

ISSN 1816-0301 (Print)  
ISSN 2617-6963 (Online)

**БИОИНФОРМАТИКА**  
**BIOINFORMATICS**

УДК 51-76:577.322:539.19:004.94:004.021  
<https://doi.org/10.37661/1816-0301-2020-17-2-44-53>

Поступила в редакцию 24.04.2020  
Received 24.04.2020

Принята к публикации 07.05.2020  
Accepted 07.05.2020

## Предсказание структуры гомодимерных белковых комплексов на основе глубокой нейронной сети

А. Ю. Хадарович<sup>1,2✉</sup>, А. А. Калиновский<sup>1</sup>, А. В. Тузиков<sup>1</sup>

<sup>1</sup>Объединенный институт проблем информатики  
Национальной академии наук Беларуси, Минск, Беларусь  
✉E-mail: [ahadarovich@gmail.com](mailto:ahadarovich@gmail.com)

<sup>2</sup>Белорусский государственный университет, Минск, Беларусь

**Аннотация.** Предсказание структур белковых комплексов имеет важные приложения в таких областях, как моделирование биологических процессов и разработка лекарственных средств. Гомодимеры (комплексы, состоящие из двух одинаковых белков) являются наиболее распространенным типом белковых комплексов в природе, но до сих пор нет универсального алгоритма для предсказания их трехмерных структур. Экспериментальные методы для определения структур белковых комплексов требуют значительных затрат времени и ресурсов и имеют свои ограничения. Глубокие нейронные сети позволили предсказать структуры отдельных белков, значительно превзойдя по точности другие алгоритмические подходы. Опираясь на идею этого подхода, в статье разработан алгоритм для моделирования трехмерной структуры гомодимерных комплексов на основе глубокого обучения. Он состоит из двух основных этапов: на первом этапе прогнозируется карта контактов белкового комплекса при помощи глубокой сверточной нейронной сети, а на втором предсказывается трехмерная структура гомодимера на основе полученной карты контактов и процедуры оптимизации. Предложенный подход был протестирован и проверен на наборе белковых гомодимеров из базы данных белков PDB (Protein Data Bank). Разработанная методика может быть использована для оценки моделей белковых гомодимеров в качестве одного из этапов разработки лекарственных соединений.

**Ключевые слова:** методы глубокого обучения, нейронные сети, моделирование белковых комплексов, гомодимеры, оптимизация

**Для цитирования.** Хадарович, А. Ю. Предсказание структуры гомодимерных белковых комплексов на основе глубокой нейронной сети / А. Ю. Хадарович, А. А. Калиновский, А. В. Тузиков // Информатика. – 2020. – Т. 17, № 2. – С. 44–53. <https://doi.org/10.37661/1816-0301-2020-17-2-44-53>

---

## Protein homodimers structure prediction based on deep neural network

Anna Y. Hadarovich<sup>1,2✉</sup>, Alexander A. Kalinouski<sup>1</sup>, Alexander V. Tuzikov<sup>1</sup>

<sup>1</sup>The United Institute of Informatics Problems of the National Academy  
of Sciences of Belarus, Minsk, Belarus  
✉E-mail: [ahadarovich@gmail.com](mailto:ahadarovich@gmail.com)

<sup>2</sup>Belarusian State University, Minsk, Belarus

**Abstract.** Structural prediction of protein-protein complexes has important application in such domains as modeling of biological processes and drug design. Homodimers (complexes which consist of two identical proteins) are the most common type of protein complexes in nature but there is still no universal algorithm to

predict their 3D structures. Experimental techniques to identify the structure of protein complex require enormous amount of time and resources, and each method has its own limitations. Recently Deep Neural Networks allowed to predict structures of individual proteins greatly prevailing in accuracy over other algorithmic approaches. Building on the idea of this approach, we developed an algorithm to model the 3D structure of homodimer based on deep learning. It consists of two major steps: at the first step a protein complex contact map is predicted with the deep convolutional neural network, and the second stage is used to predict 3D structure of homodimer based on obtained contact map and optimization procedure. The use of the neural network in combination with optimization procedure based on gradient descent method allowed to predict structures for protein homodimers. The suggested approach was tested and validated on a dataset of protein homodimers from Protein Data Bank (PDB). The developed procedure could be also used for evaluating protein homodimer models as one of the stages in drug compounds developing.

