2. *Ширяев А.Н.* Основы финансовой стохастической математики. В 2 т. / Фазис М., 1998. Т. 1: Факты. Модели.
3. *Никифоров А.Ф., Уваров В.В.* Основы теории специальных функций. / М.: Наука, 1974.

ДВУХКРИТЕРИАЛЬНАЯ ЗАДАЧА СРАВНЕНИЯ СТРУКТУР БЕЛКОВ

Т. В. Кирис

Все биологические процессы осуществляются при непременном участии белков. Белок состоит из линейной полипептидной цепи, которую можно изобразить линейной последовательностью аминокислотных остатков. Вторичной структурой белка называют пространственное расположение атомов главной цепи молекулы на отдельных ее участках. К основным регулярным вторичным структурам относятся α-спирали и β-складки (рис. 1, *б*). Третичной структурой называют распределение в пространстве всех атомов белковой молекулы (рис. 1, *а*).

Несмотря на наличие различных алгоритмов [1-4] сравнения белков, остается открытым вопрос об определении лучшего сравнения, так как нет корректного стандартного сравнения.

ПОСТАНОВКА ЗАДАЧИ

Пусть два белка представлены координатами C_α-атомов в пространстве и известны их элементы вторичной структуры. Задача состоит в нахождении евклидова совмещения белков как множеств точек в пространстве, которое минимизирует среднеквадратичное отклонение. Первичная структура белка при этом не учитывается.

Предлагаемый в докладе алгоритм IPA состоит из двух шагов. На первом шаге находятся начальные приближения на основе элементов вторичной структуре, которые улучшаются на втором шаге с помощью алгоритма поиска ближайшей точки (FICP)[5]. Алгоритм поиска ближайшей точки состоит из трех шагов: каждой точке из одного множества ставят в соответствие ближайшую точку из другого множества; вычисляют движение, минимизирующее среднеквадратичное отклонение; применяют это движение к первому множеству. Эти три шага повторяются. В зависимости от преследуемых целей, результатом может быть не одно, а несколько совмещений белков.

Процедура повторяется для различных начальных приближений. Далее формируется множество решений, более предпочтительных относительно среднеквадратичного отклонения и количества совмещенных точек.

Данный алгоритм был реализован на C++ с использованием программного пакета MATLAB под управлением Windows.

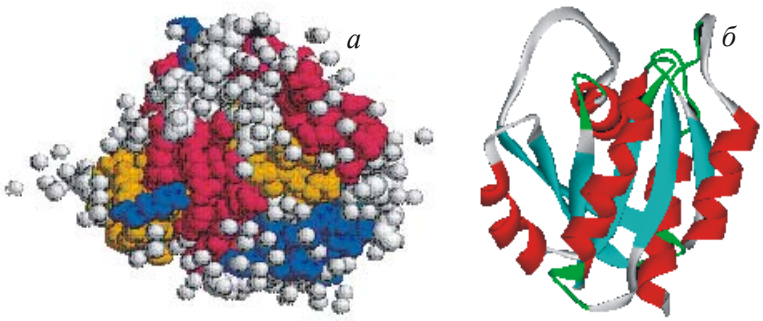

Рис. 1 Представлени белка 1fs0:
а) третичная структура белка;
б) элементы вторичной структуры,
α-спирали (темные) и *β*-складки (светлые)

Для близкородственных белков получены результаты, сопоставимые с известными. В настоящее время алгоритм SSM[6] является наиболее эффективным по сравнению с другими алгоритмами сравнения пространственных структур белков.

В проведенных экспериментах в большинстве случаев наш алгоритм сопоставляет меньшее число остатков, чем в SSM, однако при этом величина среднеквадратичного отклонения также меньше. Алгоритм дает выигрыш во времени, в среднем он сходится за 20 итераций, каждая из которых занимает несколько секунд. Для достижения лучших результатов интересно использование данного алгоритма совместно с другими. Например, если стоит задача сравнения белка с базой данных, то этот алгоритм может быть использован для нахождения близкородственных белков, мера сходства которых потом будет уточняться другими алгоритмами.

МНОГОКРИТЕРИАЛЬНАЯ ОПТИМИЗАЦИЯ

Дальнейшие исследования связаны с использованием многокритериальной оптимизации для решения исходной задачи. Действительно, для сравнения двух пространственных структур белков необходимо максимизировать число сопоставленных остатков и минимизировать среднеквадратичное отклонение. При максимизации числа сопоставленных остатков среднеквадратичное отклонение увеличивается, а при минимизации среднеквадратичного отклонения число сопоставленных остатков уменьшается. Нами предлагаются два пути решения этой задачи.

