

Применение нейронных сетей для распознавания конформационных изменений в структуре белка по рентгеновским дифрактограммам его одиночных молекул на примере фотоцикла бактериородопсина

Г. А. Армеев¹, М. П. Кирпичников², Г. М. Кобельков³,
А. В. Кудрявцев⁴, М. А. Ложников⁵, В. Н. Новоселецкий⁶,
А. К. Шайтан⁷, К. В. Шайтан⁸

Лазеры на свободных электронах становятся всё более мощными и доступными установками для определения структуры бел-

¹ *Армеев Григорий Алексеевич* — в.н.с. каф. биоинженерии биологического ф-та МГУ, e-mail: satory@yandex.ru.

Armeev Grigoriy Alekseevich — leading researcher, Lomonosov Moscow State University, Faculty of biology, Chair of bioengineering

² *Кирпичников Михаил Петрович* — заведующий каф. биоинженерии биологического ф-та МГУ, академик, e-mail: kirpichnikov@inbox.ru.

Kirpichnikov Mikhail Petrovich — head of the chair, academician, Lomonosov Moscow State University, Faculty of biology, Chair of bioengineering

³ *Кобельков Георгий Михайлович* — заведующий каф. вычислительной математики мех.-мат. ф-та МГУ, профессор, e-mail: george.kobelkov@yandex.ru.

Kobelkov Georgiy Mikhailovich — head of the chair, professor, Lomonosov Moscow State University, Faculty of Mechanics and Mathematics, Chair of Computational Mathematics

⁴ *Кудрявцев Александр Викторович* — аспирант каф. биоинженерии биологического ф-та МГУ, e-mail: aleks.kudryavcev@gmail.com.

Kudryavtsev Aleksandr Viktorovich — PhD student, Lomonosov Moscow State University, Faculty of biology, Chair of bioengineering

⁵ *Ложников Михаил Андреевич* — ассистент каф. вычислительной математики мех.-мат. ф-та МГУ, e-mail: lozhnikovma@gmail.com.

Lozhnikov Mikhail Andreevich — assistant, Lomonosov Moscow State University, Faculty of Mechanics and Mathematics, Chair of Computational Mathematics

⁶ *Новоселецкий Валерий Николаевич* — доцент каф. биоинженерии биологического ф-та МГУ, e-mail: valery.novoseletsky@yandex.ru.

Novoseletsky Valeriy Nikolaevich — associate professor, Lomonosov Moscow State University, Faculty of biology, Chair of bioengineering

⁷ *Шайтан Алексей Константинович* — доцент каф. биоинженерии биологического ф-та МГУ, e-mail: shaytan_ak@mail.bio.msu.ru.

Shaitan Alexey Konstantinovich — associate professor, Lomonosov Moscow State University, Faculty of biology, Chair of bioengineering

⁸ *Шайтан Константин Вольдемарович* — профессор каф. биоинженерии биологического ф-та МГУ, e-mail: shaytan49@yandex.ru.

Shaytan Konstantin Voldemarovich — professor, Lomonosov Moscow State University, Faculty of biology, Chair of bioengineering

ков. Конформационные перестройки характерны для подавляющего большинства белков, однако скоротечность превращений зачастую не допускает получение кристаллов достаточного размера. Выходом могло бы стать получение картин рентгеновской дифракции от одиночных молекул, однако в этом случае нет возможности определять их ориентацию, что делает невозможным восстановление структуры современными методами. В работе рассмотрена применимость ряда архитектур нейронных сетей к распознаванию конформационного состояния белка по данным дифракции на одиночных молекулах бактериородопсина.