**Keywords:** deep learning, neural network, modeling of protein complexes, homodimers, optimization

**For citation.** Hadarovich A. Y., Kalinouski A. A., Tuzikov A. V. Protein homodimers structure prediction based on deep neural network. *Informatics*, 2020, vol. 17, no. 2, pp. 44–53 (in Russian). <https://doi.org/10.37661/1816-0301-2020-17-2-44-53>

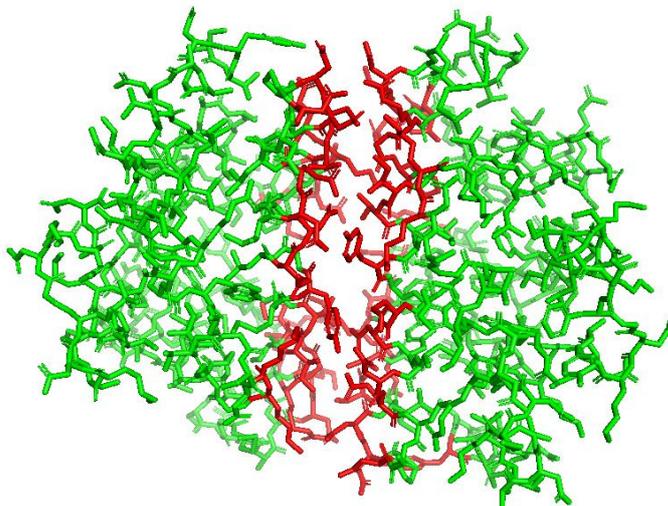
**Введение.** Большинство процессов в клетке базируется на специфическом взаимодействии белков, для которого необходима определенная пространственная структура. Поэтому биологическая функция белков тесно связана с существованием их в виде трехмерных структур. Даже небольшие изменения этих структур часто ведут к утере или резкому изменению активности белков. Знание пространственной организации белковых молекул не только является ключом к пониманию их функций и механизма работы, но и служит основой для разработки эффективных и безопасных лекарственных средств. Общеизвестно утверждение, что сама аминокислотная последовательность белка (при подходящей температуре и pH воды) определяет ее пространственную структуру [1].

В последние годы набирает популярность подход к машинному обучению, называемый глубоким обучением [2]. В частности, в 2018 г. был предложен алгоритм на основе глубокого обучения для свободного (ab initio) моделирования структуры белка [3, 4], превосходящий по точности предсказания другие ab initio алгоритмы [5]. Подходы, основанные на машинном обучении, и раньше использовались в задаче предсказания структуры белка, однако позволяли с достаточной точностью предсказать не структуру белка, а только ее локальные особенности: карты контактов [6–8], элементы вторичной структуры [9], интерфейсные остатки и т. д. [4, 6, 10]. Успехи методов глубокого обучения обусловлены тем, что количество накопленных данных позволило с достаточной точностью тренировать разработанные модели.

Вступая во взаимодействие, белки образуют белок-белковый комплекс. Задача нахождения трехмерной структуры комплекса, образованного при взаимодействии белков, называется белковым докингем. Так же как и в случае индивидуальных белков, экспериментальные методы могут использоваться в ограниченном количестве случаев и требуют длительного времени, а в массовом порядке и вовсе неприменимы. Поэтому ускоренное развитие получили вычислительные методы, которые позволяют предсказывать трехмерные структуры комплексов быстро, используя в качестве входных данных трехмерные структуры белков, составляющих комплекс. Наибольшее значение при этом имеет область связывания белков, или интерфейс белкового комплекса, поскольку именно в данной области находятся аминокислоты, являющиеся необходимыми для образования комплекса и выполнения соответствующих функций в организме.

Пара множеств аминокислотных остатков  $\{(A_1, \dots, A_N), (B_1, \dots, B_M)\}$  называется интерфейсом (или областью связывания), если выполняется следующее условие: для любого остатка  $A_i$  из белка  $A$  найдется по крайней мере один остаток  $B_j$  из белка  $B$ , такой, что расстояние между данными остатками не превосходит некоторого порога (величина порога, как правило, варьируется от 6 до 12 Å). На рис. 1 изображены белок-белковый комплекс 1jzk из базы данных PDB [11], состоящий из двух одинаковых белков с выделенным интерфейсом, и аминокислотная последовательность белка, образующего комплекс. Красным цветом выделен интерфейс комплекса, а зеленым цветом – аминокислотные остатки, не входящие в интерфейс. На нижней

панели отображена аминокислотная последовательность белка, образующего комплекс, в которой красным цветом обозначены аминокислоты, принадлежащие интерфейсу, а черным цветом – все остальные аминокислоты.



