

КОНФОРМАЦИОННАЯ СТАБИЛЬНОСТЬ/ЛАБИЛЬНОСТЬ ПЕПТИДНЫХ ФРАГМЕНТОВ В КОНТЕКСТЕ ПОСЛЕДОВАТЕЛЬНОСТИ АМИНОКИСЛОТ

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

*Вычислительный центр им. А.А. Дородницына Федерального исследовательского центра
«Информатика и управление» РАН, 119333, Москва, ул. Вавилова, 40*

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

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

****Институт молекулярной биологии им. В.А. Энгельгардта РАН, 119991, Москва, ул. Вавилова, 32
E-mail: ty135@yahoo.com*

Поступила в редакцию 14.11.18 г.

После доработки 11.12.18 г.

Принята к публикации 14.12.18 г.

Сформулированы критерии оценки конформационной стабильности/лабильности пептидных фрагментов, входящих в состав структур белков. С использованием предложенных оценок проведен статистический анализ тетрапептидных фрагментов (их конформаций и последовательностей) в выборке из 25121 структуры белковых цепей из банка данных белков PDB. В результате анализа показано, что тетрапептидные фрагменты существенно различаются по значениям оценки конформационной стабильности/лабильности с точки зрения предложенных статистических критериев. Результаты расчетов тетрапептидов методом денатурирующей молекулярной динамики использованы в качестве независимого подхода для оценки стабильности/лабильности пептидных фрагментов. Продемонстрирована корреляция между оценками конформационной лабильности, полученными на основании статистического анализа ансамблей конформаций пептидов, наблюдаемых в экспериментально определенных структурах белков, и оценками конформационной лабильности/стабильности, вычисленными на основании траекторий молекулярной динамики. Выделены подгруппы более «конформационно-стабильных пептидов», которые характеризуются преимущественно α -спиральной конформацией. Сформулированы с использованием комплексных критериев консенсусные списки тетрапептидов, характеризующихся наименьшей конформационной лабильностью (наибольшей конформационной стабильностью). Выделены пептиды, обладающие повышенной конформационной лабильностью. Таким образом, среди всех комбинаторно возможных тетрапептидов имеет место существенная доля тетрапептидов, характеризующихся определенностью конформационного состояния. Охарактеризовано соотношение между степенью конформационной определенности пептида и его участием в первичной структуре белка. При этом оказалось, что роль конформационно-стабильных пептидов в формировании структуры белка может быть достаточно велика, так как они составляют, в среднем, около 10% аминокислотной последовательности. На примерах реальных растворимых пептидов показана возможность оценки конформационной стабильности любой наперед заданной аминокислотной последовательности на основе разработанных критериев конформационной лабильности тетрапептидов.

Ключевые слова: конформационно стабильные/лабильные сегменты белка, локальная структура белка, конформационный анализ, статистический анализ, молекулярная динамика.

DOI: 10.1134/S0006302919020054

Задача определения структуры белка по его последовательности остается в списке важнейших вызовов современной биологии. Результа-

ты проведенных ранее исследований свидетельствуют о чрезвычайной сложности задачи, и представляется разумным сформулировать некоторые подзадачи в рамках этой общей задачи, в частности изучить взаимосвязь «последовательность–структура» как функцию длины пеп-

Сокращение: МД – молекулярная динамика.

тида. Из априорных соображений следует, что структурная определенность (или «конформационная стабильность») должна расти с удлинением последовательности. Действительно, разница в энергии (свободной энергии) конформаций короткого пептида существенно меньше, чем для длинного пептида. В принципе возможна ситуация, когда полная структурная однозначность наступает только при достижении некоторой длины аминокислотной последовательности (хотя даже отдельные аминокислотные остатки проявляют тенденцию концентрироваться в определенных областях карты Рамачандрана [1]).

Возникает закономерный вопрос, имеет ли место соответствие между последовательностью и структурой на уровне последовательностей более коротких, чем нативная цепь белка. Особенно многообещающей исследовательской задачей является анализ локальных структур, соответствующих коротким участкам полипептидной цепи (содержащими две, три или четыре аминокислоты) и установление возможных закономерностей связи между конформацией таких пептидных фрагментов и их аминокислотными последовательностями. Представляет интерес именно исследование коротких сегментов, поскольку в этом случае все возможные последовательности могут быть перечислены. В самой общей постановке задача может быть рассмотрена как классификация объектов по двум группам существенно разнородных признаков: аминокислотной последовательности и пространственной конформации.

Для решения такой исследовательской задачи можно, во-первых, изучать общие конформации для сходных последовательностей аминокислот. Во-вторых, можно на основе классификации наблюдаемых конформаций постараться определить набор аминокислотных последовательностей, характерных для конформации того или иного типа. Подход, заключающийся в систематизации олигопептидов с конкретной последовательностью по тому, какую структуру или конформацию (по критерию значений двугранных углов в основной цепи) они принимают, был применен для анализа и предсказания вторичной структуры белка [2–4] и распространен на конформации общего типа. Можно привести работы [5,6] как пример удачного варианта сопоставления конформаций тем или иным аминокислотным последовательностям.

При использовании альтернативной методологии [6] систематизация фрагментов ведется на структурном уровне по признаку их структурного сходства с выделением структурных

кластеров. Каждому структурному кластеру приписываются определенные аминокислотные последовательности, и в конце концов структуру канонического фрагмента удастся устанавливать только на основе информации о последовательности. Работу [7] можно привести как пример комбинированного подхода, когда учитываются разнонаправленные связи конформации и последовательности.

Мы допускаем в качестве рабочей гипотезы существование достаточно коротких фрагментов полипептидной цепи, характеризующихся некоторой предпочтительной конформацией. Мы определяем конформационно-стабильный олигопептид как сегмент полипептидной цепи с определенной последовательностью аминокислотных остатков, который проявляет явные конформационные предпочтения в достаточно большой выборке структур белков. Очевидно, что подобное качественное определение нуждается в количественных критериях оценки «явности» конформационных предпочтений.

Поиск такого рода «конформационно-стабильных» тетрапептидов был начат в работе [8]. Конформационно-стабильные пептиды выделяли по постоянству конформации одной и той же последовательности в разных белках. Конкретно, конформационно-стабильными считались те олигопептиды, для которых более 65% наблюдаемых конформационных состояний были близки друг к другу. Степень близости конформаций в пределах этого базового конформационного типа также должна была быть достаточно велика (отличие по двугранным углам менее 10°). В итоге в работе [9] был получен список, включающий 900 α -спиральных и около 50 не α -спиральных конформационно-стабильных пептидов.

Использованные в работе [9] критерии отбора «конформационно-стабильных» пептидов нуждаются в дополнительной аргументации в смысле выбора численных значений критериев. Один путь уточнения численных критериев отбора конформационно-стабильных пептидов может основываться на оценке статистики распределения по выбранным параметрам. Другой подход, реализуемый в настоящей работе, сводится к введению целого ряда показателей конформационной стабильности (или, что удобнее, лабильности) пептидных фрагментов, носящих более общий характер, для исследования феномена встречаемости «конформационно-стабильных» пептидов в структурах белков. Показатели такого рода были введены в настоящей работе и применены к анализу представительной выборки белков из PDB.

МАТЕРИАЛЫ И МЕТОДЫ

Выборка данных. Структуры белков были взяты из банка данных белков PDB [10], использован релиз 2017 г. С использованием ресурса VAST [11] из PDB были выбраны файлы с разрешением не хуже 2,0 Å и с попарной идентичностью последовательностей, не превышающей 50%. В результате была сформирована выборка из структур 25121 белковых цепей. Для каждой из цепей для аминокислотных остатков с известными координатами каждого неводородного атома главной цепи были рассчитаны значения углов φ и ψ, и проведено разбиение на тетрапептиды. В анализ были включены только те тетрапептидные фрагменты, для которых известны все восемь двугранных углов.

Методы оценки конформационной лабильности и стабильности пептидных фрагментов на основе анализа множеств наблюдаемых конформаций пептидов с определенной последовательностью. Пусть P – m-пептид (m = 2, 3, 4, ...) с заданной аминокислотной последовательностью. Будем описывать произвольную конформацию m-пептида P посредством задания набора углов φ и ψ, упорядоченного в виде вектора $\vec{p}_k \in R^{2m}$ с размерностью 2m, $\vec{p}_k = (p_1^k, \dots, p_i^k, \dots, p_{2m}^k)$, $p_i \in R$, где компоненты p_i с нечетными значениями индекса i соответствуют углам φ, а компоненты p_i с четными значениями i – углам ψ.