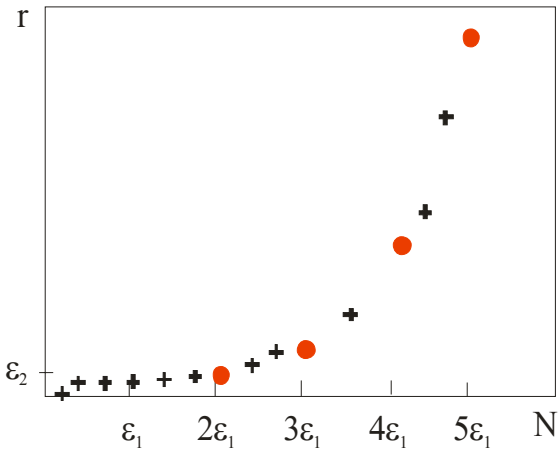

Рис. 2. Пример построения
ε -Аппроксимации множества Парето

Решение исходной задачи сведем к минимизации среднеквадратичного отклонения r при заданном числе сопоставленных остатков N_{align} :

$$\begin{cases} N_{align} = N_i \\ rmsd \to \min \end{cases} \quad (1)$$

или же можно решать задачу максимизации числа сопоставленных остатков при заданном уровне среднеквадратичного отклонения:

$$\begin{cases} r <= r_i \\ N_{align} \to \max \end{cases} \quad (2)$$

Для решения задачи (1) можно использовать алгоритм FICP [5] с фиксированной долей точек, а для решения задачи (2) в алгоритме FICP долю точек будем выбирать такой, чтобы выполнялось первое неравенство системы.

Второй подход состоит в использовании ε-аппроксимации Парето оптимальных решений в определении доли точек в алгоритме FICP. Для этого введем несколько понятий.

Для каждой точки $k_i \in K$ выбираем ближайшую точку $q_i \in Q$ и вычисляем расстояние между ними *d(i)*. Сортируем эти расстояния по неубыванию и строим множество $U = \{(i, r(i)) : i = 1,..., N\}$, где N - минимальное число остатков в двух белках, а *r(i)* – среднеквадратичное отклонение для *i* остатков.

Множество Парето оптимальных решений обозначается P_U и определяется следующим образом

$P_U = \{(i^*, r^*) \in U$: *не существует таких* $(i, r) \in U$, *что* $i^* \le i$ *и* $r^* \ge r\}$.

Пусть $\varepsilon = (\varepsilon_1, \varepsilon_2)$, $\varepsilon_1 \ge 0$, $\varepsilon_2 \ge 0$. Множество P_U^ε недоминированных пар *(i, r)* назовем ε*-аппроксимацией множества Парето*, если для любой

пары $(i, r) \in P_U$ существует пара $(i', r') \in P_U^\varepsilon$ такая, что $i \le i' + \varepsilon_1$ и $r' \le r + \varepsilon_2$ (см. Рис. 3). В качестве решения рассматриваемой задачи строиться множество P_U^ε.

В дальнейшем планируется использование первичной структуры белка, а также модификация данного алгоритма для увеличения области его применения.

Работа частично поддержана фондом INTAS (проект 04-77-7178) и БФФИ Ф06-132.

Литература

1. Chang P.-K., Chen C.-C., Ouhyoung M. A tool for structure alignment of molecules // Proc. of the IEEE Sixth International Symposium on Multimedia Software Engineering (ISMSE'04), 2004. − P. 354-361.

2. Holm L. and Sander C. Protein structure comparison by alignment of distance matrices // J. Mol. Biol., Vol. 233, 1993. − P.123-138.
3. Kawabata T., Nishikawa K. Protein structure comparison using the Markov transition model of evolution. Proteins, Vol. 41, 2000. − P. 108-122.
4. Siew N., Elofsson A. MaxSub: an automated measure for the assessment of protein structure prediction quality // Bioinformatics, Vol. 16, 2000. − P. 776-785.
5. Phillips J.M., Liu R., Tomasi C. Outlier robust ICP for minimizing fractional RMSD /J. Phillips// ArXiv Computer Science e-prints [Electronic resource]. − Mode of access: http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=2006cs.....6098P&db_key=PRE. − Date of access: june 2006.
6. http://www.ebi.ac.uk/msd-srv/ssm/.

ПОВЕДЕНИЕ ТВЕРДЫХ ТЕЛ ПОД ВОЗДЕЙСТВИЕМ ВОЛН ДЕФОРМАЦИЙ

А. В. Круподеров

1. ОБЩИЕ ПОЛОЖЕНИЯ

Рассмотрим массивы горных пород в случае возникновения в них, вследствие силового или кинематического воздействия, волн деформаций и напряжений. При этом, если в массиве образуются области потери устойчивости или разрушения, то такие волны будем называть волнами разрушения.

Как известно, последствия возникновения в массивах горных пород волновых процессов могут быть разнообразными по их проявлениям. Следует отметить, что сегодня проблема моделирования и изучения процессов распространения волн разрушения и инициированного ими движения больших объемов горных пород, приводящего, в свою очередь, к возникновению динамических срывов различной интенсивности, представляет собой одно из наиболее актуальных и важных направлений построения общей теории катастрофических геомеханических явлений [1, 2]. При этом спектр задач, связанных с изучением устойчивости и разрушения массивов горных пород вследствие воздействия волн разрушения, весьма широк.

Путем искусственного возбуждения динамики горного давления до необходимого уровня импульсами ограниченной мощности можно создать ситуации направленного управления состоянием массива горных пород. Преимущества таких механических возмущений в направленном изменении состояния и свойств массива очевидны, так как они малоэнергоемки и достаточны просты в управлении. Это направление реализуется, например, в практике ведения горных работ в способах виброобработки массива [3].