

3D-ВИЗУАЛИЗАЦИЯ МАКРОМОЛЕКУЛ В БИОИНФОРМАТИКЕ: ЭПИСТЕМОЛОГИЧЕСКИЙ АСПЕКТ

М. Ю. Волошин

Московский государственный университет имени М. В. Ломоносова, Россия
allrour95@rambler.ru

Исследование выполнено при поддержке Междисциплинарной научно-образовательной школы Московского университета «Сохранение мирового культурно-исторического наследия»

Биоинформатики часто описывают собственную научную деятельность как практику работы с большими объемами данных с помощью вычислительных устройств. Существенной частью этого самоопределения является создание способов визуального представления результатов такой работы, некоторые из которых направлены на построение удобных репрезентаций данных и демонстрацию закономерностей, присутствующих в них (графики, диаграммы, графы). Другие являются способами визуализации объектов, непосредственно не доступных человеческому восприятию (микрофотография, рентгенограмма). И создание визуализаций, и особенно создание новых компьютерных методов визуализации рассматриваются в биоинформатике как значимые научные достижения.

Репрезентации трехмерной структуры белковых молекул занимают особое место в деятельности биоинформатиков. 3D-визуализация макромолекулы, с одной стороны, является, подобно графику, представлением результатов компьютерной обработки массивов данных, полученных материальными методами, – данных о взаимном расположении элементов молекулы. С другой стороны, подобно микрофотографии, такие 3D-структуры должны служить точными отображениями конкретных научных объектов. Это приводит к параллельному существованию двух противоречивых эпистемических режимов: творческий произвол в создании удобных, коммуникативно успешных моделей сочетается с верностью объекту «как он есть на самом деле».

Парадокс усиливается тем, что научное исследование репрезентируемых объектов (определение свойств структуры, ее функций, сравнение с другими структурами) посредством компьютеров само по себе вообще не требует визуализации. Ее очевидно высокая ценность для биоинформатики не выглядит оправданной, если иметь в виду значительную искусственность и художественность получаемых изображений.

Однако статус этих изображений становится яснее при соотнесении с более ранними представлениями о роли визуального в научном поиске. Высокая оценка визуализации как итогового результата научного исследования была характерна для науки эпохи Возрождения. Художественная репрезентация идеальных существенных свойств вместо строгого соответствия конкретному биологическому объекту – эпистемическая добродетель, типичная для натуралистов XVII–XVIII веков. И то и другое предполагало тесное сотрудничество ученого с художником; и стандарты визуализации макромолекул в биоинформатике вырастают из аналогичного сотрудничества (рисунки Гейса). Стремление же к максимальной точности и детализации наследует регулятиву «механической объективности» (как определяли это Л. Дастон и П. Галисон), для которого важным оказывается и устранение субъекта из процесса производства изображения (в биоинформатике – передача этих функций компьютерным программам). Таким образом, 3D-визуализация белковых структур несет на себе следы исторически разных ценностных ориентиров, но научная практика XX–XXI веков, дополненная компьютерными технологиями, позволяет им сочетаться в конкретных дисциплинарных единствах.

Ключевые слова: эпистемология, визуализация, научный объект, биоинформатика, анализ данных.

3D-VISUALIZATION OF MACROMOLECULES IN BIOINFORMATICS: AN EPISTEMOLOGICAL ASPECT

Mikhail Yu. Voloshin

Lomonosov Moscow State University, Moscow, Russian Federation
allrour95@rambler.ru

Bioinformatics scientists often describe their own scientific activities as the practice of working with large amounts of data using computing devices. An essential part of their self-identification is also the development of ways to visually represent the results of this work. Some of these methods are aimed at building convenient representations of data and demonstrating patterns present in them (graphics, diagrams, graphs). Others are ways of visualizing objects that are not directly accessible to human perception (microphotography, X-ray). Both the construction of visualizations and (especially) the creation of new computer visualization methods are considered in bioinformatics as significant scientific achievements. Representations of the three-dimensional structure of protein molecules play a special role in the inquiries of bioinformatics scientists. 3D-visualization of a macromolecule, on the one hand, is, like a graph, a representation of the results of computer processing of data arrays

obtained by material methods – spatiotemporal coordinates of structural elements of the molecule. On the other hand, like microphotography, these 3D structures should serve as accurate representations of specific scientific objects. This leads to the parallel existence of two contradictory epistemic regimes: creative arbitrariness in making convenient, communicatively successful models, is combined with commitment to the object “as it really is”. The paradox is reinforced by the fact that the scientific study of objects in question (determining the properties of the structure, its functions, comparison with other structures) by means of computers does not require visualization at all. Its obviously high value for bioinformatics does not look justified if we take into account the prominent artificiality and artistry of the resulting images. However, the status of these images becomes clearer if we relate them to earlier notions of the role of the visual in scientific discovery. The highest estimation of visualization as the final result of scientific research was characteristic of Renaissance science. The artistic representation of ideal essential properties, instead of a strict correspondence to a particular biological object, is an epistemic virtue typical of the naturalists of the 17th and 18th centuries. Both suggested a close collaboration between the scientist and the artist; and standards for visualizing macromolecules in bioinformatics grow out of a similar collaboration (Geis’ drawings). The desire for maximum accuracy and detail inherits the regulation of “mechanical objectivity” (as Daston and Galison put it into words), for which it is also important to eliminate humans from the image production process (in bioinformatics, to transfer these functions to computer programs). Thus, 3D-visualization of protein structures bears traces of historically different value orientations, but the scientific practice of the 20th and 21st centuries, supplemented by computer technologies, allows them to be intertwined in particular disciplinary units.

