

# Как предсказать неизвестную структуру белка



*Если геном — это список инструментов в оркестре, то протеом — это оркестр, играющий симфонию.*

Р.Симпсон

А.В.Попинако, О.С.Соколова

Современная молекулярная биология стремительно расширяет круг знаний, область интересов перемещается от одиночных молекул к более сложным молекулярным машинам, растет потребность в структурной информации о комплексах нанобиообъектов. Прошлый век стал свидетелем революционного прорыва в биологии. Обнаружение совершенно нового мира с его неожиданными явлениями — событие редкое для современной науки. Расшифровка генома дала огромную статистическую информацию о потенциальной жизни конкретного организма. Однако по сравнению с количеством прочтенных пар оснований количество известных структур белков ничтожно. До сих пор в базах данных остаются неизвестными функции множества белков. Сегодня, после расшифровки генома, основной задачей исследователей стало подробное изучение протеома — совокупности всех белков организма.

Исследование протеома чем-то напоминает разгадывание древних текстов: буквы известны, но как толковать письма? Разгадать тайный смысл этих вечных текстов и помогают структурные исследования. Открытый доступ к полной информации о геномных последовательностях переориентировал направление научных исследований в биологии. Знание структуры каждого конкретного



**Анна Владимировна Попинако**, аспирант кафедры биоинженерии биологического факультета Московского государственного университета им.М.В.Ломоносова. Область научных интересов — протеомика и биоинформатика.



**Ольга Сергеевна Соколова**, кандидат биологических наук, доцент той же кафедры. Занимается электронной микроскопией и структурной биологией.

белка позволяет понять механизм его функционирования и взаимодействия с другими молекулами.

Соотнесение структур белков с продуктами определенных генов — первый шаг к пониманию того, каким образом организмы реализуют свою геномную информацию. При переходе к белковым структурам остается позади «одномерность» нуклеотидных и аминокислотных последовательностей — открывается трехмерный (3D) мир пространственных молекулярных структур.

В настоящее время для получения 3D-структур белковых молекул широко используются методы ядерного магнитного резонанса (ЯМР), спектроскопии и рентгеноструктурного анализа. К 2012 г. в мире расшифровано более 70 тыс. атомных структур белков, в их числе аквапорины, рибосомальные субъединицы, РНК- и ДНК-полимеразы и др., но для многих белков их четвертичная структура неизвестна: ионные каналы, транспортеры, крупные мультидоменные комплексы с трудом поддаются кристаллизации.

Для получения полной структурной информации о неизвестном белке (когда определена только последовательность нуклеотидов)

© Попинако А.В., Соколова О.С., 2012

приходится решать уравнение с двумя, а то и тремя неизвестными. Какие же методы существуют для определения третичной и четвертичной структур?

### Моделирование по гомологии

*Сама по себе аминокислотная последовательность еще мало о чем говорит, вот две гомологичные последовательности уже что-то нащепывают, ну а когда их много, они орут востор!*  
А.Леск

Эволюция белковых структур идет таким образом, что для произвольного белка существует лишь единственный — термодинамически наиболее выгодный (нативный) — способ укладки основной цепи. Но мутации ДНК часто порождают изменчивость в белковой структуре и функциях, что сказывается на жизнеспособности организма и, соответственно, в естественном отборе.

Вариации аминокислотных замен в гомологичных белках показывают, как структура приспосабливается к изменениям. Остатки на поверхности белка, не влияющие на его функцию, свободно мутируют. Механизмы, стабилизирующие структуру белка, ограничивают конформационные изменения, а функциональные требования накладывают дополнительные ограничения на структурные искажения. Таким образом, семейства родственных белков имеют тенденцию к сохранению общих

паттернов укладки. Поиск гомологий — это мост между аминокислотной последовательностью и функцией. Сегодня уже известны многие (но еще не все) способы укладки белков (фолды), поэтому предсказание структуры белка, основанное на сходстве с уже известными структурами, весьма продуктивно. Такой подход называется моделированием структуры по гомологии.

Рассмотрим, к примеру, кодируемый 18-й хромосомой человека белок FHOD3, для которого известна только аминокислотная последовательность. С помощью программы SWISS-MODEL в базе структурных данных (pdb.org) удалось найти родственников этого белка — гомологов с известной структурой. Ими оказались эукариотические формины mDia1, mDia2, Bni1, Daam1, которые инициируют и пролонгируют рост актиновых филаментов [1, 2]. Предполагается, что в клетке большинство форминнов находятся в автоингибированном состоянии и не могут спонтанно регулировать образование надмолекулярных структур актиновых филаментов. Активность форминнов блокируется за счет N-концевого DID-домена, который напрямую взаимодействует с DAD-доменом, находящимся на C-конце молекулы [3]. Отдельные фрагменты молекулы формина (C-концевой FH2-домен и N-концевой DID-домен) получены в кристаллическом состоянии [4, 5], но до настоящего времени информации о структуре целого белка и его конформационных перестройках при активации нет.