SVYDAAAQLTADVKKDLRDSWKVIGSDKKGNGVALMTTLFADNQETIGYFKRLGNV SQGAND-  
KLRGHSITLMYALQNFDQLDNPDDLVCVVEKFAVNHITRKISAAEFGKFNPIKVKVLASKNFGDKYANAWAKLVAVVQAAL

Рис. 1. Белок-белковый комплекс 1jzk

В настоящей статье поставлена задача разработки алгоритма предсказания трехмерной структуры комплексов белковых гомодимеров, используя в качестве входных данных трехмерные структуры белков, составляющих комплексы. Данная задача является актуальной, поскольку гомодимеры – это наиболее часто встречающийся тип белковых комплексов [12]. Для ее решения требуется найти взаимное расположение двух белков в пространстве, оптимизирующее выбранную целевую функцию. В качестве критерия оптимизации обычно используют минимизацию энергии связывания белков в комплексе. Это вычислительно сложная оптимизационная задача в шестимерном пространстве параметров, задающих расположение белков в пространстве друг относительно друга. В работе предлагается алгоритм для ее решения, основанный не на выборе конкретной целевой функции, а на машинном обучении. Предложенный алгоритм состоит из следующих этапов:

- формирование набора данных белковых гомодимерных комплексов из базы данных белков, отвечающих заданным требованиям: высокое разрешение экспериментально полученной структуры, отсутствие неточностей в файле описания структуры;
- фильтрация набора данных по площади интерфейса для исключения белковых гомодимеров, которые были ошибочно аннотированы как комплексы;
- построение архитектуры глубокой сверточной нейронной сети для предсказания карты контактов белковых гомодимерных комплексов;
- обучение нейронной сети;
- оценка результатов обучения нейронной сети;
- построение алгоритма предсказания трехмерной структуры гомодимерного комплекса на основе карты интерфейсных контактов и трехмерной структуры входящих в комплекс белков;
- оценка полученных результатов.

**Глубокая нейронная сеть для предсказания карты контактов белковых гомодимерных комплексов.** Для предсказания контактов (близкого расположения) между атомами двух белковых цепей в составе комплекса была разработана глубокая нейронная полносверточная сеть

с разреженной сверткой и остаточными блоками [13–16]. Архитектура построенной нейронной сети показана на рис. 2.

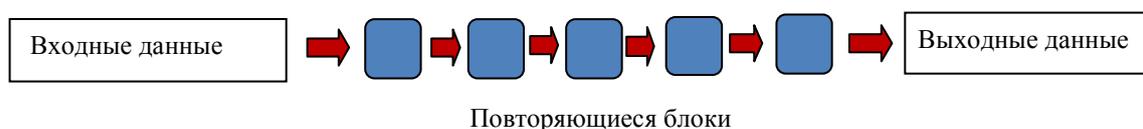


Рис. 2. Схема глубокой нейронной сети для предсказания карты контактов гомодимерного комплекса

Схема повторяющихся блоков, обозначенных на рис. 2 синим цветом, изображена на рис. 3.

