

$$c_n = \frac{1}{\pi} \int_0^{2\pi} \varphi(t) a_n^\rho(e^{it}) dt - \frac{1}{\pi} \int_0^{2\pi} \varphi_\rho(t) a_n^\rho(\zeta_0 + \rho e^{it}) dt ,$$

$$\tilde{c}_n = \frac{1}{\pi} \int_0^{2\pi} \varphi(t) \tilde{a}_n^\rho(e^{it}) dt - \frac{1}{\pi} \int_0^{2\pi} \varphi_\rho(t) \tilde{a}_n^\rho(\zeta_0 + \rho e^{it}) dt ,$$

$$d_n^\rho = \frac{1}{\pi} \int_0^{2\pi} \varphi(t) b_n^\rho(e^{it}) dt - \frac{1}{\pi} \int_0^{2\pi} \varphi_\rho(t) b_n^\rho(\zeta_0 + \rho e^{it}) dt ,$$

$$\tilde{d}_n^\rho = \frac{1}{\pi} \int_0^{2\pi} \varphi(t) \tilde{b}_n^\rho(e^{it}) dt - \frac{1}{\pi} \int_0^{2\pi} \varphi_\rho(t) \tilde{b}_n^\rho(\zeta_0 + \rho e^{it}) dt .$$

Здесь при $n=2^j+k$ ($0 \leq k < 2^j$)

$$a_n^\rho = 2^{-(j-1)/2} \sum_{\nu \in Z_+} \widehat{\Theta}_\varepsilon \left(\frac{\nu}{2^{j+1}} \right) \cos \frac{2\pi\nu(k+1/2)}{2^{j+1}} \frac{|z|^\nu}{1-\rho^{2\nu}} \cos i x ,$$

$$\tilde{a}_n^\rho = 2^{-(j-1)/2} \sum_{\nu \in Z_+} \widehat{\Theta}_\varepsilon \left(\frac{\nu}{2^{j+1}} \right) \cos \frac{2\pi\nu(k+1/2)}{2^{j+1}} \frac{|z|^\nu}{1-\rho^{2\nu}} \sin i x .$$

Аналогично вводятся функции $b_n^\rho(\zeta)$ и $\tilde{b}_n^\rho(\zeta)$.

Литература.

1. Субботин Ю.Н., Черных Н. И. Гармонические всплески и асимптотика решения задач Дирихле в круге с малым отверстием. // Математическое моделирование, 2002 год, том 14, номер 5, стр. 17 – 30.
2. Субботин Ю.Н., Черных Н.И. Всплески в пространствах гармонических функций. // Известия РАН: серия математическая, 2000, том 64, номер 1, стр. 145 – 174.
3. Фрейзер М. Введение в вейвлеты в свете линейной алгебры. Москва: «БИНОМ. Лаборатория знаний», 2008. – 487 с.
4. Заренок М.А. Ортогональные вейвлеты на плоскости // Сборник работ 64-ой научной конференции студентов и аспирантов Белгосуниверситета (Минск, 15-18 мая 2007 г.), в 3-х частях. Ч. 1. С. 159-163.

ПРЕДСКАЗАНИЕ ВЗАИМОДЕЙСТВИЯ БЕЛКОВ ПО ГОМОЛОГИИ

Т. В. Кирис

Взаимодействия белков составляют основу жизненных процессов на молекулярном уровне. Это сложный процесс, определяемый различными структурными и физико-химическими факторами. Последние достижения решения этой проблемы основаны на использовании геометрических подходов и процедур ограниченного конформационного поиска [1,2].

Форма и другие физические характеристики поверхности белков в наибольшей степени определяют природу молекулярных взаимодействий. Во многих случаях пространственные структуры связанных компонент комплексов достаточно близки к структурам свободных белков. Таким образом, геометрическое сопоставление играет важную роль в определении структуры белковых комплексов.

Все методы можно условно разделить на две категории: методы, основанные на структурной информации о белках [3, 4] и методы, основанные на аминокислотных последовательностях белков [5].

