

## ПРЕДОПРЕДЕЛЕННЫЕ КОНФОРМАЦИИ В ИЗГИБАХ ПОЛИПЕПТИДНЫХ ЦЕПЕЙ. ГЕОМЕТРИЧЕСКИЙ АНАЛИЗ

© 2019 г. Л.А. Урошлев, И.Ю. Торшин\*, А.В. Батяновский\*\*,  
Н.Г. Есипова\*\*\*, В.Г. Туманян\*\*\*

*Институт общей генетики им. Н.И. Вавилова РАН, 119991, Москва, ул. Губкина 3*

*\*Вычислительный центр им. А.А. Дородницына ФИЦ «Информатика и управление» РАН,  
119333, Москва, ул. Вавилова, 40*

*\*\*Институт биофизики клетки и клеточной инженерии НАН Беларуси, 220072, Минск, ул. Академическая, 27*

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

*E-mail: leoniduroshlev@gmail.com*

Поступила в редакцию 07.05.18 г.

После доработки 29.01.19 г.

Принята к печати 04.02.19 г.

$\beta$ -Изгибы – типовые локальные структуры полипептидной цепи, широко представленные в белках, играют важную структурную и функциональную роль. Априори следует ожидать большой структурной определенности  $\beta$ -изгиба, т.е. малого числа конформаций  $\beta$ -изгиба. Вследствие замыкания псевдоциклов через водородную связь в основаниях  $\beta$ -изгибов число независимых параметров должно уменьшиться настолько, что в этом случае конформационный анализ может быть сведен к геометрическому анализу. Тем самым определение конформационного набора  $\beta$ -изгиба может быть выполнено без привлечения силовых полей с большой точностью и достоверностью. В работе проведен конформационный анализ  $\beta$ -изгибов основных типов (I, I', II, II') с использованием независимых методов: оригинальной процедуры геометрии расстояний и процедуры перебора конформаций с последующей оптимизацией. Показано, что геометрический анализ в разработанном нами варианте достаточен для конформационного анализа  $\beta$ -изгибов: число геометрически состоятельных конформаций  $\beta$ -изгиба сводится к двум. Первое из найденных решений совпадает с экспериментальными данными рентгеноструктурного анализа. Второе решение, хотя и геометрически корректно, но энергетически маловероятно, так как соответствующие значения двугранных углов приходятся на строго запрещенные области карты Рамачандрана, вследствие недопустимого сближения атомов и экспериментально не наблюдаются. Таким образом выяснена природа формирования конформаций основных типов  $\beta$ -изгибов, в том числе содержащих так называемые «запрещенные» конформации.

*Ключевые слова:*  $\beta$ -изгиб, конформационный анализ, запрещенные конформации.

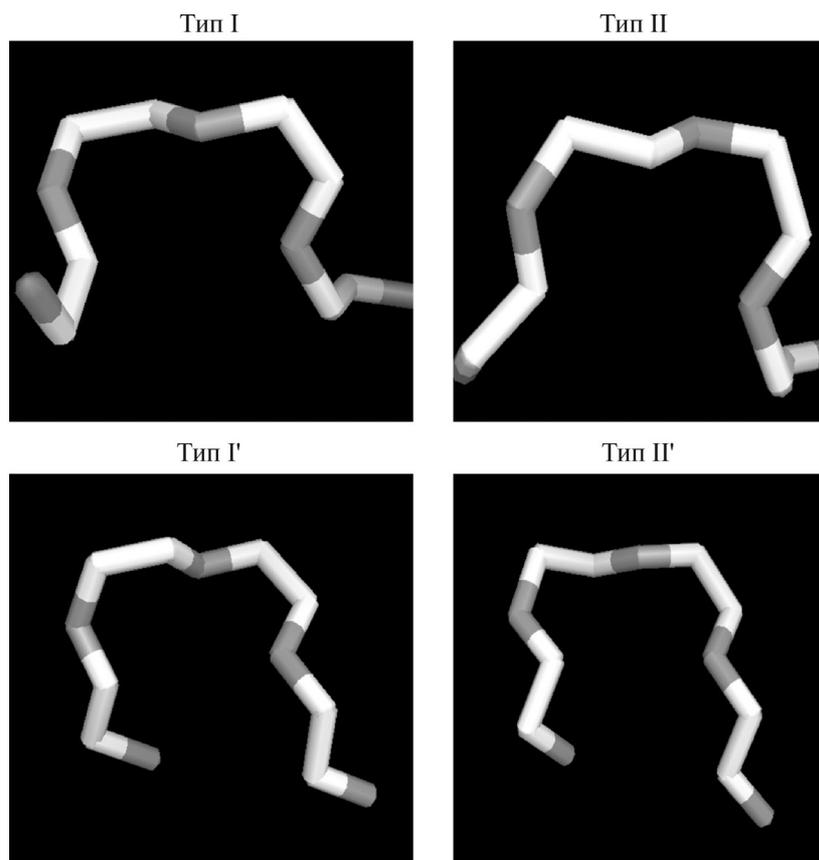
**DOI:** 10.1134/S0006302919020066

$\beta$ -Изгибы – один из самых распространенных структурных элементов в белках. Очевидна их роль в формировании структур белков и в самом процессе фолдинга белка. Хотя в изучении и предсказании  $\beta$ -изгибов достигнут определенный прогресс, вопрос о генезисе этого типа структур остается открытым.  $\beta$ -Изгибы – уникальный объект, поскольку для их конформационного анализа можно применить чисто геометрический метод геометрии расстояний. В свою очередь, это связано с тем, что конформация изгиба определяется малым числом независимых параметров.

