



УДК 577.112.012:681.3

НЕСТАТИСТИЧЕСКИЙ ПОДХОД К ПРЕДСКАЗАНИЮ РЕГУЛЯРНЫХ СТРУКТУР В БЕЛКАХ НА ПРИМЕРЕ α -СПИРАЛЕЙ

© 2002 г. Г. Т. Килосанидзе, А. С. Куценко, Н. Г. Есипова, В. Г. Туманян[#]*Институт молекулярной биологии им. В.А. Энгельгардта РАН, 119991, Москва, ГСП-1, ул. Вавилова, 32*

Поступила в редакцию 14.09.2001 г. Принята к печати 18.01.2002 г.

Предложен новый подход для анализа стандартных конформаций в белках на основе метода молекулярной механики. Метод использует информацию только о последовательности аминокислотных остатков. Для моделирования α -спиральной конформации применена программа молекулярной механики ISM. Для восьми белков из банка данных пространственных структур белков построены энергетические профили последовательностей в α -спиральной конформации, охватывающей всю полипептидную цепь. Участки каждого профиля, соответствующие минимумам энергии, отвечают α -спиральным участкам реальной пространственной структуры белка. Из 25 спиралей четко выявляются 24, то есть вероятность предсказания достаточно велика. Энергетические профили позволяют также выделять короткие участки, соответствующие 3/10-спиралям и те повороты, которые включают α -спиральную конформацию. В отличие от известных способов предсказания, которые являются статистическими, в рамках предложенного метода становится возможным выявить физические принципы формирования α -спиральной конформации.

Ключевые слова: белки, предсказание структур, вторичная структура, молекулярная механика.

ВВЕДЕНИЕ

Вторичная структура белка остается неисчерпаемой темой для исследований в области протеомики. Достигнут существенный прогресс в предсказании вторичной структуры как методами, использующими физико-химические подходы, так и, особенно в последнее время, статистическими методами [1–3]. Тем не менее до конца не ясными остаются физические основы формирования структур данного типа, что ограничивает возможности методов предсказания (см. [4]).

В большинстве случаев набор конформаций глобулярного белка можно представить как ансамбль известных стандартных конформаций, таких, как регулярные спиральные или иные структуры. Для решения вопроса о роли локальных и дальних взаимодействий в стабилизации стандартных наборов конформаций полипептидной цепи прежде всего необходимо определить, имеют ли данные конформации на том участке полипептидной цепи, где они находятся, меньшую конформационную энергию по сравнению с соседними или нет. Иными словами, если для данной конформации место ее нахождения в реальной структуре белка в цепи подчиняется балансу ближних и дальних взаимодействий, то оно должно соответствовать глобальному минимуму энергии всей макромолекулы. Выявление роли взаи-

модействий того или иного типа можно было бы осуществить путем расчета оптимальной энергии (при варьировании параметров боковых радикалов) для основных регулярных структур, используя современные усовершенствованные методы молекулярной механики. Между тем, анализ данного типа не проделан даже для α -спирали – наиболее энергетически выгодной структуры остова полипептидной цепи. При этом именно для α -спиральных белков вопрос о роли локальных и дальних взаимодействий в связи с последовательностью аминокислот стоит достаточно остро [5–7].

В данной работе мы использовали α -спиральную конформацию как модельную конформацию целой полипептидной цепи с последовательностью аминокислот реальных глобулярных белков. Далее метод молекулярной механики был использован для того, чтобы оптимизировать модельную структуру, не нарушая симметрии α -спирали, то есть не давая двугранным углам идеальной α -спирали (φ , $\psi = -57^\circ$, -47°) отклоняться от указанных значений более чем на 10° . Анализируя далее распределение разных компонент энергии полученного оптимизированного конформационного набора вдоль последовательности реального белка, можно будет ответить одновременно на два вопроса: соответствуют ли реальным участкам α -спиралей в пространственных структурах глобулярных белков значения минимумов полной энергии или ее компонент и соот-

[#] Автор для переписки (эл. почта: tuman@imb.imb.ac.ru; тел.: (095) 135-60-00).

ветственно можем ли мы предсказать данным методом участки α -спиралей в глобулярных белках.

Основная цель моделирования – найти конформацию полипептидной цепи, отвечающую минимуму энергии молекулы в гипотетическом состоянии: “весь белок– α -спираль” с оптимальной для α -спирали конформацией боковых цепей аминокислотных остатков. Идея метода заключается в том, что для участков полипептидной цепи, которые в реальном белке действительно свернуты в α -спираль, конформация боковых групп является самой оптимальной из возможных, что диктуется соответствующей аминокислотной последовательностью. С другой стороны, для участков, не являющихся α -спиралями в реальном белке (β -структуры, петли и другое), конформация боковых цепей, отвечающая α -спиральному состоянию основной цепи, в реальных глобулярных белках не оптимальна среди допустимого набора конформаций для данных аминокислотных остатков.