Будем считать, что в векторном пространстве R^{2m} определено скалярное произведение, т.е. для произвольных $\vec{p}_1, \vec{p}_2 \in R^{2m}$ определен скаляр $\langle \vec{p}_1, \vec{p}_2 \rangle = \sum_{i=1}^{2m} p_i^1 p_i^2$. Соответственно,

для произвольной пары векторов $\vec{p}_1, \vec{p}_2 \in R^{2m}$ определены *расстояние* $\text{dist}(\vec{p}_1, \vec{p}_2) = \sqrt{\langle \vec{p}_1 - \vec{p}_2, \vec{p}_1 - \vec{p}_2 \rangle}$ и *угол* $\text{ang}(\vec{p}_1, \vec{p}_2) = \arccos\left(\frac{\langle \vec{p}_1, \vec{p}_2 \rangle}{\sqrt{\langle \vec{p}_1, \vec{p}_1 \rangle \langle \vec{p}_2, \vec{p}_2 \rangle}}\right)$.

Для удобства последующего изложения введем несколько дополнительных обозначений. Пусть задано конечное множество чисел $A = \{a_1, a_2, \dots, a_i, \dots, a_n\}$, $a_i \in R$. Назовем $\hat{\phi}(x)$ оператором формирования эмпирической функции распределения чисел во множестве A, $\hat{\phi}(x)A = \sup\{B \subseteq A \mid \forall a \in B : a \leq x\} / |A|$, $x \in R$. Для краткости $\hat{\phi}(x)A$ будем также записывать как $\hat{\phi}A$.

Назовем \hat{z} оператором формирования множества значений множества A (например, $A \in R$), $\hat{z}A = B \subseteq A \mid \forall a \in A : a \in B, \forall a, b \in B : a \neq b, \hat{v}^+$ – оператором упорядочения множества по возрастанию, $\hat{v}^+A = (a_{I(1)}, a_{I(2)}, \dots \mid a_{I(1)} \leq a_{I(2)} \leq \dots \leq a_{I(n)}, \forall a_i \in A \exists j \in N[1..n] : i = I(j))$, где $I(j) : N[1..n] \rightarrow N[1..n]$ – подходящая функция перенумерации элементов (эномератор) A. Тогда j-й элемент упорядоченного множества \hat{v}^+A обозначим $\hat{v}^+(j)A = a_{I(j)}$.

Тогда, определим $\hat{\mu}$ – оператор вычисления математического ожидания значения $x \in A$ по эмпирической функции распределения $\hat{\phi}A$ как

$$\hat{\mu}\hat{\phi}A = \frac{1}{|\hat{z}A|} \sum_{j=1}^{|\hat{z}A|} x_j (\hat{\phi}(x_j)A - \hat{\phi}(x_{j-1})A), \text{ где } x_j = \hat{v}^+(j)\hat{z}A, \text{ а произвольное } x_0 < \inf(A), x_0 \in R. \text{ Оператор } \hat{\sigma}, \text{ вычисляющий стандартное отклонение } x \in A \text{ по } \hat{\phi}A, \text{ определим как } \hat{\sigma}\hat{\phi}A = \sqrt{\frac{1}{|\hat{z}A|} \sum_{j=1}^{|\hat{z}A|} (x_j - \hat{\mu}\hat{\phi}A)^2 (\hat{\phi}(x_{j-1})A - \hat{\phi}(x_j)A)}.$$

Пусть для m-пептида P с фиксированной аминокислотной последовательностью задано n_1 конформаций, каждой из которых соответствует вектор $\vec{p}_k \in R^{2m}$, $k = 1, \dots, n_1$. Назовем множество векторов $C(P) = \{\vec{p}_k\}$, $k = 1, \dots, n_1$, *множеством наблюдаемых конформаций* пептида P. Пусть $A_1 \subset R$ – множество попарных расстояний между элементами множества $C(P)$: $A_1(C(P)) = \{\text{dist}(a, b) \mid a, b \in C(P), a \neq b\}$, а множество $A_2 \subset R$ – множество попарных значений углов между элементами множества $C(P)$: $A_2(C(P)) = \{\text{ang}(a, b) \mid a, b \in C(P), a \neq b\}$.

Пусть задан алфавит $B = \{b_1, \dots, b_{|B|}\}$ для описания вторичной структуры отдельной аминокислоты. Таковым может являться, например, алфавит $B = \{\alpha, \beta, C\}$, где α соответствует α -спирали, β – β -структуре, C – неструктурированным фрагментам (coil). Определим функцию $\Phi : R^2 \rightarrow B$, ставящую в соответствие значениям пары углов φ и ψ символ из алфавита B (такая функция может быть определена, например, в соответствии со стандартом DSSP).

Для заданного пептида P множеству наблюдаемых конформаций $C(P) = \{\vec{p}_k\}$ однозначно соответствует множество литер вторичной структуры $L(C(P)) = \{\Phi(p_i^k, p_{i+1}^k) \mid i = 2j + 1, j = 0, \dots, m - 1, k = 1, \dots, n_1\}$. Тогда для произвольной литеры $b \in B$ определим частоту

встречаемости v_b как $v_b(L(C(P))) = |\{a \in L(C(P)) | a = b\}|/|L(C(P))|$. Для алфавита $B = \{\alpha', \beta', \gamma\}$, таким образом, для каждого пептида P определены частоты встречаемости соответствующих литер $v_\alpha, v_\beta, v_\gamma$.

В настоящей работе исследуется конформационная стабильность пептидных фрагментов. Практически, однако, на основе множеств векторов $C(P) = \{\vec{p}_k\}$ для m -пептида P с заданной аминокислотной последовательностью гораздо удобнее оценивать конформационную лабильность пептидов. Тогда после упорядочения рассматриваемых m -пептидных фрагментов в соответствии с числовыми значениями оценок конформационной лабильности верхняя часть такого списка будет соответствовать более «конформационно-стабильным» m -пептидам, а нижняя часть списка – более «конформационно-лабильным» m -пептидам.

С использованием введенных выше обозначений сформулируем несколько способов оценки конформационной лабильности пептидных фрагментов, представленных в структурах белков. В настоящей работе было использовано пять оценок конформационной лабильности произвольного m -пептида P :

$s_1(P) = \hat{\mu}\hat{\phi}A_1(C(P))$, т.е. среднее расстояние между элементами множества $C(P)$;

$s_2(P) = \hat{\sigma}\hat{\phi}A_1(C(P))$, т.е. стандартное отклонение расстояния между элементами множества $C(P)$;

$s_3(P) = \hat{\mu}\hat{\phi}A_2(C(P))$, т.е. средний угол между элементами множества $C(P)$;

$s_4(P) = \hat{\sigma}\hat{\phi}A_2(C(P))$, т.е. стандартное отклонение угла между элементами множества $C(P)$.

$s_5(P) = 1 - \max(v_\alpha(L(C(P))), v_\beta(L(C(P))), v_\gamma(L(C(P))))$, т.е. единица минус максимальная из частот $v_\alpha, v_\beta, v_\gamma$ для пептида P .

В работе также использовали *композиционные оценки конформационной лабильности m -пептидов*, т.е. оценки, представляющие собой композиции сформулированных выше оценок s_1 – s_5 . Рассмотрим пример построения композиции из двух оценок x, y . Пусть между оценками x, y существуют корреляции, описываемые функциями $y = f(x)$ и $x = f^{-1}(y)$. Конкретный вид функций $f(x)$ и $f^{-1}(y)$ устанавливается в результате проведения соответствующих регрессионных анализов. Тогда композиции оценок x и y вводятся посредством корреляционных преобразований, т.е. как $s_{xy} = x + f^{-1}(y)$ или как $s_{yx} = y + f(x)$. Подобным образом получаются оценки s_{125}, s_{345} и др.

Способ оценки конформационной лабильности и стабильности пептидных фрагментов на основе денатурирующей молекулярной динамики. Процедура денатурирующей молекулярной динамики (МД), в варианте, подробно описанном в работе [12], может быть применена для оценки стабильности конформации при нагреве системы. Для проведения МД был использован программный пакет ESMMS [13] с силовым полем UFF [14]. Координаты атомов водорода были рассчитаны исходя из стандартной геометрии. «Денатурирующая» молекулярная динамика была проведена с использованием модифицированного алгоритма скачка [15] в сочетании с термостатом (400, 600 и 1000 К), без модели растворителя (с целью увеличения денатурирующего эффекта) и без явного потенциала для описания водородных связей.

Набор декартовых координат исходной конформации исследуемого пептида P выбирали на основании некоторой медианной конформации, вычисленной на основании значений углов ϕ и ψ . В терминах введенных ранее обозначений пептиду P соответствует множество наблюдаемых конформаций $C(P) = \{\vec{p}_k\}, \vec{p}_k \in R^{2m}, k = 1, \dots, n_1$. Медианной конформацией пептида P будем называть такой вектор $\vec{a} \in C(P)$, для которого сумма расстояний до остальных элементов $C(P)$ минимальна, т.е. $\vec{a} = \{\vec{p}_k \in C(P) | \sum_{j=1 \dots n_1} \text{dist}(\vec{p}_k, \vec{p}_j) \rightarrow \min\}$.