$\beta$ -Изгибы состоят из четырех аминокислот, которые принято обозначать как  $i$ ,  $i + 1$ ,  $i + 2$ ,  $i + 3$ . Собственно изгиб приходится на позиции  $i + 1$ ,  $i + 2$ .

Мы исходим из обычно применяемого определения  $\beta$ -изгиба, включающего требование формирования водородной связи типа «остов–остов» между  $i$ -м и  $(i + 3)$ -м остатками, точнее, между СО-группой остатка  $i$  и NH-группой остатка  $i + 3$ .

Все  $\beta$ -изгибы можно разделить на несколько категорий, согласно возможным значениям углов  $\phi$  и  $\psi$  для аминокислот, находящихся в



**Рис. 1.** Изображение  $\beta$ -изгибов различных типов. Верхние рисунки слева-направо I-тип и II-тип; нижние рисунки слева-направо I'-тип и II'-тип.

центре изгиба. В данной работе рассматриваются четыре базовых класса  $\beta$ -изгибов полипептидной цепи. Согласно работе [1] они имеют обозначение I, I', II и II' (рис. 1). Класс I от I', так же как класс II от II', отличаются изменением знака у всех двугранных углов. Соответствующие пары структур основной цепи могут рассматриваться как энантиомеры.

Надо отметить, что конформационный переход от изгиба типа I к изгибу типа II (или изгиба типа I' к изгибу типа II') сводится к повороту (почти перевороту) пептидной группы между остатками  $i + 1, i + 2$  (при этом сильно изменяются только двугранные углы  $\psi_{i+1}$  и  $\phi_{i+2}$ ). Это явление получило название флиппинг пептидной группы [2].

Хотя в задаче предсказания конформационных элементов и, в частности,  $\beta$ -изгибов был достигнут определенный прогресс, большая часть методов предсказания построена по принципу «черного ящика» и, следовательно, не дает никакой информации о механизмах образования и устойчивости локальных конформаций данного типа.

Естественно полагать, что выяснение физических закономерностей формирования  $\beta$ -изгибов будет способствовать созданию более эффективных и физически обоснованных методов их предсказания.

Отдельной и весьма актуальной проблемой в задаче предсказания вторичной структуры является уточнение используемых определений вторичной структуры («альфа-спиралей», «бета-стрэндов», «левых спиралей типа полипролин II» и др.).

В настоящей работе для анализа структуры  $\beta$ -изгибов мы применяем подход, основанный на метрической геометрии, также известной как «геометрия расстояний» (от англ. «distance geometry»). Оптимизация функционалов оценки соответствия конформации заданному набору геометрических ограничений является существенным отличием геометрического подхода от повсеместно используемых физических подходов, в которых оптимизируются те или иные функционалы для оценки энергии внутримолекулярных взаимодействий.

Обычно в качестве критерия реалистичности той или иной конформации используется тот или иной способ оценки потенциальной энергии и ее компонентов. При этом считается, что нативная структура белка соответствует некоторому минимуму используемого функционала для оценки энергии (в идеале – свободной энергии). Критерий минимума энергии позволяет объяснить, почему цепь укладывается именно так, а не иначе. Например, при анализе «запрещенных» конформаций в  $\beta$ -изгибе типа II' [3] «энергетическое» объяснение заключается в том, что в системе « $\beta$ -изгиб +  $\beta$ -шпилька», несмотря на проигрыш энергии в  $\beta$ -изгибе (из-за напряженных ван-дер-ваальсовых контактов), имеет место выигрыш энергии в структуре в целом.

Для выяснения конкретного конформационного механизма, который бы отвечал такому распределению энергии, мы выдвинули идею «топологического замка» [4], заключающуюся в том, что фиксация концевых остатков  $\beta$ -шпильки, примыкающих к  $\beta$ -изгибу, посредством формирования псевдоцикла через водородную связь резко ограничивает число вариантов конформации для аминокислотных остатков, входящих в  $\beta$ -изгиб. При этом остается только вариант, соответствующий «запрещенной» конформации, в то время как принятие изгибом незапрещенной конформации потребовало бы перестройки всей шпильки.

Таким образом,  $\beta$ -изгибы с запрещенными конформациями являются следствием топологических ограничений: их следует считать «навязанными» со стороны  $\beta$ -шпилек, к которым примыкают  $\beta$ -изгибы. Более конкретно это определяется тем, что при замыкании псевдоцикла ряд конформационных параметров становятся зависимыми. Из-за этого число параметров, определяющих структуру, значительно уменьшается, и проблему можно рассматривать с точки зрения чисто геометрического анализа. Как правило, в роли конформационных параметров выступают двугранные углы, валентные углы и расстояния между валентно связанными атомами. Это так называемые «естественные параметры»; возможно использование других параметров, например координат атомов.