Keywords: epistemology, visualization, scientific object, bioinformatics, data analysis.

DOI 10.23951/2312-7899-2021-4-12-35

Учебник Певзнера

Третье издание учебника Джонатана Певзнера «Биоинформатика и функциональная геномика» содержит более 1 100 страниц [Pevsner 2015]. Слова «визуализация», «визуализировать» и производные от них встречаются 126 раз, то есть в среднем каждые 9 страниц. Среди вхождений этих слов много концептуально несущественных случаев (например, конструкция «визуализация чего-то» в подписях к иллюстрациям), но есть как минимум одно, игнорировать которое невозможно. Это цитируемое Певзнером определение

биоинформатики с сайта Национального института здоровья (National Institutes of Health, NIH): «Биоинформатика – это исследование, развитие и применение вычислительных средств и подходов для расширения использования биологических, медицинских, поведенческих данных или данных о здоровье, включая средства получения, хранения, организации, анализа и визуализации таких данных» [Pevsner 2015, 3].

Это определение носит перечислительный характер, и следует сделать два тривиальных замечания к нему: во-первых, визуализация в явном виде рассматривается как одна из существенных задач биоинформатики, а во-вторых, биоинформатика к ней не сводится. Второе особенно важно в связи с тем, что далее Певзнер делит упомянутые в определении методы и подходы на два большие группы по использованию компьютерных интерфейсов: интерфейс командной строки (command-line) и графический пользовательский интерфейс (graphical user interface, GUI) [Pevsner 2015, 10–11]. На уровне командной строки работа биоинформатика визуально мало чем отличается от работы программиста; учебники по биоинформатике часто включают в себя (или требуют в качестве базовых входных компетенций) знания основ программирования. GUI характерен для баз данных и визуальных способов репрезентации этих данных, для работы с которыми требуется только point-and-click (тоже выражение из Певзнера). Но и в этих случаях форма визуального представления данных внешне (для непосвященного пользователя) может не слишком отличаться от кода (например, длинные ряды текстовых символов).

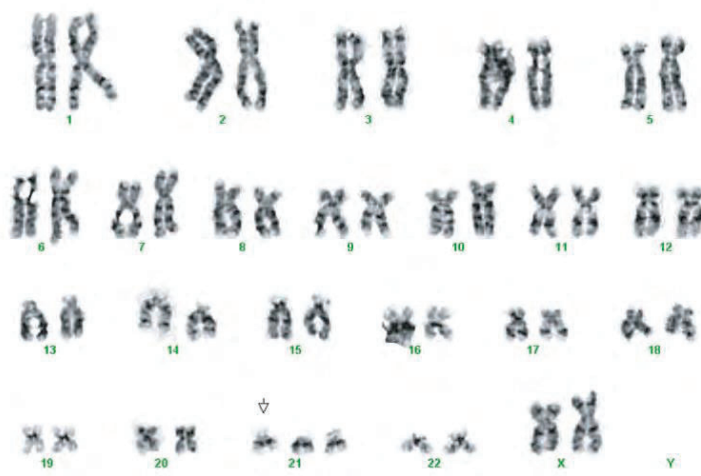
Мы приводим это замечание по двум причинам: во-первых, чтобы подчеркнуть, что «существенная задача биоинформатики» – визуализация – выполняется не всегда и не всеми вовлеченными в исследовательские практики; а во-вторых, чтобы указать на некорректность замечания историка биоинформатики Халлама Стивенса. Он связывал один из ключевых этапов становления этой дисциплины с появлением программ с «дружественным интерфейсом» на персональных компьютерах в лаборатории [Stevens 2013, 27–30]. Однако в 2015 году биоинформатиков все еще требуется учить программировать в командной строке. Говоря совсем строго, GUI также является визуализацией результата работы программного кода – визуализацией, естественно, значимой, но не конститутивной для биоинформатики как дисциплины.

Но вернемся к тексту Певзнера. В подавляющем большинстве случаев термин «визуализация» относится к репрезентации результатов

компьютерного анализа больших массивов данных: привычным многим научным дисциплинам графикам и гистограммам, а также графам (главным образом филогенетические деревья или сети, отображающие функциональные взаимосвязи фрагментов генома или протеома организма). Однако можно выделить три группы случаев, представляющих особый интерес.

Во-первых, это заголовки статей: в конце каждой главы учебника приводится небольшая библиография, и многие статьи специально посвящены программному обеспечению для визуализации данных. Например: VISTA: Visualizing global DNA sequence alignments of arbitrary length; Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration; DRAGON and DRAGON view: methods for the annotation, analysis, and visualization of large-scale gene expression data; Circular genome visualization and exploration using CGView и т. д. Мы можем принять в качестве не слишком сильного допущения, что научная статья отражает результат некоторого завершенного этапа исследовательской деятельности в рамках предметной области дисциплины; в таком случае можно утверждать, что сама возможность визуальной репрезентации рассматривается как значимый тип результата.

Во-вторых – редкие случаи, когда термин в учебнике Певзнера применяется не к созданному на компьютере изображению – о «визуализации» говорят в контексте флюоресцентного окрашивания, например, хромосом (ил. 1).