Таким образом, локальная энергия на участках модельной структуры, соответствующая участкам отдельно α -спиралей и не α -спиральных структур, либо будет отвечать возможному минимуму энергии на данном участке (для α -спиралей), либо нет (для не α -спиралей). Разница в энергиях участков позволит выявить α -спиральные области.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Апробация предлагаемого нами подхода к нестатистическому предсказанию α -спиральных конформаций проводилась с использованием рентгеноструктурных данных по восьми белкам, содержащим α -спирали. Все отобранные белки (коды 1pdp, 1baz, 1igd, 1ptf, 1pdo, 2chs, 1opc, 1bm9 в международном банке трехмерных структур: www.rcsb.org/pdb [8]) удовлетворяли следующим условиям: а) их пространственная структура получена с разрешением менее 2.4 Å (более низкое разрешение не позволяет точно определять участки регулярной структуры); б) отсутствуют дисульфидные мостики, способные вносить дополнительный вклад в стабилизацию регулярных структур; в) предпочтительно отсутствие лигандов и ионов, взаимодействующих с участками спиралей.

Для каждой аминокислотной последовательности из ранее отобранного набора структур строилась модель “весь белок– α -спираль”, путем задания значений соответствующих двугранных углов основной цепи $\varphi = -57^\circ$, $\psi = -47^\circ$. Оптимальные конформации боковых цепей моделировались в программе ICM [9, 10]. Также для всех исследуемых аминокислотных последовательностей строилась реперная (полностью развернутая) структура, получаемая из оптимизированной мо-

дельной α -спиральной конформации изменением углов φ и ψ до значений -160° и 160° соответственно.

Для полипептидных цепей всех рассмотренных белков были построены энергетические профили аминокислотных последовательностей исследуемых белков в α -спиральной и реперной конформациях. Рассчитывались полные энергии перекрывающихся пентапептидов с первого по N-4, где N – длина цепи, с шагом в один остаток. Разбиение последовательностей белков на пентапептиды позволило выявить все локальные взаимодействия, а также обеспечить формирование водородных связей, характерных для α -спиралей. Все детали расчета см. в разделе “Расчетная часть”.

Анализ энергетических профилей. Рисунки 1а, б иллюстрируют суть метода построения энергетического профиля на примере полипептидной цепи ДНК-связывающего регуляторного белка (код в международном банке трехмерных структур 1pdp). Этот двухдоменный белок принадлежит структурному классу “только α ”. Анализировался второй домен белка длиной 54 а.о. (фрагмент 71–124). Его вторичная структура образована тремя α -спиралями – остатки 80–86, 96–105 и 117–123, и не содержит β -тяжей (расположение α -спиралей схематически показано на рис. 1в).

Энергетический профиль первичной структуры белка (рис. 1б) образуется из разности профилей модельной α -спиральной конформации данной полипептидной цепи и реперной структуры, отвечающей развернутой конформации этой же цепи (рис. 1а). Как видно из графиков, форма результирующего профиля в основном определяется видом энергетического профиля α -спиральной конформации. Участки профиля, соответствующие реальным α -спиралям, отвечают меньшим значениям энергии. То есть формирование α -спиралей происходит на тех участках аминокислотной последовательности, для которых нахождение в α -спиральном состоянии при данных боковых радикалах является энергетически наиболее выгодным.

С другой стороны, боковым радикалам аминокислот на участках, расположенных вне α -спиральных фрагментов белка, невыгодно находиться в состоянии, соответствующем α -спиральной конформации, поэтому на энергетическом профиле в этих областях действительно наблюдаются подъемы и максимумы.

Выявление α -спиралей на энергетических профилях. На рис. 1б, в приведены графики полной энергии и отдельно энергии ван-дер-ваальсовых взаимодействий для всех пентапептидов полипептидной цепи того же белка, полученные с использованием модельной α -спиральной конформации. Из сопоставления энергетических профилей со

схемой расположения пентапептидов, включающих полностью или частично участки α -спиралей в реальной третичной структуре данного белка (рис. 1в; см. подпись), видно, что участки профиля с наиболее выраженными понижениями энергии соответствуют α -спиральным участкам реальной пространственной структуры данной полипептидной цепи. На обоих графиках трем α -спиральным участкам белка соответствуют три последовательных минимума энергии.

В то же время участки профиля, не соответствующие α -спиралям в реальном белке – фрагменты β -структуры (для других рассмотренных белков), петли и некоторые β -повороты, – соответствуют подъемам и максимумам в энергетическом профиле модельной α -спиральной конформации.

Минимумы полной энергии характеризуются значениями от -50 до -70 ккал/моль. Важной особенностью профиля энергии ван-дер-ваальсовых взаимодействий является возможность выделить реперное значение энергии, служащее границей между склонными и не склонными к нахождению в α -спиралях участками профиля. Для всех рассмотренных белков это значение составило $27-28$ ккал/моль по абсолютной величине. Участки профиля, лежащие ниже этой линии, во всех случаях совпадали с α -спиральными участками реальных структур.