Важность анализа структуры бета-изгибов обусловлена также и тем, что от учета таких структурных деталей существенно зависит точность (и обобщающая способность) алгоритмов предсказания вторичной структуры.

Первые работы по статистическому анализу и предсказанию  $\beta$ -изгибов появились еще в конце 80-х годов XX века [5]. В настоящий момент

для предсказания  $\beta$ -изгибов используется весь арсенал биоинформационных и статистических методов, таких как метод  $k$ -ближайших соседей [6], предсказания с помощью машины опорных векторов [7] и так называемых искусственных нейронных сетей [8]. Для моделирования  $\beta$ -изгибов в конкретных структурах часто используются классические методы моделирования структуры *in silico*, такие как молекулярная динамика или метод Монте-Карло.

Эти методы особенно актуальны для исследования  $\beta$ -изгибов в амилоидных белках, для которых затруднено получение данных такими экспериментальными методами, как рентгеноструктурный анализ и ЯМР. В группе теоретических и расчетных методов, моделирование структуры производится с использованием энергетического критерия.

Ряд работ в последние годы был посвящен анализу «запрещенных» II'-изгибов в SH3-доме, включая статистический анализ их распределения на карте Рамачандрана. Это связано с тем, что локальные конформации (углы  $\phi$  и  $\psi$ ) аминокислотных остатков, принадлежащих II'-изгибу, на карте Рамачандрана находятся в зоне, разрешенной только для глицина. Однако в SH3-доме  $\alpha$ -спектрина в области  $\beta$ -изгиба II'-типа на карте парадоксальным образом достаточно часто находятся другие аминокислоты.

## ПОСТАНОВКА ЗАДАЧИ

Основной задачей конформационного анализа является описание и объяснение конформационного набора соединения, в нашем случае фрагмента полипептидной цепи. В качестве критерия реализации той или иной конформации используется энергия и ее компоненты. Считается, что нативная структура белка соответствует минимуму свободной энергии, или в первом приближении – энергии; этот принцип распространяется и на отдельные фрагменты структуры. Полное решение задачи предполагает объяснение формирования определенного типа структуры, определенной конформации, соответствующей энергетическому минимуму.

Для случая  $\beta$ -изгибов конкретного объяснения генезиса такого рода структуры до сих пор не было. Для частного случая  $\beta$ -изгиба типа II', характеризующегося запрещенной конформацией (с напряженными ван-дер-ваальсовыми контактами), как мы уже упоминали, объяснение на уровне энергии заключается в том, что в суммарной системе « $\beta$ -изгиб + массивная  $\beta$ -шпилька», несмотря на проигрыш энергии в

**Таблица 1.** Среднее и дисперсия двугранных углов для изгибов разных типов

Тип изгиба	Среднее по $\phi_{i+1}$	Среднее по $\psi_{i+1}$	Среднее по $\phi_{i+2}$	Среднее по $\psi_{i+2}$	Дисперсия по $\phi_{i+1}$	Дисперсия по $\psi_{i+1}$	Дисперсия по $\phi_{i+2}$	Дисперсия по $\psi_{i+2}$
Тип I	-61,12	-29,35	-89,321	-0,189	22,096	30,284	29,464	29,668
Тип I'	57,181	31,93	87,214	-0,413	19,621	27,735	24,938	28,393
Тип II	-58,038	125,05	81,187	0,52	23,882	16,365	27,858	28,851
Тип II'	58,287	-123,99	-82,447	-1,883	25,054	20,605	28,903	26,849

$\beta$ -изгибе, имеет место выигрыш энергии в структуре в целом. Но вопрос заключается в том, какая конкретная геометрическая реализация стоит за этими общими энергетическими соотношениями.

В идее топологического замка [4] содержится механизм формирования геометрии (конформации)  $\beta$ -изгиба, состоящий в том, что фиксация концевых остатков  $\beta$ -шпильки, примыкающих к  $\beta$ -изгибу, ограничивает число конформационных вариантов для аминокислотных остатков, входящих в  $\beta$ -изгиб так, что остается только вариант, соответствующий «запрещенной» конформации. Принятие изгибом другой (незапрещенной) конформации потребовало бы перестройки всей шпильки.

Предложенный механизм носит общий характер и может быть применен для объяснения формирования всех  $\beta$ -изгибов, а не только содержащих запрещенную конформацию. Таким образом,  $\beta$ -изгибы являются следствием топологических ограничений: их следует считать навязанными со стороны  $\beta$ -шпилек, к которым примыкают  $\beta$ -изгибы. Говоря более конкретно, это определяется тем, что при замыкании (через водородные связи) псевдоцикла ряд конформационных параметров становятся зависимыми. Из-за этого число параметров, определяющих структуру, значительно уменьшается, и проблему можно рассматривать с точки зрения чисто геометрического анализа.

