

Математическое моделирование линейного докинга.

I. Определение участков связывания между белковыми молекулами

© К.Г. Куликов,¹ Т.В. Кошлан²

¹ Санкт-Петербургский политехнический университет Петра Великого,
195251 Санкт-Петербург, Россия

² Санкт-Петербургский государственный университет,
199034 Санкт-Петербург, Россия
e-mail: kulikov.kirill.g@gmail.com

(Поступило в Редакцию 29 ноября 2017 г.)

Представлены результаты математического моделирования взаимодействия различных аминокислотных последовательностей белков Mdm2, P53 и Nap1, согласно разработанным алгоритмам, с помощью которых были идентифицированы участки связывания различных белков при образовании ими биологических комплексов: P53–Mdm2, Mdm2–Mdm2, Nap1–Nap1. Разработанный подход позволит определять активные участки связывания полипептидных цепей различных белков, а также подбирать и синтезировать высокоселективные пептиды, которые будут связываться в активном центре белка и приводить к его активированию или ингибированию и блокировке его биологических функций.

DOI: 10.21883/JTF.2018.08.46301.2573

Введение

Для современной протеомики исследование и предсказание белковых взаимодействий являются задачами огромной важности, поскольку они определяют функцию белков на уровнях начиная от клетки и заканчивая уровнем всего организма. Для белков, структура которых известна, поиск межмолекулярных взаимодействий по известным данным о строении их третичной структуры сводится к задачам поиска геометрической комплементарности участков двух взаимодействующих молекулярных поверхностей и моделирования их контактов — так называемого молекулярного докинга [1].

Проблема молекулярного докинга является задачей алгоритма конформационного поиска, который сводится к перебору конформационного пространства, образованного биологическим комплексом за счет варьирования торсионных углов белковых молекул и перемещения такой белковой молекулы, как целого, относительно неподвижной структуры другого белка-мишени.

Современные алгоритмы конформационного поиска в большинстве случаев находят конформации, которые в целом близки к экспериментально найденным структурам, за сравнительно небольшое время. Тем не менее есть факторы, которые также оказывают существенное влияние на успешность докинга, часто не учитываемые в стандартных алгоритмах. Одним из таких факторов является конформационная подвижность белка-мишени. Диапазон подвижности может быть разным — начиная с небольшой „подстройки“ боковых цепей и заканчивая масштабными доменными движениями [2].

Эти движения играют существенную роль. На первый взгляд самым логичным решением этой проблемы является учет подвижности белка в программе докинга. К сожалению, современные вычислительные средства

не позволяют проводить такое моделирование за приемлемое время, так как молекула белка очень велика и учет подвижности по всем степеням свободы может привести к так называемому „комбинаторному взрыву“ (астрономическому увеличению числа возможных вариантов). Лишь в некоторых программах предусмотрена ограниченная подвижность сайтов связывания белка (как правило, на уровне небольшой адаптации конформаций боковых цепей остатков активного центра). Другой подход к этой проблеме состоит в осуществлении докинга с несколькими разными конформациями одного и того же белка с последующим выбором лучших решений из каждого запуска докинга. Третий подход — найти некоторую универсальную структуру белка-мишени, с участием которой докинг давал бы достаточно хорошие результаты для различных классов лигандов. При этом уменьшается число „пропущенных“ (но правильных) решений, однако сильно также возрастает и число неверных вариантов [3].

В отличие от приведенных выше подходов компьютерного моделирования в настоящей работе разработаны оригинальные математические алгоритмы, которые позволяют определять нахождение активных участков белков и выявлять стабильность различных участков белковых комплексов (линейный докинг) путем анализа матрицы потенциальной энергии электростатического взаимодействия между различными сайтами биологического комплекса. Такими биологическими комплексами выступают гомодимеры белков Nap1–Nap1, Mdm2–Mdm2 и гетеродимеры белков P53–Mdm2.

Статья состоит из нескольких частей. В первом, втором и третьем разделе дано описание структур и функций белков P53, Mdm2 и Nap1. Четвертый раздел посвящен разработанным алгоритмам. В пятом и шестом

разделах приводятся численные расчеты и их анализ. В приложении дана физическая постановка задачи.

Аминокислоты последовательности исследуемых белков P53, Nap1, Mdm2 брались из [4] с номерами P04637, P25293, Q00987 соответственно.

1. Структура и функции белка P53

Белок P53 был идентифицирован в 1979 г. Данный белок называют „стражем генома“, поскольку он регулирует огромное количество клеточных процессов, включая арест клеточного цикла, репарацию ДНК, апоптоз. В ответ на сигналы об отклонениях от нормы клеточных процессов и о наличии повреждений генетического аппарата P53 активируется, что приводит либо к ускорению процессов репарации, либо к остановке клеточного цикла и при сильном стрессовом стимуле — к апоптозу. Таким образом, этот белок предотвращает деление потенциально онкогенных клеток [5,6].

Биохимические реакции с участием P53 нарушаются при онкологических заболеваниях вследствие инактивации белка P53 или мутации TP53-гена, что приводит к накоплению поврежденных клеток и прогрессии онкологических заболеваний. Уровень и активность P53 напрямую регулируются через пост-трансляционную модификацию в нормальных клетках. Белок P53 является короткоживущим белком, который в дальнейшем подвергается убиквинтированию и протеасомной деградации [6].

Дикий тип P53 содержит 393 аминокислотных остатка (а.а.) и включает в себя несколько функциональных доменов. Такие домены представлены на рис. 1. Существуют два домена транскрипции TAD1 и TAD2 (1 а.а.—60 а.а.). Основная область включает специфический участок связывания с ДНК — домен DBD (102 а.а.—292 а.а.), в то время как С-концевая область содержит последовательности, связанные с ядерной локализацией и домен тетрамеризации TET [7].

Во время клеточного стресса происходит фосфорилирование белка P53, и его уровень концентрации в клетке возрастает. Это приводит к активированию генов белков, которые вовлечены в клеточный апоптоз, таких как p21 и Вах [8], а также белка Mdm2, который вовлечен в отрицательную регуляцию белка P53 [9].

Белок Mdm2 связывается с доменом транскрипции белка P53, находящимся на гибком N-конце белка P53 [10], что способствует регуляции количества белка P53 в клетке и приводит к его последующей деградации [11].

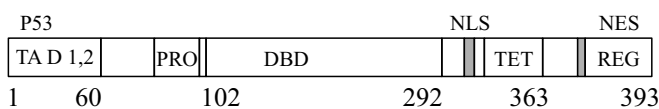


Рис. 1. Схема белка P53.

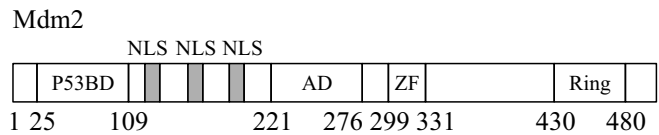


Рис. 2. Схема белка Mdm2.

2. Структура и функции белка Mdm2

Белок Mdm2 (Human Murine double minute 2) представляет собой фосфопротеин и является онкогеном. Ингибирование белка P53 происходит посредством связывания с белком Mdm2 и осуществляется двумя путями. Первый путь приводит к блокированию транскрипционного домена белка P53, второй путь — одновременно с блокированием функций транскрипционного домена происходит экспорт белка P53 из ядра к разрушению белка в протеасомах [9].

Человеческий белок Mdm2 имеет несколько доменов [12,13]. Первый домен P53BD со стороны N-конца Mdm2 содержит примерно 100 а.а., которые связываются с α-спиралью N-конца транскрипционного домена P53. Другие участки включают в себя сигнал ядерной локализации (NLS), сигнал ядерного экспорта (NES), центральный кислый домен AD, содержащий 40% остатков глутаминовой и аспарагиновой кислот. Далее расположен домен цинковый палец (ZF), за ним следует богатый цистеином домен Ring [9,14–15], который функционирует в качестве E3 лигазы (рис. 2).

Детальное изучение структуры связывания белков Mdm2 и P53 показало, что N-конец домена Mdm2 формирует глубокий гидрофобный желобок, с которым связывается транскрипционный домен белка P53. Прямое взаимодействие между двумя белками было локализовано относительно небольшого количества гидрофобного кармана на N-конце MDM2 и 15 а.а. амфипатического участка на N-конце P53 [16].

Белок P53 связывается с белком Mdm2 области аминокислотных остатков 18–110 белка Mdm2 и 15–29 белка P53 [17].

В ранее выполненных работах были исследованы механизмы образования гетеро- и гомодимеров Mdm2—Mdm2 и Mdm2—MdmX. В своей активной форме Mdm2 формирует гомодимеры или гетеродимеры с MdmX (Murine double minute 4), которые связываются при участии своих С-концевых Ring доменов. Последующий анализ выявил, что С-концевой участок Mdm2 отвечает как за связывание с Mdm2, так и за взаимодействие с MdmX [11]. Как было выявлено, при гомодимеризации учувствуют С-концевой и центральный кислый домен Mdm2, Ring-домен необходим для образования гомодимеров Mdm2—Mdm2.