Следует также отметить, что границы α -спиральных областей не всегда резко выражены на энергетическом профиле. Это объясняется тем, что для четырех пентапептидов модельной α -спиральной конформации, последовательность которых включает границу реальной α -спирали, вклад в энергию стабилизации вносят только 4, 3, 2 или 1 а.о. Остальные остатки из пяти вносят менее благоприятный вклад, поскольку для них α -спиральная конформация не является оптимальной.

Вклад ван-дер-ваальсовых взаимодействий в полную энергию α -спиральной конформации. Из сопоставления профилей полной энергии (рис. 1б) и энергии ван-дер-ваальсовых взаимодействий (рис. 1в) видно, что последние вносят основной вклад в энергию по сравнению с энергией электростатических взаимодействий или с энергией водородных связей. При этом важно отметить, что энергия Ван-дер-Ваальса, отнесенная к молекулярной массе (выраженная в ккал/моль), в общем не зависит от аминокислотного состава, так как общее число взаимодействий данного типа пропорционально числу атомов. Таким образом, энергия Ван-дер-Ваальса вносит решающий вклад в стабилизацию α -спиральной конформации. Это характеризует в свою очередь локальную природу сил, лежащих в основе формирования α -спиралей.

Также, как видно из рис. 1, профиль энергии Ван-дер-Ваальса позволяет более точно идентифицировать границы α -спиральных участков.

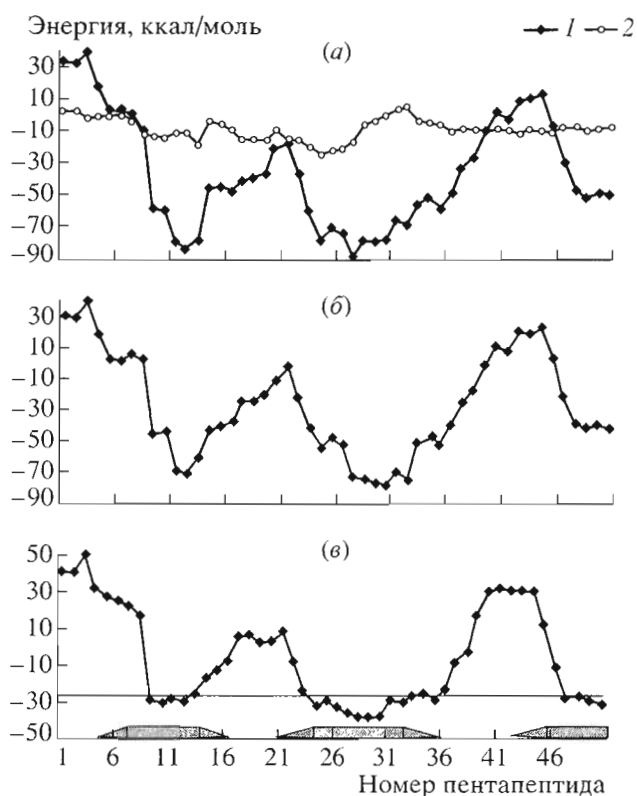


Рис. 1. Построение энергетического профиля полипептидной цепи домена (остатки 71–124) ДНК-связывающего регуляторного белка (Iрdп); (а) модельная α -спиральная конформация (1) и реперная конформация (2); (б) профиль полной энергии. Энергия в каждой точке рассчитывалась как сумма энергий Ван-дер-Ваальса, водородных связей, электростатического взаимодействия и торсионной энергии; (в) профиль энергии ван-дер-ваальсовых взаимодействий. Приведена схема расположения α -спиральных участков вдоль полипептидной цепи белка Iрdп. Трапециям соответствуют участки, где в пентапептидах присутствуют остатки, находящиеся в экспериментальной структуре в α -спиральном состоянии. Верхнему основанию трапеции соответствуют пентапептидные фрагменты последовательности, целиком состоящие из α -спиральных остатков. Пентапептидные фрагменты, соответствующие боковым сторонам трапеций, частично образованы из α -спиральных остатков (4, 3, 2 или 1 остаток в α -спирали по мере удаления от ее границы) и частично из не α -спиральных остатков (соответственно 1, 2, 3 или 4 остатка).

Каждой i -й точке на графиках соответствует энергия пентапептида с $(70 + i)$ -го по $(70 + i + 4)$ -й остаток. Графики содержат N-4 точек по оси абсцисс, отвечающих N-4 пентапептидам для данной первичной структуры, где N – длина полипептидной цепи.

Выявление β -поворотов на энергетических профилях. Отличительной особенностью предлагаемого подхода для предсказаний вторичной структуры является чувствительность метода к β -поворотам. Причем метод позволяет выявлять те β -повороты, которые включают остатки, находящиеся в α -спиральной конформации. К ним



Рис. 2. Энергетические профили энергии ван-дер-ваальсовых взаимодействий (1) и полной энергии (2) для фрагмента полипептидной цепи (1–61 а.о.) иммуноглобулинсвязывающего белка (Igd). Каждой i -й точке на графике соответствует энергия пентапептида с i -го по $(i + 4)$ -й остаток. На схеме обозначены β -повороты (показаны в виде стрелок длиной восемь пентапептидов). Два пентапептида, которые в данный β -поворот входят полностью, расположены в середине, пентапептиды, в которые частично входят остатки β -поворота, расположены по краям. Другие обозначения на графике см. в подписи к рис. 1.

