32, Москва, 119991, Россия

<sup>#</sup>E-mail: [tty135@yahoo.com](mailto:tty135@yahoo.com)

Поступила в редакцию 14.11.2022 г.

После доработки 11.12.2022 г.

Принята к публикации 14.12.2022 г.

Для оценки характера связи между интегральной конформационной стабильностью тетрапептидов и главными типами  $\beta$ -изгибов (также являющихся тетрапептидами) рассчитаны спектр-диаграммы, асимметрия распределения конформационно-стабильных и конформационно-нестабильных тетрапептидов. Показано, что  $\beta$ -изгибы типов I', II и II' состоят из преимущественно конформационно-лабильных пептидов, что согласуется с контекстно-предопределенным характером их структуры. Поскольку, как мы показали ранее, в этом случае конформация навязывается внешними условиями (конкретно, замыканием цикла), превалирование конформационно-лабильных пептидов облегчает формирование структуры за счет внешних факторов. Тип I  $\beta$ -изгиба представляет собой исключение: в нем пептиды с разной конформационной лабильностью распределены достаточно равномерно. Можно предположить, что формирование  $\beta$ -изгиба I-го типа не является контекстно-обусловленным.

*Ключевые слова:* конформационно стабильные/лабильные сегменты белка, локальная структура белка, конформационный анализ,  $\beta$ -изгибы, статистический анализ, молекулярная динамика.

DOI: 10.31857/S0006302923010XXX, EDN:

Общие представления о  $\beta$ -изгибах сводятся к тому, что это — типовой структурный элемент белка, представляющий интерес как сам по себе, так и в качестве составной части определенного уровня иерархии структуры белка. Эта иерархия, вообще говоря, не может не отражать процесса складывания белка в глобулу (так называемый фолдинг) и комплексы глобул. Значение проблемы  $\beta$ -изгибов подчеркивается тем обстоятельством, что в некоторых белках их доля достигает 80%, т.е. они встречаются чаще  $\alpha$ -спиралей и  $\beta$ -тяжей. По оценке, сделанной в работе [1], каждый четвертый остаток входит в состав  $\beta$ -изгиба. Зачастую  $\beta$ -изгибы являются компонентом более сложного структурного элемента, прежде всего так называемых  $\beta$ -шпилек [2, 3].

Для  $\beta$ -изгиба характерна четко очерченная локальная структура (иначе говоря, определенный тип вторичной структуры), причем определенного размера. В отличие от  $\alpha$ -спирали,  $\beta$ -структуры или левой спирали РРП (при исследовании которых встает вопрос об их длине и о критериях, используемых для определения длины), длина  $\beta$ -изгиба — строго четыре аминокислотных остатка. Важно и то, что  $\beta$ -изгиб — небольшой структурный мотив, конформация которого определяется малым числом степеней свободы, что делает возможным применение переборных алгоритмов. Среди структурных функций  $\beta$ -изгиба самая очевидная связана с изменением хода полипептидной цепи на обратный, что отражено в альтернативном названии  $\beta$ -изгиба — «обратный поворот»

(reverse turn). Такого рода изменение хода полипептидной цепи способствует ее компактизации и формированию белковой глобулы.

Таким образом, доля  $\beta$ -изгибов в белках достаточно велика, их структурное и функциональное значение вполне очевидно. Между тем в имеющихся представлениях о  $\beta$ -изгибах до сих пор присутствуют существенные пробелы и противоречия. В частности, аннотация этих типов локальной структуры и способы их прогнозирования, исходя из аминокислотной последовательности, нуждаются в дальнейшем совершенствовании. Действительно, хотя в прогнозировании  $\beta$ -изгибов наметился заметный прогресс, он связан не с углублением наших представлений об этом типе структуры, а с применением более эффективных методов интеллектуального анализа данных и так называемого машинного обучения (метод поддерживающих векторов, сверточные нейронные сети и др. [4, 5]). Дальнейший рост качества прогнозирования ограничен не только трудностями в исходной аннотации  $\beta$ -изгибов, но и тем, что фактически приходится решать две различные задачи: нахождение места  $\beta$ -изгиба в полипептидной цепи и определение типа изгиба. Следует добавить, что чисто биоинформационный подход, при котором исследуется схожесть аминокислотных последовательностей, имеет ограниченную область применения. Дальнейшее углубление понимания проблемы  $\beta$ -изгибов следует искать на пути анализа физико-химических свойств этих локальных структур.