На основании медианной конформации \vec{a} из PDB брали соответствующую цепь и декартовы координаты, соответствующие \vec{a} , использовали как исходные координаты моделируемого пептида. Исходные значения скоростей назначали в соответствии с распределением Максвелла для заданной температуры. Для каждого пептидного фрагмента вычисляли траектории продолжительностью 100 пс.

Конформационную лабильность исследуемого пептида характеризовали посредством сравнения конформации пептида в данный момент времени симуляции и схожей конформацией ($t = 0$). Различия в конформации оценивали посредством среднеквадратичного отклонения атомов главной цепи («*rmsd*») и коэффициента корреляции «*rmap*» между значениями элементов матриц попарных расстояний. Показатель *rmap* характеризует схожесть структуры в заданный момент времени с начальной, исходной конформацией пептида (с точностью до трансляции и ротации). В результате МД-симуляции для каждого пептида были получены кривые *rmsd(t)* и *rmap(t)*.

Таблица 1. Корреляционный анализ оценок конформационной лабильности $s_1 - s_5$.

Оценка	s_1	s_2	s_3	s_4	s_5
s_1	1,00	0,67	0,93	0,72	0,63
s_2		1,00	0,75	0,88	0,17
s_3			1,00	0,70	0,62
s_4				1,00	0,20
s_5					1,00

Примечание. В таблице приведены значения коэффициента корреляции. Матрица корреляций симметрична, поэтому приведены значения только для $n(n + 1)/2$ независимых элементов матрицы.

В качестве интегральных мер конформационной лабильности/стабильности, вычисляемых на основании данных МД-траекторий пептидов, были использованы величины, которые мы обо-

значили как AUC_{rms} и AUC_{map} . $AUC_{rms} = \int_{t=0}^{t_{MD}} rms(t)$

представляет собой площадь под кривой $rmsd(t)$ и характеризует степень отклонения атомов главной цепи пептидного фрагмента от исходной конформации за время моделирования t_{MD} (100 пс), т.е. конформационную лабильность пептидного фрагмента. Соответственно,

$AUC_{map} = \int_{t=0}^{t_{MD}} rmap(t)$ – площадь под кривой

$rmap(t)$ и, наоборот, характеризует стабильность конформации пептида (более высокие значения соответствуют большей стабильности). Для каждого пептида проводили вычисление десяти МД-траекторий с различными значениями начальных скоростей, на основании которых вычисляли средние значения AUC_{rms} и AUC_{map} для данного пептида и стандартные отклонения этих величин (σAUC_{rms} и σAUC_{map} соответственно).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Вычисление значений оценок конформационной лабильности $s_1 - s_5$ для 135000 тетрапептидов, которые были найдены в PDB (при комбинаторно возможном числе тетрапептидов $1,6 \cdot 10^5$), показало значительную корреляцию между этими оценками (табл. 1, рис. 1). Наиболее выражены линейные корреляции между оценками s_1 и s_3 , s_2 и s_4 . Очевидны выраженные степенные корреляции между оценками s_1 и s_2 , s_3 и s_4 (т.е. между средним значением величины и ее стандартным отклонением), описываемые

выражениями вида $y = A\sqrt{x}$. Существование таких корреляций между средним значением величины (μ) и стандартным отклонением величины (σ) вполне очевидно как следствие асимптотической аппроксимации гипергеометрического распределения [16], при котором $\sigma \approx \sqrt{\mu}$.

Приведенные выше результаты корреляционного анализа предложенных в настоящей работе оценок конформационной лабильности $s_1 - s_5$ показали, что оценки s_1 , s_2 и s_5 наименее скоррелированы друг с другом (см. табл. 1) и, с известными оговорками, могут считаться независимыми. Соответственно, для дальнейшего анализа использовались преимущественно именно эти три оценки, наряду с композитной оценкой s_{125} (см. раздел «Материалы и методы»), полученной на основании регрессионных уравнений (примеры таковых приведены на рис. 1).

При сравнении конформаций молекул, в том числе белков, обычно используется такая мера схожести конформаций двух фрагментов белковых структур как среднеквадратичное отклонение (root mean square deviation, RMSD) [17]. В случае белковых структур показатель $RMSD$ рассчитывается, как правило, на основании декартовых координат всех атомов (за исключением атомов водорода) и/или только координат атомов главной цепи. В то же время при анализе структур белковых молекул также используются так называемые «внутренние» или «естественные» координаты, т.е. наборы длин связей, валентных и двугранных углов, которые однозначно пересчитываются в декартовы координаты и наоборот [18].

В случае молекул белков физически обоснованной и повсеместно применяемой моделью внутренних координат является использование двугранных углов ϕ и ψ , на основе значений которых строятся карты Рамачандрана [19].

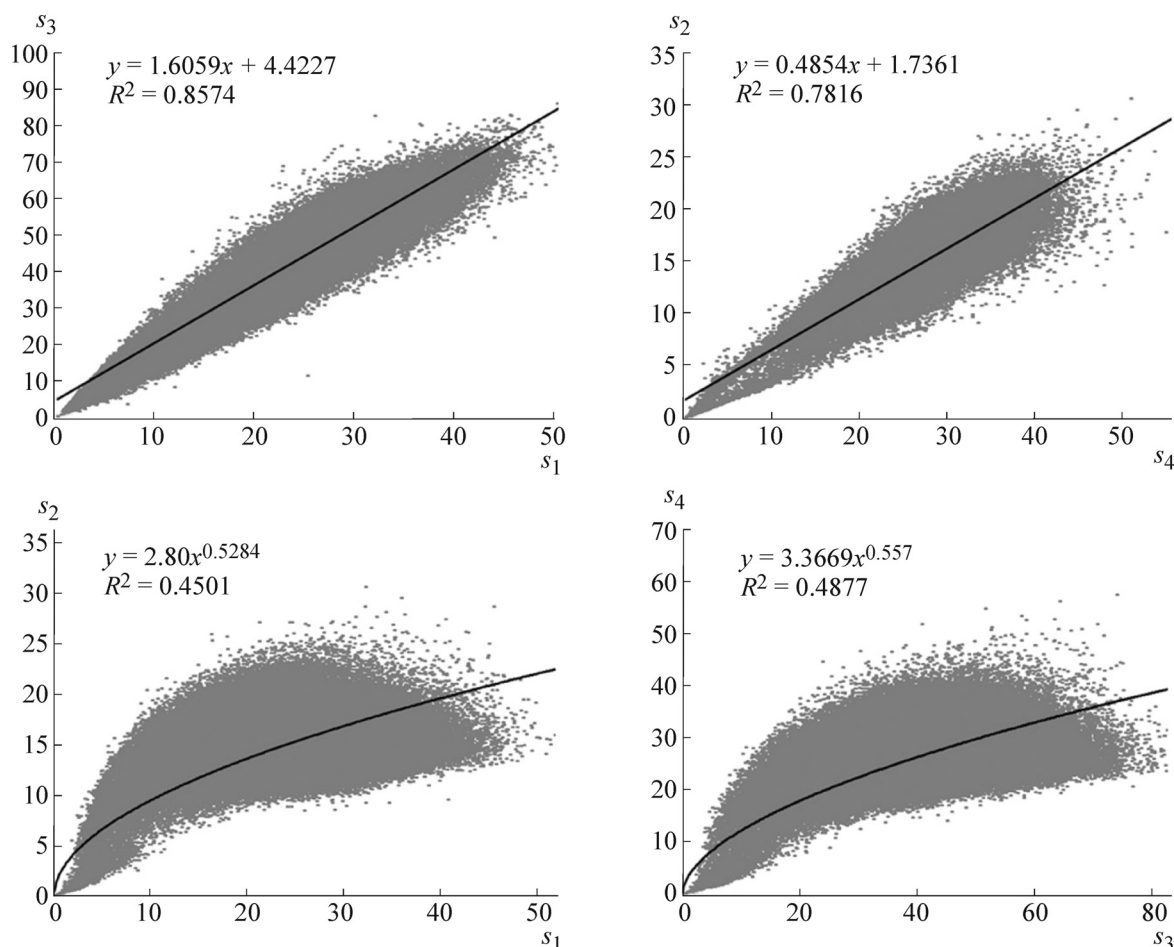


Рис. 1. Примеры корреляций между оценками конформационной лабильности. Каждая точка на диаграмме соответствует тетрапептиду с определенной аминокислотной последовательностью. Приведены значения квадрата коэффициента корреляции (R^2) и уравнения, полученные в результате проведения регрессионного анализа.

Принимая во внимание высокую степень планарности плоскости пептидной связи (значения угла ω лежат в достаточно узком диапазоне, $180^\circ \pm 5^\circ$), использование только наборов углов ϕ и ψ в качестве внутренних координат пептидного фрагмента является приближением, вполне допустимым для целей настоящей работы. Предлагаемая нами оценка конформационной лабильности s_1 основана, таким образом, на вычислении *RMSD* (только не на наборах декартовых координат, а на наборах углов ϕ и ψ). Другие меры конформационной лабильности (s_2 – s_4) основаны на стандартных статистических подходах [20] и представляют собой средние величины и их стандартные отклонения (см. раздел «Материалы и методы»). Оценка s_5 была использована в предыдущих работах [8,9] для оценки конформационной лабильности/стабильности *n*-пептидов.

