

**Математические
модели в биологии
и медицине**

Рассмотрены преимущества использования карт Рамачандрана в основанных на методах минимизации энергии подходах к предсказанию структуры белков. Приведены результаты сравнительного анализа разных биофизических фильтров, используемых для сокращения времени компьютерных подсчетов путем исключения из рассмотрения тех вариантов белковых структур, которые не схожи со структурами существующих в природе белков.

© А.В. Быць, 2010

УДК 519.21

А.В. БЫЦЬ

**ИСПОЛЬЗОВАНИЕ КАРТ
РАМАЧАНДРАНА И БИОФИЗИЧЕСКИХ
ФИЛЬТРОВ ПРИ ПРЕДСКАЗАНИИ
ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ
БЕЛКОВ**

Введение. Молекула белка представляет собой линейную последовательность аминокислот. Для подавляющего большинства встречающихся в живых организмах белков каждый белок может иметь одно определенное расположение своей аминокислотной цепочки в трехмерном пространстве. И только у относительно небольшой части встречающихся в живых организмах белков отсутствует устойчивая пространственная структура или возможны несколько вариантов пространственной структуры.

Определение пространственной структуры белка в том случае, когда известна только линейная последовательность его аминокислот, – очень важная с практической и теоретической точек зрения задача, решение которой актуально для фармакологии, медицины и молекулярной биологии.

Участки аминокислотных последовательностей разных белков могут иметь одинаковые пространственные расположения. Некоторые непрерывные участки аминокислотных последовательностей образуют так называемые вторичные структуры, характеризующиеся тем, что взаимное расположение входящих в них аминокислот определяется только типом этой вторичной структуры и практически не зависит от самой аминокислотной последовательности. Наиболее распространенными типами вторичных структур являются β -тяжи и правозакрученные α -спирали. Они присутствуют

в подавляющем большинстве белков, а остальные типы вторичных структур или встречаются в белках очень редко, или мало отличаются от двух вышеупомянутых

Пример пространственной конфигурации молекулы белка схематически показан на рис. 1. Цифрами отмечены вторичные структуры: 1 – правозакрученные α -спирали, 2 – параллельные β -тяжи, 3 – антипараллельные β -тяжи.

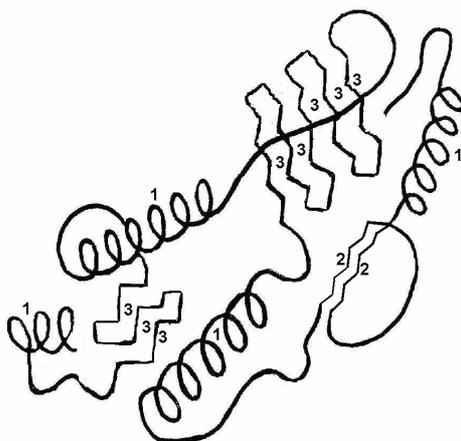


РИС. 1

Поскольку пространственную структуру белка можно определить только дорогостоящими методами рентгеноструктурного анализа и ядерного магнитного резонанса, очень важным представляется развитие математических методов предсказания пространственной структуры белка по известной аминокислотной последовательности.

Предсказание пространственной структуры белка методами минимизации энергии основано на гипотезе о том, что при естественном свернутом состоянии белка свободная энергия системы, состоящей из белка и окружающих его молекул, минимальна [1, 2]. Существует два различных подхода к оценке значения этой энергии для того или иного белка. Первый основывается на рассмотрении различных сил взаимодействия между атомами и молекулами, а второй – на уже имеющихся данных о пространственной структуре ранее исследованных белков.

При предсказании пространственной структуры белков методами минимизации энергии для известной аминокислотной последовательности генерируют разные варианты ее пространственной структуры и проверяют для каждого из них значения свободной энергии. В рассматриваемый текущий вариант структуры вносят некоторые изменения, получая новый вариант пространственной структуры. Если для него значение функционала энергии меньше, чем у предыдущего варианта, то новый вариант становится текущим. Потом в текущий вариант структуры снова вносят некоторые изменения и т. д.

Поскольку вычисление функционала энергии требует больших затрат машинного времени, то еще до этапа подсчета энергии желательно исключить из рассмотрения те варианты структуры, в которых нарушаются физические законы или которые имеют значительные отличия от известных пространственных структур природных белков. Эти варианты определяют по наличию запрещенных значений торсионных углов на картах Рамачандрана, а также с помощью так называемых биофизических фильтров.

