

УДК 57:007:001.891.57

Т.В. Кирис, А.В. Тузиков

АЛГОРИТМ ПРЕДСКАЗАНИЯ ВЗАИМОДЕЙСТВИЯ БЕЛКОВ НА ОСНОВЕ СТРУКТУРНОЙ ГОМОЛОГИИ

Предлагается оригинальный алгоритм предсказания взаимодействия белков, основанный на структурной схожести с экспериментально определенным интерфейсом белкового комплекса, отобранным из базы интерфейсов белковых комплексов. Пара свободных белков выравнивается на интерфейс из библиотеки белок-белковых интерфейсов. Выравнивание выполняется с помощью метода динамического программирования путем максимизации корреляции между матрицами расстояний отрезков интерфейса и белка.

Введение

Задача предсказания взаимодействия белков является центральной в компьютерной биологии. Знания о взаимодействиях белков в организме являются необходимыми при создании лекарств и вакцин, а также для предсказания функций белков.

Несмотря на разнообразие алгоритмов белок-белкового докинга [1], задача предсказания комплекса белков по их свободным структурам остается нерешенной [2]. В большинстве алгоритмов белок-белкового докинга можно выделить три стадии: генерирование конформаций комплекса, их улучшение, выбор лучшей модели.

Генерация конформаций комплекса происходит либо на основе свободного докинга, когда структура одного белка перемещается вокруг структуры другого белка [3–6], либо на основе шаблонов (гомологии) [7–12]. Различные методы отличаются способами наложения структур белков на шаблоны, улучшения начальных моделей и целевой функцией. Значение методов на основе шаблонов возрастает с увеличением количества экспериментально полученных структур белок-белковых комплексов. Данные методы верно предсказывают комплекс белков при наличии схожего закристаллизованного комплекса.

Различные техники улучшения используют метод Монте-Карло, имитационный отжиг и др. [13–20]. Выбор лучшей модели осуществляется на основе целевой функции [2, 13, 18, 21–24]. Существуют целевые функции на основе статистических потенциалов [23, 25, 26], а также энергетические [2, 21, 27], которые могут учитывать электростатику, взаимодействия ван дер Ваальса, десольватации и др.

Поскольку белки при взаимодействии часто претерпевают конформационные изменения [28], которые включают изменения основной и боковых цепей, другим важным аспектом докинга является учет гибкости белка [15, 20, 29, 30]. Обзор существующих методов гибкого докинга содержится, например, в работах [29, 31].

Включение гибкости структур белка в алгоритмы докинга значительно усложняет этот процесс из-за увеличения числа степеней свободы, что приводит к значительному увеличению времени работы и большому числу ложноположительных результатов.

1. Определения и методы

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

В интерфейсе B выделены последовательные отрезки непрерывной первичной цепи B_j , $j=1, \dots, k$, длиной не меньше l остатков:

$$B_1 = \{b_{11}, \dots, b_{1j_1}\}, \dots, B_k = \{b_{k1}, \dots, b_{kj_k}\}.$$

Объединение таких отрезков будем обозначать $\bar{B} = \bigcup_{i=1}^k B_i$.

Матрица расстояний M_A белка A представляет собой матрицу $n \times n$, где элемент этой матрицы $m(i, j)$ равен евклидову расстоянию в пространстве между атомами a_i и a_j .

Сходство двух множеств одинаковой мощности $A = \{a_1, a_2, \dots, a_n\}$ и $C = \{c_1, c_2, \dots, c_n\}$ будем оценивать через корреляцию их матриц расстояний и обозначать $\text{Correlation}(A, C)$. Корреляция матриц M_A и M_C размерности $n \times n$ вычисляется по следующей формуле:

$$\text{Correlation}(A, C) = \frac{\sum_{i=1}^n \sum_{j=1}^n m_A(ij) \cdot m_C(ij)}{\sqrt{\sum_{i=1}^n \sum_{j=1}^n (m_A(ij))^2} \cdot \sqrt{\sum_{i=1}^n \sum_{j=1}^n (m_C(ij))^2}} \quad (1)$$