Таким образом, корреляционный анализ указал на существование корреляций между

оценками конформационной лабильности s_1 – s_5 . Это позволяет предположить, что при упорядочении множества всех тетрапептидов в список по возрастанию значения любой из этих оценок на вершине списка могут оказываться одни и те же пептиды.

Возьмем, например, 10% (т.е. 13500) наименее конформационно-лабильных пептидов из такого списка, упорядоченного в соответствии с оценкой s_1 , и 10% пептидов из списка, упорядоченного в соответствии с s_2 . При пересечении этих двух множеств было установлено, что 3349 тетрапептидов (т.е. 2,09% ото всех тетрапептидов) входили в оба списка, т.е. являлись «конформационно-стабильными» сразу по двум оценкам, s_1 и s_2 . Аналогичным образом могут быть получены множества пептидов, которые «конформационно-стабильны» в соответствии с другими оценками конформационной лабильности (s_3 , s_4 , s_5).

Таблица 2. Примеры наиболее конформационно-стабильных (или наименее конформационно-лабильных) тетрапептидных фрагментов

Пептид	n	s_1	s_2	s_3	s_4	s_5	В.С.
WEYC	5	21,067	22,34	31,86	35,24	0	Hhhh
MSWV	6	28,401	23,22	45,02	38,61	0	Hhhh
LYYC	7	11,726	10,34	14,26	15,48	0	Bbbb
MAVQ	7	1,696	0,518	4,985	1,562	0	Hhhh
VYYC	15	3,98	1,458	4,448	1,851	0,014	Bbbb
CLAM	7	30,756	22,4	50,72	34,79	0,036	Hhhh
VMTI	7	6,761	2,632	7,47	3,019	0,036	Bbbb
CFIT	6	6,886	2,056	6,968	1,676	0,042	Bbbb
EMMS	6	4,514	1,984	13,000	6,164	0,042	Hhhh
VYYC	35	3,98	1,458	4,448	1,851	0,014	Bbbb
FHWG	9	24,577	16,62	27,86	19,56	0,028	Bbbb
IVCN	12	18,433	15,74	22,49	19,77	0,042	Bbbb
TYYC	23	7,904	6,28	8,737	8,529	0,054	Bbbb
YIYV	14	7,411	2,40	8,44	2,67	0,054	Bbbb
DYYC	13	4,792	2,243	5,369	2,797	0,058	Bbbb
IMIT	12	14,908	11,14	18,53	15,67	0,063	Bbbb
MPTF	12	8,695	8,204	11,36	11,64	0,063	Bbbb

Примечание. Представлены пептиды с наименьшими значениями оценки s_5 , упорядоченные по s_5 ; n – число раз, которое пептидный фрагмент встретился в исследованной выборке белков. В.С. – вторичная структура (h – α -спираль, b – β -структура).

При пересечении таких множеств тетрапептидов, полученных с использованием каждой из пяти оценок s_1 – s_5 , был получен список из 1034 тетрапептидов, каждый из которых входил в 10% наименее конформационно-лабильных пептидов при использовании любой из оценок s_1 – s_5 . В табл. 2 приведены примеры некоторых из этих наименее конформационно-лабильных (или наиболее конформационно-стабильных) тетрапептидных фрагментов, полученных с использованием оценок s_1 – s_5 .

Приведенные в табл. 2 примеры показывают, что конформационно-стабильные (конформационно-нелабильные) пептиды, в соответствии с выводами предыдущих работ, могут соответствовать α -спиральным конформациям («hhhh»); некоторые соответствуют β -структурным конформациям («bbbb»). Представляет интерес рассмотреть, насколько предпочтителен каждый из этих типов вторичной структуры для «конформационно-стабильных» пептидов.

Для этого мы провели расчет эмпирических функций распределения значений оценок конформационной лабильности s_1 , s_5 и композитной оценки s_{125} для отдельных типов вторичной структуры (рис. 2–4).

Анализ эмпирических функций распределения оценок конформационной лабильности s_1 , s_5 , s_{125} для отдельных типов вторичной структуры показал, что при всех исследованных оценках наиболее стабильными являлись α -спиральные конформации. Данное отличие между эмпирическими функциями распределения для α -спиралей и других типов вторичной структуры было достоверно в соответствии с критерием Колмогорова–Смирнова.

Например, среднее значение оценки s_1 для α -спиральных конформаций составило $s_1 = 24 \pm 5$, а для β -структур и неструктурированных фрагментов (coil) – $s_1 = 27 \pm 4$, что является статистически достоверным отличием ($P < 0,001$). Аналогичные результаты были получены

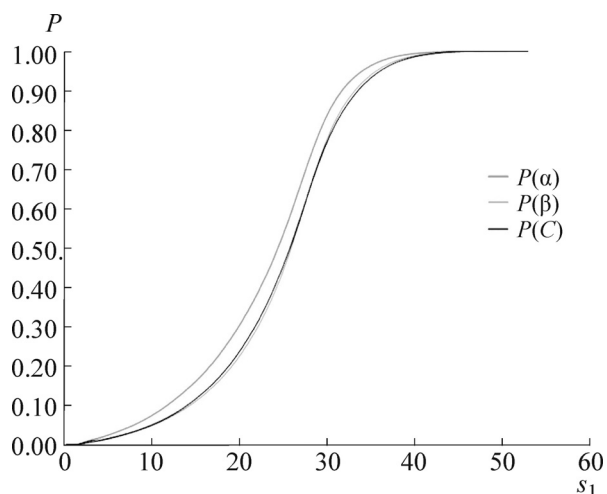


Рис. 2. Эмпирическая функция распределения значений оценки конформационной лабильности s_1 для отдельных типов вторичной структуры.

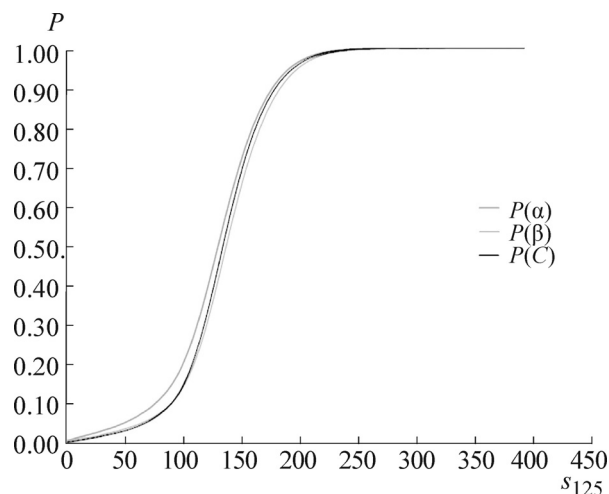


Рис. 4. Эмпирическая функция распределения значений композитной оценки конформационной лабильности s_{125} для отдельных типов вторичной структуры.

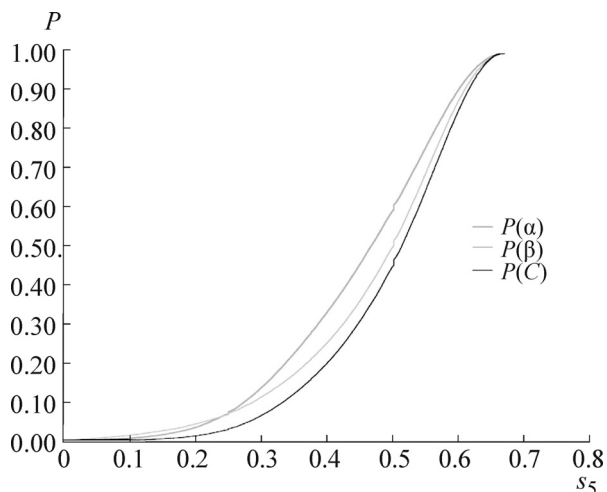


Рис. 3. Эмпирическая функция распределения значений оценки конформационной лабильности s_5 для отдельных типов вторичной структуры.

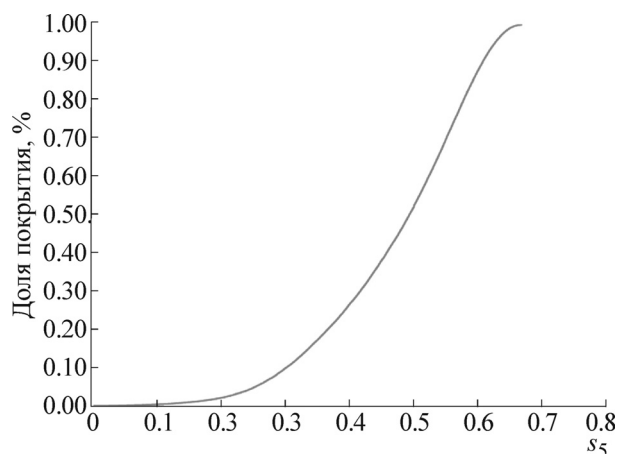


Рис. 5. Доля покрытия аминокислотных последовательностей белков тетрапептидами с заданным уровнем конформационной лабильности s_5 .