Ил. 1. Результат окрашивания хромосом красителем Райта; кариотип человека с синдромом Дауна (трисомия по 21 хромосоме). [Pevsner 2015, 313]

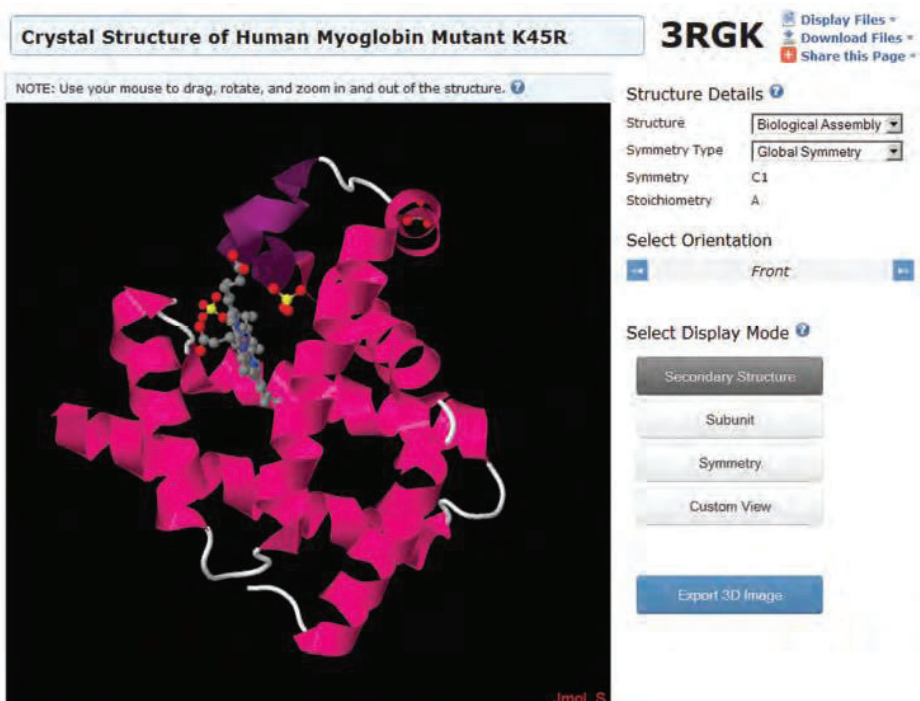
Эти случаи являются предметом цитологии, а не биоинформатики. Поэтому сходство способа употребления термина при явном различии семантики резко бросается в глаза: ведь речь идет не о репрезентации *данных*, а о репрезентации *объекта*. Сам термин «хромосома» означает «окрашенное тело», то есть такое тело, которое становится заметным и различимым при определенном способе цитологической работы. И существование хромосом, и их роль в клеточной жизни, и их количество¹ устанавливаются благодаря окрашиванию (и другим приемам подготовки материала). В то время как графики, гистограммы, графы (и филогенетические деревья) демонстрируют статистически значимые корреляции данных о природе, «визуализация» хромосом должна репрезентировать сам объект – так, как он выглядит *на самом деле*. В этом различии можно усмотреть функциональное разграничение «мокрых» и «сухих» лабораторий: если в первых работают с природными объектами, хотя и специальным образом подготовленными, то во вторых компьютеры ведут расчеты, выполняют алгоритмы и визуализируют результаты их работы. Все вышесказанное (включая определение биоинформатики и приведенные названия статей) должно вести к логичному заключению: биоинформатика – это то, что происходит в лабораториях второго типа.

Но это не совсем так.

Фигура умолчания

Глава 13 учебника [Revsner, 2015] содержит третью группу «странных» случаев использования термина «визуализация», случаев, которые преимущественно и будут нас здесь интересовать. Речь идет о программных средствах визуализации трехмерных объектов, в частности пространственных конфигураций белков (ил. 2).

¹ До 1956 года ошибочно полагали, что число хромосом в диплоидном наборе человека – 48. Джо Тью и Альберт Леван показали, что корректное число – 46. Для этого им пришлось усовершенствовать метод разграничения хромосом, изобретенный известным цитогенетиком Т. Хсю: по мнению авторов, «хотя [существующий метод] представляет собой значительный прогресс в разведении хромосом, их внешние границы часто оказываются размытыми и нечеткими» [Tjio, Levan 1956, 1].

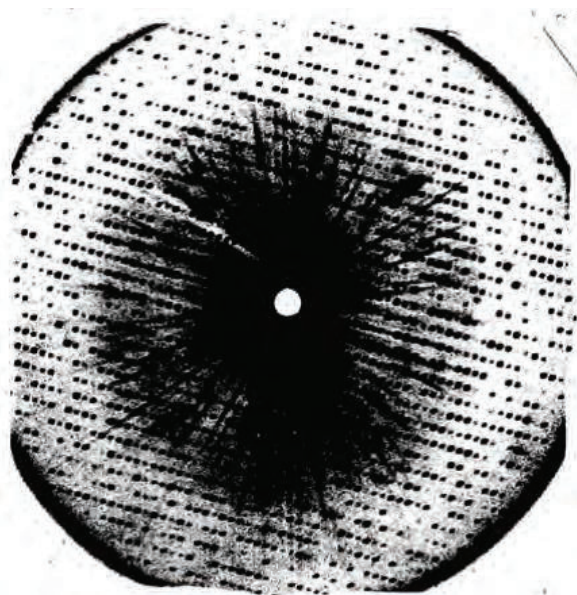


Ил. 2. Визуализация миоглобина человека с помощью программы Jmol, встроенной в базу данных Protein Data Bank. [Pevsner 2015, 607]