В работе [18] было выявлено, что С-концевой фрагмент, содержащий Ring-домен, ведет себя как высокомолекулярный олигомерный комплекс.

3. Структура и функции белка Nap1

Белок Nap1 (Nucleosome assembly protein 1) является интегральным компонентом в создании и динамике эукариотического хроматина. Белок Nap1 перемещает гистоны в ядро, участвует в сборке нуклеосомы и способствует подвижности хроматина [19]. Nap1 широко используется в исследованиях по сборке хроматина, для структурных и функциональных исследований. Помимо его роли в содействии экспериментальному исследованию нуклеосом, высоко консервативный белок Nap1 является одним из наиболее изученных членов группы гистоновых шаперонов.

Рассмотрим более подробно структуру белка гистонового шаперона Nap1. В [10] было обнаружено, что белок Nap1 имеет хорошо структурированный центральный участок, тогда как концевые регионы N-конца и C-конца были в основном неупорядоченные. Центральная область представляет собой участок с 74 по 365 а.а.

Структуру белка Nap1 можно разделить на несколько разделов: неструктурированные N- и C-концы, домен I и домен II. Рассмотрим более подробно домен I, который отвечает за димеризацию Nap1. На рис. 3 представлено схематическое представление структур домена I и домена II. Домен I включает в себя $\alpha 1$ -, $\alpha 2$ -, $\alpha 3$ -спирали, которые отвечают за димеризацию белка Nap1. Домен II охватывает аминокислотные остатки с 181 по 370 а.а. Nap1 [10].

Во время процесса димеризации происходит взаимодействие в области домена I между длинными $\alpha 2$ -спиралями двух белков в противоположных направлениях. Далее димер стабилизируется петлями $\alpha 2-\alpha 3$, $\alpha 3$ -спиралью и петлями $\alpha 3-\alpha 4$, которые обертываются вокруг $\alpha 2$ -спирали партнера по димеризации. Схема димеризации двух белков Nap1–Nap1 показана на рис. 4. Интерфейс димера характеризуется главным образом гидрофобными взаимодействиями по всей длине вовлеченной аминокислотной последовательности $\alpha 2$ -спирали.

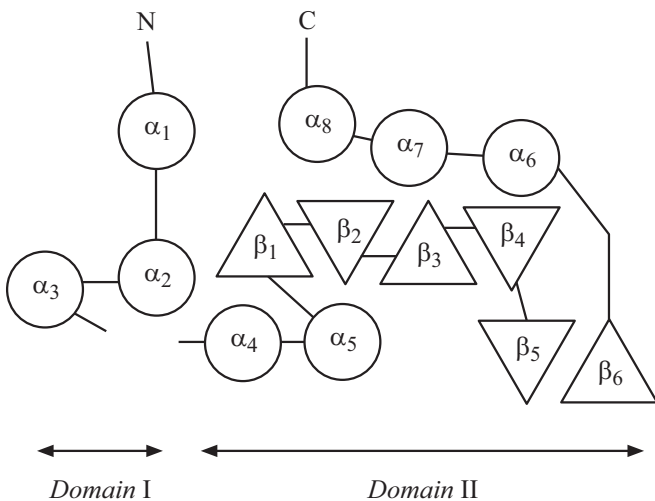


Рис. 3. Структура белка Nap1.

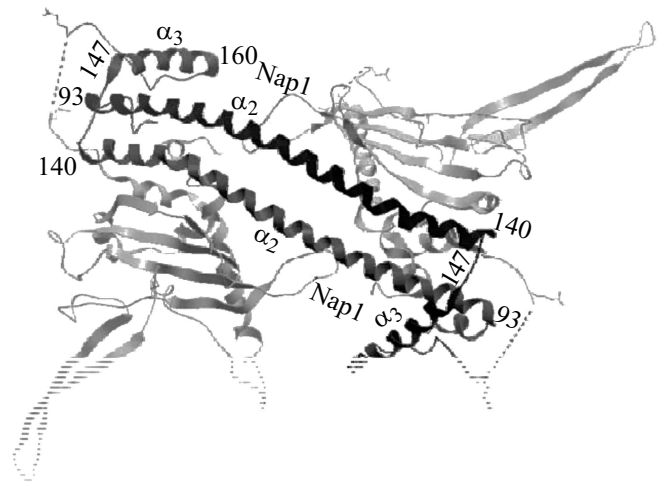


Рис. 4. Схема формирования гетеродимера Nap1–Nap1 двумя белками гистонового шаперона. Связывания двух мономерных белков происходит путем образования широкой области димеризации, охватывающей весь димер по диагонали. Два мономера Nap1 связываются в противоположных направлениях в области домена I, который представлен на рисунке $\alpha 2$ -спиралью и $\alpha 3$ -спиралью.

4. Описание алгоритмов

4.1. Алгоритм 1

Этот алгоритм был разработан для поиска белковых сайтов, ответственных за белковые взаимодействия.

При разработке этого алгоритма мы сделали следующие предположения:

- мы знаем короткую аминокислотную последовательность одного из белков, которая принимает активное участие в связывании с другим белком, с образованием большого количества близко расположенных взаимодействующих аминокислотных остатков, например формирование гомодимера Nap1–Nap1,

- мы не знаем активный сайт второго белка, который отвечает за связывание с короткой полипептидной последовательностью.

Таким образом, используя алгоритм 1, мы находим активный сайт на полипептидной последовательности второго белка.

Алгоритм 1 (рис. 5) представлен в виде двух массивов: одномерный массив 1 „DTEVNKL“ и одномерный массив 2 „ADHVNRTYOIK“, которые представляют собой аминокислотные последовательности белков P_1 и P_2 соответственно. Одномерный массив белка P_1 имеет меньшее количество аминокислотных остатков в своей полипептидной последовательности, чем одномерный массив белка P_2 . Более короткая аминокислотная последовательность одномерного массива P_1 движется вдоль более длинной аминокислотной последовательности одномерного массива белка P_2 с некоторым шагом, в нашем примере шаг равен одной аминокислоте. На каждом

этапе происходит образование короткого участка аминокислотной последовательности белка P_2 , равная длине более короткого одномерного массива белка P_1 . Каждый новый отрезок одномерного массива P_2 соответствует длине полипептидной цепи одномерного массива P_1 . Пять пар одномерных массивов P_1 и P_2 были успешно сформированы, когда одномерный массив P_1 сдвигался на один аминокислотный остаток вдоль одномерного массива P_2 .

$$\begin{aligned}
 P_{11} &= \text{DTEVNKL} & P_{21} &= \text{ADHVNRT}, \\
 P_{12} &= \text{DTEVNKL} & P_{22} &= \text{DHVNRTY}, \\
 P_{13} &= \text{DTEVNKL} & P_{23} &= \text{HVNRTYO}, \\
 P_{14} &= \text{DTEVNKL} & P_{24} &= \text{VNRTYOI}, \\
 P_{15} &= \text{DTEVNKL} & P_{25} &= \text{NRTYOIK}.
 \end{aligned}$$

Заметим, что одномерный массив P_1 остается неизменным во всех образованных парах одномерных массивов, т.е. $P_{11} = P_{12} = P_{13} = P_{14} = P_{15}$. Найдя все участвующие пары одномерных массивов, построим матрицу потенциального электростатического взаимодействия между их аминокислотными остатками. Эти матрицы будут иметь квадратную форму. Далее из каждой из этих матриц мы вычисляем значение $\lg(\text{cond}(W))$ и строим график зависимости $\lg(\text{cond}(W))$ в зависимости от порядкового номера аминокислотного остатка одномерного массива P_2 , где $(\text{cond}(W))$ — число обусловленности.