и для оценок s_5 (α -спирали – $0,45 \pm 0,06$, β -структуры – $0,48 \pm 0,05$) и s_{125} (α -спирали – 124 ± 12 , β -структуры – 135 ± 14). Иначе говоря, α -спиральные конформации менее конформационно-лабильны, т.е. более конформационно-стабильны, чем другие разновидности конформаций тетрапептидов. При этом, в соответствии с оценкой s_1 , конформационная лабильность β -структур и неструктурированных фрагментов в среднем одинакова. Оценка s_5 , основанная на анализе частот встречаемости типов вторичной структуры (см. раздел «Материалы и методы»), показывает несколько меньшую конформационную лабильность β -структур по

сравнению с неструктурированными фрагментами.

Мы провели анализ процента покрытия аминокислотных последовательностей тетрапептидами с заданным уровнем конформационной лабильности по оценке s_5 (рис. 5). Очевидно, что тетрапептиды с низкой конформационной лабильностью ($s_5 < 0,10$, т.е. более 90% наблюдаемых конформаций принадлежат к одному типу вторичной структуры) покрывают менее 1% (0,07%) от произвольной аминокислотной последовательности. Пептиды с несколько большей конформационной лабильностью ($s_5 < 0,20$) покрывают 1,0% произвольной аминокислотной последовательности, а

Таблица 3. Сравнение оценок конформационной лабильности для «пептидов-хамелеонов» из базы данных CHSEQ [21]

Множество пептидных фрагментов	s_1	s_2	s_3	s_4	s_5
Пептиды-хамелеоны	$25,67 \pm 6,67$	$15,25 \pm 2,54$	$46,34 \pm 11,22$	$27,98 \pm 4,52$	$0,47 \pm 0,11$
Не-хамелеоны	$21,91 \pm 8,29$	$14,67 \pm 3,97$	$38,88 \pm 14,43$	$26,49 \pm 7,26$	$0,45 \pm 0,13$
P	$< 10^{-250}$	$4 \cdot 10^{-238}$	$< 10^{-250}$	$< 10^{-250}$	$3 \cdot 10^{-219}$
Конформационно стабильные ($s_5 < 0,35$)	$20,46 \pm 6,62$	$13,90 \pm 3,75$	$39,14 \pm 11,15$	$25,60 \pm 6,73$	$0,29 \pm 0,06$
P	$< 10^{-250}$	$< 10^{-250}$	$< 10^{-250}$	$< 10^{-250}$	$< 10^{-250}$

Примечание. *P* – статистическая достоверность по тесту Колмогорова–Смирнова.

пептиды с $s_5 < 0,35$ – уже 11% от длины последовательности. Таким образом, тетрапептиды, достаточно конформационно-стабильные (т.е. с $s_5 < 0,35$), могут составлять в среднем заметную долю произвольной аминокислотной последовательности.

Следует отметить, что определенность конформаций пептидного фрагмента растет с увеличением длины сегмента. Ранее было обнаружено, что только 29 из 8000 комбинаторно возможных последовательностей трипептидов обладают предпочтительными конформациями (т.е. 0,36%) [8]. Используя схожие критерии «конформационной стабильности», в настоящей работе было идентифицировано как относительно стабильные уже ~3300 из 160000 комбинаторно возможных последовательностей тетрапептидов (т.е. 2,1%). Возрастание доли «конформационно-стабильных» пептидов при увеличении длины пептидного фрагмента находится в согласии с тем, что определенность конформации пептидного фрагмента возрастает с увеличением длины фрагмента от трех до четырех остатков.

Интересно отметить, что так называемые «пептиды-хамелеоны» (сегменты полипептидной цепи, которые в одних белках входят в состав α -спирали, а в других – β -структуры при одной и той же последовательности) [21] могут рассматриваться как антипод конформационно-стабильным пептидам. Действительно, сравнительный анализ значений оценок конформационной лабильности s_5 – s_5 показал, что «пептиды-хамелеоны» характеризуются достоверно более высокими значениями оценок конформационной лабильности, чем все остальные пептиды, в том числе «конформационно-стабильные» ($s_5 < 0,35$, табл. 3).

Какова физическая природа существования таких «конформационно-стабильных» пепти-

дов? Можно предположить, что различия в конформационной стабильности/лабильности пептидов (т.е. в значениях оценок s_1 – s_5) обусловлены различиями в «гибкости» или подвижности атомов главной цепи, которые, в свою очередь, приводят к различиям в конформационной лабильности, вычисляемой на основании наблюдаемых множеств конформаций пептида.

Для тестирования этой гипотезы мы провели МД-моделирование отдельных пептидов. Методом денатурирующей МД были проведены расчеты показателей AUC_{rms} и AUC_{rmap} , характеризующих конформационную лабильность/стабильность пептидов на основании МД-траекторий. На ряде примеров показано, что для более «конформационно-стабильных» пептидов характерна более высокая устойчивость конформации по данным денатурирующей МД по сравнению с пептидами с более высокими оценками конформационной лабильности.

Для МД-моделирования были выбраны примеры «стабильных» тетрапептидов на основании результатов, полученных с помощью кластерного анализа конформаций 135000 тетрапептидов из списка комбинаторно-возможных тетрапептидов. При этом критериями отбора тетрапептидов являлась не максимально возможная частота встречаемости, а близость выбранного пептида к «центрам» полученных кластеров. Дополнительными критериями отбора «конформационно-стабильных» пептидов служили следующие пороговые значения оценок конформационной лабильности: $s_1 \leq 20$, $s_2 \leq 15$ и $s_5 \leq 0,35$. Данные критерии носят, конечно же, чисто эмпирический характер и были выбраны на основе экспертного анализа кривых, представленных на рис. 2–5. В результате были отобраны четыре репрезентативных «конформационно стабильных» тетрапептида, представ-

Таблица 4. Тетрапептиды, для которых было проведено моделирование посредством молекулярной динамики

Пептид	N	s_1	s_2	s_3	s_4	s_5	s_{125}	AUC_{rmsd}	σAUC_{rmsd}	AUC_{rmap}	σAUC_{rmap}
Более «конформационно-стабильные» тетрапептиды											
HEAA	25	19,59	14,90	29,07	20,48	0,26	47,6	0,11	0,02	0,096	0,017
MELI	21	20,00	12,86	31,87	21,13	0,34	49,7	0,13	0,04	0,098	0,019
NIQK	14	19,65	13,81	32,34	21,25	0,28	46,2	0,14	0,03	0,099	0,021
EAAV	126	15,54	14,34	24,52	22,89	0,33	48,7	0,13	0,02	0,092	0,024
Менее «конформационно-стабильные» тетрапептиды											
GGGG	137	50,64	12,85	89,57	22,44	0,55	101	0,21	0,06	0,093	0,021
EEAI	108	27,07	15,88	26,45	31,05	0,39	71	0,25	0,07	0,094	0,019
VVAV	126	24,04	15,91	36,16	30,18	0,40	69	0,92	0,19	0,085	0,022
AIKE	99	23,90	16,37	29,91	30,34	0,38	69	0,18	0,03	0,093	0,016

Примечание. AUC_{rmsd} (Å·нс) и AUC_{rmap} (нс) – меры конформационной лабильности, вычисленные на основании МД-траекторий пептидов; σAUC_{rmsd} и σAUC_{rmap} – стандартные отклонения (см. раздел «Материалы и методы»). Более «конформационно-стабильные» тетрапептиды были выбраны с соответствии с критериями $s_1 \leq 20$, $s_2 \leq 15$ и $s_5 \leq 0,35$.

Таблица 5. Различия в средних значениях оценок конформационной лабильности/стабильности между группой из четырех более «конформационно-стабильных» ($s_1 \leq 20$, $s_2 \leq 15$ и $s_5 \leq 0,35$) и группой четырех менее «конформационно-стабильных» тетрапептидов

Оценка	Более «конформационно-стабильные» тетрапептиды	Менее «конформационно-стабильные» тетрапептиды	P
s_1	18,70 ± 2,11	31,41 ± 12,90	0,011
s_2	13,98 ± 0,87	15,25 ± 1,62	0,015
s_3	29,45 ± 3,59	45,52 ± 29,64	>0,1
s_4	21,44 ± 1,03	28,50 ± 4,06	0,018
s_5	0,30 ± 0,04	0,43 ± 0,08	0,021
s_{125}	48,12 ± 1,52	77,88 ± 15,71	0,016
AUC_{rmsd}	0,13 ± 0,03	0,39 ± 0,35	0,012
AUC_{rmap}	0,10 ± 0,02	0,09 ± 0,02	0,053

Примечание. P – статистическая значимость (по критерию Колмогорова–Смирнова).