Два белка или интерфейса называются пространственно выровненными, если они совмещены в пространстве с выбранным соответствием между остатками (C_α -атомами) совмещенных структур. Пара соответствующих остатков в разных структурах называется выровненной. Сходство пространственно выровненных интерфейсов вычисляется с помощью меры TM-score [32]:

$$\text{TM-score} = \frac{1}{L_I} \sum_{i=1}^{L_M} \frac{1}{1 + \left(\frac{d_i}{d_0}\right)^2}, \quad (2)$$

где L_I – количество остатков в интерфейсе; L_M – количество выровненных остатков; d_i – расстояние между i -й парой выровненных остатков. Коэффициент нормализации d_0 вычисляется по формуле

$$d_0 = 1,24 \sqrt[3]{L_N - 15} - 1,8,$$

где L_N – количество остатков в наименьшей из двух структур.

Библиотека белок-белковых интерфейсов, использованная в данной работе для тестирования алгоритма, была построена на основе эталонного набора комплексов белков [33]. Остатки интерфейса определялись как остатки, у которых имеется по крайней мере один атом другого белка в комплексе на расстоянии не более 12 \AA . Библиотека состоит из 99 интерфейсов белок-белковых комплексов. Пример интерфейсов комплексов белков показан на рис. 1.

В среднем интерфейс белка состоит из 62 остатков (минимальный интерфейс содержит 17 остатков, максимальный – 219), содержит четыре отрезка непрерывной полипептидной цепи длиной более шести остатков (минимальный – 1, максимальный – 13), пять отрезков длиной более пяти остатков (минимальный – 1, максимальный – 14), пять отрезков длиной более четырех остатков (минимальный – 1, максимальный – 15) и шесть отрезков непрерывной полипептидной цепи длиной более трех остатков (минимальный – 1, максимальный – 20).

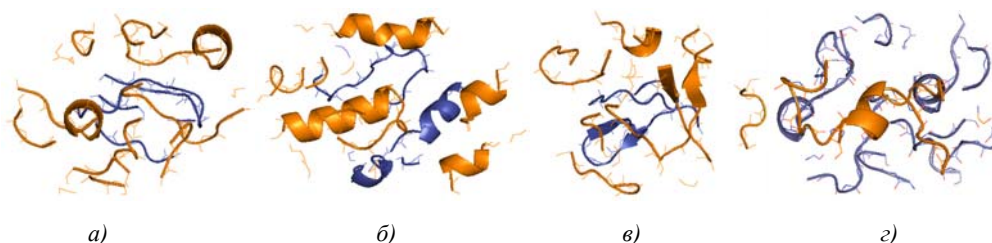


Рис. 1. Интерфейсы комплексов белков: а) протеазы (оранжевая) и эглина С (синий), pdb-код 1cse; б) оксидоредуктазы (оранжевый) и адренодоксина (синий), pdb-код 1e6e; в) трипсина (оранжевый) и ингибитора (синий), pdb-код 1rpe; г) овомукоида (оранжевый) и протеазы (синий), pdb-код 1r0r.

Непрерывные сегменты первичной последовательности, содержащие не меньше трех остатков, выделены жирными линиями

Качество предсказанного комплекса оценивается с помощью меры i -RMSD [1], которая используется в международном соревновании по предсказанию комплексов белков CAPRI [34] и вычисляется следующим образом: большие по числу аминокислотных остатков белки (из двух белков в комплексе) предсказанного и экспериментально определенного комплекса совмещаются, а затем вычисляется среднеквадратичное отклонение (RMSD) между C_{α} -атомами интерфейсов меньших белков. Остатки интерфейса определяются как остатки, у которых имеется по крайней мере один атом другого белка в комплексе на расстоянии не более 6 Å.