Ключевые слова: рентгеновская дифракция, нейронные сети, конформационные состояния белка, бактериородопсин

1. Введение

Лазер на свободных электронах (ЛСЭ, англ. XFEL — X-ray Free Electron Laser) представляет собой комплекс из источника электронов, их ускорителя, магнитного ондулятора и ловушки электронного пучка. Проходя через ускоритель, электроны разгоняются до высоких энергий и релятивистских скоростей, после чего попадают в ондулятор, образованный последовательностью периодически расположенных дипольных магнитов, которые заставляют электроны двигаться вдоль синусоидальных траекторий. При этом каждый отдельный электрон испускает рентгеновское излучение, которое постепенно усиливается, причем особенности взаимодействия электронов с излучением приводят к формированию его чрезвычайно коротких и интенсивных когерентных вспышек [1]. На данный момент в мире работает или готовятся к вводу в эксплуатацию несколько установок ЛСЭ: LCLS (США, 2009), SACLA (Япония, 2011), PAL-XFEL (Республика Корея, 2016), SwissFEL (Швейцария, 2016), EuXFEL (Германия, 2016), SHINE (Китай, 2025). ЛСЭ обладают возможностью генерировать вспышки излучения с очень высокой пиковой яркостью [2], а различие между установками состоит в частоте генерации вспышек излучения и максимальной энергии ускоряемых электронов, что влияет на длину волны и количество фотонов во вспышке.

Конечное применение получаемого рентгеновского излучения происходит на экспериментальных станциях [3], ключевыми параметрами которых являются технология фокусировки пучка излучения, технологии подачи образца и параметры детектора дифракционной картины. Так, например, на экспериментальной станции SPB/SFX установки EuXFEL предполагаются следующие параметры [4]: фокусировка пучка до диаметра 1 мкм или 100 нм, подача образца несколькими разными способами (в струе жидкости, с помощью газодинамической форсунки, в виде аэрозоля и с помощью закреплением образца), детектор размером 1024x1024

пикселя с размером пикселя 200x200 мкм на минимальном расстоянии 129 мм от образца.

С точки зрения определения структуры биологических макромолекул перспективным является режим дифракции от одиночных частиц, который позволил бы отказаться от стадии выращивания кристаллов. Однако интенсивность рассеяния в таком режиме сильно меньше, чем при рассеянии на кристаллах или нанокристаллах. Это означает, что для работы в таком режиме необходима высокая частота импульсов, генерируемых установкой, а также стабильность работы всех ее компонент, поскольку для реконструкции необходимы миллионы изображений. Кроме того, особое внимание должно быть уделено способам доставки одиночных молекул в пучок излучения [5].

Ввиду специфики экспериментов на ЛСЭ, особенностей получаемых данных, методы их обработки, а также методы реконструкции по этим данным структуры объектов экспериментов нуждаются в активном развитии [6].

Бактериородопсины (БР) — это семейство мембранных белков, использующих энергию света для конформационных перестроек, приводящих в итоге к однонаправленному переносу протона через мембрану. Со структурной точки зрения БР представляют собой пучки из семи трансмембранных альфа-спиралей, окружающих ковалентно связанный хромофор (ретиаль) в полностью распрямлённой конформации. После поглощения кванта света хромофор переходит в возбуждённое состояние и изомеризуется, изменяя конформацию. Это приводит к ряду структурных перестроек в окружающих его атомах белка, что в итоге приводит к переносу протона, после чего хромофор и белок в целом возвращаются в исходные состояния — этот процесс называется фотоциклом [7].

Вот уже на протяжении 50 лет механизм функционирования БР привлекает к себе неустанное внимание исследователей. Ввиду быстроты происходящих превращений, изучение структурных особенностей этих конформаций до недавнего времени было доступно лишь с помощью методов молекулярного моделирования. Однако в последние годы в связи с развитием ЛСЭ стало возможным экспериментальное наблюдение конформационных перестроек, что позволяет сопоставить экспериментальные и вычислительные методы изучения конформационной подвижности белков. Применительно к ретиаль-содержащим белкам последние данные, которые удалось получить с помощью ЛСЭ для БР, наиболее простого модельного объекта из ретиаль-содержащих белков, предлагают наносекундное разрешение [8].