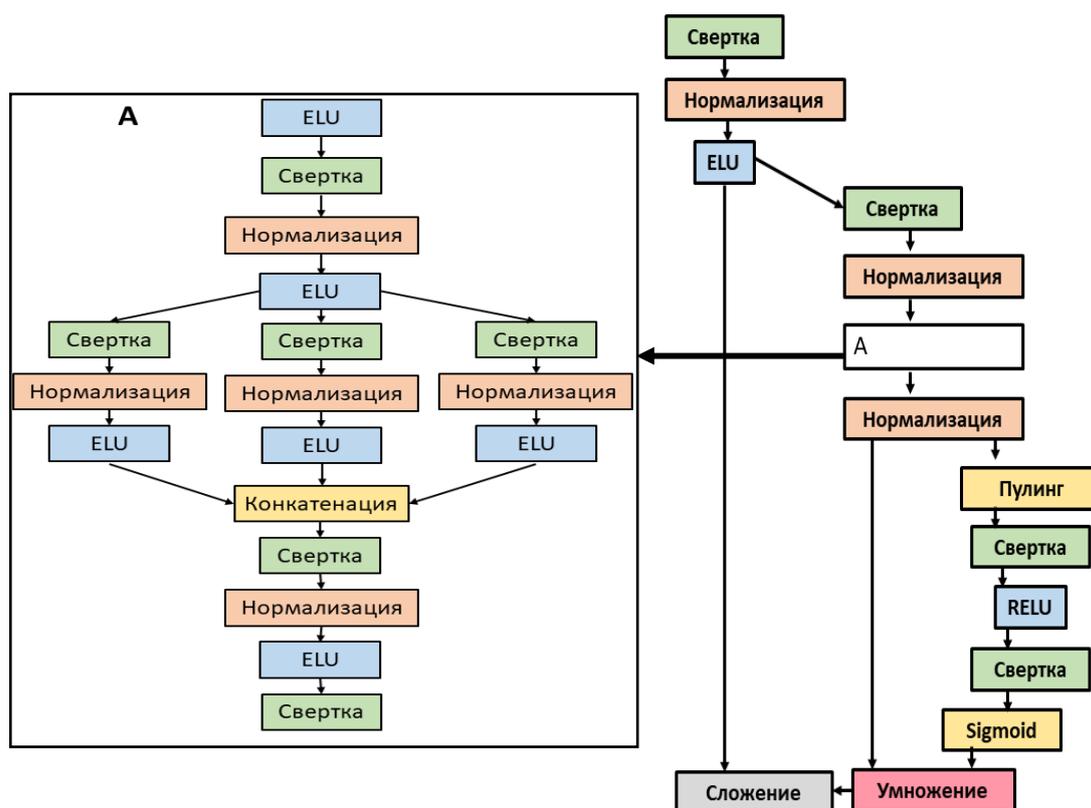


Рис. 3. Схема блока глубокой нейронной сети для предсказания карты контактов гомодимерного комплекса

Блоки ELU, RELU и Sigmoid обозначают выполнение соответствующих функций активации нейронов, блоки умножения и сложения подразумевают поэлементные операции.

На слоях нейронной сети используется активационная функция RELU [1]

$$f(x) = \begin{cases} 0 & \text{при } x < 0; \\ x & \text{при } x \geq 0 \end{cases}$$

и сигмоидная активационная функция, обозначенная на схеме как Sigmoid [2],

$$f(x) = \frac{1}{1 + e^{-x}}.$$

На промежуточных слоях нейронной сети используется активационная функция ELU [3]

$$f(x) = \begin{cases} (e^x - 1) & \text{при } x < 0; \\ x & \text{при } x \geq 0. \end{cases}$$

Преимуществом использования функции ELU является то, что она может принимать отрицательные значения, которые помогают сети изменять веса (значения параметров), выполнять смещения в правильных направлениях, а также производить активации весов нейронной сети при расчете градиента и избегать так называемой dead RELU проблемы, когда большинство компонентов сети не обновляется.

В глубокой нейронной сети выделяют две основные операции: свертку и пулинг. Свертка – это процесс применения фильтра к каждому участку слоя. Данная операция может быть задана формулой [4]

$$x_{i,j}^l = f \left( \sum_m \sum_n x_{i+m,j+n}^{l-1} \cdot w_{m,n}^l + b^l \right),$$

где  $x^l$  – выход слоя  $l$  в виде матрицы ( $x$  – входные данные в случае  $l = 0$ );  $f()$  – функция активации;  $b^l$  – коэффициент сдвига слоя  $l$ ;  $w_{m,n}^l$  – элемент матрицы весовых коэффициентов фильтра для получения слоя  $l$ .

Пулинг – это процесс сжатия (уменьшения размеров) входных данных путем некоторой обработки значений блоков данных, который позволяет снизить размерность данных, ускоряя время работы сети. Пулинг «сжимает» карты признаков, полученные на предыдущем сверточном слое, с помощью выбранного преобразования (обычно используется функция выбора максимума значений признаков).

Следует отметить, что некоторым слоям подавались на вход нормализованные данные (на рис. 2 это блоки «Нормализация»), т. е. предварительно обработанные и имеющие нулевое математическое ожидание и единичную дисперсию.

В качестве функции потерь использовалась бинарная кросс-энтропия, вычисляемая по формуле

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)),$$

где  $y_i$  – наличие контакта между атомами ( $y_i = 1$ , если контакт есть,  $y_i = 0$  в противном случае);  $p(y_i)$  – предсказанная вероятность наличия контакта;  $N$  – количество атомов [5].

В качестве входных данных использовались атомы СА (альфа-углерода) и СВ (бета-углерода) аминокислотных остатков. Атом СА в аминокислоте означает первый атом углерода, который крепится к функциональной группе, а второй атом углерода называется атомом СВ. На рис. 4 изображен белок гомодимера с выделенными атомами СА и СВ.