2. Алгоритм

Идея предсказания взаимодействия белков A и B состоит в поиске белковых интерфейсов A' B' в базе данных, таких, что интерфейсы белков A и B подобны интерфейсам A' и B' соответственно (рис. 2). Предсказанный комплекс AB получается при наложении белка A на интерфейс A' и белка B на интерфейс B' . Такой комплекс будем называть комплексом AB , предсказанным на основе интерфейса $A' B'$.

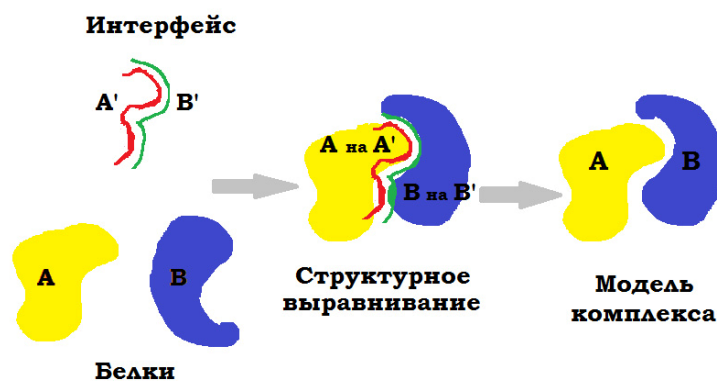


Рис. 2. Идея предсказания взаимодействия белковых комплексов на основе поиска гомологичных интерфейсов. Модель комплекса AB , предсказанная на основе интерфейса $A'B'$, получается путем выравнивания белка A на интерфейс A' и белка B на интерфейс B'

В интерфейсе A' выделяются последовательные непрерывные по первичной последовательности отрезки A'_i длиной не меньше $l=3$ остатков, $\overline{A'} = \bigcup_i A'_i$.

Для нахождения наилучшего выравнивания белка A с интерфейсом A' ищется такой отрезок $A_i \subset A$, что длина отрезка A_i равна длине отрезка A'_i и корреляция матриц расстояний между отрезками максимальна:

$$\text{Correlation}(M_{\bigcup_{i=1}^k A_i}, M_{\bigcup_{i=1}^k A'_i}) \rightarrow \max.$$

2.1. Выравнивание интерфейсов

Для нахождения отрезков A_j , $j=1, 2, \dots, k$, соответствующих отрезкам A'_j длины l_j , $j=1, 2, \dots, k$, с максимальной корреляцией их матриц расстояний используется метод динамического программирования (ДП). Идея ДП состоит в разбиении задачи на подзадачи, нахождении решений подзадач рекурсивно и, далее, использовании полученных решений подзадач для конструирования решения исходной задачи.

Для нахождения k отрезков A_j , $j=1, 2, \dots, k$, соответствующих отрезкам A'_j , $j=1, 2, \dots, k$, с максимальной корреляцией их матриц расстояний, будем последовательно искать сопоставле-

ние одного отрезка A'_1 , двух отрезков A'_1, A'_2 , трех отрезков A'_1, A'_2, A'_3 и т. д. с отрезками из A с максимальной корреляцией их матриц расстояний:

$$\text{Correlation} \left(M_{\bigcup_{j=1}^i A_j}, M_{\bigcup_{j=1}^i A'_j} \right) \rightarrow \max, i = 1, 2, \dots, k.$$

Отрезки $A'_j, j=1, 2, \dots, k$, будут сопоставляться с отрезками из A последовательно по первичной цепи, т. е. между отрезками $A_j, j=1, 2, \dots, k$, могут быть пропуски, но начало отрезка A_j расположено в линейной цепи после конца отрезка $A_{j-1}, j=2, \dots, k$. Отрезки $A_j, j=1, 2, \dots, k$, не пересекаются.