Таким образом, каждое результирующее значение $\lg(\text{cond}(W))$ будет соответствовать строго определенному сегменту одномерного массива P_2 . Стоит отметить, что на графике это значение откладывается напротив

D	T	E	V	N	K	L				
A	D	H	V	N	R	T	Y	O	I	K

$$\begin{aligned}
 P_{11} &= \begin{array}{|c|c|c|c|c|c|c|} \hline \text{D} & \text{T} & \text{E} & \text{V} & \text{N} & \text{K} & \text{L} \\ \hline \end{array} \\
 P_{21} &= \begin{array}{|c|c|c|c|c|c|} \hline \text{A} & \text{D} & \text{H} & \text{V} & \text{N} & \text{R} & \text{T} \\ \hline \end{array}
 \end{aligned}$$

$$\begin{aligned}
 P_{12} &= \begin{array}{|c|c|c|c|c|c|c|} \hline \text{D} & \text{T} & \text{E} & \text{V} & \text{N} & \text{K} & \text{L} \\ \hline \end{array} \\
 P_{22} &= \begin{array}{|c|c|c|c|c|c|} \hline \text{D} & \text{H} & \text{V} & \text{N} & \text{R} & \text{T} & \text{Y} \\ \hline \end{array}
 \end{aligned}$$

$$\begin{aligned}
 P_{13} &= \begin{array}{|c|c|c|c|c|c|c|} \hline \text{D} & \text{T} & \text{E} & \text{V} & \text{N} & \text{K} & \text{L} \\ \hline \end{array} \\
 P_{23} &= \begin{array}{|c|c|c|c|c|c|} \hline \text{H} & \text{V} & \text{N} & \text{R} & \text{T} & \text{Y} & \text{O} \\ \hline \end{array}
 \end{aligned}$$

$$\begin{aligned}
 P_{14} &= \begin{array}{|c|c|c|c|c|c|c|} \hline \text{D} & \text{T} & \text{E} & \text{V} & \text{N} & \text{K} & \text{L} \\ \hline \end{array} \\
 P_{24} &= \begin{array}{|c|c|c|c|c|c|} \hline \text{V} & \text{N} & \text{R} & \text{T} & \text{Y} & \text{O} & \text{I} \\ \hline \end{array}
 \end{aligned}$$

$$\begin{aligned}
 P_{15} &= \begin{array}{|c|c|c|c|c|c|c|} \hline \text{D} & \text{T} & \text{E} & \text{V} & \text{N} & \text{K} & \text{L} \\ \hline \end{array} \\
 P_{25} &= \begin{array}{|c|c|c|c|c|c|} \hline \text{N} & \text{R} & \text{T} & \text{Y} & \text{O} & \text{I} & \text{K} \\ \hline \end{array}
 \end{aligned}$$

Рис. 5. Схема алгоритма 1.

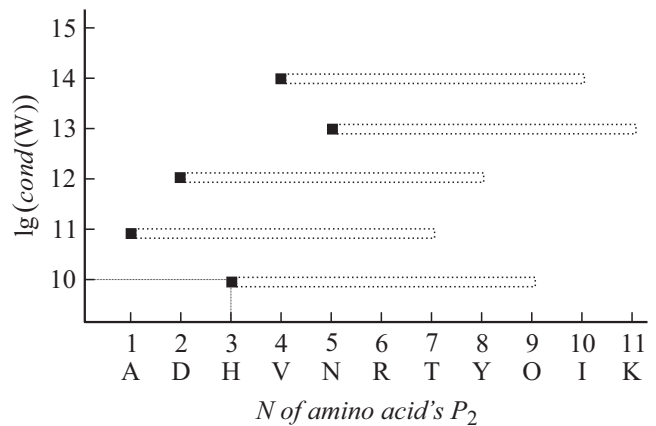


Рис. 6. График зависимости значений величины $\lg(\text{cond}(W))$ от порядкового номера одномерного массива P_2 .

первого аминокислотного остатка сегмента $P_{21}, P_{22}, P_{23}, P_{24}$ или P_{25} одномерного массива P_2 .

На рис. 6 приведен график зависимостей величин $\lg(\text{cond}(W))$ от порядкового номера отрезка полипептидной цепи участвующих одномерных массивов P_1 и P_2 . В данном примере, как мы видим из графика, наименьшее значение величины $\lg(\text{cond}(W))$ отвечает взаимодействию массивов $P_{13} = \text{DTEVNKL}$ и $P_{23} = \text{HVNRTYO}$. Пунктиром обозначен каждый из образованных участков массива P_2 , которые принимают участие в образовании биологического комплекса с массивом P_1 . Значение величины $\lg(\text{cond}(W))$ ставится напротив первого аминокислотного остатка отрезка массива P_2 .

Для обработки полученных данных будем предполагать, что за наиболее устойчивое положение взаимодействующих участков будет отвечать наибольшее количество близлежащих точек величины $\lg(\text{cond}(W))$ с его минимальным значением, при этом данный участок мы будем именовать кластером.

4.2. Алгоритм 2

Был разработан второй алгоритм обнаружения взаимодействующих областей белковых молекул. Схема взаимодействия показана на рис. 7. В этом алгоритме исследуются целые аминокислотные последовательности двух белков P_1 и P_2 с целью определения участков их взаимодействия. Для выбора взаимодействующих сайтов мы смещаем рамку определенного размера вдоль двух одномерных массивов белков P_1 и P_2 .

Согласно алгоритму 2, мы ищем участки взаимодействия для следующих пар белков: Nap1–Nap1, Mdm2–Mdm2.

Стоит отметить, что алгоритм 2 может быть использован для анализа взаимодействия двух белков, которые имеют идентичные участки взаимодействия при образовании димера.

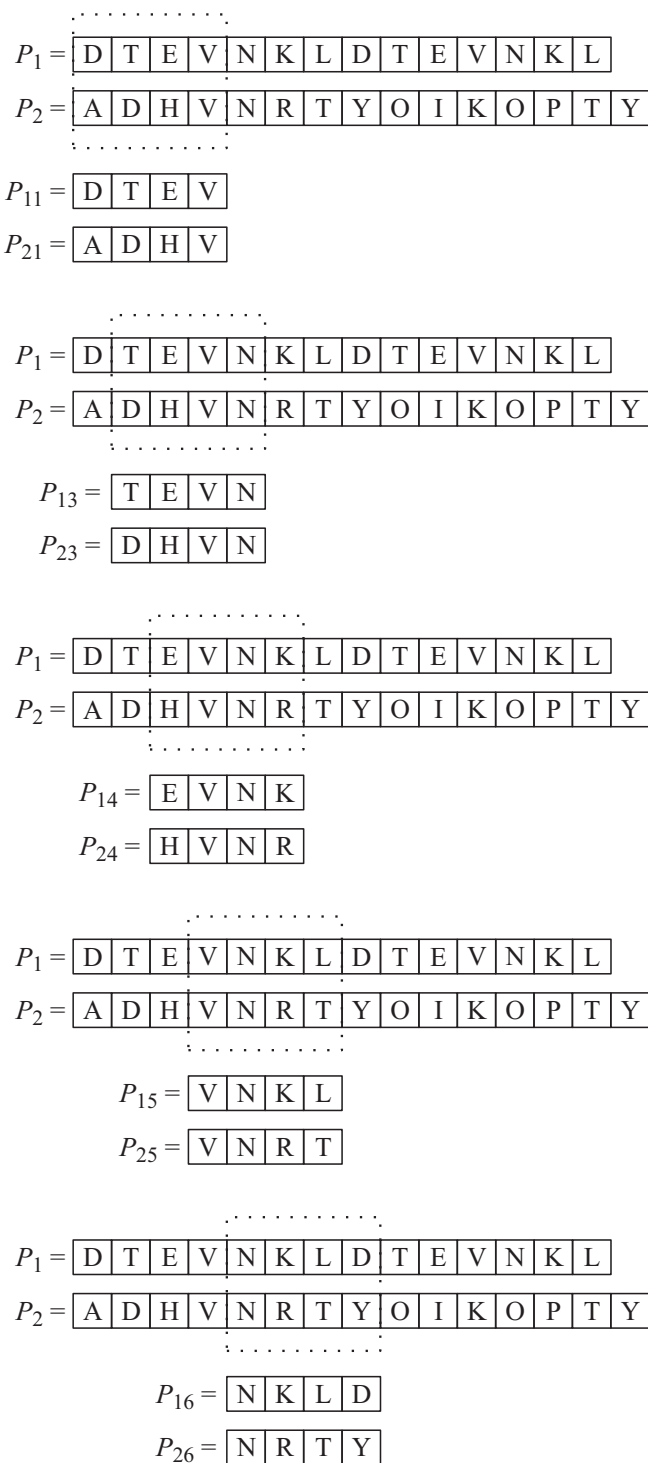


Рис. 7. Схема алгоритма 2. Пунктиром обозначена рамка, которая сдвигается вдоль полипептидных последовательностей, формируя одномерные массивы.

Найдя все участвующие пары одномерных массивов, мы строим матрицы потенциальной энергии электростатических взаимодействий. Эти матрицы также будут иметь квадратную форму. Далее, для всех полученных матриц вычисляется значение $\lg(\text{cond}(W))$ и строится

график зависимости $\lg(\text{cond}(W))$ от порядкового номера аминокислотного остатка участвующих одномерных массивов.

Для обработки полученных данных будем предполагать, что за наиболее устойчивое положение взаимодействующих участков будет отвечать наибольшее количество близлежащих точек величины $\lg(\text{cond}(W))$ с его минимальным значением.

5. Численное моделирование образования гетеродимеров и гомодимеров согласно алгоритму 1

В данном разделе были промоделированы взаимодействия коротких аминокислотных последовательностей белков, которые отвечают за образование гомодимеров и гетеродимеров с полноразмерными белками согласно разработанному алгоритму 1. Таким образом, целью данного раздела является возможность определения наиболее активных участков взаимодействия короткой полипептидной последовательности с полноразмерными белками.

Для анализа были выбраны следующие белковые пары: $P53_{(11-29)}-Mdm2$, $Mdm2_{(436-482)}-Mdm2$, $Nap1_{(81-140)}-Nap1$.