Двугранный (торсионный) угол. Для описания взаимного расположения атомов линейной четырехатомной молекулы или входящего в состав более сложной молекулы линейного четырехатомного фрагмента А–В–С–D используется такой геометрический параметр как двугранный (иначе – торсионный) угол. Это угол между плоскостью, в которой лежит фрагмент А–В–С, и плоскостью, в которой лежит фрагмент В–С–D. На рис. 2, а двугранный угол обозначен Ψ , а величины плоских углов трехатомных фрагментов А–В–С и В–С–D обозначены соответственно γ_1 и γ_2 . Наиболее наглядно двугранный угол может быть изображен при рассмотрении А–В–С–D вдоль направления связи В–С. В этом случае двугранный угол будет виден как угол между связями А–В и С–D (рис. 2, б).

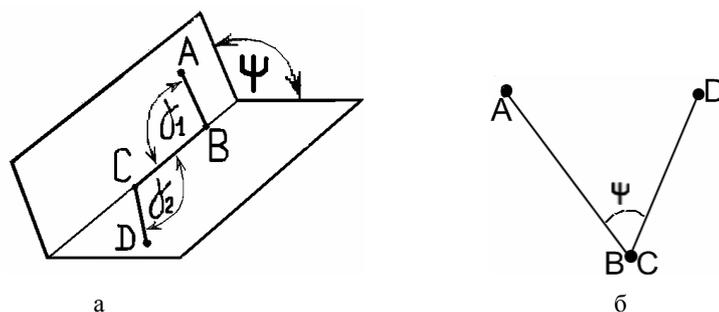


РИС. 2

Двугранные углы межатомных связей белковой цепи. Линейная аминокислотная последовательность белка может быть представлена как проходящая через все аминокислотные остатки этого белка линейная последовательность атомов, в которую из каждого аминокислотного остатка входят три атома: атом азота N_i и два атома углерода C_i^α и C_i , где i – номер аминокислотного остатка в аминокислотной последовательности белка.

На рис. 3 показан участок аминокислотной последовательности белка, в который полностью входит один аминокислотный остаток. Границы между аминокислотными остатками обозначены пунктирными линиями. Торсионный угол ϕ_i определяется четырехатомным фрагментом $C_{i-1}-N_i-C_i^\alpha-C_i$, угол ψ_i – фрагментом $N_i-C_i^\alpha-C_i-N_{i+1}$, а угол ω_i – последовательностью $C_i^\alpha-C_i-N_{i+1}-C_{i+1}^\alpha$. При этом угол ω_i обычно принимает значения только близкие к 0° или 180° (это связано с тем, что химическая связь C_i-N_{i+1} является частично двойной), причем в большинстве случаев углы ω_i близки к 180° .

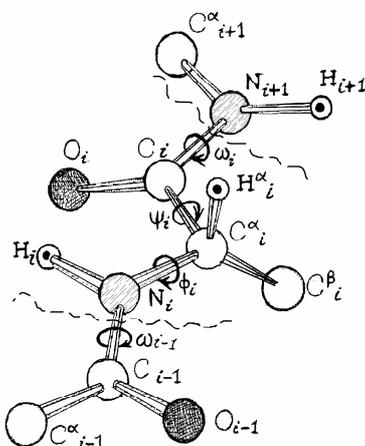


РИС. 3

Использование данных о запрещенных двугранных углах при предсказании структуры белков методами минимизации энергии. С помощью квантово-химических расчетов с учетом известных ограничений на расстояния между атомами могут быть определены допустимые и запрещенные области значений двугранных углов ϕ_i и ψ_i аминокислот [3]. С другой стороны, экспериментально, методами рентгеноструктурного анализа или ядерного магнитного резонанса белков, могут быть получены частоты встречаемости значений углов ϕ_i и ψ_i в каждом виде аминокислот существующих в природе белков, а также в том или ином типе вторичной структуры [4].