Введем функцию $\text{align}(i, j)$, которая обозначает максимальную корреляцию при сопоставлении первых i отрезков $A'_q, q=1, \dots, i$, с отрезками из A так, что конец последнего отрезка A'_i сопоставляется с j -й аминокислотой a_j белка A . Сопоставление, на котором достигается эта корреляция, будет записываться в матрицу Path размерности $k \times n$ в позицию (i, j) , $1 \leq i \leq k$, $1 \leq j \leq n$, т. е.

$$\text{Path}(i, j) = \bigcup_{k=1}^i A_k.$$

Для $i=1$ и $|A'_1|=l_1$ значение $\text{align}(1, j) = \text{Correlation}(A'_1, (a_{j-l_1+1}, \dots, a_j))$, для $i=2$ значение $\text{align}(2, j)$ равно максимальной корреляции $\text{Correlation}(A'_1 \cup A'_2, A_1 \cup A_2)$, когда конец отрезка A'_2 сопоставляется с j -й аминокислотой a_j белка A , а A_1 пробегает по всем возможным позициям от начала последовательности до пересечения с A_2 .

Максимальной корреляцией отрезков $A'_q, q=1, 2, \dots, k$, с отрезками из A будет $\max_j \text{align}(k, j)$. Поскольку эта задача является NP-трудной [35], предлагается эвристика, которая находит субоптимальное решение. Рекуррентное соотношение для вычисления максимальной корреляции при сопоставлении первых i отрезков $A'_q, q=1, 2, \dots, i$, с отрезками из A определяется как максимальная корреляция найденного сопоставления $i-1$ отрезков и i -го отрезка A'_i :

$$\text{align}(i, j) = \max_{j_1 < j} \left\{ \text{Correlation} \left(\bigcup_{q=1}^i A'_q, \text{Path}(i-1, j_1) \right) \right\}. \quad (3)$$

Введем массив Begin. Его i -й элемент $\text{Begin}[i]$ – это номер остатка белка A , начиная с которого может сопоставляться конец i -го отрезка $A'_i, l_i = |A'_i|$, т. е.

$$\text{Begin}[1] = l_1,$$

а для i -го отрезка

$$\text{Begin}[i] = \text{Begin}[i-1] + l_i.$$

Алгоритм поиска сопоставления отрезков, при котором достигается максимальная корреляция их матриц расстояний, состоит из следующих шагов:

1. for ($j = \text{Begin}[1]; j \leq n; j++$).
2. $\text{align}(1, j) = \text{Correlation}(A'_1, (a_{j-l_1+1}, \dots, a_j))$.
3. for($i = 2; i \leq k; i++$).

4. for ($j = \text{Begin}[i]; j \leq n; j++$).
5. $\text{align}(i, j) = \max_{j_1 < j} \text{Correlation}(\bigcup_{q=1}^i A'_q, \text{Path}(i-1, j_1) \bigcup A_i)$.
6. $\text{Path}(i, j) = \text{Path}(i-1, \arg \text{align}(i, j)) \bigcup A_i$.
7. $\text{Path}(k, \arg \max_i \text{align}(k, i))$.

Алгоритм определяет отрезки $A_q, q=1, 2, \dots, k$, соответствующие отрезкам $A'_q, q=1, 2, \dots, k$, корреляция матриц расстояний которых максимальна. В строке 2 вычисляются значения корреляции между отрезком A'_1 и отрезками из A . В следующем цикле по формуле (3) вычисляется значение корреляции сопоставления первых i отрезков $A'_q, q=1, 2, \dots, i$, и отрезков из A (строка 5) и определяются отрезки $A_q, q=1, 2, \dots, i$, соответствующие отрезкам $A'_q, q=1, 2, \dots, i$, на которых достигается максимальная корреляция (строка 6). В строке 7 определяется сопоставление между отрезками $A'_q, q=1, 2, \dots, k$, интерфейса A' и отрезками $A_q, q=1, 2, \dots, k$.

После того как найдено сопоставление между отрезками, вычисляется оптимальное движение, минимизирующее среднеквадратичное отклонение между всеми сопоставленными остатками.