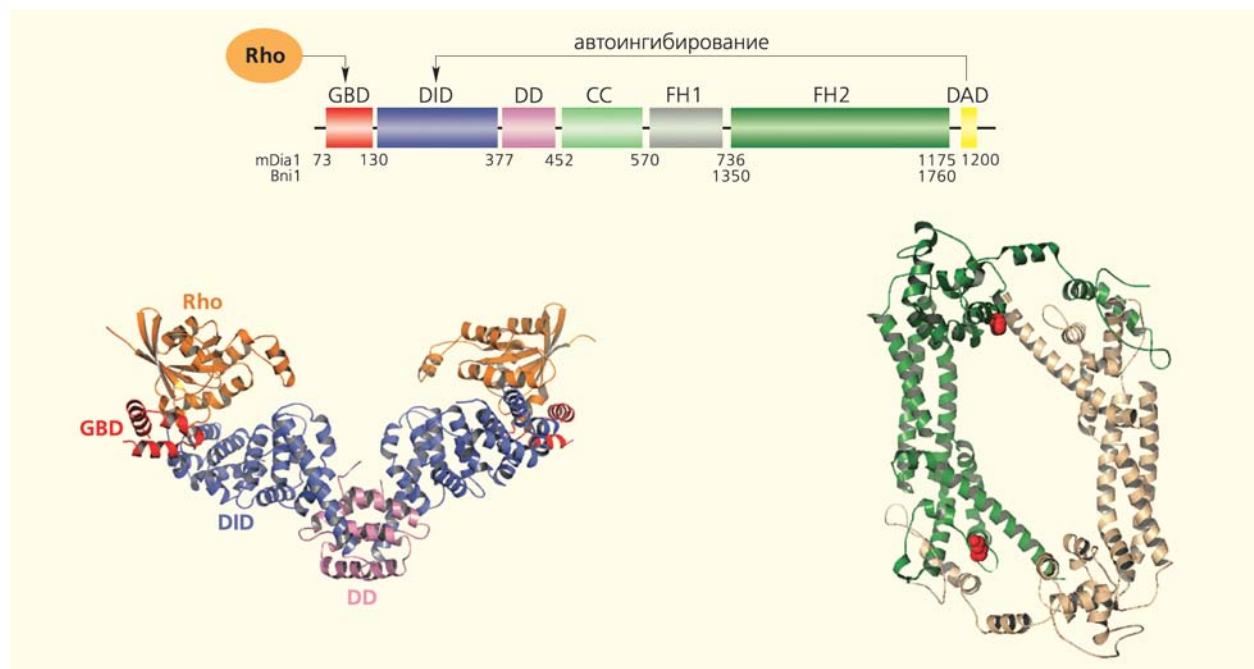


Рис.1. Схема доменной структуры мономера формина mDia (вверху); отмечены домены DAD, FH1, FH2, спиральный участок CC, домены DD, DID и домен, ответственный за связывание ГТФазы (GBD). Стрелка показывает взаимодействие доменов DD-DID-GBD в димерном формине. Внизу слева — предполагаемая ориентация мономеров в димерном формине. Каждый мономер связан с активатором ГТФазой (Rho). Отмечены также домены GBD, DID и DD. Внизу справа — кристаллическая структура C-концевого димера доменов FH2 [2].