Методы, основанные на структуре белка, используют информацию об известном пространственном расположении атомов в белке. К данной задаче примыкает задача о докинге белков. Зачастую требуется не только получить ответ на вопрос о взаимодействии белков, но и узнать, как именно, какими сайтами они взаимодействуют. Для этого можно попытаться оптимизировать пространственное расположение белков, так чтобы полученная связь обладала некоторыми оптимальными характеристиками, например, минимальной энергией или максимальной площадью взаимодействия. Данный класс задач составляет отдельную область биоинформатики.

БЕЛКОВЫЕ ИНТЕРФЕЙСЫ

Увеличивающееся число белковых комплексов, структуры которых были определены экспериментально предоставляет большое количество белковых интерфейсов. База данных интерфейсов предоставляет нам возможность определения областей поверхности заданной молекулы, которая близка по некоторой локальной структуре и структуре остатков. Если обе несвязанные компоненты предполагаемого белкового комплекса проявляют структурные сходства с известным интерфейсом, то несвязанные структуры могут быть совмещены согласно известному интерфейсу.

Основой предсказания является база данных известных домен-доменных интерфейсов (ДДИ), которые были извлечены из базы данных PDB. Идея состоит в поиске интерфейсов в обоих несвязанных частях комплекса по локальному структурному сходству. Каждый ДДИ представляется атомами главной цепи. Для решения задачи нужен структурный метод выравнивания, который не принимает во внимание первичную и вторичную структуру белков.

Для выделения ДДИ из структур белков в PDB использовались только элементы с разрешением меньше 3.5 \AA ⁰ и содержащие, по крайней мере, один SCOP домен. ДДИ состоит из остатков двух доменов, которые на-

ходятся в контакте друг с другом. Остаток представляется всему атомами основной цепи. Каждый ДДИ состоит из всех остатков домена, которые расположены на расстоянии $8 \overset{0}{A}$ от домена партнера.

СОПОСТАВЛЕНИЕ ИНТЕРФЕЙСОВ

Пусть данный белок A и интерфейс B представлены координатами C_α -атомов аминокислот в пространстве. Пусть $A=(a_1, a_2, \dots, a_n)$, где a_i – координаты C_α -атома i -ой аминокислоты, состоит из n аминокислот. Интерфейс $B=(b_1, b_2, \dots, b_m)$ состоит из m аминокислот.

Задача состоит в нахождении евклидова совмещения белка и интерфейса как множеств точек в пространстве, которое минимизирует среднеквадратичное отклонение между ними.

Предлагаемый нами алгоритм состоит из двух шагов. На первом этапе находятся начальные приближения, используя метод динамического программирования для матриц расстояний, которые улучшаются на втором шаге с помощью оптимальных движений.

Поскольку в интерфейсах, как правило, присутствуют несколько отрезков непрерывной полипептидной цепи (для краткости будем называть их отрезками), то если для некоторого белка в нем есть интерфейс, схожий с данным, то найдутся соответствующие подобные отрезки и в белке. Таким образом, если сравнивать матрицы расстояний для данного интерфейса и белка, то в матрице расстояний белка можно выделить схожие подматрицы соответствующих отрезков интерфейсов.

Шаг 1. В интерфейсе B выделяем k отрезков B_j непрерывной первичной цепи длины не меньше $l=5$;

Шаг 2. Строим матрицы расстояний M_B и M_A .

Шаг 3. Находим сопоставление выделенных отрезков B_j и отрезков A таких, что корреляция их матриц расстояний максимальна.

Для этого будем использовать метод динамического программирования (ДП).

Шаг 4. По найденному сопоставлению выполнить движение, минимизирующее среднеквадратичное отклонение (см. Алгоритм поиска ближайшей точки [6]);

ВЫВОДЫ

Предложенный алгоритм был реализован на C++ и протестирован на эталонном множестве, предоставленном Центром биоинформатики Университета Канзаса и базе данных белковых интерфейсов.

Результаты экспериментов.