Как для всего белка, так и для  $\beta$ -изгибов одной из важнейших проблем является поиск взаимосвязей между структурой изгибов, аминокислотными последовательностями собственно  $\beta$ -изгибов («текст») и последовательностями фланкирующих областей, играющих роль «контекста». В настоящей работе поставлена и получила положительное решение оригинальная задача о взаимосвязи основных структурных черт конкретного типа  $\beta$ -изгиба и конформационной подвижности/лабильности составляющих его сегментов. Такая взаимосвязь может дать возможность прогнозирования  $\beta$ -изгибов посредством оценки конформационной стабильности/лабильности его сегментов.

### О КОНТЕКСТЕ $\beta$ -ИЗГИБОВ В СТРУКТУРАХ БЕЛКОВ

Проблемы взаимосвязи структуры и последовательности аминокислот, стабилизации структуры, роли «текста» и «контекста» в случае  $\beta$ -изгибов сводятся к решению конкретных задач оптимизации по геометрическим и энергетическим критериям. Подчеркнем, что формирование всех типов  $\beta$ -изгибов (а не только изгибов II' типа) сопряжено с преодолением стереохимических пре-

пятствий. Наличие таких стереохимических трудностей становится очевидным, в частности, из рассмотрения карт Рамачандрана для  $\beta$ -изгибов: точки на картах для остатков  $i+1$  и/или  $i+2$  в изгибах сконцентрированы в запрещенных областях карты Рамачандрана (см., например, работу [6]). Более или менее «свободной» конформационной картой характеризуется только  $\beta$ -изгиб типа I. Возникает вопрос: каким образом  $\beta$ -изгиб приобретает конформацию, которая представляется энергетически «напряженной»? Второй вопрос: коль скоро энергетически «напряженная» структура уже образовалась, как компенсируется «неблагоприятная» энергия, свойственная этой структуре?

Ответ на первый вопрос был исследован нашим коллективом путем привлечения геометрических представлений об уменьшении числа независимых конформационных параметров в цикле, образованном олигопептидом [7]. Вследствие замыкания цикла (через водородную связь между первым и последним остатками  $\beta$ -изгиба) число независимых конформационных параметров уменьшается настолько, что определение конформации может быть сведено к чисто геометрическому расчету. Такой геометрический анализ был проведен в работе [8] для основных типов  $\beta$ -изгибов с использованием независимых методов – оригинальной процедуры геометрии расстояний (distance geometry) и процедуры перебора конформаций. Было продемонстрировано, что число геометрически возможных конформаций в этом случае сводится к двум: первое из найденных решений совпадает с данными рентгеноструктурного анализа, а второе хотя и геометрически корректно, но приходится на абсолютно запрещенные области карты Рамачандрана (т.е. соответствует крайне выраженным стерическим затруднениям). Таким образом, вследствие замыкания водородной связи образуется своеобразный «топологический замок» и этот фактор контекста приводит к формированию конформации  $\beta$ -изгиба.

Что касается ответа на второй вопрос, то очевидно, что даже при фиксированной геометрии  $\beta$ -изгиба энергия соответствующей локальной структуры не может быть слишком «неблагоприятной», иначе произойдет кардинальная перестройка и реализуется существенно иная структура. Наличие конформационно нестабильных сегментов позволило бы неблагоприятной энергии «рассосаться» по многим степеням свободы, и, тем самым, лишнее напряжение было бы снято. Включение конформационно лабильных сегментов увеличивает число степеней свободы и эффекта цикла не наблюдается. Итак, сегменты с вынужденной (навязанной) конформацией, взятые сами по себе, *a priori* должны были бы быть конформационно лабильными.