## МАТЕРИАЛЫ И МЕТОДЫ

Для определения конформации с помощью геометрического анализа необходимо определиться с теми конформационными параметрами, которые полагаются фиксированными, и с параметрами, которые варьируются в ходе перебора. Как правило, в роли конформационных параметров выступают двугранные углы  $\phi$  и  $\psi$  или расстояния между атомами. В свою очередь, конформационные параметры можно разделить на независимые и зависимые.

Преимущество геометрических алгоритмов анализа полипептидов состоит в том, что результат не зависит от использованного силового поля и носит, в известной мере, абсолютный характер. Трудности же состоят в том, что число параметров для варьирования не должно быть сколько-нибудь велико, иначе потребуются непомерные вычислительные ресурсы, и не должно быть слишком мало, иначе решения может просто не быть.

К основным методам геометрического анализа следует отнести метод перебора конформационных параметров (например, таких как двугранные углы) и метод геометрии расстояний.

Метод перебора конформационных параметров, элементарный в своей сути (см., например, работу [9]), позволяет определять конформационное состояние остова для случая небольшого полипептида. В случае, когда полипептид состоит более чем из пяти-шести аминокислот, алгоритм требует слишком большого количества машинного времени. К тому же количество необходимого процессорного времени растет с уменьшением шага перебора. Но метод перебора можно значительно ускорить, заменив прямой перебор конформационных параметров каким-нибудь безградиентным методом оптимизации.

Методы геометрии расстояний намного более быстрые. Кроме того, в отличие от метода перебора, число переменных может быть сделано достаточно большим. В задачах геометрии расстояний необходимо установить координаты группы атомов, на основе задаваемого пользователем, множества расстояний между ними.

Количество  $\beta$ -изгибов в банке данных PDB постоянно растет вместе с увеличением числа структур белков. Улучшается и качество данных, поскольку все большее число структур делается с высоким разрешением.

В табл. 1 приведены средние значения и величины дисперсии двугранных углов для изгибов разных типов. Исходные для анализа данные получены с помощью программы

PDBselect (использован релиз банка данных PDB от февраля 2016 г.). В качестве критерия, определяющего  $\beta$ -изгиб, служило определение из работы [10]. Допускалось отклонение от приведенных там значений на  $10^\circ$ .

Как можно видеть из таблицы, совпадение между структурами, относящимися к одному типу изгиба, достаточно строгое.

Геометрическая задача нахождения координат атомов молекулы решается одним из нескольких методов, являющихся изоморфными. Напомним наиболее известный метод Эйринга. Для каждой тройки атомов можно выбрать локальную прямоугольную координатную систему, связанную с данной группой атомов. При этом один из атомов выбирается за нуль системы координат. Первая ось (ось  $x$ ) проводится вдоль ковалентной связи между нулем системы координат и следующим атомом. Вторая ось (ось  $y$ ) проводится в плоскости, проходящей через три последовательных атома, в сторону третьего атома. Третья ось (ось  $z$ ) достраивается по правилу правой тройки ортогональных векторов.

Между такими координатными системами можно выполнять переходы с помощью соотношения матриц, соответствующих вращениям, и векторов, соответствующих трансляциям (см. например, работу [9]).

Таким образом, задавая значения двугранных и валентных углов и длин связей, можно вычислить положение атомов в единой (первой) системе координат. Отсюда легко получить простой алгоритм для определения положения атомов в замкнутом цикле, т.е. в структуре, положение атомов на концах которой строго фиксировано.

При построении геометрической модели  $\beta$ -изгиба надо определиться с параметрами, которые необходимо варьировать, и пределами их варьирования. Очевидно, что увеличение числа параметров ведет к резкому увеличению времени расчета, так как очевидно, что при увеличении количества параметров увеличивается количество вложенных циклов при переборе. В этой связи необходимо оценить, насколько именно конформации пептидных групп в  $\beta$ -изгибах отличаются от стандартной «плоской» конформации и насколько правомерно считать угол  $\omega$  фиксированным при значении  $180^\circ$ .

**Предварительная оценка степени варьирования углов  $\omega$  в пептидных связях.** Для выбора пределов варьирования, были построены гистограммы для значений  $\omega$ -углов всех рассмат-

риваемых типов изгибов в реальных белковых структурах (рис. 2).

Как видно из приведенных гистограмм, распределение во всех случаях близко к нормальному, математическое ожидание во всех случаях близко к  $180,0^\circ$ . Поэтому перебор по углам  $\omega$  необходимо и достаточно осуществлять в малых пределах ( $180^\circ \pm 5^\circ$ ). Полученные значения согласуются с вариациями соответствующих углов согласно оценке, сделанной в работе [11].