относятся β -повороты I типа, у которых $i + 1$ -й и $i + 2$ -й остатки имеют значения углов ϕ и ψ , соответствующие α -спиральной конформации, а также IV типа, которые в некоторых случаях могут иметь значения углов ϕ и ψ , характерные для α -спиралей [11].

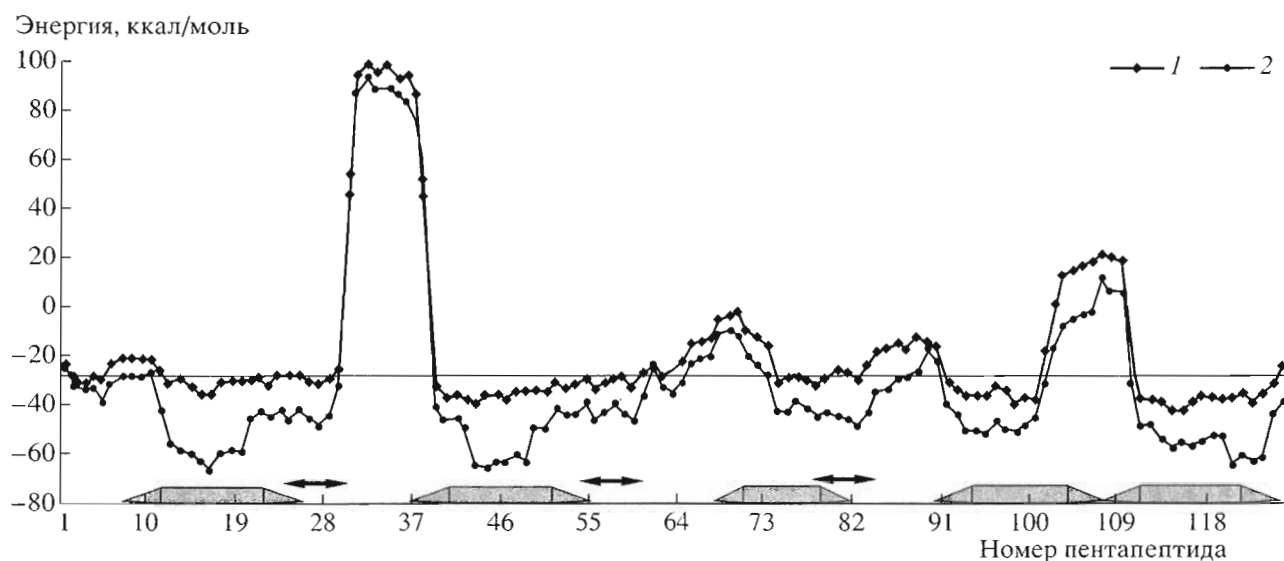


Рис. 3. Энергетические профили энергии ван-дер-ваальсовых взаимодействий (1) и полной энергии (2) для полипептидной цепи (остатки 2–129) фосфотрансферазы (Irho). Каждой i -й точке на графике соответствует энергия пентапептида с $(1 + i)$ -го по $(1 + i + 4)$ -й остаток. Обозначения на графике см. в подписях к рис. 1 и 2.

На рис. 2 приведена иллюстрация того, как с помощью модельной α -спиральной конформации проявляются β -повороты α -спирального типа. В структуре иммуноглобулинсвязывающего белка (код Igd, длина 61 а.о.), относящегося к классу “ $\alpha + \beta$ ”, включающего одну α -спираль и четыре β -тяжа, имеются три β -поворота, расположенные по разные стороны от спирали. Все три β -поворота, один изолированный и два перекрывающихся по первичной структуре (фрагменты 14–17, 51–54 и 52–55), относятся к I типу. Как видно из энергетического профиля, все они обнаруживают себя на графике понижением энергии. Однако их минимумы не столь глубоки, как у α -спиралей. Расположенная в середине полипептидной цепи (остатки 28–41) единственная спираль занимает существенно более глубокий минимум со значением до -70 ккал/моль, в то время как для β -поворотов минимумы лежат вблизи значения для полной энергии -40 ккал/моль. Для энергии Ван-дер-Ваальса поворот, расположенный слева от α -спирали, хотя и имеет энергии, близкие к спиральным, но лежит выше линии реперной энергии.