Белки	l_{rms}	f_{ni}
1avw	0.84	0.85
1b6c	0.66	0.89
1bvn	0.96	0.83
1dfj	1.03	0.94
1k93	1.05	0.97
1ma9	0.88	0.96
1nbf	1.25	0.89
1ppf	1.07	0.75
1shw	1.45	0.84
1sq0	2.00	0.92
1t6g	0.63	0.96
1tmq	0.50	1.00
1ugh	1.68	0.82
1yvb	2.30	0.71
2a42	0.47	1.00
2a5t	0.69	1.00
2bkr	0.98	0.96
1bth	1.35	0.88

В настоящее время соревнование по Критической оценке предсказания взаимодействия (Critical Assessment of PRedicted Interactions (CAPRI) [7] является широко распространенным и его оценки признаются эталонными для установления качества предсказания белковых взаимодействий.

В Таблице 1 приведены результаты предсказания белковых взаимодействий белковых комплексов, взятые из эталонного множества Университета Канзаса, по оценочным параметрам CAPRI. В графе l_{rms} указано среднеквадратичное отклонение лигандов (меньших белков в комплексе) после совмещения рецепторов (больших белков в комплексе), в графе f_{ni} – отношение природных остатков в предсказанном интерфейсе белкового комплекса относительно настоящего интерфейса белкового комплекса.

В проведенных экспериментах в большинстве случаев наш алгоритм дает хорошие результаты в соответствии с оценочными параметрами CAPRI. Алгоритм достаточно быстрый, в среднем он работает в течение нескольких секунд. Поиск схожести непрерывных отрезков интерфейса адекватно отображают природу белковых взаимодействий и дает выигрыш во времени.

В дальнейшем планируется модернизация алгоритма, учитывающая первичную структуру белка и нежесткий характер белков и белковых взаимодействий.

Литература

1. *Mintseris J., Janin J., Chen, R. and Weng Z.* A protein-protein docking benchmark // *Proteins*, Vol. 52, 2003. – P. 88–91.
2. *May P., Hoppe A., Frommel C., Gunther, S. and R. Preissner.* Docking without docking: ISEARCH - prediction of interactions using known interfaces // *Proteins*, Vol. 60, 2007. – P. 150–159.
3. *Alexov E., Kundrotas P.J.* Predicting 3d structures of transient protein–protein complexes by homology // *Biochimica et Biophysica Acta*, Vol. 1764, 2006. – P.1498–1511.
4. **Grimm V., Krakaki A., Szilagy A. and Skolnik J.** Prediction of physical protein–protein interactions // *Phys. Biol*, Vol. 2, 2005. – P. S1–S16.
5. *Matthews L. et al.* Identification of potential interaction networks using sequencebased searches for conserved protein–protein interactions or ‘interologs’ // *Genome Res.*, Vol. 11, 2001. – P. 2120–2126.
6. *T. Kirys, S. Feranchuk, A. Tuzikov, J. Rocha.* Iterative protein alignment algorithm (IPA) // *Proc. of the 3-rd Moscow conference on computational molecular biology.* – 2007. – P. 145–147.
7. *Leplae R., Lensink M., Menndez, R. and Wodak S.* Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures // *Proteins*, Vol. 60, 2005. – P. 150–169.

ЗАДАЧА НАХОЖДЕНИЯ ХЕЛЛИЕВОЙ РАЗМЕРНОСТИ ГРАФА

Е.В. Крылов

1. ВВЕДЕНИЕ

Понятие хеллиевой размерности введено и рассмотрено в работе [1]. Оно напрямую связано с классом графов r -мино – классом, в котором каждая вершина входит не более чем в r максимальных клик. Этот класс интересен по следующим причинам: при фиксированном r для r -мино известны эффективные алгоритмы построения структур данных, оптимальных по используемой памяти и позволяющих определять смежность вершин за время $O(1)$ [2]. Также известен полиномиальный алгоритм, который для фиксированного r определяет, является ли данный граф r -мино и в случае, если это так, строит список всех максимальных клик графа [3]. Класс r -мино совпадает с классом реберных графов от хеллиевых гиперграфов ранга не более r . Очевидно, что краусова размерность графа не может превосходить хеллиеву размерность графа [1]. Однако, в [1] показано, что разность между этими двумя размерностями может быть сколь угодно большой.

В [1] поставлен вопрос о сложности задачи нахождения хеллиевой