## КОНФОРМАЦИОННАЯ СТАБИЛЬНОСТЬ β-ИЗГИБОВ И КОНФИГУРАЦИИ ТЕТРАПЕПТИДОВ

В связи с изложенным мы пришли к необходимости оценить конформационную стабильность тетрапептидов, образующих β-изгибы, с помощью разработанных нами ранее оценок конформационной лабильности/стабильности тетрапептидов (см. работу [9]). Напомним, что интерес к β-изгибам как к особым фрагментам вторичной структуры связан с тем, что, с одной стороны, конформация β-изгиба может быть навязана контекстом в структуре белка. С другой стороны, имеющиеся алгоритмы прогнозирования β-изгибов позволяют успешно проводить такое прогнозирование (с аккуратностью распознавания 75–82%) при рассмотрении достаточно небольшого контекста в последовательностях вокруг β-изгибов (5–7 остатков). Поэтому представляет интерес изучить, как именно соотносится наличие структуры β-изгибов с конформационной стабильностью/лабильностью составляющих их тетрапептидов («текстов») и фланкирующих областей («контекстов») и есть ли вообще такое соотношение.

Как показано в нашей предыдущей работе по мерам конформационной стабильности/лабильности  $s_1$ - $s_5$  [9], в файлах из банка данных PDB найдено 122100 различных тетрапептидов, которые встречались не менее пяти раз в различных структурах. При более редкой встречаемости пептидов (менее пяти случаев по всему банку данных PDB) использование разработанных мер конформационной лабильности  $s_1$ - $s_5$ , основанных на сборе статистики конформаций, не вполне обосновано статистически.

Нахождение в PDB всего 122100 из ~160000 комбинаторно возможных тетрапептидов указывает, скорее всего, на особенности аминокислотных последовательностей кристаллизуемых белков, которые составляют 90% от объема данных в PDB (остальные 10% структур определены с применением методов ЯМР или электронной микроскопии). В подтверждение этому можно отметить, что частоты встречаемости одних и тех же три- и тетрапептидов существенно разнятся в базах данных PDB и UNIPROT [10].

Для каждого из тетрапептидов, найденных в файлах PDB, были вычислены меры конформационной лабильности  $s_1$ - $s_5$  [9]. Затем, для каждого  $i$ -го тетрапептида была вычислена интегральная оценка конформационной стабильности  $b_i$  по формуле  $b_i = 5 \cdot \sum_{k=1,5} \omega_k s_k(i)$ , где веса  $\omega_k$  для каждой из мер конформационной лабильности  $s_k$  определялись так, чтобы значение  $b_i$  произвольного тетрапептида лежало в диапазоне значений от 1 до 5

(1 – минимальная конформационная стабильность, 5 – максимальная конформационная стабильность).