В задачах классификации картин рассеяния довольно активно применяются алгоритмы распознавания изображений, в том числе методы, основанные на нейронных сетях. Так, например, была предложена глубо-

кая нейронная сеть для классификации кристаллов по данным порошковой рентгеновской дифракции [9]. Другим примером является использование нейронной сети типа AlexNet в задачах классификации дифрактограмм [10]. Кроме того, в этих задачах довольно популярными являются нейронные сети типа VGG [11]. Однако в последнее время в области распознавания изображений были представлены новые алгоритмы, которые пока не набрали популярности в задачах обработки дифракционных картин. Одним из таких алгоритмов является остаточная нейронная сеть типа ResNet. Эта нейронная сеть показывает более высокие результаты, чем сети типа VGG в задачах классификации изображений, например, в таких тестах как CIFAR-10 и ImageNet.

2. Архитектура и топология нейросетей

В данной работе произведено сравнение сетей типа VGG [12] и ResNet [13] в задачах обработки дифрактограмм. В частности, были рассмотрены следующие варианты нейронных сетей, которые были использованы для обучения, и проведён подбор параметров для их эффективной работы.

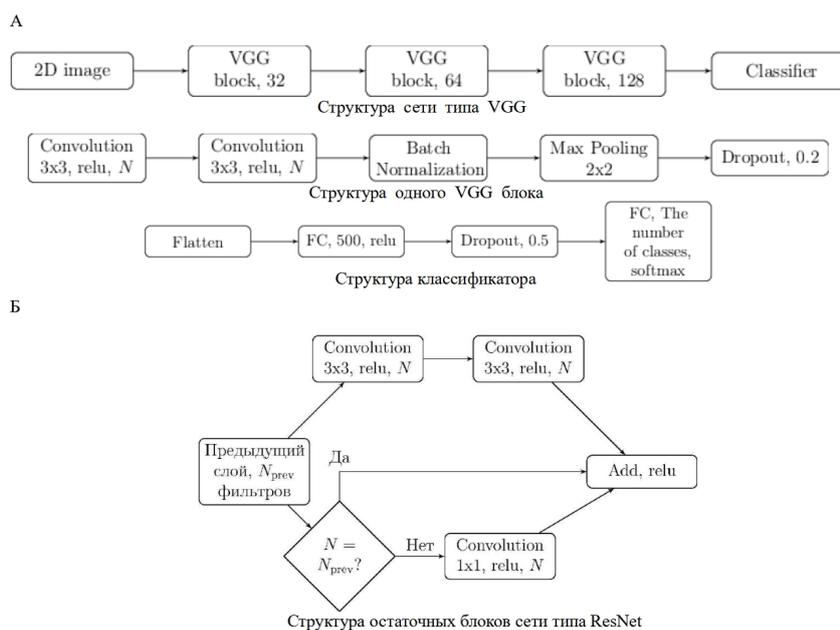


Рис. 1. Архитектуры нейронных сетей, используемые в данной работе.

А. Нейронная сеть типа VGG состоит из нескольких VGG блоков и классификатора (рис. 1, А). В настоящей работе параметры VGG се-

ти были выбраны следующими: использовалось 3 VGG блока глубины 32, 64 и 128 соответственно со значением dropout 0.0, 0.3 и 0.4 соответственно. Классификатор включал в себя полносвязный скрытый слой из 128 нейронов и dropout слой с параметром 0.5.

Б. Нейронная сеть типа ResNet состоит из остаточных блоков (рис. 1, Б), расположенных в стеке, после которых находится классификатор. В работе использовались остаточные нейронные сети ResNet20, ResNet32 и ResNet44 из работы [13] общей глубины в 20, 32 и 44 слоя соответственно.

В. Модификация остаточной нейронной сети, предложенная в работе [14]. Эта нейронная сеть показывает несколько лучшие результаты при классификации изображений, чем оригинальная остаточная нейронная сеть. В настоящей работе использовалась модификация ResNet20 общей глубины в 20 слоёв.