Карты Рамачандрана, называемые также графиками или диаграммами Рамачандрана, конформационными картами, ϕ/ψ -картами, представляют собой способ визуализации разрешенных и запрещенных значений двугранных углов ϕ_i и ψ_i аминокислот. На рис. 4 показана полученная с помощью квантово-химических расчетов карта Рамачандрана для всех аминокислот (кроме глицина), на которой белым показаны запрещенные значения углов, темным – разрешенные, а полутемным – разрешенные при принятии менее жестких ограничений на допустимые расстояния между атомами. На оси абсцисс показаны углы ϕ_i , на оси ординат – ψ_i .

При использовании методов минимизации энергии на этапе генерации вариантов пространственной структуры в каждой аминокислоте генерируются только разрешенные значения торсионных углов ϕ_i и ψ_i . Варианты структуры, имеющие запрещенные торсионные углы, не генерируются. Такое применение данных о разрешенных и запрещенных значениях двугранных углов ϕ_i и ψ_i аминокислот эффективно уменьшает количество вариантов белковых структур, для которых нужно вычислять свободную энергию.

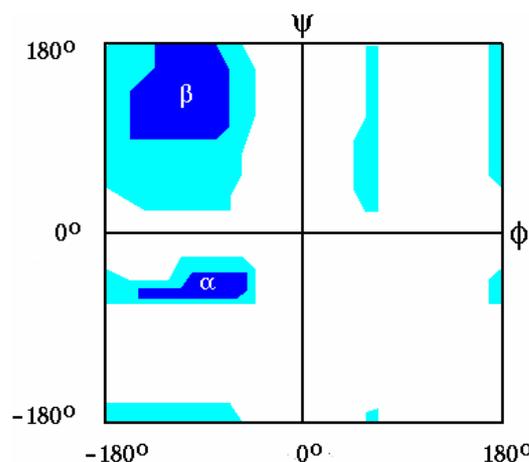


РИС. 4

Биофизические фильтры. После генерации каждого варианта пространственной структуры он может быть проверен так называемыми биофизическими фильтрами, с помощью которых у проверяемого варианта можно обнаружить несвойственные существующим в природе белковым структурам черты. Если такие черты найдены, то данный вариант структуры исключается из дальнейшего рассмотрения, и подсчет значения свободной энергии для него не осуществляется.

Применяются различные биофизические фильтры, в частности:

- фильтр по длине протяженности (Persistence length filter);
- фильтр по радиусу инерции (Radius of gyration filter);
- фильтр по пропорции гидрофобности (Hydrophobicity ratio filter);
- фильтр по упакованной фракции (Packing fraction filter).

Эти, а также другие используемые биофизические фильтры основаны на различных физических и геометрических характеристиках белковых структур.

По сравнению с другими биофизическими фильтрами наиболее эффективными в выявлении структур, непохожих на природные, оказались фильтры по длине протяженности и по радиусу инерции [5, 6, 8]. Причем наиболее эффективным для всех белков, кроме малых, оказалось использование сначала фильтра по длине протяженности, а затем фильтра по радиусу инерции [7].

Длина протяженности – это максимальная длина по прямой линии между участками непрерывной аминокислотной цепи. Для глобулярного белка ее значения варьируют от 15 до 60 Å со средним значением около 40 Å [5]. Этот интервал используется как порог, чтобы выделить конформации, не схожие с природными.

Радиус инерции белка определяется как среднее квадратичное расстояний между каждым атомом белка и общим центром масс всех атомов. Радиус инерции глобулярного белка пропорционален $n^{3/5}$ (где n – количество аминокислот) и для природных белковых структур должен удовлетворять следующему равенству:

$$R_g = \alpha n^{3/5} + \beta,$$

где $\alpha = 0.359$, а значения β находятся в промежутке от 4,257 до 11,257 [5, 7].

Применение биофизических фильтров в основанных на методах минимизации энергии подходах к предсказанию структуры белков позволяет значительно сократить время компьютерных расчетов благодаря уменьшению количества вариантов белковых структур, для которых нужно вычислять свободную энергию [7].

Заключение. Использование карт Рамачандрана в основанных на методах минимизации энергии подходах к предсказанию структуры белков при генерировании конформации каждой следующей аминокислоты аминокислотной цепи эффективно уменьшает количество вариантов белковых структур, которые нужно рассматривать и для которых нужно вычислять функционал энергии. Применение биофизических фильтров в основанных на методах минимизации энергии подходах к предвидению структуры белков дает большие преимущества из-за сокращения времени компьютерных подсчетов. Биофизические фильтры по длине протяженности и по радиусу инерции показывают наибольшую эффективность при отбраковке сгенерированных белковых структур, несхожих с существующими в природе белковыми пространственными структурами, на этапе, предшествующем вычислению энергии сгенерированной белковой структуры.