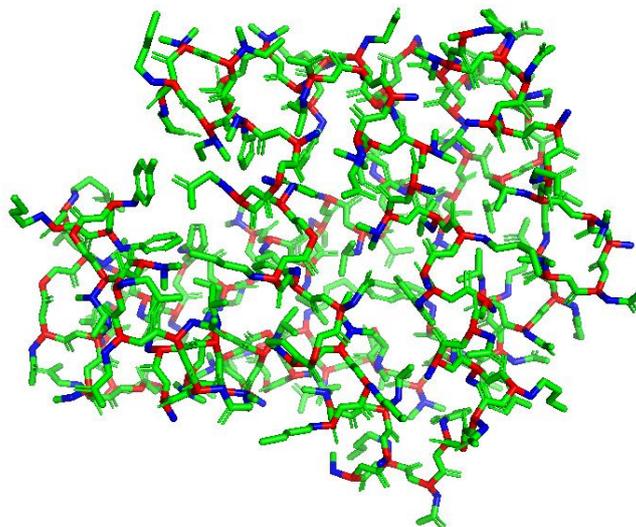


Рис. 4. Белок с выделенными атомами СА (обозначены красным цветом) и СВ (обозначены синим цветом)

Сеть принимает на вход матрицу расстояний между атомами СВ одного белка (матрица расстояний для второго белка будет такой же, поскольку белки одинаковые) и аминокислотную последовательность первой белковой цепи, а в качестве выходных данных выдает бинарную матрицу размером  $N \times N$  контактов между атомами СВ двух белков в комплексе, который они составляют ( $N$  – количество аминокислотных остатков в одной белковой цепи). Атомы СВ двух белков находятся в контакте, если евклидово расстояние между ними не превосходит  $8 \text{ \AA}$ .

Для обучения нейронной сети использовались различные параметры, но наибольшую точность показали следующие:

количество эпох (итераций в процессе обучения, включающих предъявление всех примеров из обучающего множества и проверку качества обучения на контрольном множестве) – 500;

количество итераций на одну эпоху – 200;

скорость обучения – 0,0001;

размер входа сети – 128;

оптимизатор – метод Adam [17].

Следует отметить, что авторы статьи предпринимали попытку разработать архитектуру нейронной сети, способную напрямую предсказывать матрицу расстояний между контактами белкового комплекса, однако сделали вывод, что на текущий момент для предсказания расстояний с достаточной точностью не хватает исходных данных.

**Оптимизационный алгоритм для предсказания трехмерной структуры белкового гомодимерного комплекса.** На вход алгоритма подаются трехмерные структуры двух белков, составляющих гомодимерный комплекс, и предсказанная на предыдущем шаге карта контактов, представляющая собой бинарную матрицу  $A = [a(i, j)]$  размером  $N \times N$ . Элементы матрицы вычисляются по формуле [6]

$$a(i, j) = \begin{cases} 1, & \text{если } d(i, j) < 8 \text{ \AA}, \\ 0 & \text{в противном случае.} \end{cases}$$

Выходом алгоритма является матрица аффинного преобразования  $T$  размером  $4 \times 4$  [7]:

$$T = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

где элементы  $r_{ij}$  задают параметры матрицы поворота, а  $t_{ij}$  – параметры сдвига, которые применяются к структуре первого белка для получения координат расположения в пространстве второй белковой цепи относительно первой.

Как и в случае стандартного подхода свободного моделирования димерного комплекса, одна белковая структура считается фиксированной в пространстве и ищется расположение другой белковой структуры относительно нее. Это расположение задается шестью параметрами, отвечающими за шесть степеней свободы: три параметра задают поворот и три параметра – сдвиг. При начальной инициализации предполагается, что начальное положение в пространстве второй белковой структуры совпадает с расположением первой, т. е. все шесть параметров равны нулю.

Алгоритм является итерационным, и на каждой итерации выполняются следующие шаги:

- вычисление матрицы расстояний между атомами СВ первой и второй белковой цепи;
- вычисление функции потерь;
- вычисление производной функции потерь;
- преобразование параметров, отвечающих за поворот и сдвиг; минимизация функции потерь в направлении, противоположном градиенту функции;
- вычисление матрицы преобразований для новых координат атомов;
- вычисление новых координат структуры второго белка по полученной матрице.

На последнем этапе шесть параметров преобразуются в матрицу аффинного преобразования, которое затем применяется к первой белковой цепи и задает положение в пространстве второго белка гомодимера. Таким образом, результатом работы алгоритма являются координаты всего гомодимерного белкового комплекса.

**Создание и фильтрация набора данных.** Набор данных белковых гомодимерных комплексов был получен из базы данных белков PDB [11]. Файлы описания комплексов предоставляются в специальном формате, предусматривающем структурную аннотацию каждой белковой цепи: секция АТОМ файлов с расширением PDB описывает координаты в трехмерном евклидовом пространстве и тип атома для каждой аминокислоты в белковой цепи.