Задача определения трехмерной структуры белка возникла еще до появления биоинформатики как дисциплины. Ее биологическое значение состоит в том, что роли и функции белков в жизни клетки определяются их пространственной конформацией, а не напрямую последовательностью аминокислот, из которых они состоят. Хотя с середины XX века известно, что уникальная последовательность аминокислот в естественных условиях порождает уникальную структуру, обладающую локальным минимумом свободной энергии (так называемый «постулат Анфинсена»), сам процесс сворачивания белка (фолдинг) остается во многих случаях неясным. Кроме того, сходными структурами (и, соответственно, сходными функциями) могут обладать белки с существенными различиями в исходных последовательностях аминокислот.

Выявление пространственной конформации белка – задача, требующая работы как «мокрых», так и «сухих» лабораторий. Первое решение этой проблемы (исследование структуры миоглобина кашалота Джоном Кендрю и Максом Перутцем) было удостоено

Нобелевской премии по химии 1962 года. Сильно упрощая, процедуру решения проблемы можно описать следующим образом: с помощью рентгеновских лучей делают снимок кристалла белка, получая двумерное изображение; зная принципы дифракции лучей, можно интерпретировать полученное изображение как проекцию трехмерной структуры; изменяя угол «съемки» и увеличивая количество итераций, можно строить все более точные предположения относительно реального взаимного расположения атомов кристалла. Обработка массивов данных двумерных координат и их «перевод» в предполагаемую трехмерную структуру осуществляется компьютером. В 1957 г. Кендрю, используя EDSAC I (ламповый компьютер с максимальной скоростью вычислений до 15 000 в секунду), смог обработать около 400 изображений (одно из них представлено ниже; ил. 3) и получить структуру миоглобина в разрешении 6 Å. Позднее на основе 10 000 изображений ему удалось увеличить разрешение до 2 Å [Kendrew 1964, 678–679].



Ил. 3. Рентгеновский снимок миоглобина кашалота. [Kendrew 1964, 679]

Подобно Джеймсу Уотсону и Фрэнсису Крику несколькими годами ранее, Кендрю и Перутц собрали материальную модель на основе полученных трехмерных координат, «визуализировав» таким образом свои результаты (ил. 4).



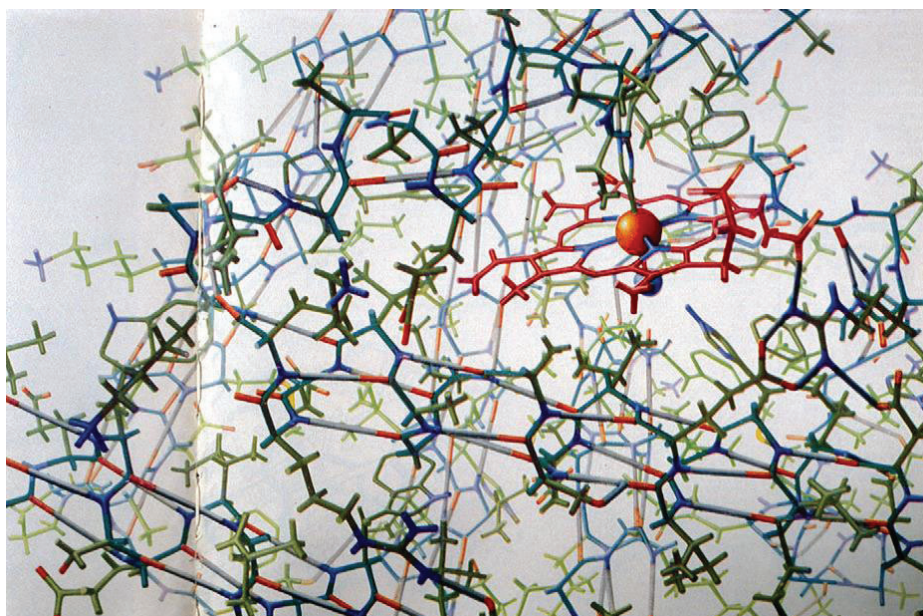
Ил. 4. Джон Кендрю и трехмерная модель миоглобина.
Фотография Макса Перутца. [Pevsner 2015, 588]



Ил. 5. Первая в истории трехмерная репрезентация структуры белка (Джон Кендрю, 1957–58 г.) Science Museum / Science & Society Picture Library. Image No. 10321094