5.1. Численные расчеты взаимодействия $P53_{(11-29)}$ с $Mdm2$

Был взят короткий участок белка $P53_{(11-29)}$, который равномерно с шагом 1 а.а. сдвигался вдоль длинной полипептидной последовательности белка $Mdm2$.

При каждом таком сдвиге формировались 2 коротких одномерных массива, один из которых был аминокислотной последовательностью $P53_{(11-29)}$, а второй — аминокислотной последовательностью белка $Mdm2$ с длиной, равной первому массиву.

Из полученных одномерных массивов формировалась матрица потенциальной энергии электростатического взаимодействия и вычислялась величина $\lg(\text{cond}(W))$. Результаты численного моделирования представлены на рис. 8.

Как видно из представленного графика, полипептидная последовательность $Mdm2$ имеет основную область связывания с короткой последовательностью $P53_{(11-29)}$ со стороны N-конца, образуя небольшой кластер из минимальных значений величины $\lg(\text{cond}(W))$, с минимальным значением 17.884, которое приходится на 92 а.а. белка $Mdm2$. Два последующих минимальных значения величины $\lg(\text{cond}(W))$ приходятся на 90 и 85 а.а. и составили 17.916 и 17.999 соответственно.

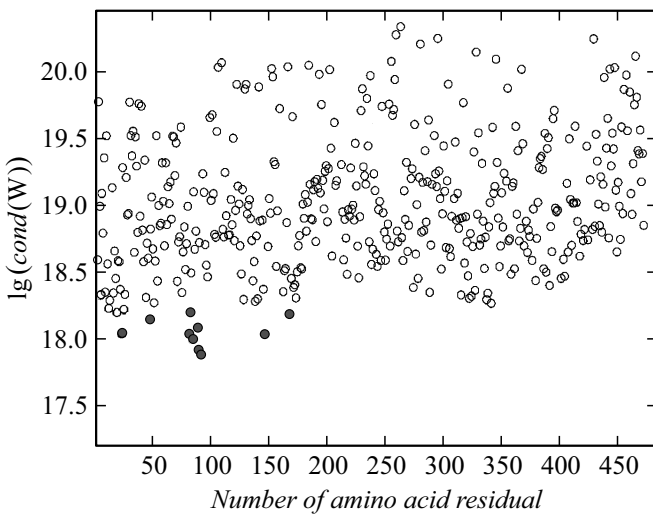


Рис. 8. Результаты численного расчета взаимодействия P53_(11–29) и Mdm2 согласно алгоритму 1.

Ниже приведен перечень аминокислотных последовательностей:

P53_(11–29)

[EPPLSQETFSDLWKLLPEN],

Mdm2_(92–110)

[SVKEHRKIYTMIRNLVVV],

Mdm2_(90–108)

[SFSVKEHRKIYTMIRNLV],

Mdm2_(85–103)

[LFGVPSFSVKEHRKIYTM].

Образовавшийся кластер при взаимодействии P53_(11–29) и белка Mdm2 в области N-конца представляет собой наибольшее количество скоплений точек с минимальными значениями величины $\lg(\text{cond}(W))$ с 85 по 110 а.а.

Введенное понятие „кластер“ интерпретируется нами, как образование стабильного биологического комплекса выбранными аминокислотными последовательностями на узком участке полипептидной цепи.

Полученный в нашей работе кластер наименьших значений величин $\lg(\text{cond}(W))$ в области с 85 по 110 а.а. белка Mdm2 при взаимодействии аминокислотной последовательности белка P53_(11–29) с белком Mdm2 является участком P53-связывающего домена Mdm2_(25–109), который играет ключевую роль в образовании димера P53–Mdm2 [9,16,17,20].

Таким образом, зная активную аминокислотную последовательность одного белка, отвечающую за образование димера, можно найти область полипептидной

цепи другого белка, с которым данная аминокислотная последовательность образует устойчивый биологический комплекс.

5.2. Численные расчеты взаимодействия Nap1_(81–140)–Nap1

В данном разделе рассмотрим результаты численного моделирования взаимодействия короткого участка белка Nap1_(81–140) с целой аминокислотной последовательностью белка гистонового шаперона Nap1.

Был выбран короткий участок белка Nap1_(81–140), который принимает активное участие в димеризации белка гистонового шаперона Nap1 и проведен численный расчет взаимодействия этого участка с полипептидной последовательностью целого белка Nap1.

Короткая последовательность Nap1_(81–140) перемещалась вдоль полипептидной последовательности белка Nap1 с шагом, равным 1 а.а.

При этом формировались одномерные массивы длиной 60 а.а., один из которых был представлен в виде одномерного массива, характеризующего аминокислотную последовательность Nap1_(81–140), а второй — одномерный массив с поочередно меняющимися аминокислотными остатками белка Nap1.

Для каждой пары полученных одномерных массивов формировалась матрица потенциальной энергии электростатического взаимодействия, вычислялась величина $\lg(\text{cond}(W))$. Результаты численного моделирования представлены на рис. 9.

Как видно из представленного графика, полипептидная последовательность белка гистонового шаперона Nap1 имеет несколько участков связывания с короткой последовательностью Nap1_(81–140). Наименьшее значение величины $\lg(\text{cond}(W))$ приходится на 134 а.а. и составило 18.625. При этом найденному минимальному

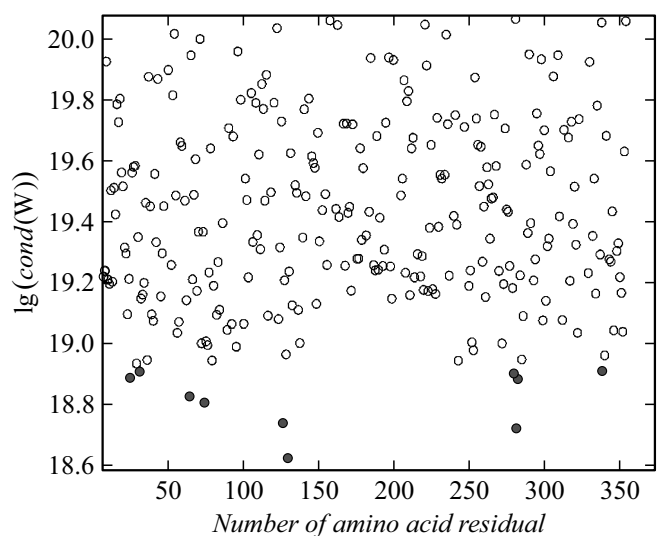


Рис. 9. Результаты численного расчета взаимодействия Nap1_(81–140) и Nap1 согласно алгоритму 1.

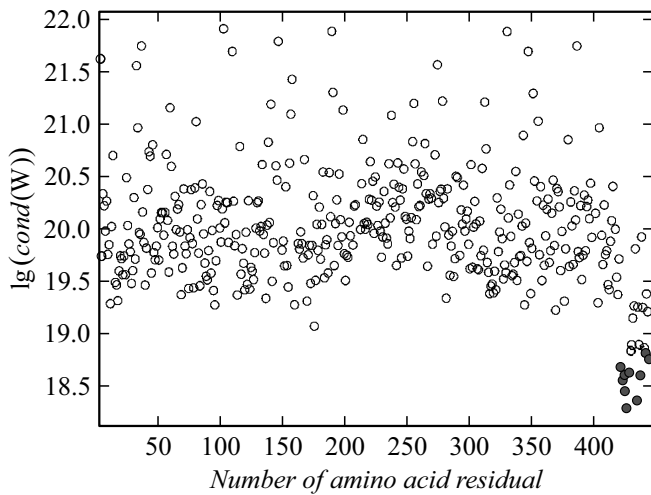


Рис. 10. Результаты численного расчета взаимодействия Mdm2_(436–482) и Mdm2 согласно алгоритму 1.

значению величины $\lg(\text{cond}(W))$ отвечают взаимодействующие отрезки белка:

Nap1_(81–140)
 [DSGYVGGLPKNVKEKLLSLKTLQSELFVEVEKEFQV
 EMFELENKFLQKYKPIWEQRSRIIS],

Nap1_(134–193)
 [QRSRIISGQEQPKPEQIAKQGEIVESLNETELLVDEE
 EKAQNDSEEEQVKGIPSWLTAL].

Отметим, что полученный участок полипептидной последовательности белка гистонового шаперона Nap1_(134–193) является участком домена I, который принимает участие в димеризации белка Nap1–Nap1 [10].

Таким образом, полученный в нашей работе участок взаимодействия с Nap1_(134–193) соответствует экспериментально найденному участку взаимодействия домену I, который играет главную роль в димеризации белка Nap1 [10].

5.3. Численный расчет взаимодействия Mdm2_(436–482) с Mdm2

Был выполнен численный расчет взаимодействия короткой последовательности домена Ring белка Mdm2_(436–482) с полной аминокислотной последовательностью белка Mdm2.