К полученным данным была применена процедура фильтрации: файлы с ошибочной информацией были исключены, отфильтрованы файлы с числом цепочек, отличным от двух, а также удалены комплексы с различным количеством аминокислотных остатков в одном файле (такая ситуация может возникнуть, если во время эксперимента не удалось определить все остатки в белковой цепи). Площадь скрытой поверхности была рассчитана с помощью библиотеки FreeSASA [18] для того, чтобы гарантировать, что выбранные комплексы имеют интерфейсы разумного размера (более  $250 \text{ \AA}^2$ ) [19]. Комплексы с меньшей площадью скрытой поверхности рассматриваются как артефакты эксперимента, а файлы, содержащие такие комплексы, не являются достоверным источником информации о белковом комплексе.

Окончательный набор данных содержал около 10 000 структур. Он был разделен на обучающую и тестовую части в пропорции 4 : 1 соответственно. Процедуры фильтрации реализовывались на языке программирования Python с помощью свободно распространяемой библиотеки для обработки биологических данных BioPython [20].

**Результаты работы алгоритма.** На рис. 5 изображены графики функции потерь для обучающего и тестового наборов данных. Они демонстрируют скорость обучения модели, а также дают ее качественную оценку: модель обучается с увеличением количества эпох, в то же время не наблюдается переобучения, поскольку значения функции потерь для тестового набора данных незначительно отличаются от значений для обучающего набора.

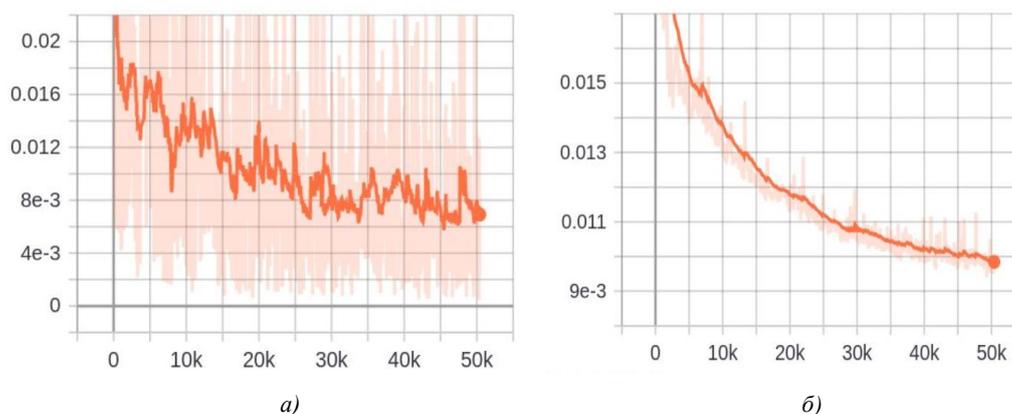


Рис. 5. Значения функции потерь (вертикальная ось) в зависимости от количества итераций (горизонтальная ось) на обучающем (а) и тестовом (б) наборах данных

Пример входных и выходных данных для первого этапа алгоритма представлен на рис. 6. На левой панели показана карта расстояний (евклидово расстояние между атомами СВ) одной белковой цепи в гомодимере, которая использовалась как инвариантное к сдвигу и повороту представление белка для нейронной сети. На средней и правой панелях показаны карты реальных и предсказанных контактов соответственно, где синие пиксели означают отсутствие контакта, а желтые – его наличие. На нижней панели указана цветовая шкала для визуализации расстояний.

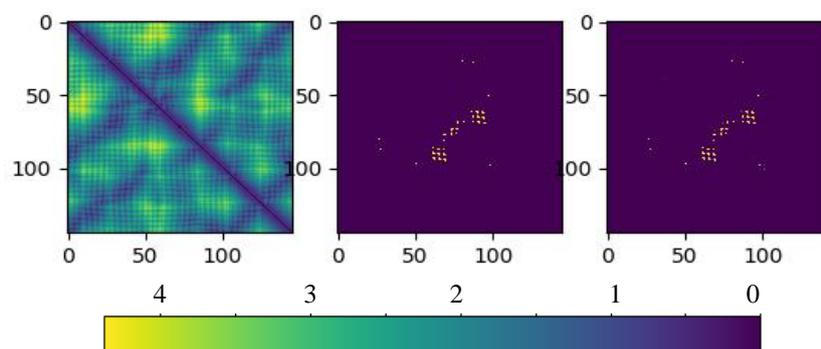


Рис. 6. Пример успешного предсказания карты контактов для гомодимера Ijzk

Для более половины протестированных комплексов удалось предсказать большую часть интерфейсных остатков с точностью более 50 %, из них для четверти комплексов – с точностью более 70 %. Предсказанные интерфейсные остатки значительно сокращают процедуру поиска трехмерной структуры гомодимера, поскольку область связывания известна. Это позволяет наложить ограничения на область поиска и существенно сократить количество возможных ориентаций структур белков в комплексе относительно друг друга. Для многих алгоритмов докинга точность предсказания структуры возрастает в несколько раз, если на вход подается информация о предполагаемой интерфейсной области.