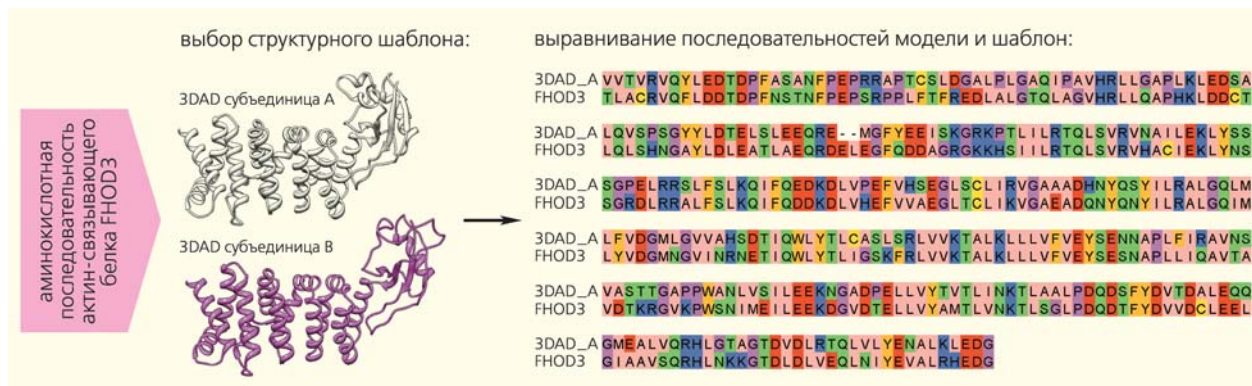


Рис.2. Выравнивание гомологичных последовательностей.

Основой для моделирования стала структура формина 3DAD — DID-домен субъединицы А (рис.1) с наибольшим процентом схожести. Известно, что полнота и качество модели зависят от схожести последовательностей. Если последовательности родственных белков схожи на 50%, то их конформация аналогична с вероятностью 90% [6].

Моделирование по гомологии начинается с выравнивания аминокислотных последовательностей шаблона и модели (рис.2). Мы проводили его с помощью программы T-COFFEE [7] и подтвердили высокую степень гомологии между фрагментами последовательностей формина и FHOD3: она составила 63.5%.

Для непосредственного моделирования по гомологии мы использовали программу MODELLER [8]. Построение гомологичной модели заключается в «натягивании» первичной последовательности моделируемого белка на трехмерный «остов» шаблона согласно исходному выравниванию. Наша программа построила десять 3D-моделей N-концевого домена белка FHOD3, среди которых мы выбрали одну — с наименьшим значением энергетической характеристики (рис.3). Поскольку мо-

дель, построенная по шаблону-структуре, имела высокое разрешение (2.3 Å) и большую степень гомологии, содержащую в ней информацию о пространственном расположении функционально важных аминокислотных остатков в белке можно использовать для проектирования новых экспериментов. Оптимизированная модель формина FHOD3 участвовала в дальнейших компьютерных исследованиях. В частности, методом молекулярной динамики изучалось взаимодействие гомологичных доменов белка FHOD3 (модели) и формина mDia1 (структуры) с C-концевым DAD-доменом формина и активатором Rho. Обе системы показали схожий характер взаимодействия. Дизайн точечной мутации DID-домена белка FHOD3 (модели) подтвердил роль некоторых заряженных аминокислотных остатков в связывании с Rho. Эти результаты доказывают пригодность полученной модели для структурных исследований.

Таким образом, метод моделирования по гомологии весьма полезен при изучении структуры и функции белка, его динамики, взаимодействия с лигандами и другими белками, а также при создании лекарственных препаратов.

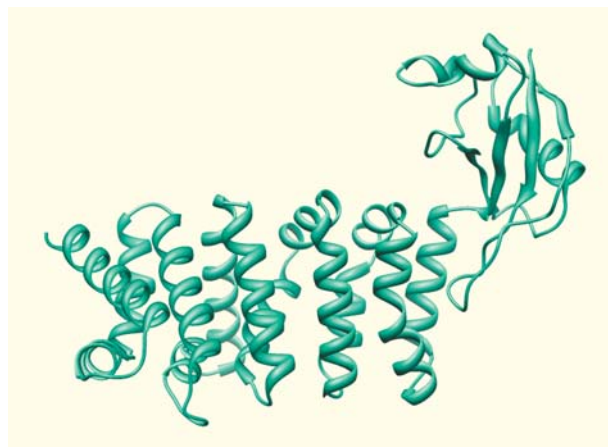


Рис.3. Финальная модель N-концевого домена белка FHOD3.

## Электронная микроскопия

*Мы не должны обижаться на тех ученых, которые вначале не поверили в электронный микроскоп.*  
Э.Руска. Нобелевская лекция. 1986 г.

В последнее десятилетие просвечивающая электронная микроскопия (ПЭМ) превратилась в полноправный метод, применяемый в структурной биологии. Это стало возможным в связи со значительным улучшением методик электронной микроскопии и повышением возможностей компьютеров, применяемых для расчетов трехмерных структур. Рассмотрим, какие принципы лежат в основе формирования электронно-микроскопического изображения.