Белок A выравнивается на интерфейс A' по найденному движению. Аналогичным образом находится сопоставление между отрезками $B'_q, q=1, 2, \dots, k$, интерфейса B' и отрезками $B_q, q=1, 2, \dots, k$, белка B , и белок B выравнивается на интерфейс B' по движению, минимизирующему среднеквадратичное отклонение между сопоставленными остатками.

2.2. Вычисление оптимального движения

Для двух равномогных множеств с известным соответствием между точками множеств существует оптимальное движение, минимизирующее среднеквадратичное отклонение между ними [36].

Пусть даны два равномогных множества точек в пространстве $P, M \subset \mathbf{R}^3$, их мощность $|P| = |M| = N$, и задано соответствие между точками множеств.

Вначале оба множества переносятся таким образом, чтобы их центры оказались в начале координат. Далее задача сводится к поиску матрицы поворота R , минимизирующей среднеквадратичное отклонение между соответствующими точками множеств:

$$\sum_{i=1}^N |p_i - Rm_i|^2 \rightarrow \max.$$

Левую часть этого выражения можно переписать в виде

$$\sum_{i=1}^N |p_i|^2 - 2\text{tr}(R^T \sum_{i=1}^N p_i m_i^T) + \sum_{i=1}^N |m_i|^2.$$

Матрица $K = \sum_{i=1}^N p_i m_i^T$. Задачу можно переформулировать так: дана матрица K , найти матрицу поворота R , такую, что

$$\text{tr}(R^T K) \rightarrow \max.$$

Если $K = U\Sigma V^T$ – разложение по сингулярным значениям, то значение $\text{tr}(R^T K)$ максимизируется матрицей $R = UV^T$.

Лемма [37]

Если K – действительная матрица $l \times n$, то существуют ортогональные матрицы

$$U = [u_1 \dots u_l] \subset \mathbf{R}^{l \times l};$$

$$V = [v_1 \dots v_n] \subset \mathbf{R}^{n \times n},$$

такие, что

$$U^T K V = \Sigma = \text{diag}(\sigma_1, \dots, \sigma_p) \in \mathbf{R}^{l \times n}, \quad (4)$$

где $p = \min(l, n)$ и $\sigma_1 \geq \dots \geq \sigma_p \geq 0$, или

$$K = U \Sigma V^T. \quad (5)$$

Определим ортогональную матрицу $Z = U^T R^T V$. Получим $\text{tr}(R^T K) = \text{tr}(R^T V (V^T K U) U^T) = \text{tr}(R^T V \Sigma U^T) = \text{tr}(Z \Sigma) = \sum_{i=1}^p z_{ii} \sigma_i$. В связи с тем что Z – ортогональная матрица, выражение максимумно при единичной матрице. Следовательно, $R = U V^T$.

Итак, для нахождения оптимального движения необходимо:

- перенести множества таким образом, чтобы их центры оказались в центре координат;
- вычислить матрицу корреляции между точками множества;
- разложить ее по сингулярным значениям;
- вычислить матрицу поворота.

3. Экспериментальные результаты

Предложенный алгоритм был протестирован на эталонном наборе комплексов белков [33]. В данный набор входят 99 белок-белковых комплексов, для которых известны структуры обоих белков в свободном состоянии. Тестирование было проведено на структурах белков в свободном состоянии.

Пара свободных белков выравнивалась на каждый интерфейс из библиотеки белок-белковых интерфейсов. Схожесть выравнивания белков A и B на интерфейс $A' B'$ вычислялась как сумма TM-score (2) между интерфейсом A и A' и интерфейсом B и B' . Для каждой пары свободных белков строилось 10 моделей комплекса на основе интерфейсов с наибольшим значением TM-score.

Качество предсказанных комплексов оценивалось с помощью i -RMSD [2]. Согласно правилам международных соревнований по предсказанию комплексов белков CAPRI [34] комплекс считается «хорошо» предсказанным, если значение i -RMSD не превышает 5 Å хотя бы для одной из 10 построенных моделей.