При этом короткая последовательность белка Mdm2_(436–482) сдвигалась вдоль длинной последовательности белка Mdm2 с шагом 1 а.а.

В результате для каждой пары полученных одномерных массивов формировалась матрица потенциальной энергии электростатического взаимодействия, вычислялась величина $\lg(\text{cond}(W))$. Результаты численного расчета представлены на рис. 10.

Как видно из представленного графика, совокупность минимальных значений образует кластер со стороны С-конца белка Mdm2 с минимальным значением величины $\lg(\text{cond}(W))$, равным 18.295, приходящимся на 427 а.а. белка Mdm2. Два последующих наименьших значения величины $\lg(\text{cond}(W))$ приходятся на 435 а.а. и на 426 а.а. и равняются 18.365 и 18.454 соответственно. Приведем данные аминокислотные последовательности, отвечающие минимальным значениям величины $\lg(\text{cond}(W))$ в кластере

Mdm2_(427–473) :
 [ESSLPLNAIEPCVICQGRPKNGCIVHGKTGHLMACFT
 CAKCLKKRNK],

Mdm2_(435–481) :
 [IEPCVICQGRPKNGCIVHGKTGHLMACFTCAKCLKK
 RNKPCPVCQRQ],

Mdm2_(426–472) :
 [VESSLPLNAIEPCVICQGRPKNGCIVHGKTGHLMACF
 TCAKCLKKRN],

а также аминокислотную последовательность Mdm2_(436–482), являющуюся Ring-доменом:

[EPCVICQGRPKNGCIVHGKTGHLMACFTCAKCLKK
 RNKPCPVCQRPI].

Таким образом, можно сделать вывод, что последовательность Mdm2_(436–482), которая играет активную роль в димеризации белка Mdm2, наиболее склонна образовывать стабильные биологические комплексы в области С-конца белка Mdm2. Данный результат находится в хорошем согласии с [11,18].

6. Численное моделирование образования гомодимеров согласно алгоритму 2

В данном разделе были выполнены численные расчеты с целью определения активных участков белков Mdm2 и Nap1 при образовании гомодимеров Nap1–Nap1 и Mdm2–Mdm2.

Мы брали целые аминокислотные последовательности одинаковых белков и сдвигали рамку заданного размера, равную 40 а.а., вдоль двух одномерных массивов, описывающих одинаковые аминокислотные последовательности Nap1 и Nap1, Mdm2 и Mdm2 с шагом, равным 1 а.а., с целью выявить наиболее активные участки взаимодействия исследуемых белков.

6.1. Численный расчет взаимодействия полипептидных последовательностей белка Nap1

В данном разделе рассмотрим результаты выполненного численного расчета для двух одинаковых полипеп-

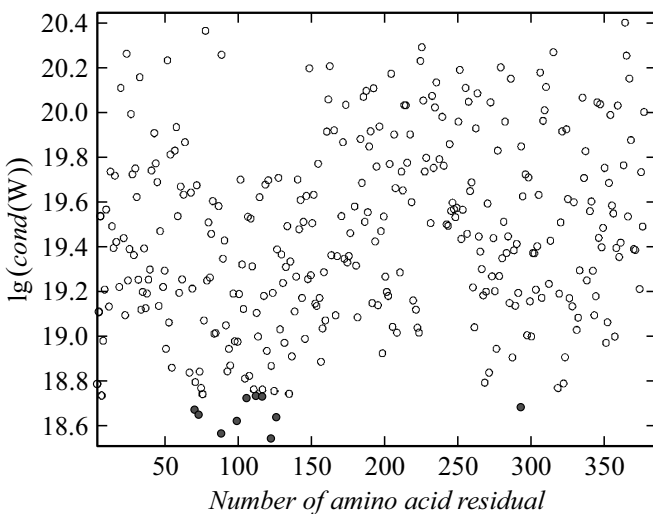


Рис. 11. Результаты численного расчета взаимодействия двух белков Nap1 согласно алгоритму 2.

тидных цепей белка Nap1 в направлении от N-конца к C-концу согласно алгоритму 2 для рамки сдвига, равной 40 а.а., вдоль двух полипептидных последовательностей белка Nap1 с шагом, равным 1 а.а.

Отметим, что образование гомодимера Nap1–Nap1 осуществляется за счет связывания двух одинаковых участков в области домена I каждого белка с 74 по 180 а.а. Главный участок димеризации двух белков принадлежит α 2-спирали с 90 по 140 а.а. каждого белка гистонового шаперона Nap1.

Как видно из графика на рис. 11, в области с 88 по 161 а.а. образуется кластер с наименьшими значениями $\lg(\text{cond}(W))$, при этом наименьшее из этих значений соответствует величине 18.541, приходящееся на 122 а.а. Два других наименьших значения величины $\lg(\text{cond}(W))$ приходятся на 88 и на 99 а.а. со значениями 18.559 и 18.619 соответственно.

Ниже приведен перечень аминокислотных последовательностей с тремя наименьшими значениями в кластере значений величины $\lg(\text{cond}(W))$:

Nap1_(122–161) :
[NKFLQKYKPIWEQRSRIISGQEQPKPEQIAKGQ
EIVESLN],

Nap1_(88–127) :
[LPKNVKEKLLSLKTLQSELFEVEKEFQVEMFE
LENKFLQK],

Nap1_(99–138) :
[LKTQSELFEVEKEFQVEMFELENKFLQKYKP
IWEQRSRI].

Полученный в нашей работе кластер минимальных значений в области с 88 по 161 а.а. полипептидной

последовательности белка Nap1 принадлежит домену I и перекрывает главный участок димеризации белка Nap1_(90–140) [10].

Таким образом, полученные численные результаты находятся в хорошем согласии в ранее проведенной экспериментальной работой [10], в которой указывалось, что домен I, находящийся с 74 по 180 а.а., связывается с таким же доменом I другого белка.

6.2. Численный расчет взаимодействия двух полипептидных цепей белка Mdm2

В этом разделе был выполнен анализ численного расчета образования гомодимера Mdm2–Mdm2 согласно алгоритму 2.

Мы брали аминокислотные последовательности белков Mdm2 и Mdm2, формировали соответствующие их аминокислотным последовательностям одномерные массивы и сдвигали рамку заданного размера, равную 40 а.а., вдоль двух одномерных массивов с шагом, равным 1 а.а., с целью выявить наиболее активные участки взаимодействия исследуемых белков. Результаты полученных численных расчетов представлены на рис. 12.

Как мы видим из графика на рис. 12, результатом численного расчета взаимодействия двух полипептидных цепей белка Mdm2, согласно алгоритму 2, является кластер наименьших значений в области C-конца двух белков Mdm2. Наименьшее значение величины $\lg(\text{cond}(W))$ приходится на 425 а.а. и составило 18.232. Два последующих наименьших значения $\lg(\text{cond}(W))$ пришлись на 436 и 450 а.а. и составили 18.248, 18.272 соответственно.

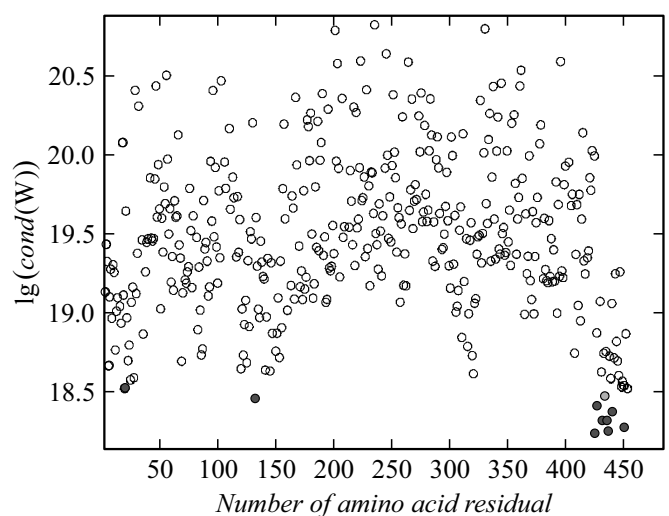


Рис. 12. Результаты численного расчета взаимодействия двух белков Mdm2 согласно алгоритму 2.

Приведем перечень аминокислотных последовательностей:

Mdm2_(425–464) :
[SVESLPLNAIEPCVICQGRPKNGCIVHGKTG
HLMACFTC],

Mdm2_(436–474) :
[EPCVICQGRPKNGCIVHGKTGHLMACFTCAKK
LKKRNKPC],

Mdm2_(450–489) :
[IVHGKTGHLMACFTCAKKLKKRNKPCPVCQRQ
IQMIVLTY].

Отметим, что полученный кластер наименьших значений в ходе численного расчета перекрывает домен Ring белка Mdm2, ответственного за гомодимеризацию белка

Mdm2_(436–482) :
[LDQDSVSDQFSVEFEVESLSEEDYSLSEEGQELSD
EDDEVYQVTVYQ].

Данный результат находится в хорошем согласии с ранее проведенными экспериментами [11,18].