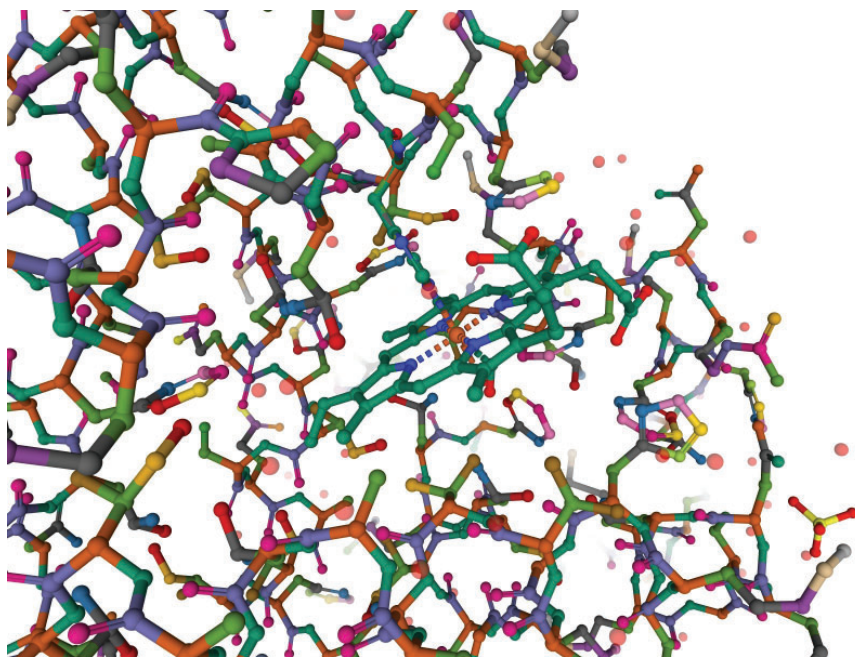
Это не первая модель миоглобина, построенная Кендрю. В 1975 году он отправил в Лондонский Музей науки пластилиновую модель, вероятно, являющуюся исторически первой моделью трехмерной структуры белка (ил. 5). Сам Кендрю был недоволен этой моделью, а увидевший ее в журнале *Nature* кристаллограф сказал Кендрю в личном письме: «Это ужасный объект, но прекрасная работа» [Chadarevian 2018, 1140].

Легко понять, что смутило и Кендрю, и кристаллографа; но и стержневая модель (см. ил. 4) далека от совершенства: находясь в непосредственной близости с ней, можно понять, как устроен миоглобин, но уже приведенная фотография не дает возможности хорошо различать расположения атомов. Для журнала *Nature* в 1958 году была использована перекрашенная в белый цвет версия «колбасной» модели (см. ил. 5), но для *Scientific American* в 1961 году художник Ирвинг Гейс перерисовал стержневую модель от руки, проведя рядом с материальным образцом около полугода (ил. 6). Итогом этой кропотливой работы стал не только рисунок для *Scientific American*: способ визуализации, использованный Гейсом, стал парадигмальным для многих компьютерных программ [Chadarevian 2018, 1141].



Ил. 6. Ирвинг Гейс: изображение структуры миоглобина на основе стержневой модели Кендрю и Перутца. [Chadarevian 2018, 1141]. Оранжевый шар чуть правее центра – атом железа, практически неразличимый на фотографиях материальных моделей (как стержневой, так и «колбасной»)

Современные инструменты биоинформатики предлагают множество вариантов визуализации. Тринадцатая глава учебника Певзнера, “*Protein Structure*”, по преимуществу представляет собой описание возможностей различных программных пакетов и баз данных, а также принципов их работы. Помимо *Protein Data Bank (PDB)* и используемого им *Jmol* (примеры которого приведены на ил. 2 и 7), это *Molecular Modeling DataBase (MMDB)*, используемая Национальным центром биотехнологической информации (NCBI), а также встроенные в MMDB инструменты визуализации (*Cn3D*) и сравнения структур (*Vector Alignment Search Tool, VAST*), инструменты поиска сходных топологий фолдинга (*CATH*)² и др. Иллюстрации, которые приводит Певзнер в главе 13, также показательны. В большинстве случаев это интерфейс соответствующего программного обеспечения: например, типичное отображение результатов поиска по базе данных³ или даже вид меню опций программы⁴.



Ил. 7. Миоглобин кашалота в базе PDB. Один из вариантов репрезентации.
В центре, окруженный пунктирными связями - оранжевый атом железа.
PDB ID: 1SPE

² Эта аббревиатура указывает на классификацию выделяемых для сравнения уровней структуры белка: *Class, Architecture, Topology, Homologous superfamily*.

³ Figures 13.9, 13.10, 13.12, 13.16 [Pevsner 2015, 605, 608, 613].

⁴ Figure 13.11b [Pevsner 2015, 607].

Чтение Певзнера неизбежно создает ощущение, что биоинформатику принципиально важно владеть навыками компьютерной визуализации, уметь манипулировать ее виртуальными результатами, выбирать уместный в каждом конкретном случае программный инструмент и понимать принципы его работы. Однако ни в главе «Структура белка», ни вообще в учебнике Певзнер не дает ответа на, казалось бы, естественный вопрос: *а зачем вообще биоинформатике визуализировать белки?*

Визуализация как цель и как средство

Молчание Певзнера можно понять. Вероятно, этот вопрос показался бы биоинформатику столь же странным, как вопрос астроному о том, зачем открывать новые звезды. Однако этот вопрос приобретает фундаментальную эпистемологическую значимость – не только для биоинформатики, но и для науки в целом – если обратить внимание на амбивалентность «визуализации», проявляющуюся в данном случае.