Пучок ускоренных электронов, сформированный осветительной системой микроскопа, взаи-





Рис.4. Распределение энергии пучка ускоренных электронов, прошедшего сквозь пленочный объект, и возникающие при этом вторичные излучения.

модействует с исследуемым объектом. Некоторые электроны, проходя через вещество объекта, рассеиваются, т.е. меняют свои траектории (рис.4), а в результате и амплитуды и фазы электронных волновых функций. Линза объектива собирает прошедшие через образец электроны в плоскости изображения, где формируется картина, которая содержит искомую информацию о структуре объекта.

Реконструкция трехмерной структуры биомолекул становится возможной, потому что на препарате молекулярные комплексы находятся в виде изолированных частиц, как правило, отличающихся своей ориентацией. В процессе реконструкции множественные изображения отдельных идентичных наночастиц, полученные с помощью электронного микроскопа, подвергаются компьютерной обработке и классифицируются. В итоге, используя специальные алгоритмы, рассчитывают трехмерную структуру исходной молекулы. Изображений большого количества молекул в разных ориентациях бывает достаточно для получения трехмерной структуры с молекулярным разрешением. Обычно для этого хватает небольшого количества микрофотографий — при условии, что на одной из них содержится около 500 частиц.

Мы изучали трехмерное строение молекулы формина mDia1. Для микроскопии применили негативное контрастирование образцов водным раствором уранилацетата. Суть метода контрастирования состоит в том, что на сетку с подложкой, содержащей нанесенные исследуемые частицы, капает раствор электронноплотного вещества (контрастер). Он равномерно заливает всю площадь подложки, кроме микроучастков, заня-

тых частицами, которые выглядят светлыми на темном фоне (отсюда и название метода). В результате контрастирования электроны больше рассеиваются в области сосредоточения контрастера, а не образца.

Изображения получали в Центре коллективного пользования Института кристаллографии им.А.В.Шубникова РАН на электронном микроскопе FEI Tecnai G2 Spirit с ускоряющим напряжением 120 кВ в условиях низкой дозы облучения (для предотвращения повреждения образцов под воздействием электронного пучка). На углеродной подложке молекулы формина находились в различных ориентациях, что позволило использовать для получения трехмерных реконструкций «метод обратных проекций» [9]. Одиночные изображения mDia1 с микрофотографий собирались на компьютере в полуавтоматическом режиме по программе Signature [10]. Шаблоном служила вращательная сумма небольшого набора одиночных изображений, выбранных вручную. Последующая обработка в программе IMAGIC-5 [11] позволила отфильтровать изображения отдельных молекул от шума и выровнять относительно общей суммы каждого изображения путем их перемещения и вращения.

Выровненные изображения были разбиты на классы: в один из них вошли несколько десятков изображений молекул, имеющих одинаковую ориентацию. Классы лучшего качества (рис.5,а), включающие большее количество изображений, использовались далее для построения угловой реконструкции (рис.5,б), а полученные в дальнейшем обратные проекции — для оценки корректности построения реконструкции (рис.5,в). Для этого рассчитывалась корреляция между классами, используемыми для реконструкции, и соответствующими обратными проекциями. В описываемом случае ошибка реконструкции составила 3%.

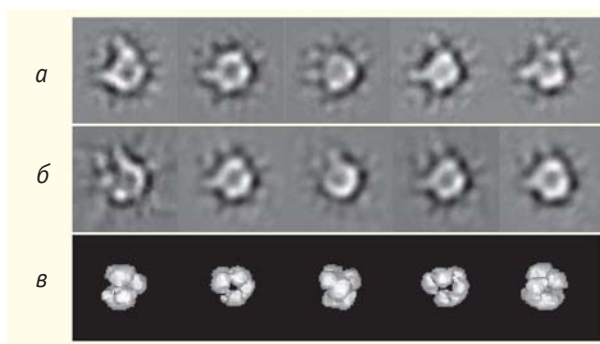


Рис.5. Реконструкция трехмерной структуры формина mDia1: а — избранные классы, полученные усреднением нескольких десятков одинаковых проекций; б — репроекции трехмерной структуры в тех же ориентациях, что и классы; в — трехмерная реконструкция в соответствующих ориентациях.