Доля «хороших» предсказаний данного алгоритма на эталонном наборе комплексов белков [33] равняется 43, 42, 39 и 42 % при выделении в интерфейсе отрезков длиной 3, 4, 5 и 6 соответственно. Предлагается выбирать отрезки длиной 3, поскольку при выборе отрезков длиной l конформации отрезков интерфейса меньшей длины не принимаются во внимание и могут остаться без сопоставления. Доля «хороших» предсказаний алгоритма ZRANK [27] на схожем эталонном множестве [38] составляет 35 %.

Время предсказания комплекса белка из 300 аминокислот на основе одного интерфейса составляет в среднем 0,3 с на компьютере 800 MHz AMD Phenom(tm) II x4.

Алгоритм предсказывает правильный интерфейс в комплексе даже при большом (≥ 3 Å) изменении основной цепи интерфейса. На рис. 3 показаны примеры комплексов, построенных на основе интерфейсов с большим изменением основной цепи. Мера TM-score комплексов 1acb/1ago равнялась 0,99, 0,87 из 2,0 (сумме TM-score обоих белков комплекса); RMSD C_α -атомов между интерфейсами белков и интерфейсами, по которым они смоделированы, равно 3,6 и 5,8 Å соответственно. Несмотря на достаточно большую разницу конформаций петель в интерфейсах, структуры комплексов были предсказаны верно.

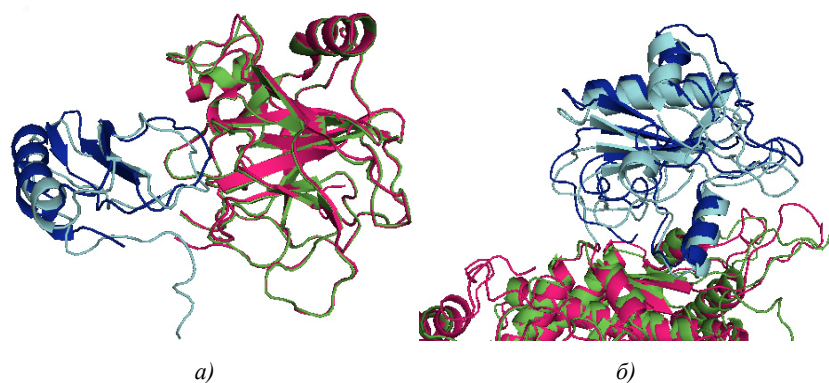


Рис. 3. Закристаллизованные и предсказанные комплексы белков: а) эглина С (синий/голубой) и альфа-химотрипсина (розовый/зеленый), pdb-код комплекса 1ac6; б) лизоцима (синий/голубой) и РНК полимеразы (розовый/зеленый), pdb-код комплекса 1ago

Заключение

Разработан алгоритм предсказания взаимодействия белков, основанный на структурной схожести с белок-белковым интерфейсом из базы данных. Пара свободных белков выравнивается на каждый интерфейс из библиотеки белок-белковых интерфейсов. Выравнивание выполняется с помощью метода динамического программирования путем максимизации корреляции между матрицами расстояний отрезков интерфейса и белка. Для каждой пары свободных белков строится 10 моделей комплекса на основе интерфейсов с наибольшей схожестью. Схожесть выравнивания белков на интерфейс вычисляется как сумма меры TM-score между интерфейсами.

Доля «хороших» предсказаний данного алгоритма на эталонном наборе комплексов белков [33] равняется 42 %. Разработанный алгоритм предсказывает правильный интерфейс в комплексе даже при некоторой гибкости конформации интерфейса.

Роль предложенного алгоритма предсказания взаимодействия белков по структурной гомологии с экспериментально определенными интерфейсами белков будет возрастать с увеличением количества экспериментально определенных комплексов белков.