Визуализация структуры белка является репрезентацией результатов обработки больших объемов данных. В этом смысле она похожа на другие режимы «визуализации» в биоинформатике, например графики. Графики наглядно демонстрируют наличие в путаной и необозримой массе данных определенных взаимосвязей, корреляций, закономерностей. График должен предъявить читателю или зрителю то, что ученый *хочет* предъявить, и сделать это удобным для участников коммуникации образом. Неудивительно поэтому, что от него требуется владение большим разнообразием инструментов, подходящих под разные задачи⁵. Ученый в известной степени произволен выбирать между вариантами репрезентации: в какие цвета окрашивать линии графика или столбцы диаграмм, как масштабировать изображение и т. д.; хотя некоторые ограничения могут существовать в силу сложившейся традиции или элементарных соображений визуальной различимости. Все это имеет место и в трехмерных компьютерных моделях белков. Положение элементов молекулы в пространстве можно изобразить множеством разных способов: окрасив в произвольные цвета, выделив

⁵ Выше мы отмечали, что создание удобного способа визуализации данных рассматривается как значимое научное достижение, что отражается в публикуемых статьях по биоинформатике. Это же отражается и в программах подготовки ученых: многие биоинформатики изучают и используют язык R и связанную с ним визуальную оболочку RStudio, поскольку R считается языком, наиболее удобным для построения графических отображений данных.

произвольные характеристики (гидрофильность / гидрофобность, заряды частиц, форму поверхности и т. д.)

С другой стороны, эти визуализации рассматриваются как прямые репрезентации соответствующих объектов – то есть того, каковы эти белки *есть*. Здесь они сближаются со снимками окрашенных хромосом: соответствующим образом подготовленный объект⁶ должен быть отображен с максимально возможной тщательностью, с минимальными помехами, наиболее *достоверно*. Такая интенция прямо противоположна свободе действий, имеющейся у научного субъекта в первом варианте визуализаций. Выбирая способ представления объекта, ученый в качестве ценностной установки имеет точность воспроизводства действительности, а не коммуникативно-риторическую эффективность⁷. Первые, еще материальные, попытки визуализации (см. ил. 5), по-видимому, еще вполне следовали этой логике: «реальные» белки больше похожи на плотные «колбаски», чем на проволочно-шариковые конструкции (см. ил. 4, 6). Кристаллография и вообще материальные методы хранят ей верность и по сей день. Кендрю и Перутц были не удовлетворены детализацией снимков в 6 Å (первые расчеты, 1957) и активно стремились увеличить разрешение; к моменту вручения Нобелевской премии (1962) они довели разрешающую способность до 2 Å и работали над разрешением 1,4 Å [Kendrew 1964, 687–689]. Примечательно, что это стремление к детализации не сопровождалось пояснениями, зачем нужна такая точность; видимо, Кендрю и Перутцу казалось очевидным, что объект должен быть изображен как можно детальнее⁸.

На данный момент в базе PDB хранится 178 747 структур, из них лишь 10 402 (около 6 %) имеют разрешение менее 1,4 Å⁹. Возможно, стандарты детализации Кендрю и Перутца в начале 1960-х годов

⁶ О трансформациях, претерпеваемых объектами в ходе разных режимов подготовки, см.: [Rheinberger 2000].

⁷ Мы выше указывали на типичные заголовки статей о методах визуализации статистических данных. Примечательно, что у Певзнера в списке литературы к главе 13, где как раз представлены методы 3D-визуализации, нет ни одной статьи с названием аналогичного типа. Вместо «визуализации структур» в заголовках употребляются «предсказание структур», «анализ» или даже «добыча» (mining) [Pevsner 2015, 628–633].

⁸ Увеличение точности сопровождается уменьшением вероятности погрешности: было обнаружено, что отдельные фрагменты и атомы после обработки большого количества снимков расположены не так, как предполагала первоначальная модель (6 Å). В то же время получение изображения в более высоком разрешении требует большего количества вычислительных итераций, что, в свою очередь, ведет к накоплению ошибок. Это противоречие между точностью и достоверностью модели сравнительно недавно стало предметом специального анализа в эпистемологии компьютерных симуляций [Lenhard 2007, Lenhard 2019].

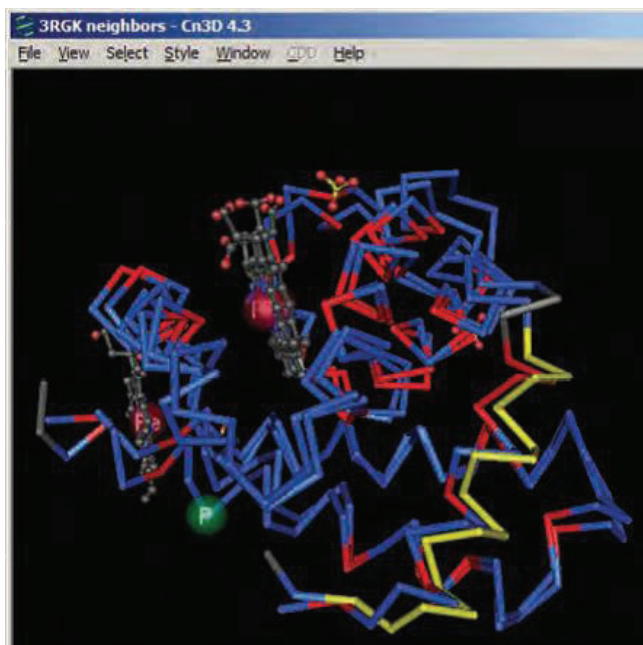
⁹ По состоянию на 14 июня 2021 года.