## РЕЗУЛЬТАТЫ

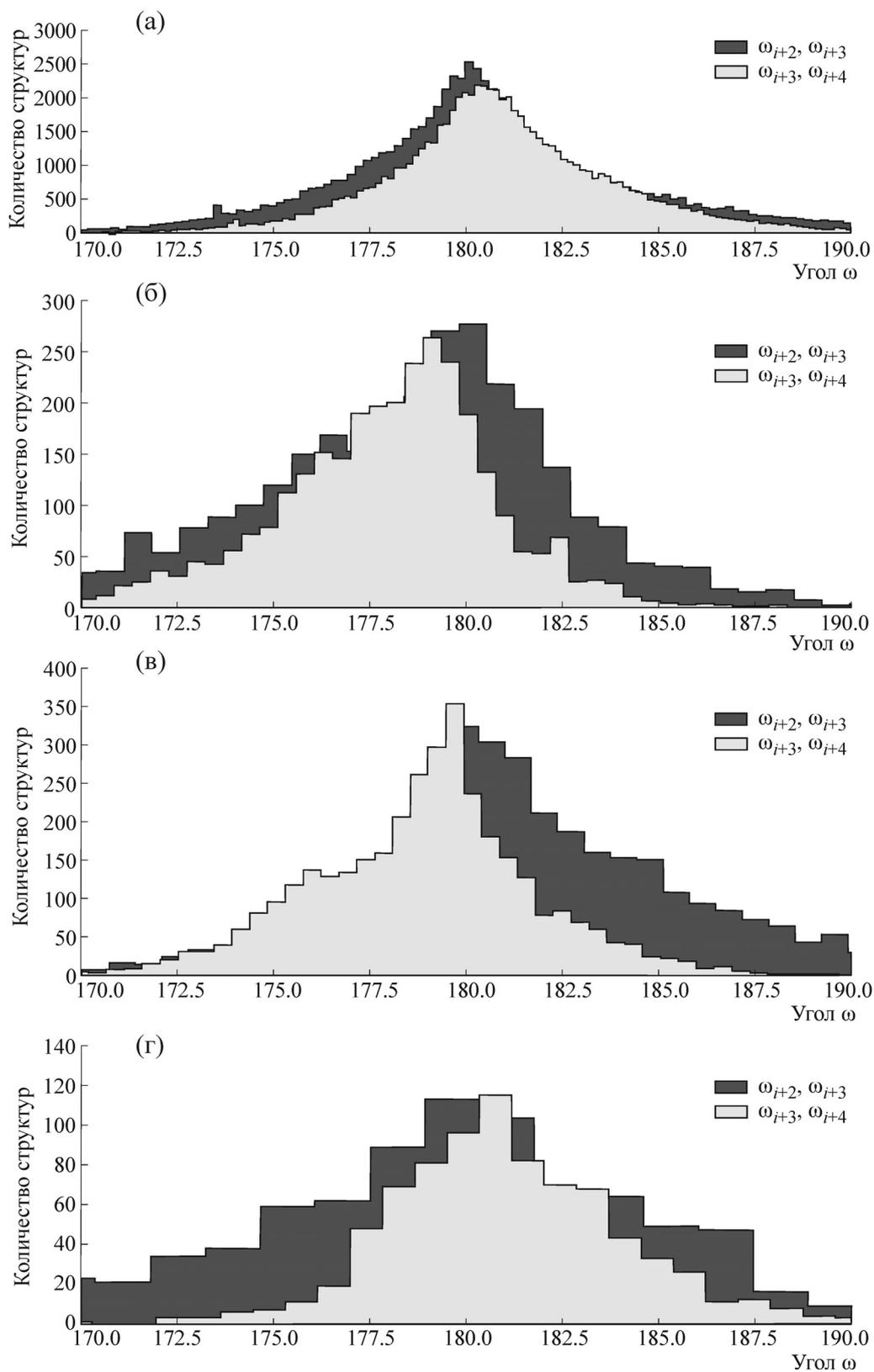
В качестве исходных данных были выбраны несколько структур с разными типами изгибов. Критерием выбора было высокое разрешение и близость к базовым значениям двугранных углов каждого типа изгиба.

При построении бета-изгибов с помощью перебора и оптимизации были жестко зафиксированы атомы, находящиеся на N- и C-конце каждого изгиба. Для них заданы расстояния до всех других атомов в изгибе. В качестве параметров, подлежащих варьированию, были выбраны четыре двугранных угла –  $\Phi_{i+1}$ ,  $\Psi_{i+1}$  и  $\Phi_{i+2}$ ,  $\Psi_{i+2}$ . Эти углы варьируют в пределах от  $-180,0^\circ$  до  $180,0^\circ$  с шагом в  $1^\circ$ . Два угла  $\omega_{i+1}$  и  $\omega_{i+2}$  были варьированы, как показано выше, в пределах  $5^\circ$  от положения, наблюдаемого в эксперименте. Шаг варьирования для углов  $\omega_{i+1}$  и  $\omega_{i+2}$  также составлял  $1^\circ$ . Конформация считалась принятой, если среднеквадратичное отклонение между рассчитанными координатами N, C $_\alpha$  и C $_\beta$  ( $i + 1$ )-й позиции и N в ( $i + 2$ )-й позиции и реальными меньше  $0,01 \text{ \AA}$ .

Как можно видеть из табл. 2, помимо конформаций, содержащихся в банке данных PDB, было получено по одному дополнительному решению. Часть дополнительных решений ранее не была описана, но найдена нами в банке данных PDB (правда, только для структур крайне низкого разрешения), а часть не описана в литературе и в банке данных не обнаруживается.