Список литературы

1. Tramontano, A. The Ten Most Wanted Solutions in Protein Bioinformatics / A. Tramontano. – Boca Raton : Chapman and Hall/CRC, 2005. – 216 p.
2. Lensink, M.F. Docking and scoring protein interactions: CAPRI 2009 / M.F. Lensink, S.J. Wodak // Proteins. – 2010. – Vol. 78, № 15. – P. 3073-3084.
3. The performance of ZDOCK and ZRANK in rounds 6–11 of CAPRI / K. Wiehe [et al.] // Proteins. – 2007. – Vol. 69. – P. 719–725.
4. Schneidman-Duhovny, D. Automatic prediction of protein interactions with large scale motion / D. Schneidman-Duhovny, R. Nussinov, H. Wolfson // Proteins. – 2007. – Vol. 69. – P. 764–773.
5. PIPER: an FFT-based protein docking program with pairwise potentials / D. Kovakov [et al.] // Proteins. – 2006. – Vol. 65. – P. 392–406.
6. Kowalsman, N. Inherent limitations in protein-protein docking procedures / N. Kowalsman, M. Eisenstein // Bioinformatics. – 2007. – Vol. 23. – P. 421–426.
7. Lu, L. MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading / L. Lu, H. Lu, J. Skolnick // Proteins. – 2002. – Vol. 49. – P. 350–364.
8. Kundrotas, P.J. Predicting 3d structures of transient protein-protein complexes by homology / P.J. Kundrotas, E. Alexov // Biochimica et Biophysica Acta. – 2006. – Vol. 1764. – P. 1498–1511.
9. Fast and accurate modeling of protein-protein interactions by combining template-interface-based docking with flexible refinement / N. Tuncbag [et al.] // Proteins. – 2012. – Vol. 80. – P. 1239 – 1249.
10. A structural perspective on protein-protein interactions / R. Russel [et al.] // Curr Opin Struct Biol. – 2004. – Vol. 14. – P. 313–324.