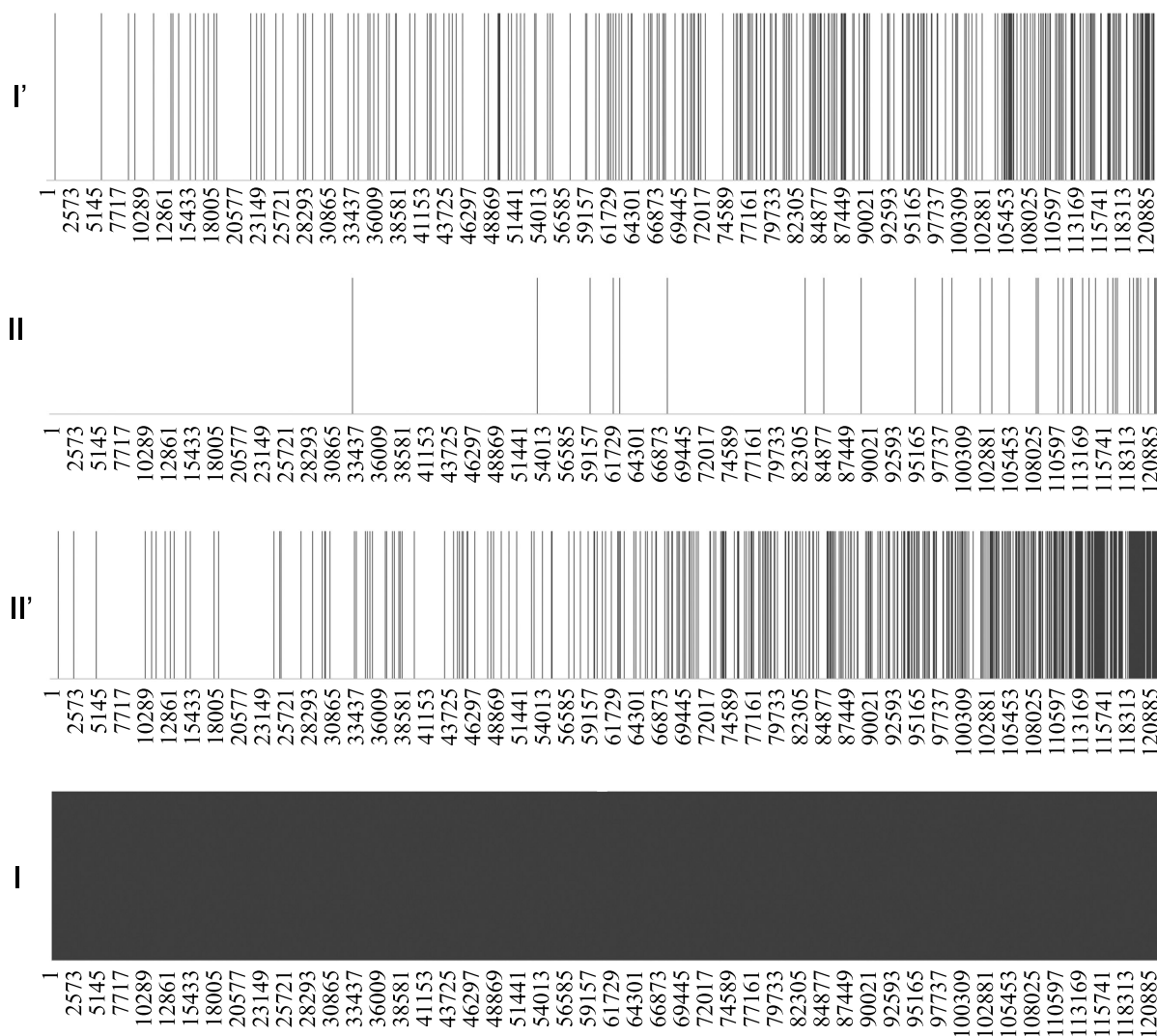
В общем списке из 122100 тетрапептидов были выделены тетрапептиды, для второго и третьего остатков которых преимущественной конформацией (т.е. конформацией, встречавшейся чаще любой другой) являлся β-изгиб одного из четырех типов (I, I', II, II'). Определения β-изгибов соответствовали критериям из работы [6]. Далее представлены результаты анализа тетрапептидов, преимущественной конформацией которых (более 50% конформаций из найденных в PDB для данного пептида) являлся один из четырех типов изгибов (I, I', II, II'). Таких тетрапептидов было найдено всего 64769.