ляющих четыре соответствующих кластера: тетрапептиды HEAA, MELI, NIQK и EAAV.

В качестве контрольных «конформационно-лабильных» пептидов были использованы пептиды GGGG, EEAI, AIKE и VVAV. Заметим, что пептиды EEAI и AIKE являются фрагментами «пептида-хамелеона» AKEEAIKE, представленного в базе данных ChSeq [21]. Пептид AKEEAIKE принимает α -спиральную конфор-

мацию в одном наборе структур (например, 2jws, 2kdl, 2lhc, 2lhg) и β -структурную конформацию в других структурах (например, 2lhd, 2kdm, 2jwu, 2lhe). Результаты МД-моделирования денатурации этих пептидов суммированы в табл. 4; в табл. 5 приведен анализ статистических различий между исследованными мерами конформационной лабильности. Примеры

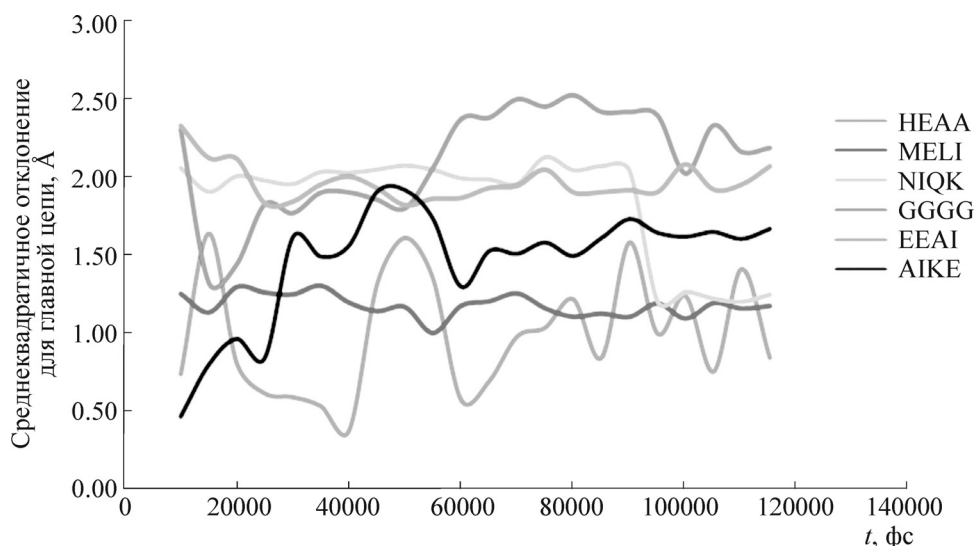


Рис. 6. Зависимость среднего квадратичного отклонения координат атомов остова полипептидной цепи от времени симуляции для конформационно стабильных пептидов HEAA, MELI, NIQK и «нестабильных» пептидов GGGG, EEAI и AIKE.

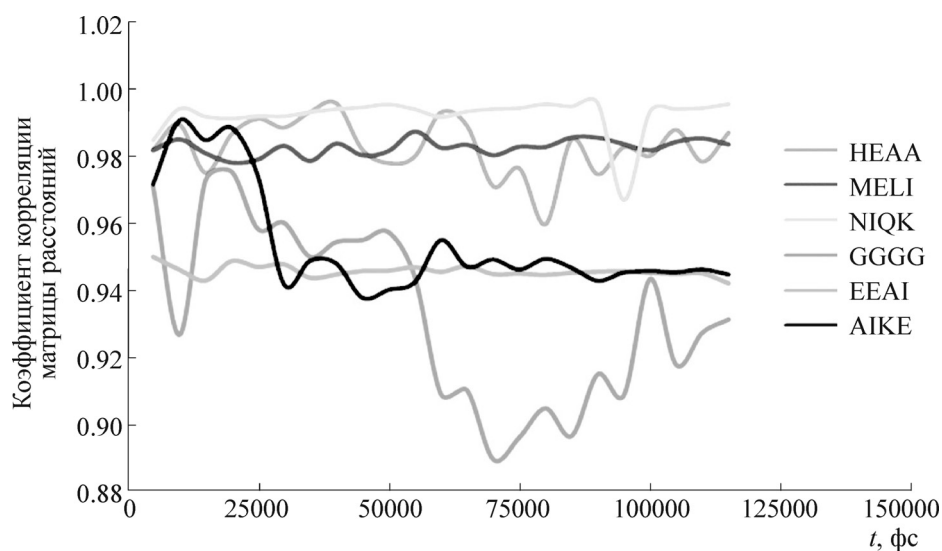


Рис. 7. Зависимость коэффициента корреляции матриц расстояний от времени симуляции конформационно-стабильных пептидов HEAA, MELI, NIQK и «нестабильных» пептидов GGGG, EEAI и AIKE.

отдельных кривых $rmsd(t)$ и $rmap(t)$ приведены на рис. 6–9.

МД-моделирование всех пептидов было выполнено в течение 100 пс при мягких денатурирующих условиях (300 К *in vacuo*, т.е. без модели растворителя – без задания координат молекул воды и без использования соответствующего термина межмолекулярного потенциала, который стабилизирует молекулу пептида). Для сравнительной оценки конформационной стабильности были рассчитаны среднее квадратичные отклонения для всех атомов главной цепи и коэффициент корреляции между значениями

элементов матриц попарных расстояний исходной структуры и структуры в заданный момент времени симуляции (1 пс, 2 пс, ..., 100 пс).

Анализ среднее квадратичных отклонений координат атомов главной цепи показал (рис. 6), что конформационно-стабильные пептиды HEAA, MELI, NIQK ко времени симуляции ~100 пс попали в соответствующие «бассейны» стабильных конформаций ($RMSD \sim 1,2 \text{ \AA}$), в то время как «нестабильные» пептиды характеризовались существенно большими значениями $RMSD$ (GGGG: 2,0...2,5 Å, EEAI: 1,8...2,0 Å и AIKE: 1,5...1,7 Å). Заметим, что на рис. 6 кривые

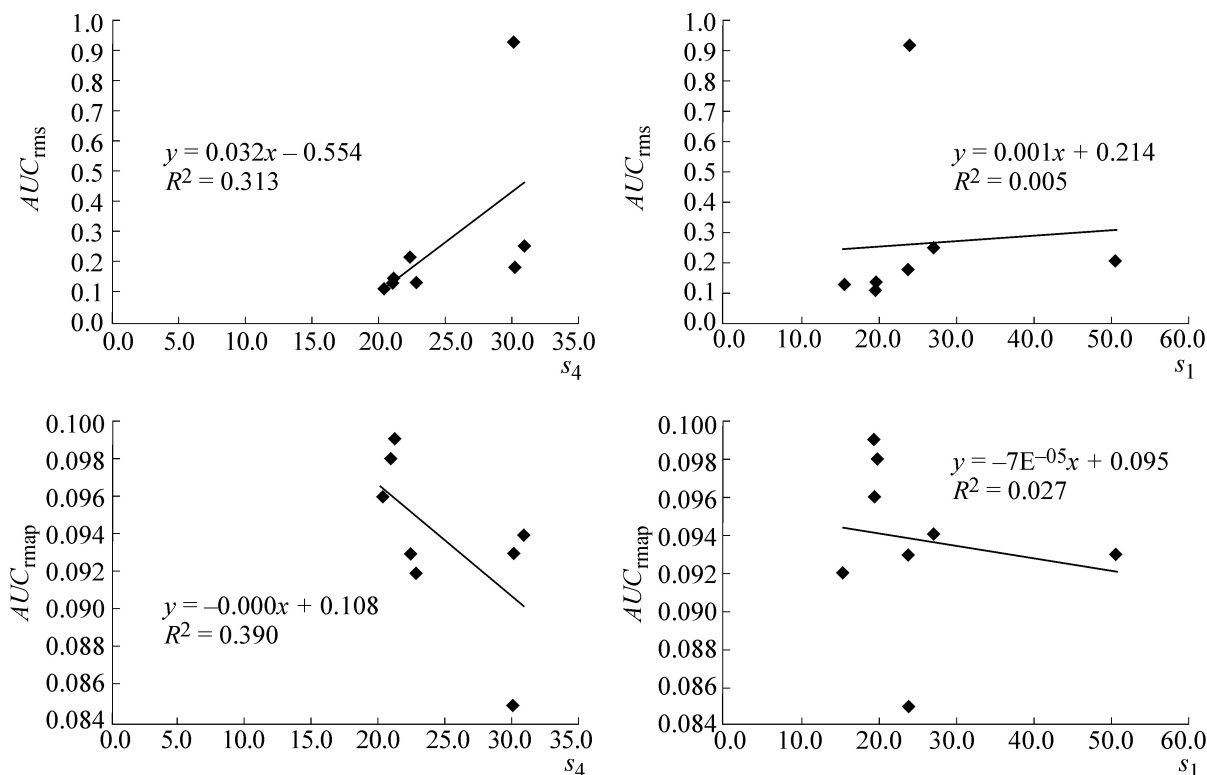


Рис. 8. Примеры корреляций между оценками конформационной лабильности, полученными на основании анализа множеств наблюдаемых конформаций тетрапептидов, и оценками конформационной лабильности/стабильности, вычисленными на основании МД-траекторий для восьми пептидов.