Заключение

В работе было выполнено математическое моделирование взаимодействия различных аминокислотных последовательностей белков Mdm2, P53 и Nap1 согласно разработанным алгоритмам 1 и 2.

Результаты расчетов, выполненные для белка Nap1 при взаимодействии с другим белком Nap1 и последующим образованием гомодимера Nap1–Nap1, согласно алгоритмам 1 и 2, продемонстрировали практически идентичные найденные полипептидные последовательности с помощью двух данных алгоритмов.

Так, с помощью алгоритма 1 было выявлено минимальное значение величины $\lg(\text{cond}(W))$ при взаимодействии аминокислотных последовательностей Nap1_(81–140)–Nap1_(134–193), при этом последовательность Nap1_(81–140) была задана изначально, а последовательность Nap1_(134–193) была найдена в ходе численного моделирования.

При выполнении численного расчета, согласно алгоритму 2 был выявлен кластер минимальных значений величины $\lg(\text{cond}(W))$ в области полипептидной последовательности белка Nap1 с 88 по 161 а.а. Таким образом, найденные аминокислотные последовательности в ходе численного расчета, согласно алгоритмам 1 и 2, являются участками домена I.

Данный результат находится в хорошем согласии с ранее выполненной экспериментальной работой, в которой указывается, что участок, ответственный за димеризацию белка Nap1–Nap1, находится в области домена I белка Nap1 с 74 по 180 а.а. [10].

Результаты, полученные в ходе выполненного численного моделирования для белка Mdm2 при его гомодимеризации Mdm2–Mdm2, согласно алгоритмам 1 и 2, демонстрируют практически идентичные найденные аминокислотные участки полипептидной цепи.

Так, было обнаружено образование кластера в области С-конца Mdm2_(426–481), согласно выполненным расчетам по алгоритму 1.

При выполнении расчетов, согласно алгоритму 2, был выявлен кластер минимальных значений в области С-конца белка Mdm2_(425–489).

Найденные полипептидные последовательности являются практически одинаковыми аминокислотными последовательностями белка Mdm2.

Установленные участки Mdm2_(426–481) и Mdm2_(425–489) являются участком домена Ring, который, как было выявлено ранее, играет существенную роль в образовании гомодимера Mdm2–Mdm2 [11,18].

Образование биологического комплекса белками P53 и Mdm2 было промоделировано согласно алгоритму 1.

При взаимодействии двух выбранных белков происходит формирование гидрофобного желобка белком Mdm2, с которым связывается короткий участок белка P53.

Результаты численного моделирования при участии P53_(11–29) и белка Mdm2 позволили идентифицировать участок связывания белка Mdm2 с выбранными аминокислотными последовательностями участка белка P53_(11–29). Таким участком является участок белка Mdm2_(85–110), который образует кластер, состоящий из наименьших значений величины $\lg(\text{cond}(W))$.

Найденный в ходе моделирования участок Mdm2_(85–110) находится в области P53-связывающего домена белка Mdm2_(25–109) [16,17].

Таким образом, в настоящей работе были разработаны два алгоритма, с помощью которых были идентифицированы участки связывания различных белков при образовании ими биологических комплексов: P53–Mdm2, Mdm2–Mdm2, Nap1–Nap1.

Для поиска новых участков связывания полипептидов в димеры необходимо проводить более масштабные вычислительные исследования по поиску активных участков связывания биологических молекул. При этом рекомендуется использовать примерно одинаковые длины коротких одномерных массивов вектора P_1 для случая алгоритма 1, а также одинаковые размеры рамки сдвига при использовании алгоритма 2 для последующего корректного сравнения полученных результатов между собой.

В перспективе это позволит определять активные участки связывания полипептидных цепей различных белков, а также подобрать и синтезировать высокоселективные пептиды, которые будут связываться в активном центре белка и приводить к его активированию или ингибированию и блокировке его биологических функций.

Приложение

П1. Описание физической модели

Опишем физическую модель попарного электростатического взаимодействия между аминокислотными последовательностями различных белков. Отметим, что электростатическое взаимодействие является основным при формировании различных биологических комплексов [21].

В данной модели белок мы представляем как свободно-сочлененную полиаминокислотную последовательность [22], а каждая аминокислота представлена в виде равномерно заряженной сферы со своим значением радиуса.

При исследовании взаимодействия заряженных белковых молекул использовался ряд приближений:

- 1) энергия взаимодействия белков определяется только силами электростатического взаимодействия;
- 2) белковая молекула моделируется как соединенные между собой аминокислотные остатки, имеющие шарообразную форму со своим радиусом;
- 3) каждый аминокислотный остаток белка представляется как равномерно заряженная сфера.

Размер радиуса сферы каждого аминокислотного остатка был взят из [23].

Отметим, что способность взаимодействовать различных аминокислотных остатков сильно зависит от их окружения, что обусловлено их полярной и неполярной частями молекулы.

Поскольку в настоящей работе мы рассматриваем задачу попарного электростатического взаимодействия между аминокислотными последовательностями различных белков, сделаем следующее предположение: разделим взаимодействия между 20 аминокислотными остатками на 5 классов от $0.15e$ до $1e$ согласно [24].

Первому классу мы присвоили одноименные заряды по $0.15e$ каждый. Второму классу мы присвоили одноименные заряды по $0.25e$ каждый. Третьему классу были присвоены разноименные заряды $-0.45e$ и $+0.45e$ каждый. Четвертому классу были присвоены разноименные заряды $-0.65e$ и $+0.65e$ каждый. Пятому классу были присвоены разноименные заряды $-1e$ и $+1e$ каждый, где e — заряд электрона.

Таким образом, мы присваиваем аминокислотным остаткам заряд равный по величине заряду электрона и заряд меньший заряда электрона. Отметим, что квадрат волновой функции, которая описывает состояние электрона в многоцентровой модели, определяет плотность вероятности нахождения электрона или плотность электронного облака, которая характеризует неодинаковую вероятность нахождения электрона в выбранном диапазоне электронного облака в многоатомном аминокислотном остатке. Отсюда следует, что перераспределение электронных зарядов в многоатомных аминокислотных остатках вполне может приводить к тому, что вероятность нахождения валентного электрона в окрестности

остатка окажется меньше единицы, а среднее значение заряда остатка — меньше, чем заряд электрона.

Расстояния между двумя взаимодействующими аминокислотными остатками соседних белков были определены из работ [23,25–27].

П2. Задача электростатического взаимодействия двух проводящих сфер

Пусть имеются две проводящие сферы с радиусами R_1 , R_2 и зарядами Q_1 и Q_2 , расположенные на расстоянии L между центрами. Тогда, следуя работам [28–30], запишем выражение для потенциальной энергии электростатического взаимодействия сфер между их центрами:

$$W_q = \left[\frac{1 + \gamma}{2\alpha} \right] \frac{\alpha^2 c_{11} - 2\alpha c_{12} + c_{22}}{c_{11}c_{22} - c_{12}^2}, \quad (1)$$

где Q_1 и Q_2 — заряды первой и второй сферы $\gamma = R_2/R_1$, $\alpha = Q_2/Q_1$, c_{11} , c_{12} , c_{22} — емкостные коэффициенты, которые определяются следующим образом [31]:

$$c_{11} = 4\pi\epsilon_0 R_1 \gamma \sinh(\beta) \sum_{n=1}^{\infty} [\gamma \sinh(n\beta) + \sinh[(n-1)\beta]]^{-1}, \quad (2)$$

$$c_{22} = 4\pi\epsilon_0 R_1 \gamma \sinh(\beta) \sum_{n=1}^{\infty} [\sinh(n\beta) + \gamma \sinh[(n-1)\beta]]^{-1}, \quad (3)$$

$$c_{12} = -4\pi\epsilon_0 R_1 \gamma \frac{\sinh(\beta)}{(1 + \gamma)h} \sum_{n=1}^{\infty} [\sinh(n\beta)]^{-1}. \quad (4)$$

При этом за единицу длины выбрана величина $(R_1 + R_2)$, так что h есть безразмерное расстояние между центрами сфер, которое определяется как $h = L/(R_1 + R_2)$, а параметр β связан с расстоянием между центрами сфер следующим образом [30]

$$\cosh(\beta) = \frac{h^2(1 + \gamma)^2 - (1 + \gamma^2)}{2\gamma}. \quad (5)$$

Отметим, что емкостные коэффициенты (2)–(4) определены в единицах R_1/k , $k = 1/4\pi\epsilon_0$, ϵ_0 — электрическая постоянная. Тогда в качестве единицы измерения энергии выберем величину $kQ_1Q_2/(R_1 + R_2)$, в этом случае для безразмерной энергии электростатического взаимодействия сфер и соответственно для емкостных коэффициентов c_{11} , c_{12} , c_{22} получим следующие выражения:

$$W_q = \left[\frac{1 + \gamma}{2\alpha} \right] \frac{\alpha^2 c_{11} - 2\alpha c_{12} + c_{22}}{c_{11}c_{22} - c_{12}^2}, \quad (6)$$

$$c_{11} = \gamma \sinh(\beta) \sum_{n=1}^{\infty} [\gamma \sinh(n\beta) + \sinh[(n-1)\beta]]^{-1}, \quad (7)$$