Среди этих тетрапептидов с преимущественной конформацией в виде одного из типов β-изгибов наиболее часто встречались пептиды с β-изгибом типа I, соответствующие (в большинстве случаев) α-спиралям (в широком смысле, включая спирали 3/10) или их «кэпирующим» фрагментам (63770 тетрапептидов). Тетрапептиды, наиболее частой конформацией которых является β-изгиб других трех типов, встречались гораздо реже: тип I' – 331 тетрапептид, тип II – 41 тетрапептид, тип II' – 627 тетрапептидов. Обращает на себя внимание, что изгибу II', самому напряженному со стереохимической точки зрения, соответствует большее количество различных аминокислотных последовательностей тетрапептидов.

Следует подчеркнуть, что приведенные выше данные являются новыми, уникальными и не имеющими мировых аналогов, так как относятся не к частоте встречаемости конформаций (которая обычно представлена в имеющейся литературе), а к тетрапептидам (сегментам длиной четыре аминокислоты с определенной последовательностью), конформации которых преимущественно ассоциированы с указанными выше типами изгибов, т.е. эти данные имеют отношение собственно к «тексту», а не к «контексту» β-изгиба.

Для анализа взаимоотношений между конформационной стабильностью тетрапептидов и β-изгибами четырех типов (I, I', II, II') использовали: 1) – спектр-диаграммы значений  $b_i$ ; 2) – эмпирические функции распределения значений интегральной оценки конформационной стабильности  $b_i$ ; 3) – анализ распределения тетрапептидов, преимущественно соответствующих β-изгибам между верхней и нижней частью общего списка из 122100 тетрапептидов, упорядоченного по убыванию значений  $b_i$ . Все три независимых метода анализа показали схожие результаты.

Спектр-диаграммы позволяют наглядно оценить различия в частоте встречаемости пептидов заданного типа (тип β-изгиба) при различных



**Рис. 1.** Спектр-диаграммы значений  $b_i$   $\beta$ -изгибов. По оси абсцисс отложен ранг стабильности тетрапептида, вычисляемый как порядковый номер значения  $b_i$  в списке тетрапептидов, упорядоченных по убыванию значений  $b_i$ . Очевидна большая частота встречаемости (более высокая плотность линий) конформационно лабильных пептидов для трех типов изгибов (I, II, II'). Спектр-диаграмма для изгиба типа I, очевидно, представляет собой равномерное распределение линий по всем значениям ранков.

значениях  $b_i$ . По оси абсцисс спектр-диаграммы отложен ранг стабильности тетрапептида, вычисляемый как порядковый номер значения  $b_i$  в списке тетрапептидов, упорядоченных по убыванию значений  $b_i$ . Визуальный анализ спектр-диаграмм значений  $b_i$  показал, что конформационно лабильные пептиды встречаются гораздо чаще для трех типов изгибов – I, II, и особенно II' (рис. 1).

Рассмотрим статистические оценки распределения пептидов с преимущественной конформацией в виде  $\beta$ -изгибов по значениям  $b_i$ . Анализ конформационно-лабильных пептидов, которые встречались гораздо чаще для трех типов изгибов (I, II и II'), показал, что данные пептиды отличались

выраженным избытком глицина и сниженной встречаемостью аланина, глутамата, фенилаланина, изолейцина, лизина, лейцина, аргинина, валина (таблица). По сравнению с встречаемостью аминокислот в среднем по банку данных PDB, достоверными по критерию  $\chi^2$  были только различия для глицина, аланина, изолейцина, лейцина и валина. Конформационно стабильные пептиды в этой подвыборке не показали достоверных различий в аминокислотном составе.

Анализ эмпирических функций распределения значений  $b_i$  (рис. 2) показал очевидные смещения кривых эмпирических функций распределения  $\beta$ -изгибов типов I, II и II' в сторону более

**Таблица 1.** Частоты встречаемости (в %) различных аминокислот в конформационно-лабильных мотивах

Выборка	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Конформационно-лабильные для типов I', II и II'	3.1	3.4	6.9	3.8	3.6	24.5	4.6	2.5	3.9	2.5	3.3	6.2	4.9	4.1	3.8	5.3	4.2	2.5	3.1	4.0
Среднее по выборке из PDB	5.9	2.8	5.6	5.7	5.0	5.9	4.1	5.5	5.4	6.0	3.3	5.1	5.2	4.9	5.4	5.6	5.5	5.8	2.8	4.8