3. Структуры бактериородопсина

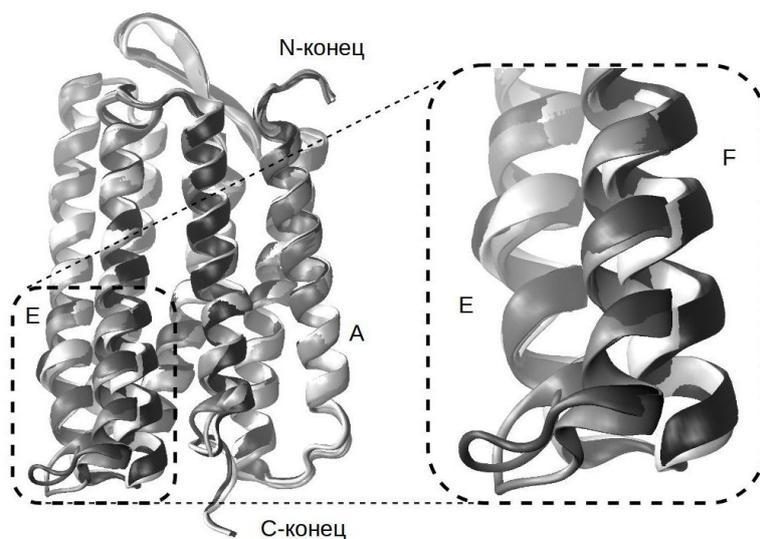


Рис. 2. Конформационные перестройки альфа-спиралей в ходе начальных стадий фотоцикла бактериородопсина (ленточное представление показывает ход основной цепи белка; атомы боковых цепей не показаны). Слева: общий боковой вид четырёх совмещённых конформаций белка, справа крупный план конца спирали E и начала спирали F — области, где различие между конформациями максимально (в конформациях, показанных черным и белым, спираль F прямая, в конформациях, показанных серым и светлосерым, изогнута влево).

Для выяснения возможности определения конформации белка по виду его дифрактограммы были рассмотрены 4 конформации бактериородопсина, соответствующие разным состояниям фотоцикла (идентификаторы в базе данных белковых структур PDB 6RNJ, 6RPH, 6RQO, 6RQP) (рис. 2) [15]. Эти конформации в целом довольно похожи друг на друга, но отличаются положением некоторого числа атомов. Наиболее заметными отличиями является положение одного из концов спирали F.

4. Набор обучающих изображений

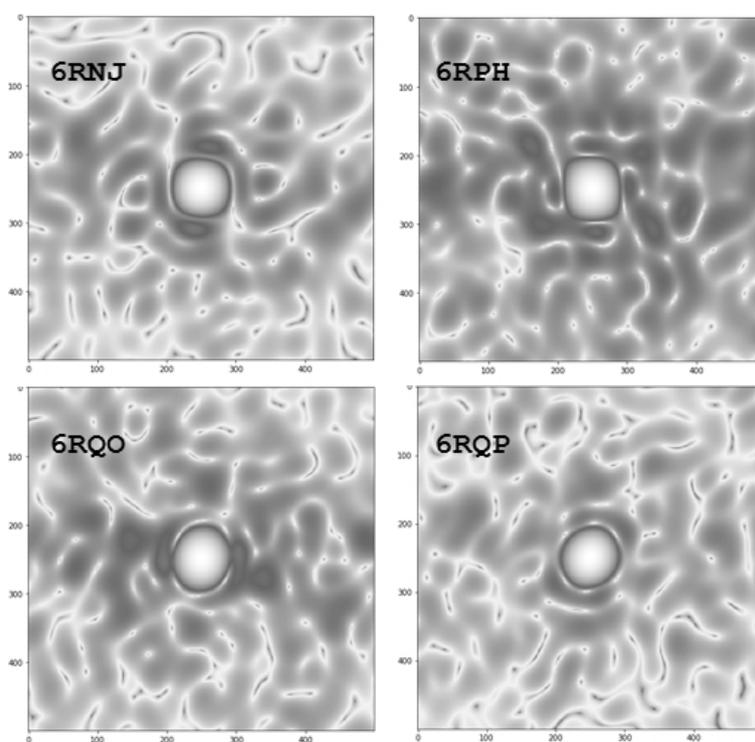


Рис. 3. Примеры расчётных дифракционных картин, полученных от структур бактериородопсина в различных состояниях фотоцикла.