11. Sinha, R. Docking by structural similarity at protein-protein interfaces / R. Sinha, P. Kundrotas, I. Vakser // *Proteins*. – 2010. – Vol. 78. – P. 3235–3241.
12. Vreven, T. Integrating atom-based and residue-based scoring functions for protein-protein docking / T. Vreven, H. Hwang, Z. Weng // *Protein Science*. – 2011. – Vol. 20. – P. 1576–1586.
13. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations / J.J. Gray [et al.] // *J. Mol. Biol.* – 2003. – Vol. 331, № 1. – P. 281–299.
14. FireDock: a web server for fast interaction re-nement in molecular docking / E. Maschiach [et al.] // *Nucleic Acids Res.* – 2008. – Vol. 36. – P. W229–W232.
15. Fernandez-Recio, J. ICM-DISCO docking by global energy minimization with fully flexible side-chains / J. Fernandez-Recio, M. Totrov, R. Abagyan // *Proteins*. – 2003. – Vol. 52. – P. 113–117.
16. Dominguez, C. HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information / C. Dominguez, R. Boelens, A. Bonvin // *J. Am. Chem. Soc.* – 2003. – Vol. 125. – P. 1731–1737.
17. Kozakov, D. Discrimination of near-native structures in protein-protein docking by testing the stability of local minima / D. Kozakov, O. Schueler-Furman, S. Vajda // *Proteins*. – 2008. – Vol. 72. – P. 993–1004.
18. Pierce, B. A combination of rescoring and refinement significantly improves protein docking performance / B. Pierce, Z. Weng // *Proteins*. – 2008. – Vol. 72. – P. 270–279.
19. Lorenzen, S. Monte Carlo refinement of rigid-body protein docking structures with backbone displacement and side-chain optimization / S. Lorenzen, Y. Zhang // *Proteins*. – 2007. – Vol. 66. – P. 2716–2725.
20. Cavasotto, C. Representing receptor flexibility in ligand docking through relevant normal modes / C. Cavasotto, J. Kovacs, R. Abagyan // *J. Am. Chem. Soc.* – 2005. – Vol. 127. – P. 9632–9640.
21. Fiorucci, S. Binding site prediction and improved scoring during flexible protein-protein docking with ATTRACT / S. Fiorucci, M. Zacharias // *Proteins*. – 2010. – Vol. 78. – P. 3131–3139.
22. Kastritis, P. Are scoring functions in protein-protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark / P. Kastritis, A. Bonvin // *J. Proteome Res.* – 2010. – Vol. 9. – P. 2216–2225.
23. A knowledge-based energy function for protein-ligand, protein-protein, and protein-DNA complexes / C. Zhang [et al.] // *J. Med. Chem.* – 2005. – Vol. 48. – P. 2325–2335.
24. Scoring by intermolecular pairwise propensities of exposed residues (SIPPER): A new efficient potential for protein-protein docking / C. Pons [et al.] // *J. Chem. Inf. Model.* – 2011. – Vol. 51. – P. 370–377.
25. Shen, M. Statistical potential for assessment and prediction of protein structures / M. Shen, A. Sali // *Protein Science*. – 2006. – Vol. 15. – P. 2507–2524.
26. Liu, S. DECK: Distance and environment-dependent, coarse-grained, knowledge-based potentials for protein-protein docking / S. Liu, I. Vakser // *BMC Bioinformatics*. – 2011. – Vol. 12. – P. 280–286.
27. Pierce, B. ZRANK: Reranking protein docking predictions with an optimized energy function / B. Pierce, Z. Weng // *PROTEINS: Structure, Function, and Bioinformatics*. – 2007. – Vol. 67. – P. 1078–1086.
28. Betts, M. An analysis of conformational changes on protein-protein association: implications for predictive docking / M. Betts, M. Sternberg // *Protein Eng.* – 1999. – Vol. 12. – P. 271–289.
29. Zacharias, M. Accounting for conformational changes during protein-protein docking / M. Zacharias // *Curr Opin Struct Biol.* – 2010. – Vol. 20, № 2. – P. 180–186.
30. Ding, F. Rapid flexible docking using a stochastic rotamer library of ligands / F. Ding, S. Yin, N.V. Dokholyan // *J. Chem. Inf. Model.* – 2010. – Vol. 50, № 9. – P. 1623–1632.
31. Principles of flexible protein-protein docking / N. Andrusier [et al.] // *Proteins*. – 2008. – Vol. 73, № 2. – P. 271–289.
32. Zhang, Y. Scoring function for automated assessment of protein structure template quality / Y. Zhang, J. Skolnick // *Proteins*. – 2004. – Vol. 57. – P. 702–710.
33. Dockground system of databases for protein recognition studies: Unbound structures for docking / Y. Gao [et al.] // *Proteins*. – 2007. – Vol. 69, № 4. – P. 845–851.

34. Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures / R. Menendez [et al.] // *Proteins*. – 2005. – Vol. 60. – P. 150–169.
35. Kolodny, R. Approximate protein structural alignment in polynomial time / R. Kolodny, N. Linial // *PNAS*. – 2004. – Vol. 101. – P. 12201–12206.
36. Kabsch, W.A. A solution for the best rotation to relate two sets of vectors / W.A. Kabsch // *Acta Cryst.* – 1976. – Vol. 32. – P. 922–923.
37. Horn, R. *Matrix Analysis* / R. Horn, C. Johnson. – Cambridge University Press, 1985.
38. Protein-protein docking benchmark 2.0: an update / J. Mintseris [et al.] // *Proteins*. – 2005. – Vol. 60. – P. 214–216.

Поступила 10.09.12

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: tatsiana.kirys@gmail.com
tuzikov@newman.bas-net.by*

T.V. Kirys, A.V. Tuzikov

**ALGORITHM OF PROTEIN INTERACTION PREDICTION
BASED ON STRUCTURAL HOMOLOGY**

An original algorithm of protein interaction prediction is suggested. It is based on structural similarity with experimentally determined protein-protein interfaces, selected from interface database. Two proteins are structurally aligned on protein-protein interface. The alignment is performed by a dynamic programming algorithm that maximizes correlation of the protein distance matrices.