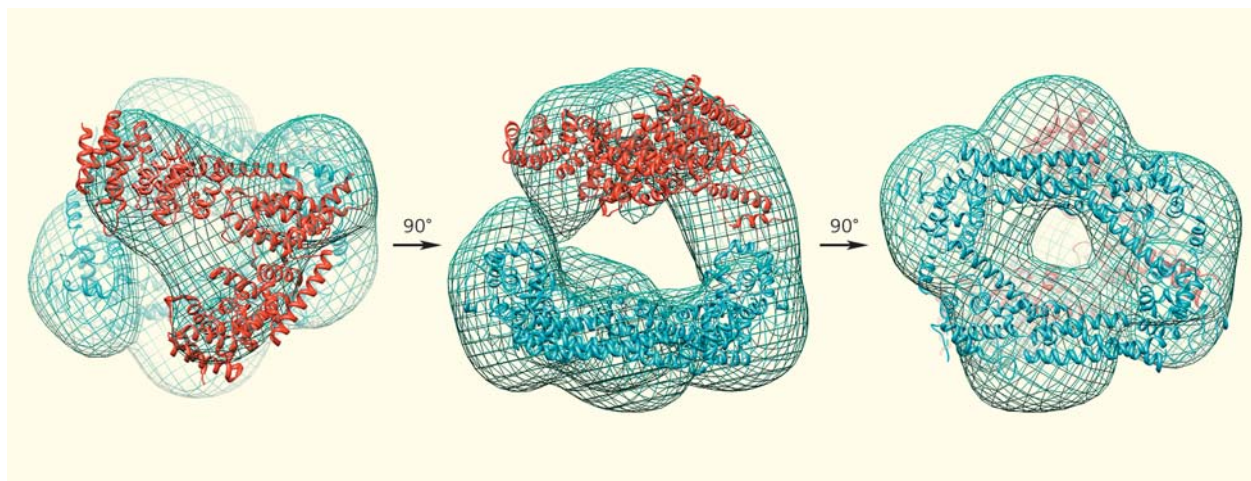


Рис.6. Трехмерная модель формина mDia1, отображенная в виде сетки с шагом 1 пх. Слева направо: вид сверху, вид сбоку и вид снизу. В верхнюю часть реконструкции проведен докинг кристаллической структуры N-концевого димера (красный цвет), в нижнюю — С-концевого димера (голубой цвет).

В ходе обработки данных на основе срезов на разном уровне объемной модели рассчитали трехмерное изображение поверхности структуры (рис.6). Итоговые 3D-структуры скорректировали согласно частотно-контрастной характеристике электронного микроскопа, используя программу FREALIGN [12]. Значительное улучшение изображений получили после 10 итераций, а затем рассчитали финальную структуру. Визуализацию и измерение полученных структур проводили с помощью программы UCSF Chimera [13]. Трехмерные структуры отображались в виде сетки, шаг которой соответствовал 1 пикселю. Выяснилось, что структура формина mDia1 образована двумя слоями: на верхнем в виде «вилки», а на нижнем — как кольцевая (рис.6).

Для дальнейших структурных исследований необходимо оценить разрешение полученной структуры. Для этого, как правило, весь массив данных разбивают на две части (например, в одной все четные частицы, а в другой — все нечетные). Затем для каждого массива рассчитывают независимую реконструкцию и с помощью объемного критерия Фурье находят корреляцию между реконструкциями [14]. Разрешение, при котором коэффициент корреляции равен 0.5, и есть искомое разрешение структуры. Существуют также другие критерии оценки разрешения, например  $3\sigma$  и  $5\sigma$ . Они показывают разрешение, при котором собрано достаточно данных, которые соответственно в три или пять раз превышают уровень фона.

При расчете трехмерной структуры формина разрешение реконструкции было 2.5 нм. По сравнению с атомным (0.1–0.3 нм) оно может считаться низким, но, тем не менее, позволяет делать структурные выводы о взаимодействии с лигандами, об олигомерном состоянии белка и его конформационных перестройках.

## Интерпретация структуры

*Уравнение может иметь как одно-единственное решение, так и бесконечное множество решений.*  
БСЭ

Следующий этап в исследовании неизвестной структуры — совмещение данных, полученных различными методами: данные ПЭМ комбинируются с данными рентгеноструктурного анализа или (при отсутствии атомной структуры) с данными молекулярного моделирования по гомологии. Обычно для интерпретации трехмерной структуры используют метод докинга, в основе которого лежит принцип взаимной корреляции. Докинг можно проводить вручную, с помощью программ для трехмерной визуализации наночастиц (например, UCSF Chimera [13]), или автоматически (программа Situs [15]).