Для обучения нейронных сетей требуется большой объём данных. В то же время, на данный момент не существует достаточно большого набора размеченных данных, полученных в результате экспериментов на рентгеновских лазерах, пригодного для обучения нейронных сетей. Поэтому было решено строить дифракционные изображения методом прямого моделирования дифракции, которое производили при помощи программы Condor [16]. Для каждой структуры было получено 4000 дифрак-

тограмм в условиях, приближенных к реальным (случайная ориентация молекулы относительно пучка, размер детектора 1 мегапиксель, размер пикселя 200 мкм, расстояние до детектора 0,7 м, энергия пучка 1 мДж) (Рис. 3). Размер обучающей выборки составил 12000 изображений, размер тестовой выборки составил 4000 изображений.

5. Результаты

Все нейросети продемонстрировали возможность распознавания состояния фотовозбуждения вне зависимости от числа мономеров в предъявленной картине рассеяния (Рис. 4, Таблица 1). Наиболее точный прогноз дают нейронные сети VGG и ResNet20. Более глубокие нейронные сети переобучаются на предложенных данных. Полученный в данном разделе работы результат говорит о перспективности применения нейросетевых алгоритмов при классификации образцов, гетерогенных по стадии фоточикла.

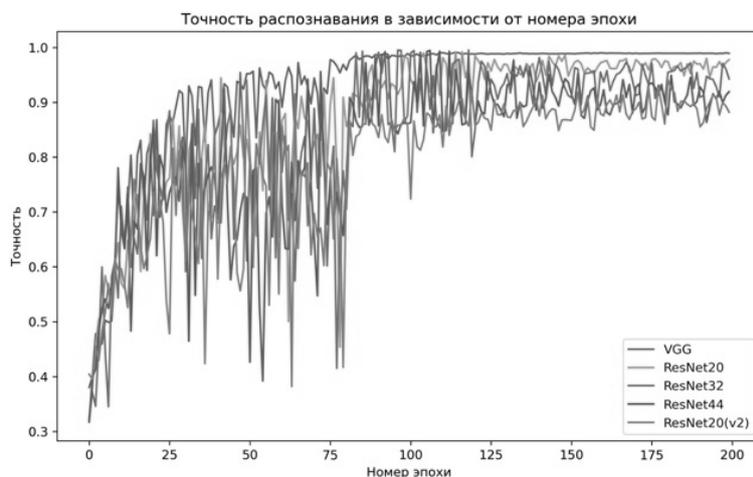


Рис. 4. График точности распознавания конформации бактериородопсина в зависимости от номера эпохи и типа использованной нейросети.

Нейронная сеть	VGG	ResNet20	ResNet32	ResNet44	ResNet20(v2)
Точность после 200 эпох обучения	0.989	0.978	0.943	0.919	0.882
Время обучения (Ч:ММ)	2:32	3:35	5:38	7:43	5:59

Таблица 1. Сводная таблица результатов обучения нейросетей.

Работа выполнена при поддержке РФФИ (№18-02-40010мега) и Междисциплинарных научно-образовательных школ Московского университета «Молекулярные технологии живых систем и синтетическая биология» и «Мозг, когнитивные системы, искусственный интеллект».