Выявленные закономерности не зависят от длины рассматриваемой полипептидной цепи, так как модельная α -спиральная структура имеет протяженную конфигурацию и на величину энергии пентапептидов влияет лишь их локальное окружение. На рис. 3 приведен пример более длинного белка – фосфотрансферазы (код Irho, длина 129 а.о.), принадлежащей к структурному классу “ α/β ” с архитектурой домена – трехслойный α/β -сандвич. Вторичная структура данного белка включает пять α -спиралей (остатки 13–25, 42–53, 72–81, 95–105 и 111–125). Все спиральные участ-

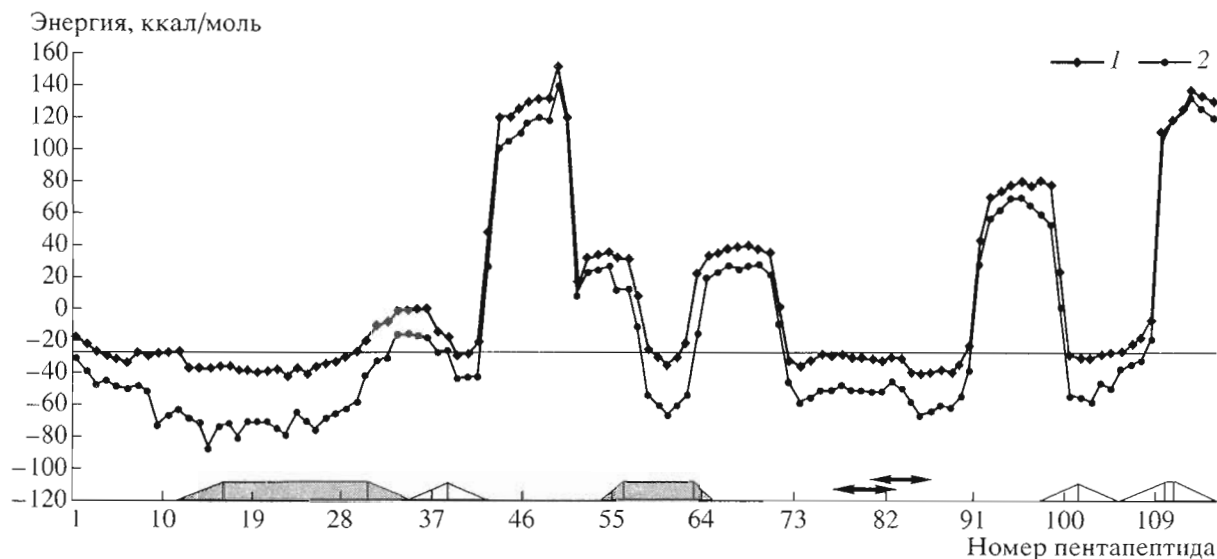


Рис. 4. Энергетические профили энергии ван-дер-ваальсовых взаимодействий (1) и полной энергии (2) для полипептидной цепи (остатки 2–115) изомеразы (2chs). Каждой i -й точке на графике соответствует энергия пентапептида с $(1 + i)$ -го по $(1 + i + 4)$ -й остаток. Обозначения на графике см. в подписи к рис. 1 и 2. Выделенные цветом трапеции соответствуют α -спиралям, незакрашенные фигуры – 3/10-спиралям.

ки, как видно из графиков полной и ван-дер-ваальсовой энергий, четко выявляются и соответствуют потенциальным ямам. Также на обоих графиках выявляются три β -поворота, один из которых I типа, а два других IV типа, которые содержат α -спиральную конформацию.

На примере данного белка (1pfo) можно увидеть искажение энергетических профилей в виде некоторого завышения барьеров, вызываемого присутствием остатка пролина. Пролин, как теперь показано [12], достаточно часто встречается в составе α -спиралей аминокислота. Он может находиться на первых нескольких позициях α -спиралей, а на первой позиции пролин – самая ожидаемая аминокислота. Находясь внутри α -спирали, остаток пролина часто вызывает искажение спиральной структуры [13]. Кроме того, боковая цепь пролина не достаточно точно моделируется в алгоритме направленного графа, применяемого, в частности, в использованной нами программе ICM и в большинстве программ молекулярного моделирования. В определенном окружении пролин может приводить к повышению энергетического профиля на участках, соответствующих левым границам α -спиралей (вторая и третья α -спирали на рис. 3). С другой стороны, и в том случае, когда пролин отсутствует в начале спирали, это не препятствует возможности выделить спираль на фоне барьеров (барьеры перед первой и последней α -спиралями на рис. 3).

Выявление 3/10-спиралей на энергетических профилях. На рис. 4 приведены энергетические профили белка изомеразы (код 2chs) класса “ α/β ”, в структуре которой имеются пять спиралей: две α

и три 3/10-спирали. Кроме одной 3/10-спирали, расположенной на C-конце цепи, все спиральные участки четко проявляются на графиках полной и ван-дер-ваальсовой энергий пентапептидов для полипептидной цепи данного белка. Две 3/10-спирали (остатки 39–41 и 101–103) имеют минимум полной энергии около -50 ккал/моль, что несколько выше, чем для двух α -спиралей (остатки 17–35 и 58–64), четко выявляемых на обоих энергетических профилях. График ван-дер-ваальсовой энергии на участке 3/10-спирали также имеет характерные черты, отличающиеся от α -спиральных, – минимумы энергии 3/10-спиралей лежат ближе к реперной линии.