уже были достаточно высоки; а возможно, само стремление получить *точную* картинку сменилось стремлением получить картинку *удобную*. В пользу второго предположения говорит и рассмотренная выше история с рисунками Гейса: переход к хорошо различимым и эстетически привлекательным, но мало похожим на «реальные» белки способам визуализации. Остается неясным: были ли, помимо эстетических, *эпистемические* основания перехода к этим новым способам? Иначе говоря, какова познавательная ценность этих визуализаций для дисциплины?

Этот вопрос не эквивалентен вопросу о том, зачем выяснять структуру белка. Певзнер описывает две стратегии работы со структурами белков, соотнося их с двумя дисциплинарными регионами: классической структурной биологией и структурной геномикой. Структурная биология работает с белками, чьи функции известны; такой белок выделяется, очищается, кристаллизуется и подвергается рентгеноструктурному анализу (или другим материальным методам, например ЯМР). Выяснив строение этого белка, можно связать элементы его структуры (например, устройство активного центра) с известными выполняемыми функциями и понять, как именно данный белок выполняет данные функции. Структурная геномика работает иначе: исходя из некоторой последовательности ДНК или РНК, устанавливают аминокислотную последовательность кодируемого белка, затем *in silico* определяют, какова возможная структура этого белка. Соотнеся полученные данные с уже известными функциями типичных структурных элементов других белков¹⁰, можно сделать предположение о том, какие функции способен выполнять изучаемый белок. Параллельно можно экспрессировать этот белок в модельном организме (в клетках *E. coli*), чтобы затем пойти по пути классической структурной биологии: получить реальный кристалл белка и исследовать его материально [Pevsner 2015, 600–601].

Обратим внимание на то, что в обоих вариантах структура белка не является чем-то самоценным: по структуре нужно понять функции. Если функция белка неизвестна, то требуется сравнить ее структуру с другими структурами, предположив, что сходные структурные элементы указывают на сходство роли в клетке. Но что значит «сравнить»? Сопоставлением структур также занимаются специальные программы. Мы упоминали наиболее популярную из них – VAST (ил. 8).

¹⁰ Существуют обширные базы данных не только для цельных структур, но и для типичных структурных элементов (например, SCOP).



Ил. 8. Сравнение структур миоглобина (PDB ID: 3GRK) и бета-глобина (PDB ID: 4HHB) с помощью VAST. [Pevsner 2015, 611]



Ил. 9. Выравнивание в CATN суперсемейства белков – множества консервативных структур, обладающих, вероятно, общим эволюционным происхождением. [Pevsner 2015, 615]

Здесь можно видеть «невооруженным взглядом», что аминокислотные цепочки двух белков практически совпадают (а заодно заметить еще один возможный стиль визуализации). В какой именно мере они совпадают и как в точных терминах определить «совпадение», остается неясным, если исходить только из картинки. К тому же Певзнер не ищет легких путей и снабжает нас еще одним выравниванием структур, теперь уже в программе SATH (ил. 9).

Что для понимания степени родства, или общности функционала, суперсемейства белков, в которые входит и миоглобин, дает это изображение? SATH самостоятельно определяет ряд структур как гомологи и описывает их как суперсемейство¹¹; SATH также может наложить их друг на друга и выдать что-то вроде ил. 9. В чем смысл второго этапа? Зачем человеку (а точнее, ученому-биоинформатику) *видеть* это изображение?

В самом деле, заметная «на глазок» степень сходства структур имеет точные математические выражения. Это, например, такие метрики различия, как расстояние Хэмминга или расстояние Левенштейна, применяемые для сравнения не только трехмерных структур, но и линейных последовательностей (нуклеотидов и аминокислот). Таким образом, один набор данных (координаты структурных элементов белка) сравнивается с другим аналогичным набором данных посредством некоторого алгоритма, результатом работы которого является оценка степени сходства [Леск 2013, 263–267]. Параллельно можно визуализировать одну и вторую молекулы и наложить их друг на друга, получив изображение, похожее на ил. 8. Таким образом, если целью работы биоинформатика в этой области является выяснение сходства структур белков для лучшего понимания их функций (а именно так формулирует задачи биоинформатики Певзнер), то эта цель может быть успешно достигнута *вообще без визуализации*.

Более того, визуализация может стать препятствием. На фоне тотального стремления к визуальной репрезентации белков довольно забавно выглядит начало главы о структуре белков в учебнике Леска 2002 года: «К настоящему моменту мы знаем структуру примерно 15 000 белков... Первая проблема, связанная с анализом этих структур, касается визуализации. [Существуют] трудности интерпретации как полностью детализированной презентации структуры,

¹¹ Следует сделать важное пояснение: хотя аминокислотная последовательность белка определяет его структуру, обратное неверно: сходство структур не обязательно означает сходство последовательностей. Поэтому адекватно определить гомологии (суперсемейства) невозможно, сравнивая только секвенированные линейные последовательности аминокислот; требуется сравнивать именно пространственные параметры белков.