Предполагается, что полученные вторые решения не могут реализоваться из-за попадания в строго запрещенные области, ранее описанные в работе [12]. Соответствия между дополнительными конформациями и строго запрещенными областями приведены в табл. 3.

**Метод геометрии расстояний.** При реализации метода геометрии расстояний необходимо найти все минимумы так называемой функции ошибки. В большинстве источников она определяется следующим образом:



**Рис. 2.** Гистограммы, показывающие распределение угла  $\omega$  между остатками  $(i + 2) - (i + 3)$ : (а) – в изгибах типа I; (б) – в изгибах типа I'; (в) – в изгибах типа II; (г) – в изгибах типа II'. Темно-серым цветом показано распределение угла  $\omega$  в первой аминокислоте изгиба, светло-серым цветом – во второй.

**Таблица 2.** Локальные конформации для  $\beta$ -изгибов различных типов, полученные с помощью методов перебора конформационных параметров

Тип изгиба	$\Phi_{i+1}$	$\Psi_{i+1}$	$\omega_{i+1}$	$\Phi_{i+2}$	$\Psi_{i+2}$	$\omega_{i+2}$	Средняя невязка по углам с экспериментом
I	-57,003	-25,534	175,31	-91,976	-5,342	-176,31	0,014800
I (дополнительное решение)	-57,003	-25,534	175,31	91,316	84,587	-176,31	–
I'	51,324	34,817	170,03	81,205	0,628	-176,11	0,307090
I' (дополнительное решение)	51,324	34,817	170,03	-156,517	-97,187	-176,12	–
II	-58,148	123,28	173,79	83,381	-3,309	175,72	0,014022
II (дополнительное решение)	-58,148	123,28	173,79	-12,595	58,537	175,72	–
II'	57,49	-101,87	-176,71	-118,3	-13,3	177,5	0,015411
II' (дополнительное решение)	57,49	-101,87	-176,71	-2,44	-83,91	177,5	–

Примечание. Невязка вычисляется между вычисленной структурой и структурой, полученной из рентгеноструктурных данных PDB.

**Таблица 3.** Соответствие дополнительных решений, полученных в результате замыкания цикла и строго запрещенных областей из работы [12]

Тип изгиба	$\Phi_{i+1}$	$\Psi_{i+1}$	$\Phi_{i+2}$	$\Psi_{i+2}$	Строго запрещенная область
I	-57,003	-25,534	91,316	84,587	D
I'	51,324	34,817	-156,517	-97,187	C'
II	-58,148	123,28	-12,595	58,537	A
II'	57,49	-101,87	-2,44	-83,91	B

$$F(p) = \sum_{i < j} \left( \min^2 \left( 0, \frac{\|p_i - p_j\|^2}{l^2(a_i, a_j)} - 1 \right) + \max^2 \left( 0, \frac{\|p_i - p_j\|^2}{u^2(a_i, a_j)} - 1 \right) \right), \quad (1)$$

где  $p_i$  – вектор координат, а  $l(a_i, a_j)$  и  $u(a_i, a_j)$  – верхняя и нижняя границы расстояний между атомом  $i$  и атомом  $j$ .

Очевидно, что минимумов для данной функции бесконечно много, так как функция инвариантна относительно переносов начала координат, а также поворотов. Поэтому первые три точки исследуемого изгиба фиксируются следующим образом: первая точка принимается за начало координат, вторая – как  $p_2 = [0, 0; 0, 0; r_{12}]$ , третья –  $p_3 = [0, 0; r_{23} \sin \alpha; r_{12} + r_{23} \cos \alpha]$ , где  $\alpha$  – валентный угол между векторами, проведенными между началом координат и первой точкой и между второй и третьей точкой, а  $r_{12}$ ,  $r_{23}$  – расстояния между первой и второй и второй и третьей точкой соответственно. От-

сюда следует первое ограничение на исходные данные – верхняя и нижняя границы расстояний между атомами 1 и 2, 1 и 3, а также атомами 2 и 3 должны быть равны. Кроме того, очень важна корректная постановка задачи – если заданные интервалы между верхней и нижней границами расстояния слишком большие, то вместо одной, заранее определенной конформации, можно получить бесконечное множество конформаций, отличающихся малыми возмущениями расстояний между атомами. Для получения однозначного решения, вышеописанная функция ошибки должна быть «сглажена», в ней не должно быть «протяженных» минимумов.

Ранее уже было известно некоторое количество программ, реализующих различные алгоритмы для решения задачи геометрии расстояний. В качестве примера можно привести программы Mджеер [13] и DGsol [14]. Однако для типичных на сегодняшний день задач имеющиеся реализации не подходят, так как эти методы устанавливают ограничения, неприем-

**Таблица 4.** Локальные конформации для  $\beta$ -изгибов различных типов, полученные с помощью метода геометрии расстояний