Таким образом, метод чувствителен к разнице между α - и 3/10-спиралями, при этом следует отметить, что распространенные статистические методы предсказания вторичных структур не делают никаких различий в предсказании α - и 3/10-спиралей. На энергетическом профиле изомеразы не наблюдается последняя в структуре 3/10-спираль (остатки 110–114), что можно объяснить ее локализацией в самом конце полипептидной цепи (ее длина для известной кристаллографической структуры составляет 115 а.о.) и влиянием в этой области краевых эффектов. В частности, вклад в стабилизацию последних пентапептидов модельной α -спиральной конформации вносит только N-концевой фрагмент аминокислотной последовательности, тогда как вклад C-концевого фрагмента отсутствует.

Предсказание α -спиралей. Всего нами было проанализировано восемь доменов из однодоменных и многодоменных белков (см. раздел “Рас-

Сопоставление предсказанных α -спиралей с экспериментальными данными

Количество доменов	Количество α -спиралей*	Выявлено α -спиралей*	Перепредсказано α -спиралей
8	25	24	1

* Приведена сумма α -спиралей и 3/10-спиралей.

четная часть”). Данные по всем белкам сведены в таблицу. Предложенным нами методом с использованием молекулярной механики не выявлена только одна спираль из 25. Это 3/10-спираль, локализованная в самом конце полипептидной цепи (рис. 4). Имеется также одно перепредсказание. Это область остатков 72–91 в изомеразе (рис. 4). Данному участку частично соответствуют два β -поворота α -спирального типа. Что касается возможности выявления границ α -спиральных участков, то, как видно, представленные данные в большинстве случаев достаточно близки к экспериментально наблюдаемым.

ЗАКЛЮЧЕНИЕ

Моделирование стандартным методом молекулярной механики с использованием исключительно информации об аминокислотной последовательности позволяет предсказать участки формирования α -спиральных фрагментов белка. Показано, что на участках модельной структуры, соответствующих α -спиралям в реальных белках, существует выигрыш в энергии при формировании α -спиральной конформации полипептидной цепи. С другой стороны, для не α -спиральных участков возможность принять α -спиральную конформацию значительно ограничена, что четко проявляется на энергетическом профиле. Поэтому описанный в данной работе метод обладает предсказательной силой и способен в принципе выявлять α -спирали для полипептидных цепей с неизвестной пространственной структурой.

Метод хорошо выявляет те участки полипептидной цепи, у которых ход основной цепи в реальной третичной структуре устойчив на протяжении нескольких остатков, отвечает α -спиральной конформации и близким ей конформациям. Более четко эта закономерность обнаруживается на профиле энергий ван-дер-ваальсовых взаимодействий. Это означает, что важнейшим условием для формирования стандартных конформаций в белке является благоприятный характер локальных взаимодействий. Следует отметить, что энергия Ван-дер-Ваальса, как показано в работе [14], является определяющей также при принятии правильной пространственной формы для данной первичной структуры. То, что в практически фибриллярной структуре воспроизводятся реальные α -спиральные сегменты, означает, что третичная структура белка не оказывает опреде-

ляющего влияния на энергетическую предпочтительность при формировании α -спиралей. В то же время метод позволяет выявить протяженные участки последовательности, не склонные входить в α -спирали. Дополнительный анализ в дальнейшем позволит выявить их локальные конформации.

Энергия стабилизации 3/10-спиралей и β -изгибов соответствует энергии стабилизации α -спиралей, но несколько выше. Это позволяет предсказывать также 3/10-спирали и β -изгибы.

Дальнейшее развитие метода требует усовершенствования процедуры моделирования путем увеличения числа варьируемых параметров. Перспективен также анализ энергетических профилей для элементарных фрагментов разной длины и для разных компонент энергии.

РАСЧЕТНАЯ ЧАСТЬ

Моделирование структур. Для моделирования структур полипептидных цепей в различных конформациях, а также для расчета энергии и ее компонент была использована программа молекулярной механики ICM [9, 10], применимая для расчета сложных белковых систем [15, 16].

В программе ICM структура биологических макромолекул рассчитывается исходя из набора “естественных переменных”, которые включают длины связей и значения валентных и двухгранных углов. Трехмерная структура биополимера представляется в виде направленного графа (дерева) без замыканий, в узловых точках которого находятся пронумерованные атомы. Алгоритм допускает фиксацию произвольного набора переменных. Таким образом, в частности, возможны конформационные расчеты в приближении жесткой геометрии, то есть при фиксированных значениях длин связей, валентных и некоторых двухгранных углов, например, двухгранных углов, отвечающих за планарность пептидной связи или колец. Использование данного подхода значительно понижает размерность пространства переменных, обеспечивая эффективность минимизации функции энергии системы.