низких значений  $b_i$ , т.е. в сторону более конформационно-лабильных пептидов. Например, среднее значение  $b_i$  для изгиба типа I составило  $2.4 \pm 0.2$  (среднее  $\pm$  стандартное отклонение), а для изгиба типа I', II' и II –  $2.22 \pm 0.2$ ,  $2.1 \pm 0.15$  и  $2.05 \pm 0.12$  соответственно.

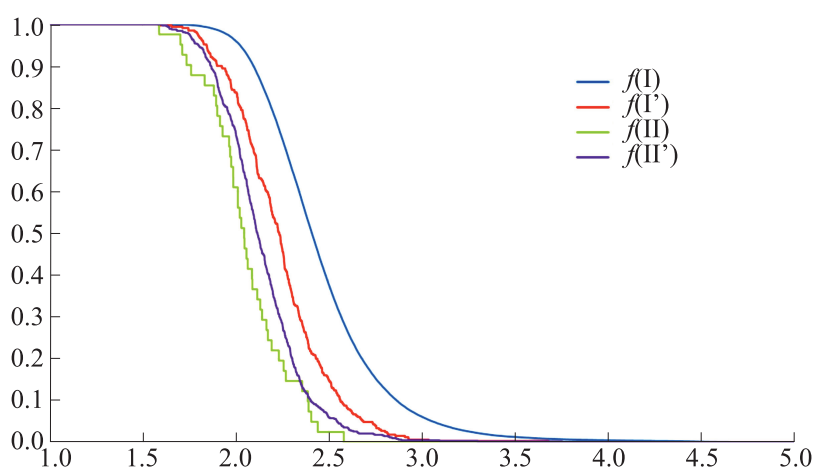
Анализ распределения соответствующих  $\beta$ -изгибам тетрапептидов между верхней и нижней частью общего списка из 122100 тетрапептидов, упорядоченного по убыванию значений  $b_i$ , подтвердил описанные выше результаты. Так, для изгибов типа I частота встречаемости в верхней и в нижней половинах списка была практически одинаковой (32955 пептидов в верхней части, 30785 – в нижней, т.е. частоты равны 0.548 и 0.512 соответственно). В случае трех других типов  $\beta$ -изгибов были отмечены статистически достоверные различия в частотах встречаемости с преобладанием конформационно лабильных пептидов: для типа I' – 74 и 256 тетрапептидов, т.е. частоты отличались более чем в три раза (0.0012 и 0.0043); для типа II – 3 и 37, частоты отличались на порядок ( $5 \cdot 10^{-5}$  и  $6 \cdot 10^{-4}$ ); для типа II' – 65 и 555, частоты отличались почти на порядок (0.0011 и 0.0092). Эти различия для  $\beta$ -изгибов типов I', II и II' были

достоверны по критерию  $\chi^2$  ( $P < 0.008$  во всех трех случаях).

Таким образом, результаты проведенного анализа показали, что среди  $\beta$ -изгибов типов I', II и II' достоверно чаще встречаются конформационно лабильные пептиды. Очевидно, что предлагаемые показатели конформационной лабильности пептидов могут быть использованы в качестве дополнительной независимой информации в вычислительных схемах для распознавания  $\beta$ -изгибов этих типов по аминокислотным последовательностям белков. Полученные данные свидетельствуют о принципиальном различии в структурообразовании типа I и остальных типов  $\beta$ -изгибов. В отличие от трех остальных типов  $\beta$ -изгибы I-го типа, по-видимому, не являются конформационно зависимыми.