Тип изгиба	$\Phi_{i+1}$	$\Psi_{i+1}$	$\omega_{i+1}$	$\Phi_{i+2}$	$\Psi_{i+2}$	$\omega_{i+2}$	Среднеквадратичное отклонение с решением, полученным в результате перебора-оптимизации
I	-57,1	-25,7	175,31	-94,1	-3,9	-176,31	0,643608
I	-57,4	-25,5	175,31	91,4	84,5	-176,31	0,111744
I'	51,3	34,9	170,03	-81,2	0,7	-176,11	0,028144
I'	51,3	34,8	170,03	-156,5	-97,2	-176,12	0,022252
II	-59,3	123,3	173,79	83,4	-3,4	175,72	0,294124
II	-54,7	123,5	173,79	-12,5	58,5	175,72	0,862898
II'	57,8	-102,3	-176,71	-117,7	-14,2	177,5	0,019276
II'	57,5	-101,8	-176,71	-2,6	-83,8	177,5	0,01134

лемые, в том числе, в контексте поставленных в данной работе задач. Mdjeer работает только когда верхняя и нижняя границы расстояний равны между собой, Dgsol же выполняет ядерное сглаживание функции ошибок, однако при этом требуются значительные вычислительные ресурсы.

В связи со сказанным была разработана собственная реализация для решения задачи геометрии расстояний. На первом шаге алгоритма определялась любая конформация, удовлетворяющая заданному набору ограничений. Для этого с помощью алгоритма SLSQP [15], определялся минимум функции ошибки (1). На втором шаге в построенной конформации вычислялись расстояния между всеми атомами. После этого для всех атомов строили систему из трех уравнений с фиксированными расстояниями, посчитанными исходя из полученной ранее конформации. Таким образом можно определить все возможные координаты и все возможные конформации, так как для каждой системы, состоящей из трех квадратичных уравнений, имеются два решения. Это следует из простых геометрических соображений: пересечение двух сфер дает окружность, пересечение окружности со сферой дает две точки в трехмерном пространстве.

Для тестирования разработанного алгоритма геометрии расстояний мы построили структуры для всех платоновых тел и фуллеренов C60 [16] и C70. При этом для тетраэдра, икосаэдра и октаэдра были заданы только длины связей между точками, для куба и додекаэдра – длины связей и длины диагоналей для квадрата и пятиугольника, являющихся соответственно гранями этих тел.

**Демонстрация согласованности перебора и геометрии расстояний.** С помощью разработанной нами программы был проведен поиск ре-

шений для  $\beta$ -изгибов всех рассматриваемых типов. В соответствии с принципом топологического замка жестко фиксировали расстояния от концов  $\beta$ -изгиба – были заданы расстояния и валентные углы между остатками  $i + 1$  и  $i + 2$ , а также между  $i + 3$  и  $i + 4$ . Для атомов, не находящихся на концах изгиба, задавали длины ковалентных связей и величины валентных углов.

В результате решения поставленной таким образом задачи геометрии расстояний с помощью описанного выше алгоритмов и соответствующих программ получается единственное решение, полностью совпадающее с наблюдаемым в эксперименте. Однако второе (дополнительное) решение, не подтвержденное экспериментом, отсутствует. После расчета многочисленных вариантов мы нашли, что если вместо расстояний между краями шпильки задать широкий интервал возможных расстояний, то второе решение появляется, причем в точности соответствующее второму решению, получаемому методом перебора и оптимизации.

По ходу расчетов для каждого из рассматриваемых типов изгибов был выбран набор из десяти структур с хорошим рентгеноструктурным разрешением (менее 2 Å). Между атомами изгиба были рассчитаны расстояния, по которым и строились изгибы. Для каждой из выбранных структур был построен набор из 16 конформаций для каждого из рассматриваемых типов изгибов. Среди этого набора присутствуют как конформация, наблюдаемая в эксперименте, так и вторая конформация, полученная в результате перебора двугранных углов.

Точность приведенных выше результатов достаточно высока при минимальном расходе процессорного времени. Для перебора значений двугранных углов потребовалось 23 ч работы при полной загрузке 60 четырехядерных про-

цессоров Intel Core i5. Все структуры, полученные с помощью методов геометрии расстояний, были рассчитаны за 5–10 мин на одном ядре процессора Intel Core i3.

Необходимо также отметить, что решения в методе геометрии расстояний порождаются парами. Конформации в одной паре представляют собой зеркальные отражения друг друга. Поэтому при получении конформации бета-изгиба типа I (II) в наборе конформаций всегда будет присутствовать изгиб типа I' (II').

Итак, в результате проведенного конформационного анализа двумя независимыми геометрическими методами установлен генезис конформаций  $\beta$ -изгибов остова полипептидной цепи. Показано, что для каждого типа изгибов существуют два решения. Одно – точно соответствует данным рентгеноструктурного эксперимента. Второе – геометрически состоятельное, но не могущее осуществиться вследствие неблагоприятных контактов атомов остова.

#### ДИСКУССИЯ

В работе изложен исключительно геометрический подход к проблеме сворачивания  $\beta$ -изгибов и обойден стороной вопрос их формирования в реальных структурах. Вероятнее всего, обнаруженное второе решение имеет чисто геометрическую природу и запрещено из физических соображений. Косвенно это подтверждается запрещенными контактами между N-атомами соседних аминокислот, стоящих в вершине шпильки.