	A	R	Q	T
V	W_{11}	W_{12}	W_{13}	W_{14}
A	W_{21}	W_{22}	W_{23}	W_{24}
R	W_{31}	W_{32}	W_{33}	W_{34}
Q	W_{41}	W_{42}	W_{43}	W_{44}

Рис. 13. Представление матрицы потенциальной энергии электростатического взаимодействия $W_{i,j}$ $i = \overline{1,4}$, $j = \overline{1,4}$ двух белков P_1 и P_2 .

$$c_{22} = \gamma \sinh(\beta) \sum_{n=1}^{\infty} [\sinh(n\beta) + \gamma \sinh[(n-1)\beta]]^{-1}, \quad (8)$$

$$c_{12} = -\gamma \frac{\sinh(\beta)}{(1+\gamma)h} \sum_{n=1}^{\infty} [\sinh(n\beta)]^{-1}. \quad (9)$$

Проведем следующие преобразования для выражений (7)–(9) по аналогии с [28,32]. Введем переменную $z = \exp(-\beta)$, тогда выражения (7)–(9) примут следующий вид:

$$c_{11} = 2\gamma \sqrt{\cosh^2(\beta) - 1} \times \sum_{n=1}^{\infty} \frac{z^n}{(1-z^{2n}) \left[(\gamma + \cosh(\beta)) - \sqrt{\cosh^2(\beta) - 1} (1+z^{2n}) / (1-z^{2n}) \right]},$$

$$c_{22} = 2\gamma \sqrt{\cosh^2(\beta) - 1} \times \sum_{n=1}^{\infty} \frac{z^n}{(1-z^{2n}) \left[(1 - \cosh(\beta)) - \gamma \sqrt{\cosh^2(\beta) - 1} (1+z^{2n}) / (1-z^{2n}) \right]},$$

$$c_{11} = -\frac{2\gamma \sqrt{\cosh^2(\beta) - 1}}{h(1+\gamma)} \sum_{n=1}^{\infty} \left[\frac{z^{2n}}{(1-z^{2n})} \right],$$

где $\cosh(\beta)$ определен выражением (5).

Полученные в результате решения этой задачи значения потенциальной энергии попарного электростатического взаимодействия между соответствующими аминокислотными остатками, которые мы представляем в виде заряженных сфер, записываются в матрицу W (рис. 13).

В нашей модели будем считать, что каждый аминокислотный остаток одной молекулы белка может взаимодействовать с любым другим аминокислотным остатком другого белка.

Для анализа биохимических процессов используем понятие числа обусловленности $\text{cond}(W)$ матрицы W . Для его вычисления используем SVD-разложение, тогда имеем [33]

$$\text{cond}(W_k) = \frac{\sigma_{\max}(W_k)}{\sigma_{\min}(W_k)},$$

где $\sigma_{\max}(W_k)$, $\sigma_{\min}(W_k)$ — наибольшие и наименьшие сингулярные числа матрицы потенциальной энергии попарного электростатического взаимодействия между пептидами W_k , k — число матриц.

В рассматриваемой физической постановке задачи число обусловленности будет характеризовать степень устойчивости конфигурации биологического комплекса. При этом для выбора более устойчивого биохимического соединения между белками мы выбираем матрицу потенциальной энергии электростатического взаимодействия с наименьшим значением числа обусловленности (см. разд. 8.3).

ПЗ. Физическая интерпретация числа обусловленности

В данном случае, когда речь идет только о физической интерпретации числа обусловленности, допустимо рассмотреть для простоты цепочки, состоящие не из сфер, а из точечных зарядов.

Потенциальная энергия взаимодействия двух зарядов q_1 и q_2 , находящихся на стационарном расстоянии h друг от друга, равна

$$W_0 = \frac{q_1 q_2}{\kappa h},$$

где $\kappa = 4\pi\epsilon_0\epsilon$ в СИ, ϵ_0 — диэлектрическая постоянная, ϵ — диэлектрическая проницаемость среды.

При внешней деформации такого парного образования будет происходить изменение расстояния δh ; положим деформации малыми: $\delta h \ll h$. Тогда изменение потенциальной энергии имеет вид

$$\delta W_0 = -\frac{q_1 q_2}{\kappa h^2} \delta h = -W_0 \frac{\delta h}{h}.$$

Опустим знак минуса: хотя в общем случае знак δW_0 определяется знаками зарядов и деформаций, но в рассматриваемом случае отклонения от стационарного состояния δW_0 всегда положителен ($\delta W_0 > 0$).

Суммарная потенциальная энергия i -й возможной конфигурации двух взаимодействующих деформированных цепочек, пересекающихся в j -х местах, станет

$$W_{\Sigma}^i = \sum_j W_{0j}^i \left[1 + \frac{\delta h_j^i}{h_j^i} \right] = W_0^i + \sum_j W_{0j}^i \frac{\delta h_j^i}{h_j^i}.$$

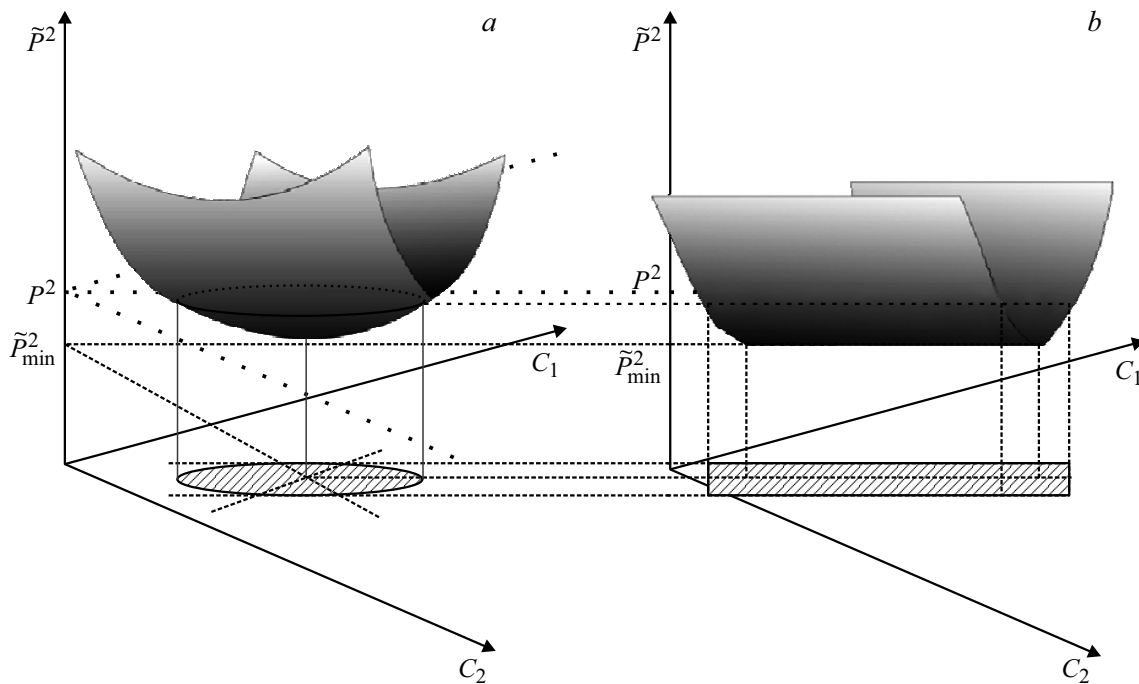


Рис. 14. Геометрическая интерпретация поведения невязки МНК в зависимости от величин параметров.

Таким образом, деформации (пока произвольные, хотя и малые) расстояний между взаимодействующими зарядами вызовут увеличение общей потенциальной энергии i -й конфигурации двух взаимодействующих цепочек, составляющих биологический комплекс,

$$\delta W_{\Sigma}^i = \sum_j W_{0j}^i \frac{\delta h_j^i}{h_j^i}, \quad \delta W_{\Sigma}^i = W_{\Sigma}^i - W_0^i. \quad (10)$$

Отсюда имеем систему уравнений

$$y_i = \sum_j c_j \varphi_j^i,$$

где $i = 1, 2, \dots, n$, n — количество возможных конфигураций двух пересекающихся цепочек, а $j = 1, 2, \dots, m$, где m — количество пересечений цепочек для одной конкретной i -й конфигурации. Из (10) вытекает, что роль c_j выполняют относительные удлинения расстояний при деформации

$$\frac{\delta h_j^i}{h_j^i},$$

роль свободных членов y_i — изменение суммарной потенциальной энергии при деформации δW_{Σ}^i , а роль матричных элементов φ_j^i — парные потенциальные энергии стационарного состояния W_{0j}^i .