Список литературы

- [1] Zhukovsky K.V., “Undulators and generation of X-ray pulses in free-electron lasers with self-amplified spontaneous emission”, *Moscow University Physics Bulletin*, **72:2** (2017), 128–143.
- [2] Oberthür D., “Biological single-particle imaging using XFELs — towards the next resolution revolution”, *IUCrJ*, **5:6** (2018), 663–666.
- [3] Tschentscher T., Bressler C., Grünert J., Madsen A., Mancuso A.P., Meyer M., Scherz A., Sinn H., Zastra U., “Photon Beam Transport and Scientific Instruments at the European XFEL”, *Applied Sciences*, **7:6** (2017), 592.
- [4] Mancuso A.P., Aquila A., Borchers G., Giewekemeyer K., Reimers N., “Scientific instrument single particles, clusters, and biomolecules (SPB)”, *XFEL. EU Technical Report*, 2013, 1–232.
- [5] Shaitan K.V., Armeev G.A., Shaytan A.K., “The dynamics of irreversible evaporation of a water–protein droplet and the problem of structural and dynamic experiments with single molecules”, *Biophysics*, **61:2** (2016), 177–184.
- [6] Шайтан К.В., Шайтан А.К., Багров Д.В., Блинов В.Н., Новоселецкий В.Н., Соколова О.С., Шуров Д.Л., Турченков Д.А., Кирпичников М.П., “Алгоритмы и методы исследования трехмерных атомистических моделей молекул белков на основе анализа картины рассеяния мощного рентгеновского лазерного излучения”, *Наноструктуры. Математическая физика и моделирование*, **9:2** (2013), 33–74.
- [7] Lueck H., Schobert B., Richter H. T., Cartailier J. P., Lanyi J. K., “Structural changes in bacteriorhodopsin during ion transport at 2 angstrom resolution”, *Science*, **286:5438** (1999), 255–260.
- [8] Nango E., Royant A., Kubo M., Nakane T., Wickstrand C., Kimura T., et al, “A three-dimensional movie of structural changes in bacteriorhodopsin”, *Science*, **354:6319** (2016), 1552–1557.

- [9] Park W.B., Chung J., Jung J., Sohn K., Singh S.P., Pyo M., Shin N., Sohn K.-S., “Classification of crystal structure using a convolutional neural network”, *IUCrJ*, **4**:4 (2017), 486–494.
- [10] Xu W., LeBeau J.M., “A deep convolutional neural network to analyze position averaged convergent beam electron diffraction patterns”, *Ultramicroscopy*, **188** (2018), 59–69.
- [11] Ke T.W., Brewster A.S., Yu S.X., Ushizima D., Yang C., Sauter N.K., “A convolutional neural network-based screening tool for X-ray serial crystallography”, *Journal of synchrotron radiation*, **25**:3 (2018), 655–670.
- [12] Simonyan K., Zisserman A., *Very deep convolutional networks for large-scale image recognition*, 2014, arXiv:1409.1556.
- [13] He K., Zhang X., Ren S., Sun J., “Deep residual learning for image recognition”, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 770–778.
- [14] He K., Zhang X., Ren S., Sun J., “Identity mappings in deep residual networks”, *European conference on computer vision*, 2016, 630–645.
- [15] Weinert T., Skopintsev P., James D., Dworkowski F., Panepucci E., Kekilli D., et al., “Proton uptake mechanism in bacteriorhodopsin captured by serial synchrotron crystallography”, *Science*, **365**:6448 (2019), 61–65.
- [16] Hantke M.F., Ekeberg T., Maia F.R.N.C., “Condor: a simulation tool for flash X-ray imaging”, *Journal of Applied Crystallography*, **49**:4 (2016), 1356–1362.

Application of neural networks for recognition of conformational changes in protein structure by X-ray diffractograms of its single molecules on the example of bacteriorhodopsin photocycle
Armeev G.A., Kirpichnikov M.P., Kobelkov G.M., Kudryavtsev A.V., Lozhnikov M.A., Novoseletsky V.N., Shaitan A.K., Shaytan K.V.

Free electron lasers are becoming increasingly powerful and affordable installations for determining the structure of proteins. Conformational rearrangements are characteristic of the vast majority of proteins, but the transience of transformations often does not allow obtaining crystals of sufficient size. The solution could be to obtain X-ray diffraction patterns from single molecules, but in this case it is

not possible to determine their orientation, which makes it impossible to restore the structure by modern methods. The paper considers the applicability of a number of neural network architectures to the recognition of the conformational state of a protein based on diffraction data on single bacteriorhodopsin molecules.