Почти для 30 % гомодимеров представленный алгоритм обеспечил предсказания структуры всего гомодимерного комплекса хорошего качества, т. е. среднеквадратическое отклонение между атомами СВ реальной и предсказанной структур составляет менее 10 Å. Пример результата работы оптимизационного алгоритма для предсказания трехмерной структуры белкового гомодимерного комплекса показан на рис. 7. Аффинное преобразование, матрица которого была получена в результате процедуры оптимизации, применялось к первому белку гомодимера (на рисунке изображен зеленым цветом), и полученное расположение второго белка (окрашен в синий цвет) сравнивалось с реальным расположением (изображен голубым цветом). Среднеквадратическое отклонение атомов между суперпозициями голубой и синей белковых цепей равно 4,3 Å.

Был также протестирован другой подход, основанный на том, что на вход нейронной сети подавались карты расстояний между атомами СА. Такой подход дал меньшую точность результатов как при предсказании карты контактов, так и при восстановлении структуры комплекса на втором этапе алгоритма. Это объясняется тем, что атомы СА принадлежат главной цепи белка и расстояние между соответствующими атомами СА двух белков больше, чем в случае с атомами СВ, расположенными на боковых цепях белка. Боковые цепи более гибкие, чем главная цепь, поэтому расстояния между атомами боковой цепи лучше фиксируют изменчивость белковых структур.

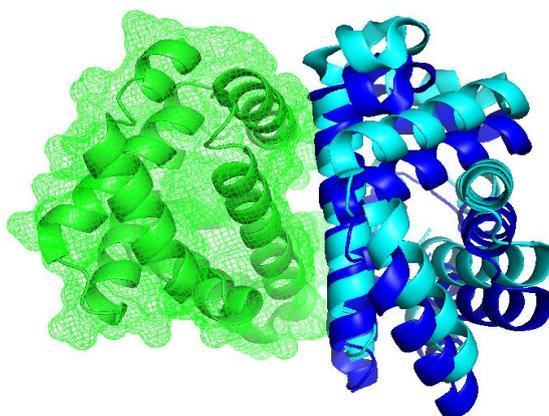


Рис. 7. Пример предсказанной гомодимерной структуры белка, построенной на основе предсказанной карты контактов белкового комплекса

**Заключение.** Анализ полученных результатов показывает, что разработанный двухэтапный алгоритм позволяет предсказывать трехмерную структуру белковых гомодимеров. Кроме того, разработанная глубокая нейронная сеть дает возможность идентифицировать пары аминокислотных остатков, находящиеся на интерфейсе, и поэтому может быть использована как промежуточный этап при оценивании построенных моделей белковых гомодимеров и фильтрации недопустимых моделей, интерфейсные контакты которых не совпадают с предсказанными.

Разработанный алгоритм восстановления структуры белкового комплекса по предсказанной карте контактов может использоваться не только для гомодимеров, но и для гетеродимеров, состоящих из двух разных белков, поскольку не зависит от количества аминокислотных остатков в одной и другой белковой цепи. Таким образом, результаты исследования свидетельствуют о том, что данный подход может быть обобщен на другие типы белковых комплексов и является перспективным для решения задачи белкового докинга.