Для целей данной работы описание биополимера во внутренних координатах обладает дополнительным преимуществом – удобством задания требуемой стартовой конформации полипептидной цепи. Чтобы получить начальную конформацию

α -спирали, достаточно задать соответствующие значения для углов ϕ , ψ основной цепи.

Расчеты были выполнены на PC Pentium II и III поколений.

Набор анализируемых структур. Для анализа отбирались белки, содержащие α -спирали. Отбор осуществлялся случайным образом из содержащих α -спирали структурных классов: “только α ”, “ α/β ” и “ $\alpha+\beta$ ”. Структурная классификация белков определялась согласно классификации в базе данных CATH (<http://www.biochem.ucl.ac.uk/bsm/cath>) [17]. В CATH приводится иерархическая классификация белков, первый уровень которой соответствует структурному классу белкового фолда. Всего использовано для расчетов и анализа восемь белков (1pdn, 1baz, 1igd, 1ptf, 1pdo, 2chs, 1opc, 1bm9). В двух случаях (1pdn и 1opc) анализировались отдельные домены белков. Размеры белков и отдельных доменов колеблются в интервале от 55 до 129 а.о. Структура белка анализировалась на предмет соответствия требованиям по структурной базе данных PDBsum (<http://www.biochem.ucl.ac.uk/bsm/pdbsum>) [18]. Трехмерные координаты пространственных структур анализируемых белков взяты из Брукхэвского банка белковых структур (www.pdb.bnl.gov) [8].

Моделирование α -спиральной конформации полипептидной цепи. На первом этапе для каждой первичной структуры, соответствующей реальному белку, была выполнена подготовительная процедура, с помощью которой получена стартовая конформация для α -спирального состояния полипептидной цепи. Сначала для каждой полипептидной цепи была построена модель структуры со стандартными (идеальными) значениями длин связей и валентных углов, затем вся цепь сворачивалась в стандартную α -спираль заданием соответствующих двухгранных углов: $\phi = -57^\circ$, $\psi = -47^\circ$ [19].

Поиск оптимальной конформации боковых цепей аминокислот, соответствующей α -спиральному состоянию основной цепи, осуществлялся методом Монте-Карло путем вариации двухгранных углов боковых цепей с применением алгоритма ограниченного перебора конформаций. Пороговое значение количества вариаций переменных в методе Монте-Карло составляло в среднем 1000 раз для всех двухгранных углов одной боковой цепи. На каждом шаге процедуры после расчета энергии системы найденная точка конформационного пространства либо принимается, либо отвергается на основании критерия Метрополиса [20] в соответствии с “температурой” системы, и процедура переходит к новой итерации (стартовая температура была взята 600 K). Подробное описание использовавшейся процедуры оптимизации, реализованной в программе молекулярной механики ISM, приведено в работе [10].

Для снятия остаточных напряжений в полученной структуре, вызываемых жесткой геометрией остова, на втором шаге проводился поиск оптимальной конформации остова полипептидной цепи в районе α -спиральной конформации. Это достигалось путем введения ограничительно-го потенциала на изменения углов ϕ , ψ , начиная с отклонения их от идеальных значений на углы, большие $\pm 10^\circ$, и последующей минимизации полной энергии системы путем варьирования двухгранных углов только основной цепи – ϕ , ψ .

Энергия внутримолекулярного взаимодействия в каждой точке конформационного пространства рассчитывалась как сумма энергий Ван-дер-Ваальса, водородных связей, электростатического взаимодействия и торсионной с использованием параметров ECEPP/3 [21–23].

Моделирование реперной структуры. В качестве реперной структуры для исключения из энергии модельной структуры постоянных и слабоменяющихся вкладов, зависящих от типа аминокислоты, рассматривалась развернутая конформация полипептидной цепи с боковыми группами, оставленными в положении, соответствующем α -спиральной конформации. Реперная структура была получена из оптимизированной модельной α -спиральной конформации изменением углов ϕ и ψ до значений -160° и $+160^\circ$ соответственно. Значения взяты из работы [24], где растянутая цепь использовалась в качестве базовой для определения свободной энергии остатков при расчетах стабильности глобул белков.

В заданной указанным способом развернутой конформации могут иметь место близкие контакты атомов, вызывающие локальные стерические напряжения. Для их снятия и нахождения локального минимума полной энергии системы вблизи данной точки конформационного пространства, проводилась дополнительная минимизация энергии с вариацией отдельно χ -углов боковых групп и ϕ -, ψ -углов основной цепи. Описанная процедура не меняла сколько-нибудь существенно конформацию системы.

Построение энергетических профилей. На заключительном этапе процедуры для каждой полипептидной цепи были получены энергетические профили полипептидной цепи в α -спиральной и реперной конформациях. Были рассчитаны полные энергии всех пентапептидов с первого по N-4, где N – длина цепи, с шагом в один остаток. Пентапептиды рассматривались в качестве минимальной структурной единицы биополимера для того, чтобы выявить все локальные взаимодействия, а также обеспечить формирование водородных связей, характерных для α -спиралей. Вместе с тем этот фрагмент достаточно велик, что должно позволить сгладить профиль энергии.