Для интерпретации полученной нами трехмерной структуры формина с разрешением 2.5 нм мы использовали находящиеся в базе данных pdb.org кристаллические структуры N- и С-концевых димеров этой молекулы [4, 5]. Проводили автоматический докинг кристаллических структур в электронную плотность с помощью программы Situs [15] и рассчитывали коэффициенты корреляции для докинга N-концевого димера в область «вилки» ( $R = 0.69$ ) и С-концевого димера в область «кольца» ( $R = 0.8$ ). Оказалось, что полученная трехмерная реконструкция может вместить не более двух N- и двух С-концевых доменов (рис.6). Значит, полноразмерный формин — димер. Наши исследования также подтвердили гипотезу, что полноразмерный формин находится в ингибированной конформации, когда актинсвязывающие сайты С-концевого домена (на вогнутой поверхности «кольца») перекрываются N-концевым доменом. Следовательно, автоингибирование формина имеет стерическую природу.

Для структур с более высоким разрешением применяют молекулярно-динамический метод «гибкой подгонки» (flexible fitting) [16]. Он основан на добавлении внешних сил пропорционально градиенту электронной плотности трехмерной структуры в процессе молекулярной динамики атомной структуры. С помощью гибкой подгонки можно уточнить четвертичную структуру молекулы, сравнив ее кристаллическую структуру с таковой в растворе.

Итак, чем выше качество модели белковой молекулы, тем шире область ее использования для разработки новых лекарств, виртуального скрининга, докинга лигандов. Но насколько верны результаты структурных исследований, которые уже сегодня находят широкое применение? Мы узнаем об этом в ближайшее время, когда появятся еще более точные методы изучения структуры макромолекул. Уравнение жизни имеет бесконечное множество решений. ■

**Работа выполнена при поддержке Российского фонда фундаментальных исследований (проекты 08-04-91759 и 08-04-91125) и Минобрнауки (ГК №16.740.11.0373).**

## Литература

1. Faix J., Grosse R. Staying in shape with formins // *Dev. Cell*. 2006. V.10. P.693–706.
2. Goode B.L., Eck M.J. Mechanism and function of formins in the control of actin assembly // *Annu. Rev. Biochem.* 2007. V.76. P.593–627.
3. Moseley J.B., Sagot I., Manning A.L. et al. A conserved mechanism for Bni1- and mDia1-induced actin assembly and dual regulation of Bni1 by Bud6 and profilin // *Mol. Biol. Cell*. 2004. V.15. P.896–907.
4. Xu Y., Moseley J.B., Sagot I. et al. Crystal structures of a Formin Homology-2 domain reveal a tethered dimer architecture // *Cell*. 2004. V.116. P.711–723.
5. Otomo T., Tomchick D.R., Otomo C. et al. Structural basis of actin filament nucleation and processive capping by a formin homology 2 domain // *Nature*. 2005. V.433. P.488–494.
6. Леск А. Введение в биоинформатику. М., 2009.
7. Cüdric Notredame, Desmond G. Higgins, Jaap Heringa. T-coffee: a novel method for fast and accurate multiple sequence alignment // *Journal of Molecular Biology*. 2000. V.302. №1. P.205–217.
8. Modeller N., Eswar M.A., Marti-Renom B. et al. Comparative protein structure modeling with modeller // *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., Supplement 15, 5.6.1–5.6.30, 2006.
9. Heel M.van. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction // *Ultramicroscopy*. 1987. V.21. P.111–123.
10. Chen J.Z., Grigorieff N. Signature: A single-particle selection system for molecular electron microscopy // *J. Struct. Biol.* 2007. V.157. P.168–173.
11. Van Heel M., Harauz G., Orlova E.V. et al. A new generation of the IMAGIC image processing system // *J. Struct. Biol.* 1996. V.116. №1. P.17–24.
12. Grigorieff N. FREALIGN: High-resolution refinement of single particle structures // *J. Struct. Biol.* 2007. V.157. P.117–125.
13. Pettersen E.F., Goddard T.D., Huang C.C. et al. UCSF Chimera — a visualization system for exploratory research and analysis // *J. Comput. Chem.* 2004. V.25. P.1605–1612.
14. Harauz G., Heel M.van. Exact filters for general geometry three dimensional reconstruction // *Optik*. 1986. V.73. P.146–156.
15. Wriggers W. Using situs for the integration of multi-resolution structures // *Biophysical Reviews*. 2010. V.2. P.21–27.
16. Yassin A.S., Agrawal R.K., Banavali N.K. Computational exploration of structural hypotheses for an additional sequence in a mammalian mitochondrial protein // *PLoS One*. 2011. V.6. P.e21771.