Keywords: X-ray diffraction, neural networks, conformational states of protein, bacteriorhodopsin

References

- [1] Zhukovsky K.V., “Undulators and generation of X-ray pulses in free-electron lasers with self-amplified spontaneous emission”, *Moscow University Physics Bulletin*, **72:2** (2017), 128–143.
- [2] Oberthür D., “Biological single-particle imaging using XFELs — towards the next resolution revolution”, *IUCrJ*, **5:6** (2018), 663–666.
- [3] Tschentscher T., Bressler C., Grünert J., Madsen A., Mancuso A.P., Meyer M., Scherz A., Sinn H., Zastra U., “Photon Beam Transport and Scientific Instruments at the European XFEL”, *Applied Sciences*, **7:6** (2017), 592.
- [4] Mancuso A.P., Aquila A., Borchers G., Giewekemeyer K., Reimers N., “Scientific instrument single particles, clusters, and biomolecules (SPB)”, *XFEL. EU Technical Report*, 2013, 1–232.
- [5] Shaitan K.V., Armeev G.A., Shaytan A.K., “The dynamics of irreversible evaporation of a water–protein droplet and the problem of structural and dynamic experiments with single molecules”, *Biophysics*, **61:2** (2016), 177–184.
- [6] Shaitan K.V., Shaytan A.K., Bagrov D.V., Blinov V.N., Novoseletsky V.N., Sokolova O.S., Shurov D.L., Turchenkov D.A., Kirpichnikov M.P., “Algorithms and methods for the study of 3D atomistic models of proteins based on the analysis of the diffraction patterns of the high-power X-ray laser radiation”, *Nanostructures. Mathematical physics and modeling*, **9:2** (2013), 33–74 (In Russian).
- [7] Lueck H., Schobert B., Richter H. T., Cartailier J. P., Lanyi J. K., “Structural changes in bacteriorhodopsin during ion transport at 2 angstrom resolution”, *Science*, **286:5438** (1999), 255–260.
- [8] Nango E., Royant A., Kubo M., Nakane T., Wickstrand C., Kimura T., et al, “A three-dimensional movie of structural changes in bacteriorhodopsin”, *Science*, **354:6319** (2016), 1552–1557.

- [9] Park W.B., Chung J., Jung J., Sohn K., Singh S.P., Pyo M., Shin N., Sohn K.-S., “Classification of crystal structure using a convolutional neural network”, *IUCrJ*, **4**:4 (2017), 486–494.
- [10] Xu W., LeBeau J.M., “A deep convolutional neural network to analyze position averaged convergent beam electron diffraction patterns”, *Ultramicroscopy*, **188** (2018), 59–69.
- [11] Ke T.W., Brewster A.S., Yu S.X., Ushizima D., Yang C., Sauter N.K., “A convolutional neural network-based screening tool for X-ray serial crystallography”, *Journal of synchrotron radiation*, **25**:3 (2018), 655–670.
- [12] Simonyan K., Zisserman A., *Very deep convolutional networks for large-scale image recognition*, 2014, arXiv:1409.1556.
- [13] He K., Zhang X., Ren S., Sun J., “Deep residual learning for image recognition”, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 770–778.
- [14] He K., Zhang X., Ren S., Sun J., “Identity mappings in deep residual networks”, *European conference on computer vision*, 2016, 630–645.
- [15] Weinert T., Skopintsev P., James D., Dworkowski F., Panepucci E., Kekilli D., et al., “Proton uptake mechanism in bacteriorhodopsin captured by serial synchrotron crystallography”, *Science*, **365**:6448 (2019), 61–65.
- [16] Hantke M.F., Ekeberg T., Maia F.R.N.C., “Condor: a simulation tool for flash X-ray imaging”, *Journal of Applied Crystallography*, **49**:4 (2016), 1356–1362.