Результирующий профиль полной энергии является разностью профилей для α -спиральной и реперной структур.

Также отдельно рассчитывались энергии только ван-дер-ваальсовых взаимодействий.

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (гранты № 00-04-4835, 00-07-90037 и 00-15-97834).

СПИСОК ЛИТЕРАТУРЫ

1. Persson B. // *EXS*. 2000. V. 88. P. 215–231.
2. Baldi P., Brunak S., Chauvin Y., Andersen C.A., Nielsen H. // *Bioinformatics*. 2000. V. 16. P. 412–424.
3. Zhang C.T., Zhang R. // *Proteins*. 2001. V. 43. P. 520–522.
4. Попов Е.М. // Проблема белка / Ред. В. Т. Иванов. М.: Наука, 1996. Т. 2. С. 230–331.
5. Lim V.I. // *J. Mol. Biol.* 1974. V. 88. P. 873–894.
6. Ptitsyn O.B., Finkelstein A.V. // *Biopolymers*. 1983. V. 22. P. 15–25.
7. Kuwajima K., Semisotnov G.V., Finkelstein A.V., Sugai S., Ptitsyn O.B. // *FEBS Lett.* 1993. V. 334. P. 265–268.
8. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. // *Nucl. Acids Res.* 2000. V. 28. P. 235–242.
9. Abagyan R., Totrov M., Kuznetsov D. // *J. Comp. Chem.* 1994. V. 15. P. 488–506.
10. Abagyan R.A., Totrov M. // *J. Comp. Phys.* 1999. V. 151. P. 402–421.
11. Wilmot C.M., Thornton J.M. // *Protein Eng.* 1990. V. 3. P. 479–493.
12. Kumar S., Bansal M. // *Proteins*. 1998. V. 31. P. 460–476.
13. Kumar S., Bansal M. // *Biophys. J.* 1998. V. 75. P. 1935–1944.
14. Petrey D., Honig B. // *Protein Science*. 2000. V. 9. P. 2181–2191.
15. Luescher I., Romero P., Kuznetsov D., Rimoldi D., Coulie P., Cerrotini J.-C., Jongeneel V. // *J. Biol. Chem.* 1996. V. 271. P. 12463–12471.
16. Anjuere F., Kuznetsov D., Romero P., Cerrotini J.-C., Jongeneel V., Luescher I. // *J. Biol. Chem.* 1997. V. 272. P. 8505–8514.
17. Orengo C.A., Michie A.D., Jones S., Jones D.T., Swindells M.B., Thornton J.M. // *Structure*. 1997. V. 5. P. 1093–1108.
18. Laskowski R.A. // *Nucl. Acids Res.* 2001. V. 29. P. 221–222.
19. Кантор Ч., Шиммель П. // *Биофизическая химия*. М.: Мир, 1984. Т. 1. 241 с.
20. Metropolis N.A., Rosenbluth A.W., Rosenbluth N.M., Teller A.H., Teller E. // *J. Chem. Phys.* 1953. V. 21. P. 1085–1090.
21. Momany F.A., McGuire R.F., Burgess A.W., Scheraga H.A. // *J. Phys. Chem.* 1975. V. 79. P. 2361–2380.
22. Nemethy G., Pottle M.S., Scheraga H.A. // *J. Phys. Chem.* 1983. V. 87. P. 1833–1851.
23. Nemethy G., Gibson K.D., Palmer K.A., Yoon C.N., Paterlini G., Zagari A., Rumsey S., Scheraga H.A. // *J. Phys. Chem.* 1992. V. 96. P. 6472–6484.
24. Pitera J.W., Kollman P.A. // *Proteins*. 2000. V. 41. P. 385–397.

A Nonstatistical Approach to the Prediction of Regular Structures in Proteins by the Example of α Helices

G. T. Kilosanidze, A. S. Kutsenko, N. G. Esipova, and V. G. Tumanyan[#]

[#] Phone: +7 (095) 135-60-00; e-mail: tuman@imb.imb.ac.ru

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

A new approach to the analysis of regular structures in proteins that is based on the method of molecular mechanics is proposed. The method uses only the information about the amino acid sequence. The α -helical conformation was simulated using the ICM program of molecular mechanics. Energy profiles of the sequences in the α -helical conformation, spanning the entire polypeptide chain, were plotted for eight proteins from the Protein Data Bank. The regions of each profile that exhibit energy minima were found to correspond to the α -helical regions of the real spatial structure of the protein. Twenty-four out of 25 helices were distinctly pronounced, which indicates a rather high accuracy of the prediction. The energy profiles also help reveal the short regions that correspond to 3/10-helices and the turns that include local α -helical conformations. Unlike the known statistical methods of prediction, this method makes it possible to establish the physical principles of the formation of α -helical conformations. The English version of the paper: *Russian Journal of Bioorganic Chemistry*, 2002, vol. 28, no. 6; see also <http://www.maik.ru>.

Key words: molecular mechanics; proteins, secondary structure, structure prediction