Дальнейшее изучение различных видов  $\beta$ -изгибов может пролить свет на механизмы сворачивания и агрегации амилоидных белков для болезни Паркинсона и Альцгеймера. Так как основным элементом амилоидных белков является мотив [17] «бета-лист–изгиб–бета-лист», изменение типа изгиба в мотиве может приводить к олигомеризации и развитию заболеваний. Однако при исследовании таких структур необходимо учитывать не только структурные элементы, но и окружающие кофакторы. Как было показано в работах [18,19], при формировании реальных структур значительное влияние оказывается внешним растворителем и разнообразными структурными кофакторами. Поэтому целесообразно разработать комплексный подход, где геометрические алгоритмы бу-

дут работать совместно с алгоритмами, основанными на молекулярно-динамическом моделировании.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проекты № 17-04-02105 и № 18-54-00037\_Бел\_а).

#### СПИСОК ЛИТЕРАТУРЫ

1. C. M. Venkatachalam, *Biopolymers* **6** (10), 1425 (1968).
2. S. Hayward, *Prot. Sci.* **10**, 2219 (2001).
3. I. Y. Torshin, N. G. Esipova, and V. G. Tumanyan, *J. Biomol. Structure and Dynamics* **32** (2) 198 (2014).
4. Л. А. Урошлев, И. Ю. Торшин, А. В. Батяновский и др., *Биофизика* **60**, 5 (2015).
5. P. Y. Chou and G. D. Fasman. *Biophys. J.* **26** (3), 36 (1979).
6. K. E. Han, C. Bystroff, and D. Baker, *Prot. Sci.* **6** (7), 1587 (1997).
7. Q. Zhang, S. Yoon, and W. J. Welsh, *Bioinformatics* **21** (10), 2370 (2005).
8. M. Kumar, M. Bhasin, N. K. Natt, and G. P. S. Raghava, *Nucl. Acids Res.* **33** (Suppl. 2), W154 (2005).
9. N. Go and H. A. Scheraga, *Macromolecules* **9** (4), 535 (1976).
10. P. N. Lewis, F. A. Momany, and H. A. Scheraga, *Biochim. Biophys. Acta. Protein Structure* **303** (2), 211 (1973).
11. R. Improta and L. Vitagliano, *PLoS One* **6** (9), e24533 (2011).
12. И. Ю. Торшин, А. В. Батяновский, Л. А. Урошлев и др., *Биофизика* **63** (2), 225 (2018).
13. A. Mucherino, L. Liberti, and C. Lavor, *Mathematical Software – ICMS 2010* (Springer, Berlin–Heidelberg, 2010).
14. L. Liberti, C. Lavor, N. Maculan, F. Marinelli, J. *Global Optimization* **43** (2–3), 207 (2009).
15. D. Kraft et al., *A Software Package for Sequential Quadratic Programming, Technical Report DFVLR-FB 88–28* (Obersaffelhofen, Germany, 1988).
16. Y. Li, Y. Huang, S. Du, and R. Liu, *Chem. Phys. Lett.* **335** (5–6), 524 (2001).
17. M. Ahmed, J. Davis, D. Aucoin, et al., *Nature Struct. Mol. Biol.* **17** (5), 561 (2010).
18. S. V. Rahmanov, I. V. Kulakovskiy, L. A. Uroshlev, and V. J. Makeev, *J. Bioinformatics Comput. Biol.* **8** (03), 427 (2010).
19. L. A. Uroshlev, I. V. Kulakovskiy, N. G. Esipova, et al., *J. Biomol. Structure and Dynamics* **36** (1), 221 (2018).

## Predetermined Conformations in the Bends of Polypeptide Chains. Geometric Analysis

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

\*Vavilov Institute of General Genetics, Russian Academy of Sciences, ul. Gubkina 3, Moscow, 119991 Russia

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

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

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

$\beta$ -bends are typical local structures of the polypeptide chain, widely represented in proteins, play an important structural and functional role. With a good degree of certainty, a priori one might expect that a structure of  $\beta$ -bending is quite predictable, for instance, there are a small number of  $\beta$ -bending conformations. Because of the closure of pseudo-cycles through a hydrogen bond in the bases of  $\beta$ -bends, the number of independent parameters should decrease to the amount that instead of the conformational analysis a geometric analysis can be used. Thus, determination of the conformational set of  $\beta$ -bending can be performed with great accuracy and reliability without any force fields. In this study we performed a conformational analysis of  $\beta$ -bends of the main types (I, I', II, II') using independent methods: the original distance geometry procedure and the conformation enumeration procedure with subsequent optimization. It is shown that the geometrical analysis in the form as we have developed it is sufficient for the conformational analysis of  $\beta$ -bends: the number of geometrically consistent  $\beta$ -bending conformations is reduced to two. The first solution found coincides with the experimental data of X-ray structural analysis. The second solution is correct based on the geometric analysis, but is improbable in terms of energy since the corresponding values of the dihedral angles fall on strictly forbidden areas of the Ramachandran map as a result of disallowed convergence of atoms and it is not experimentally observed. Thus, the nature of the formation of conformations of the main types of  $\beta$ -bends, including those containing so-called "forbidden" conformations, has been clarified.

*Keywords:*  $\beta$ -bend, conformational analysis, disallowed conformations