## ЗАКЛЮЧЕНИЕ

Основной результат работы заключается в установлении уникальных свойств  $\beta$ -изгибов типов I', II и II' в аспекте аминокислотного «текста» этих локальных структур белков. В отличие от них  $\beta$ -изгиб типа I не проявляет выраженных особен-

**Рис. 2.** Эмпирические функции распределения значений интегральной оценки конформационной стабильности  $b_i$ .

ностей аминокислотного состава, в том числе в смысле обогащенности конформационно-лабильными последовательностями. Повышенная конформационная лабильность тетрапептидов, формирующих  $\beta$ -изгибы типов I', II и II', способствует смягчению стереохимических затруднений, характерных для наблюдаемых в эксперименте конформаций этих локальных структур.

#### ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 20-04-01085).

#### КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии интересов.

#### СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая работа не содержит каких-либо исследований с участием людей или животных в качестве объектов исследований.

#### СПИСОК ЛИТЕРАТУРЫ

1. W. Kabsch and C. Sander, *Biopolymers*, **22** (12), 2577 (1983).
2. X. de la Cruz, E. G. Hutchinson, A. Shepherd, and J. M. Thornton, *Proc. Natl. Acad. Sci. USA*, **99** (17), 11157 (2002).
3. M. Kumar, M. Bhasin, N. K. Natt, and G. P. S. Raghava, *Nucl. Acids Res.*, **33**, W154 (2005).
4. M. K. Elbashir, J. Wang, F. X. Wu, et al., *Proteome Sci.*, **11** (Suppl. 1), S5 (2013).
5. C. Fang, Y. Shang, and D. Xu, *Proteins*, **88** (1), 143 (2020).
6. A. G. de Brevern, *Sci. Reports*, **6**, 33191 (2016).
7. Л. А. Урошлев, И. Ю. Торшин, А. В. Батыновский и др., *Биофизика*, **60** (1), 5 (2015).
8. Л. А. Урошлев, И. Ю. Торшин, А. В. Батыновский и др., *Биофизика*, **64** (2), 272 (2019).
9. И. Ю. Торшин, А. В. Батыновский, Л. А. Урошлев и др., *Биофизика*, **64** (2), 256 (2019).
10. Ю. И. Журавлёв, К. В. Рудаков и И. Ю. Торшин, *Труды МФТИ*, **3** (4), 45 (2011).

## Conformational Aspects in the Formation of Structures of the Backbone of Polypeptide Chains in Proteins. Relationship between Conformational Stability/Lability and $\beta$ -Turns

I.Yu. Torshin\*, I.V. Filatov\*\*, A.V. Batyanovskii\*\*\*, K.V. Smirnov\*\*\*\*, A.A. Anashkina\*\*\*\*, N.G. Esipova\*\*\*\*, and V.G. Tumanyan\*\*\*\*

\*Dorodnicyn Computing Centre, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119333 Russia

\*\*Moscow Institute of Physics and Technology, Institutskiy per. 9, Dolgoprudny, Moscow Region, 141700 Russia

\*\*\*Institute of Biophysics and Cell Engineering, National Academy of Sciences of Belarus, Akademicheskaya ul. 27, Minsk, 220072 Belarus

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

To assess the nature of the relationship between the integral conformational stability of tetrapeptides and the main types of  $\beta$ -turns (which are also tetrapeptides), calculations were performed using spectrum diagrams and asymmetry in the distribution of conformationally stable and unstable tetrapeptides. It was shown that type I', II, and II'  $\beta$ -turns are comprised of predominantly conformationally labile peptides, that is consistent with the predetermined nature of their structure as described earlier. As previously shown in our study, in this case the conformation is imposed by external conditions (namely the cycle closing) and the prevalence of conformationally labile peptides facilitates the formation of the structure by virtue of external factors. Type I  $\beta$ -turn is an exception: peptides with different conformational lability are fairly evenly distributed. Most likely, this indicates that the formation of the type I  $\beta$ -turn is not predetermined.

*Keywords: conformational stable/labile segments of protein, local protein structure, conformational analysis,  $\beta$ -turns, statistical analysis, molecular dynamics*