О.В. Биць

ВИКОРИСТАННЯ КАРТ РАМАЧАНДРАНА ТА БІОФІЗИЧНИХ ФІЛЬТРІВ ПРИ ПЕРЕДБАЧЕННІ ПРОСТОРОВОЇ СТРУКТУРИ БІЛКІВ

Вивчені можливості використання даних про заборонені значення двограних кутів між хімічними зв'язками білкових молекул в основаних на методах мінімізації енергії підходах до передбачення структури білків при генеруванні просторового положення кожної амінокислоти амінокислотного ланцюга для зменшення кількості варіантів білкових структур, які треба генерувати та обчислювати їхню енергію.

Досліджені переваги використання біофізичних фільтрів в основаних на методах мінімізації енергії підходах до передбачення структури білків. Наведено результати огляду і порівняльного аналізу різних біофізичних фільтрів, які використовуються для скорочення часу комп'ютерних підрахунків за рахунок виключення перед етапом обчислення енергії з розгляду тих варіантів білкових структур, які несхожі зі структурами існуючих у природі білків.

О. V. Byts

USE OF RAMACHANDRAN PLOTS AND BIOPHYSICAL FILTERS IN PROTEIN THREE-DIMENSIONAL STRUCTURE PREDICTION

The use of data about the prohibited values of dihedral angles between the chemical bonds in protein molecules in the energy minimization based approaches for the protein structure prediction is studied. These data are used when the conformation of each following amino acid in amino-acidic chain is generated to reduce the number of generated protein structure variants for which the energy is computed.

The benefits are investigated for using biophysical filters in the methods based on energy minimization approaches to the prediction of protein structure. A review and comparative analysis are made of different biophysical filters that are used to reduce the computation time by deleting the energy calculation stage for those variants of the protein structures that are unlike to natural protein structures.

1. Сергиенко И.В., Рязанов В.В., Белецкий Б.А., Быць А.В., Гупал А.М., Ржепецкий С.С. Методы предсказания пространственной структуры белков // Кибернетика и системный анализ. – 2010. – № 1. – С. 38–59.
2. Boas F., Harbury P. Potential energy functions for protein design // Current Opinion in Structural Biology. – 2007. – 17, N 2. – P. 199 – 204.
3. Ramachandran G.N. and Sasiskharan V. Conformation of polypeptides and proteins. // Adv. Protein Chem. – 1968. – 23. – P. 283–437.
4. Szabadka, Z. Ordog, R. Grolmusz, V. The Ramachandran Map of More Than 6,500 Perfect Polypeptide Chains // Biophysical Reviews and Letters. – 2007. – 2, N 3–4. – P. 267–272.
5. Madhu Smita, Singh H., Mitra A. Real Valued Genetic Algorithm Based Approach for Protein Structure Prediction - Role of Biophysical Filters for Reduction of Conformational Search Space // Third IAPR Intern. Conf. on Pattern Recognition in Bioinformatics PRIB, October 15 – 17 2008, Novotel St Kilda. – Melbourne, Australia, 2008.
<http://web2py.iii.as.in/publication/default/download/inproceedings.pdf.92ed272a-5c15-40b0-bcdd-2c1833c7528a.pdf>
6. Bhageerath: an energy based web enabled computer software suite for limiting the search space of tertiary structures of small globular proteins / B. Jayaram, K. Bhushan, SR. Shenoy et al. // Nucleic Acids Res. – 2006. – 34, N 21. – P. 6195–6204.
7. Narang P., Bhushan K., Bose S., Jayaram B. A computational pathway for bracketing native-like structures for small alpha helical globular proteins // Phys. Chem. Chem. Phys. – 2005. – N 7. – P. 2364–2375.
8. Madhusmita S., Singh H., Karlapalem K., Mitra A. A real valued Genetic Algorithm for generating native like structure of small globular protein // Conf. Proc. IEEE Eng. Med. Biol. Soc. – 2008. – P. 1359–62.

Получено 28.04.2010

Об авторе:

Быць Алексей Викторович,

кандидат технических наук, старший научный сотрудник
Института кибернетики имени В.М. Глушкова НАН Украины.

baleks@i.com.ua