## References

1. Anfinsen C. B. Principles that govern the folding of protein chains. *Science*, 1973, vol. 181 (4096), pp. 223–230. <https://doi.org/10.1126/science.181.4096.223>
2. Lecun Y., Bengio Y., Hinton G. Deep learning. *Nature*, 2015, vol. 521 (7553), pp. 436–444. <https://doi.org/10.1038/nature14539>
3. Senior A. W., Evans R., Jumper J., Kirkpatrick J., Sifre L., ..., Hassabis D. Improved protein structure prediction using potentials from deep learning. *Nature*, 2020, vol. 577 (7792), pp. 706–710. <https://doi.org/10.1038/s41586-019-1923-7>
4. Billings W. M., Hedelius B., Millicam T., Wingate D., Corte D. D. ProSPR: democratized implementation of alphafold protein distance prediction network. *Biorxiv*, 2019, p. 830273. <https://doi.org/10.1101/830273>
5. Kryshchak A. T., Schwede, Topf M., Fidelis K., Moult J. Critical assessment of methods of protein structure prediction (CASP) – Round XIII. *Proteins: Structure, Function, and Bioinformatics*, 2019, vol. 87 (12), pp. 1011–1020. <https://doi.org/10.1002/prot.25823>
6. Jones D. T., Kandathil S. M. High precision in protein contact prediction using fully convolutional neural networks and minimal sequence features. *Bioinformatics*, 2018, vol. 34, pp. 3308–3315.
7. Seemayer S., Gruber M., Söding J. CCMpred – fast and precise prediction of protein residue-residue contacts from correlated mutations. *Bioinformatics*, 2014, vol. 30 (21), pp. 3128–3130.
8. Jones D. T., Singh T., Kosciolk T., Tetchner S. MetaPSICOV: combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins. *Bioinformatics*, 2015, vol. 31 (7), pp. 999–1006.
9. Jiang Q., Jin X., Lee S.-J., Yao S. Protein secondary structure prediction: a survey of the state of the art. *Journal of Molecular Graphics and Modelling*, 2017, vol. 76, pp. 379–402. <https://doi.org/10.1016/j.jm gm.2017.07.015>
10. Skwark M. J., Raimondi D., Michel M., Elofsson A. Improved contact predictions using the recognition of protein like contact patterns. *PLoS Computational Biology*, 2014, vol. 10 (11), p. e1003889. <https://doi.org/10.1371/journal.pcbi.1003889>
11. Berman H. M. The protein data bank: a historical perspective. *Acta Crystallographica Section A: Foundations of Crystallography*, 2008, vol. 64 (1), pp. 88–95. <https://doi.org/10.1107/S0108767307035623>
12. Mou Y., Huang P.-S., Hsu F.-C., Huang S.-J., Mayo S. L. Computational design and experimental verification of a symmetric protein homodimer. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, vol. 112 (34), pp. 10714–10719. <https://doi.org/10.1073/pnas.1505072112>
13. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, Massachusetts, USA, 7–12 June 2015*. Boston, 2015, pp. 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
14. Fu J., Liu J., Tian H., Li Y., Bao Y., ..., Lu H. Dual attention network for scene segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, California, USA, 15–20 June 2019*. Long Beach, 2019, pp. 3141–3149. Available at: <http://arxiv.org/abs/1809.02983> (accessed 27.03.2020).
15. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, 27–30 June 2016*. Las Vegas, 2016, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
16. Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A. L. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions*

on *Pattern Analysis and Machine Intelligence*, 2018, vol. 40 (4), pp. 834–848. <https://doi.org/10.1109/TPAMI.2017.2699184>

17. Kingma D. P., Ba J. Adam: a method for stochastic optimization. *3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 7–9 May 2015*. San Diego, 2015. Available at: <https://arxiv.org/abs/1412.6980> (accessed 27.03.2020).

18. Mitternacht S. FreeSASA: an open source C library for solvent accessible surface area calculations. *F1000Research*, 2016, vol. 5, p. 189. <https://doi.org/10.12688/f1000research.7931.1>

19. Janin J., Bahadur R. P., Chakrabarti P. Protein-protein interaction and quaternary structure. *Quarterly Reviews of Biophysics*, 2008, vol. 41 (2), pp. 133–180. <https://doi.org/10.1017/S0033583508004708>

20. Cock P. J., Antao T., Chang J. T., Chapman B. A., Cox C. J., ..., de Hoon M. J. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 2009, vol. 25 (11), pp. 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>

### Информация об авторах

*Хадарович Анна Юрьевна*, научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси, Минск, Беларусь.  
E-mail: [ahadarovich@gmail.com](mailto:ahadarovich@gmail.com)

*Калиновский Александр Александрович*, научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси, Минск, Беларусь.  
E-mail: [gakarak@gmail.com](mailto:gakarak@gmail.com)

*Тузиков Александр Васильевич*, член-корреспондент, доктор физико-математических наук, профессор, генеральный директор, Объединенный институт проблем информатики Национальной академии наук Беларуси, Минск, Беларусь.  
E-mail: [tuzikov@newman.bas-net.by](mailto:tuzikov@newman.bas-net.by)

### Information about the authors

*Anna Y. Hadarovich*, Researcher, the United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Belarus.  
E-mail: [ahadarovich@gmail.com](mailto:ahadarovich@gmail.com)

*Alexander A. Kalinouski*, Researcher, the United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Belarus.  
E-mail: [gakarak@gmail.com](mailto:gakarak@gmail.com)

*Alexander V. Tuzikov*, Corresponding Member, Dr. Sci. (Phys.-Math.), Professor, General Director, the United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Belarus.  
E-mail: [tuzikov@newman.bas-net.by](mailto:tuzikov@newman.bas-net.by)