$rmsd(t)$ для пептидов EAAV и VVAV не приведены с целью повышения наглядности иллюстрации (средние значения $rmsd$ для этих пептидов существенно выше).

Коэффициент корреляции между значениями элементов матриц попарных расстояний характеризует схожесть структуры в заданный момент времени с начальной, исходной конформацией пептида (с точностью до трансляции и ротации). Вычисления МД-траекторий показали, что структуры конформационно-стабильных пептидов HEAA, MELI, NIQK характеризовались высокими значениями коэффициентов корреляции вдоль траектории (0,96...1,00), что указывает на их МД-стабильность. В то же время конформационно-нестабильный пептид GGGG характеризовался резкими изменениями конформации (что отражается скачками коэффициента корреляции в гораздо более широком диапазоне значений – от 0,88 до 0,98) в течение всего времени симуляции. После 25 пс симуляции «нестабильные» пептиды EEAI и AIKE можно характеризовать значительно более низкими значениями коэффициента корреляции (0,94), чем «стабильные» пептиды (см. рис. 7).

В целом более «конформационно-стабильные» тетрапептиды HEAA, MELI, NIQK и

EAAV отличались от менее «конформационно-стабильных» тетрапептидов GGGG, EEAI, AIKE и VVAV меньшими значениями оценок конформационной лабильности (включая показатель AUC_{rmsd} , полученный в результате денатурирующей МД, см. табл. 5). За исключением оценки s_3 , различия в значениях оценок были статистически достоверными. Показатель AUC_{rmap} (который, наоборот, оценивает конформационную стабильность пептидов) был достоверно выше у более «конформационно-стабильных» тетрапептидов.

Представляет интерес оценить, насколько оценки конформационной лабильности, полученные на основании анализа множеств наблюдаемых конформаций пептидов, коррелируют с оценками конформационной лабильности/стабильности, вычисленными на основании МД-траекторий. Результаты корреляционного анализа суммированы в табл. 6.

Данные в табл. 6 показывают, что, как и ожидалось, корреляции между оценками s_1-s_5 и оценкой конформационной лабильности AUC_{rms} были прямо пропорциональны, а между оценками s_1-s_5 и оценкой конформационной стабильности AUC_{rmap} – обратно пропорциональны. В целом значения коэффициента кор-

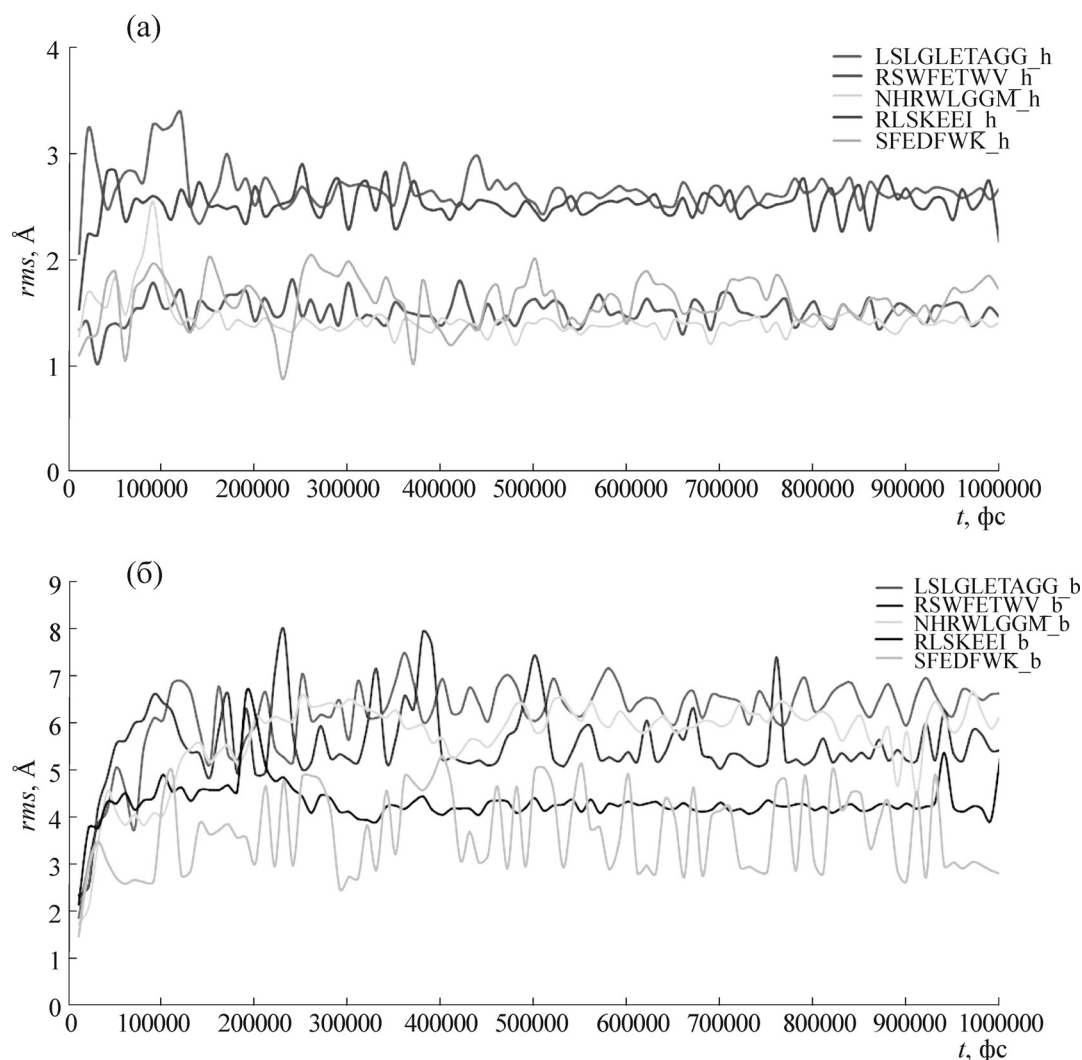


Рис. 9. МД-траектории примеров растворимых пептидов: (а) – начальная конформация – α -спираль; (б) – начальная конформация – β -структура.

реляции были достаточно невысокими (что вполне понятно, так как оценки конформационной лабильности на основе анализа структур и в соответствии с траекториями МД основаны на разных принципах).

Наиболее выражены были корреляции между оценками, полученными в результате МД-моделирования и оценками s_2 и s_4 . Напомним, что «дисперсионные» оценки s_2 и s_4 характеризуют стандартное отклонение расстояния между элементами множества $C(P)$ и стандартное отклонение угла между элементами множества $C(P)$ соответственно. Иначе говоря, с результатами МД-моделирования в большей степени коррелируют не оценки средних расстояний ($s_1(P)$) или средних углов ($s_2(P)$) между элементами множества $C(P)$ (т.е. наблюдаемыми в PDB конформациями), а оценки s_2 и s_4 , харак-

теризующие разброс значений расстояний и углов соответственно. Примеры обсуждаемых корреляций приведены на рис. 8.

Таким образом, совместный анализ данных денатурирующей молекулярной динамики и оценок конформационной лабильности s_1 – s_5 указывает на тенденцию конформационно-стабильных пептидов «удерживать» некоторую «стабильную» конформацию. В течение всего времени симуляции МД конформации исследованных «конформационно-стабильных» пептидов не претерпевали резких изменений и характеризовались высокими значениями коэффициентов корреляции по сравнению с начальной конформацией.

Молекулярно-механическое моделирование структур реальных пептидов с целью оценки их конформационной стабильности. С использова-

Таблица 6. Корреляционный анализ оценок конформационной лабильности s_1-s_{125} и оценок AUC_{rmsd} , AUC_{rmap} (нс), вычисленных на основании МД-траекторий пептидов

Оценка	AUC_{rmsd}	AUC_{rmap}
s_1	0,07	-0,16
s_2	0,42	-0,51
s_3	0,03	-0,11
s_4	0,56	-0,62
s_5	0,27	-0,43
s_{125}	0,27	-0,42

Примечание. Приведены значения коэффициента корреляции; отрицательные значения соответствуют обратным корреляциям (т.е. $k < 0$ в регрессионной формуле $kx + b$).