Таким образом, из (10) мы в состоянии рассчитать изменение общей потенциальной энергии i -й конфигурации δW_{Σ}^i биоконплекса в деформированном состоянии при произвольных величинах парных удлинений δh_j^i .

Реальные деформации связей между составляющими биоконплекса (как, собственно, и между отдельными

элементами этих составляющих) происходят вследствие хаотических тепловых соударений молекул. При температуре $T > 0$ наиболее вероятное изменение энергии биологического комплекса δW_{Σ}^i не будет произвольным; оно составит величину порядка kT (k — постоянная Больцмана, T — абсолютная температура).

Поэтому можно ставить обратную задачу: найти наиболее вероятные величины удлинений, обеспечивающие минимум уклонения δW_{Σ}^i от kT . При этом в (10) роль y_i станет играть $kT \forall i$.

Однако, во-первых, очевидно, что $n > m$ и что система линейных алгебраических уравнений (СЛАУ) (10) будет переопределенной. Во-вторых, величины

$$\frac{\delta h_j^i}{h_j^i}$$

имеют не детерминированный, а вероятностный характер, и нас интересуют наиболее вероятные (или максимально правдоподобные) их величины. Поэтому приходим к задаче метода наименьших квадратов (МНК).

Пусть даны две базисные функции φ_1, φ_2 и два параметра c_1, c_2 . Рассмотрим невязку в МНК

$$p^2(c_1, c_2) = \sum_{i=1}^k [y_i - c_1\varphi_1(x_i) - c_2\varphi_2(x_i)]^2. \quad (11)$$

Тогда величина $p^2(c_1, c_2)$ будет характеризовать потенциальную энергию деформированного биологического комплекса (конгломерата), т.е. подвергнувшегося внешнему воздействию (например, вследствие термических

столкновений при $T > 0$), а c_1, c_2 есть относительные удлинения связей.

Отметим, что в пространстве параметров c_j , где $j = \overline{1, 2}$, т.е. в координатах c_1 и c_2 функция $p^2(c_1, c_2)$ при малых воздействиях представляет собой параболоид (рис. 14, а).

Под p_{\min} мы будем понимать потенциальную энергию стационарного состояния биологического комплекса, т.е. когда реализуется минимум его потенциальной энергии, тогда под p понимается средняя потенциальная энергия биологического комплекса при наличии воздействия, т.е.

$$p - \tilde{p}_{\min} \approx kT.$$

Пусть $p \geq \tilde{p}_{\min}$ (если имеет место обратное неравенство, то это означает, что вид аппроксимационной функции выбран неудачно). Тогда сечение параболоида плоскостью $\tilde{p}^2 = p^2$ (поскольку $p = \text{const}(c_1, c_2)$) дает связь между параметрами c_1, c_2 :

$$\tilde{p}^2(c_1, c_2) = \sum_{i=1}^k [y_i - c_1\varphi_1(x_i) - c_2\varphi_2(x_i)]^2,$$

т.е. эллипс в координатах c_1, c_2 (рис. 14, а). Отношение длин полуосей этого эллипса (большой к малой) равно мере обусловленности системы линейных алгебраических уравнений вида (11) [34]. Чем хуже обусловлена матрица, тем более вытянутым является этот эллипс для больших чисел обусловленности. Таким образом, в этом случае параболоид будет похож на желоб параболического сечения с осью, направленной вдоль большой оси эллипса (рис. 14, б).

В [34] этот эллипс назван характеристическим. При этом форма характеристического эллипса при малых воздействиях не будет зависеть от kT . В направлении большой оси сопротивление растяжению наименьшее, и там наиболее вероятен разрыв связей. Следовательно, число обусловленности характеризует степень устойчивости конфигурации биологического комплекса. При этом для выбора более устойчивого биохимического соединения между белками мы выбираем матрицу потенциальной энергии электростатического взаимодействия с наименьшим значением числа обусловленности.

Авторы считают своим долгом выразить глубокую благодарность профессору кафедры „Физическая электроника“ Санкт-Петербургского политехнического университета Петра Великого доктору физ.-мат. наук А.П. Головицкому за консультации и конструктивное обсуждение результатов.

Список литературы

- [1] Системная компьютерная биология / Коллектив авторов. Монография. Новосибирск: Издательство СО РАН, 2008. 769 с.
- [2] *Betts M.J., Sternberg M.J.* // *Protein Eng.* 1999. V. 12. P. 271–283.
- [3] *Пырклов Т.В., Озеров И.В., Балицкая Е.Д., Ефремов Р.Г.* // *Биоорг. химия.* 2010. Т. 36. № 4.
- [4] Электронный ресурс. Режим доступа: <http://www.uniprot.org/> (дата обращения 05.11.2017).
- [5] *Thierry Soussi* // *EMBO Rep.* 2010. Vol. 11. N 11. P. 822–826.
- [6] *Yingjuan Q., Xinbin C.* // *Methods Mol Biol.* 2013. Vol. 965. P. 37–61.
- [7] *Zambetti G.P.* *The p53 Tumor Suppressor Pathway and Cancer.* Springer, 2005.
- [8] *Yin Y., Solomon G., Deng C., Barrett J.C.* // *Mol. Carcinog.* 1999. V. 24. N 1. P. 15–24.
- [9] *Manfredi J.J.* // *Genes. Dev.* 2010. Vol. 24. N 15. P. 1580–1589.
- [10] *Park Y.J., Luger K.* // *Proc. Natl. Acad. Sci. USA.* 2006. Vol. 103. N 5. P. 1248–1253.
- [11] *Leslie P.L., Ke H., Zhang Y.* // *J. Biol. Chem.* 2015. Vol. 290. N 20. P. 12941–12950.
- [12] *Jayne Loughery J., Meek D.* // *BioDiscovery.* Vol. 8. N 1. P. 1–20.
- [13] *Freedman Deborah A., Arnold J.L.* // *Mol. Cell. Biol.* 1998. Vol. 18. N 12. P. 7288–7293.
- [14] *Dingding Shi, Wei G.* // *Genes. Cancer.* 2012. Vol. 3. N 3–4. P. 240–248.
- [15] *Grossman S.R., Perez M., Kung A.L., Joseph M., Mansur C., Zhi-Xiong X., Sushant K., Livingston D.M.* // *Mol. Cell.* 1998. Vol. 2. N 4. P. 405–415.
- [16] *Petrenko O.* // *Mol. Cancer. Res.* 2003. Vol. 1. N 14. P. 1001–1008.
- [17] *Golestanian S., Sharifi A., Popowicz G.M., Azizian H., Foroumadi A., Szwaigierczak A., Holak T.A., Amanlou M.* // *Life. Sci.* 2016. Vol. 145. P. 240–246.
- [18] *Poyurovsky M.V., Priest C., Kentsis A., Borden K.L., Pan Z.Q., Pavletich N., Prives C.* // *EMBO J.* 2007. Vol. 26. N 1. P. 90–101.
- [19] *Zlatanova J., Seebart C., Tomschik M.* // *FASEB J.* 2007. Vol. 21. N 7. P. 1294–1310.
- [20] *Jiandong C.* // *Transl. Cancer. Res.* 2016. Vol. 5. N 6. P. 639–649.
- [21] *Fenley A.T., Adams D.A., Onufriev A.V.* // *Biophys. J.* 2010. Vol. 99. P. 1577–1585.
- [22] *Семчиков Ю.Д.* *Высокомолекулярные соединения.* М.: Академия, 2010.
- [23] *Gerstein M., Richards F.M.* *Protein Geometry: Volumes, Areas, and Distances.* Yale University, 1977.
- [24] *Biro J.C.* // *Theoretical Biology and Medical Modelling.* 2006. Vol. 3. N 15. P. 1–12.
- [25] *Almazov A.B.* *Probabilistic Methods in the Theory of Polymers* М.: Science, 1971.
- [26] *Ryskin Ya.I.* *Hydrogen Bond and Structure of Hydrosilicates.* L.: Science, 1972.
- [27] *Kulikov K.G., Koshlan T.V.* // *Technical Physics.* 2001. Vol. 61. N 10. P. 1572–1579.
- [28] *Саранин В.А.* // *УФН.* 1999. Т. 169. С. 453–458.
- [29] *Саранин В.А.* // *УФН.* 2002. Т. 172. С. 1449–1454.
- [30] *Саранин В.А.* // *УФН.* 2010. Т. 180. С. 1109–1117.
- [31] *Смайт В.* *Электростатика и электродинамика.* Перевод со второго американского издания А.В. Гапонова и М.А. Миллера. М.: ИЛ, 1954.
- [32] *Щерба Е.А., Григорьев А.И., Коромыслов В.А.* // *ЖТФ.* 2002. Т. 72. Вып. 1. С. 15–19.
- [33] *Вержебицкий В.М.* *Вычислительная линейная алгебра.* М.: Высшая школа, 2009.
- [34] *Демиденко Е.З.* *Линейная и нелинейная регрессии.* М.: Финансы и статистика, 1981.