Таблица 7. Растворимые пептиды и прогнозируемые значения оценок конформационной лабильности

Пептид	s_1	ds_1	s_2	ds_2	s_3	ds_3	s_4	ds_4	s_5	ds_5
RSWFETWV	26,2	10,9	11,2	5,0	48,0	15,6	22,0	7,1	0,49	0,13
SFEDFWK	22,8	7,7	14,6	2,4	43,0	12,5	26,3	3,2	0,47	0,11
RLSKEEI	22,7	4,0	14,4	1,4	45,7	6,6	28,1	1,4	0,42	0,07
LSLGGLETAGG	32,4	6,7	14,0	1,5	60,5	9,0	25,3	2,9	0,52	0,08
NHRWLGGM	36,7	4,3	13,5	1,4	66,4	8,5	23,2	1,6	0,58	0,05

Примечание. Оценки в колонках s_1-s_5 рассчитывали как среднее от оценок s_1-s_5 тетрапептидов, входящих в состав пептида. Величины ds_1-ds_5 – стандартные отклонения значений оценок s_1-s_5 .

нием данных о конформациях тетрапептидов, разработанных оценок конформационной лабильности s_1-s_5 и МД были исследованы примеры известных растворимых пептидов с целью оценки их конформационной стабильности. Для моделируемых пептидов были рассчитаны значения оценок конформационной лабильности как среднее от оценок s_1-s_5 тетрапептидов, входящих в состав каждого пептида. Результаты суммированы в табл. 7.

МД-моделирование проводили по описанной ранее методологии денатурирующей МД при $T = 350$ К, длина траектории – 1 нс. Поскольку пространственные структуры изучаемых пептидов неизвестны, каждый из пептидов моделировался исходя из двух стартовых конформаций (α -спираль и β -стрэнд). Примеры траекторий $rmsd(t)$ приведены на рис. 9.

Был выполнен корреляционный анализ между показателями конформационной лабильности, рассчитанными на основании тетрапептидного состава, и показателями, полученными в результате МД-моделирования пяти пептидов в двух стартовых конформациях (табл. 8). Из

результатов анализа явствует, что показатели ds_1, ds_2, ds_3, ds_5 (т.е. стандартные отклонения оценок s_1, s_2, s_3, s_5) проявили достоверные корреляции со значениями стандартных отклонений МД-траекторий, σAUC_{rmsd} , причем вне зависимости от типа стартовой конформации.

Таким образом, расчеты тетрапептидного состава и МД моделирование указывают на пептиды RSWFETWV, SFEDFWK и RLSKEEI как наиболее конформационно-стабильные. Данные результаты позволяют планировать проведение соответствующих экспериментов по анализу конформаций рассмотренных пептидов.

ВЫВОДЫ

В работе предложены и апробированы несколько способов для оценки конформационной стабильности/лабильности пептидных фрагментов, входящих в состав структур белков. Предлагаемые нами формулы для оценки конформационной лабильности пептидных фрагментов путем статистического анализа структур позволили показать, что тетрапептидные фрагменты

Таблица 8. Результаты корреляционного анализа между показателями конформационной лабильности, рассчитанными на основании тетрапептидного состава и полученными в результате МД-моделирования пептидов в двух стартовых конформациях

Оценка	Начальная конформация – β -стрэнд		Начальная конформация – α -спираль	
	AUC_{rmsd}	σAUC_{rmsd}	AUC_{rmsd}	σAUC_{rmsd}
ds_1	0,10	0,72	-0,37	0,23
ds_2	0,07	0,62	-0,50	0,10
ds_3	-0,02	0,81	-0,61	0,27
ds_5	-0,22	0,85	-0,29	0,51

существенно различаются по оценке конформационной стабильности/лабильности. В результате выделены подгруппы более «конформационно-стабильных пептидов». Последние характеризуются преимущественно α -спиральной конформацией. Сформирована база данных (список) из 1034 тетрапептидов, характеризующихся наименьшей конформационной лабильностью в соответствии со всеми предложенными оценками. В результате проведенных расчетов методом денатурирующей МД получены качественные и количественные оценки конформационной лабильности или «гибкости» некоторых тетрапептидов, согласующиеся с оценками, полученными на основании анализа конформационных наборов пептидов, наблюдаемых в экспериментально определенных структурах белков. Пептиды, характеризующиеся достаточной степенью конформационной стабильности, составляют существенную долю комбинаторно возможных пептидов (тетрапептидов). Их роль в формировании структуры белка может быть достаточно велика, так как они составляют в среднем около 10% аминокислотной последовательности. На примерах реальных растворимых пептидов показана возможность оценки конформационной стабильности произвольной аминокислотной последовательности на основе полученных критериев конформационной лабильности тетрапептидов.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проекты №№ 16-54-00219-Бел и 18-54-00037-Бел) и Белорусского республиканского фонда фундаментальных исследований (проект № Б18Р-268).

СПИСОК ЛИТЕРАТУРЫ

- N. V. Kalmankar, C. Ramakrishnan, and P. Balaram, *Proteins* **82**, 1101 (2014).
- A. Figureau, M. A. Soto, and J. Toha, *Prot. Engineer.* **16** (2), 103 (2003).
- P. K. Vlasov, A. V. Vlasova, V. G. Tumanyan, and N. G. Esipova, *Proteins* **61** (4), 763 (2005).
- C. Bystroff, K. T. Simons, K. F. Han, and D. Baker, *Curr. Opin. Biotechnol.* **7** (4), 417 (1996).
- C. Bystroff and D. Baker, *J. Mol. Biol.* **281**, 565 (1998).
- C. G. Hunter and S. Subramaniam, *Proteins* **50** (4), 572 (2003).
- O. Sunder, I. Sommer, and T. Lengauer, *BMC Bioinformatics* **7**, 14 (2006).
- A. V. Balyanovskii and P. K. Vlasov, *Biophysics* **53** (4), 264 (2008).
- A. V. Balyanovskii, I. D. Volotovskiy, V. A. Namiot, et al., *Biophysics* **60** (3), 348 (2015).
- H. M. Berman, J. Westbrook, Z. Feng, et al., *Nucl. Acids Res.* **28**, 235 (2000).
- J. F. Gibrat, T. Madej, and S. H. Bryant, *Curr. Opin. Struct. Biol.* **6**, 377 (1996).
- I. Yu. Torshin, N. G. Esipova, and V. G. Tumanyan, *J. Biomol. Struct. Dyn.* **32** (2), 198 (2014).
- I. Y. Torshin, *Sci. World J.* **4**, 228 (2004).
- A. R. Rappé and W. A. Goddard, *J. Phys. Chem.* **95** (8), 3358 (1991).
- M. A. Cuendet and van W. F. Gunsteren, *J. Chem. Phys.* **127** (18), 184102 (2007).
- K. K. Talukdar and W. D. Lawing, *J. Acoust. Soc. Amer.* **89** (3), 1193 (1991).
- A Dictionary of Physics*, 6 ed. (Oxford University Press, 2009).
- J. Parsons, J. B. Holmes, J. M. Rojas, et al., *J. Comput. Chem.* **26**, 1063 (2005).
- G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *J. Mol. Biol.* **7**, 95 (1963).
- Quadratic deviation (Encyclopedia of Mathematics)*, Ed. by M. Hazewinkel (Springer, 2001).
- L. Wenlin, L. Kinch, A. Karplus, and N. Grishin, *Prot. Sci.e* **24**, 1075 (2015).

Conformational Stability/Lability of Peptide Fragments in the Sequence Context of Amino Acids

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

*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

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

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

Criteria for evaluating the conformational stability/lability of peptide fragments referred to fragments of protein structures are formulated. Using the criteria proposed, a statistical analysis of tetrapeptide fragments (their conformations and sequences) was performed in a sample of 25121 protein chain structures from a PDB protein databank. As a result of the analysis, it was shown that tetrapeptide fragments significantly differ in the degree of the conformational stability/lability from a proposed statistical criteria point of view. The results of tetrapeptide denaturing molecular dynamics simulations were used as an independent approach to estimate the stability/lability of peptide fragments. A correlation between the estimates of conformational lability, obtained on the basis of a statistical analysis of the ensembles of peptide conformations observed in experimentally determined protein structures, and estimates of conformational lability/stability, calculated on the basis of molecular dynamics trajectories, is demonstrated; Subgroups of more “conformationally stable peptides”, characterized mainly by α -helical conformation, were obtained. Consensus tetrapeptides characterized by the lowest conformational lability (the highest conformational stability) were determined using complex criteria. Peptides with enhanced conformational lability were described. Thus, among all combinatorially possible tetrapeptides, tetrapeptides, characterized by a certainty of conformational state play a significant role. The relationship between the degree of conformational certainty of the peptide and its involvement in the primary structure of the protein has been characterized. Meanwhile, the role of conformationally stable peptides in the formation of the protein structure was great enough, since they constitute, on average, about 10% of the amino acid sequence. Using real soluble peptides as examples, the possibility of assessing the conformational stability of any preset amino acid sequence based on the developed criteria of the conformational lability of tetrapeptides has been shown.

Keywords: conformationally stable/labile protein segments, local protein structure, conformational analysis, statistical analysis, molecular dynamics