так и тех упрощенных картинок, которые выдают компьютерные программы при визуализации» [Леск 2013, 248]. Во-первых, визуализация прямо объявляется «проблемой» в первом же разделе учебника, где о ней заходит речь. Во-вторых, под «полностью детализированной» (а не упрощенной) визуализацией имеется в виду формат, введенный Гейсом (см. ил. 6, 7); такие изображения, естественно, тоже «выдают компьютерные программы», и они тоже являются «упрощенными». В-третьих, Леск предлагает довольно интересное решение трудностей интерпретации: изобразив на следующей странице несколько вариантов визуализации, он предлагает «попытаться мысленно наложить различные презентации друг на друга, как бы поворачивая страницу на 90°, и представить себе объемные структуры (не слишком при этом напрягаясь)» [Леск 2013, 250]¹².

Можно усмотреть в этом трудность технического характера: в 2002 году в PDB еще не было функционала виртуальной 3D-репрезентации; сейчас каждый может получить объемное изображение белка, покрутить его, приблизить или отдалить, перекрасить или изменить стилистику. Первым читателям Леска приходилось поворачивать страницу туда-сюда. Однако если риск ошибочной интерпретации довольно высок (на что и указывал Леск), а ценность воображаемой структуры для научного поиска как минимум не ясна, возникает ощущение, что такая визуализация носит, скорее, развлекательный характер. Любопытно, что Кендрю и Перутц в начале 1960-х годов тоже озаботились проблемой адекватного трехмерного представления своей молекулы для зрителя: поскольку множить материальные модели было дорого и неудобно, а их фотографии (см. ил. 4) были трудно интерпретируемы, они подготовили слайды со стереоизображениями рисунка Гейса и раздали зрителям в аудитории стереоочки [Chadarevian 2018, 1141].

Эта невыносимая жажда визуального, стремление «увидеть своими глазами», не находит оправдания в рациональной логике оптимизации научного поиска. Если координаты белковых структур нужны для того, чтобы соотнести их с координатами других структур, то их визуализация – лишний промежуточный шаг, включающий в себя затраты вычислительных ресурсов (в условиях их постоянной нехватки для решения биоинформатических задач). Однако этот «лишний» шаг оказывается для биоинформатики настолько существенным, что словарь терминов в конце одного из учебников,

¹² Автор этих строк, не являющийся биоинформатиком, может подтвердить, что напрягаться действительно не приходится.

определяя «моделирование», выдает следующее: «(в биоинформатике) обозначает молекулярное моделирование, процесс, в рамках которого трехмерная архитектура биологических молекул интерпретируется (или предсказывается), визуально репрезентируется и подвергается манипуляциям, чтобы определить их молекулярные свойства. (в общем) серия математических уравнений или процедур, которые имитируют реальный процесс, имея набор допущений, граничных параметров и начальных условий» [He, Petoukhov 2011, 283]. Именно так: пока «моделирование вообще» связано с решением уравнений, «моделирование в биоинформатике» визуализирует трехмерные структуры, при том что ответами на существенные исследовательские вопросы все равно будут решения уравнений.

Ученый и художник

В то же самое время, когда Ирвинг Гейс работал над цветным изображением стержневой модели (ил. 4 и 6), Морис Мерло-Понти писал свое знаменитое эссе о феноменологии восприятия живописи – «Око и дух». Оно начинается с критической оценки возможностей научного познания в том, что касается доступа к действительному миру; вместо погружения в него наука – особенно современная наука – конструирует произвольные модели его фрагментов, обладающие автономностью как по отношению к миру, так и по отношению друг к другу [Мерло-Понти 1992, 8–9]. Так, в 1960 году Мерло-Понти высказал то, что позже станет общим местом для сторонников прагматического подхода в философии науки (Н. Картрайт, М. Моррисон, М. Морган, Я. Хакинг и др.). Но его пессимизм в отношении научного мышления в сочетании с оптимизмом в отношении познавательных средств живописи, на наш взгляд, не нашел бы понимания у этих эпистемологов. Автономность научных моделей для прагматического подхода означает, помимо прочего, разнообразие практик их конструирования, в том числе – с использованием художественных средств выражения [Morrison, Morgan 1999, 10]. Иначе говоря, произведение живописи вполне может *одновременно* быть инструментом научного познания; эпистемические достоинства первого компенсируют недостатки пропозициональных форм второго.

Мерло-Понти пишет, например, о живописи: «Она дает видимое бытие тому, что обычное, заурядное зрение полагает невидимым,

она делает так, что нам уже не нужно 'мышечного чувства', чтобы обладать объемом мира» [Мерло-Понти 1992, 18]. Конечно, под «заурядным зрением» имеется в виду не диапазон разрешающей способности человеческого глаза, которому недоступна молекулярная структура белка. Но вспомним и о том, что виртуальная визуализация не является увеличенной копией объекта: она прорисовывается, окрашивается, вращается в зависимости от того, какое из «невидимых» свойств молекулы следует сделать различимым. Цвет может говорить о вещах не меньше, чем их пространственная протяженность. Критикуя картезианскую концепцию живописи, Мерло-Понти говорит, что, если бы Декарт обратил должное внимание на цвета, он «обнаружил бы, что не существует регулярного, или проективного, отношения между ними и подлинными свойствами вещей и что тем не менее их свидетельство нам понятно» [Мерло-Понти 1992, 27]. Если пространственную структуру виртуальной модели можно интерпретировать как абстракцию действительного расположения элементов молекулы (шарик – там, где находится атом железа), то достаточно произвольные – но при этом *все равно понятные* – стандарты ее окраски не связаны с действительностью никаким структурным подобием (почему этот шарик оранжевый?)

Вероятно, Мерло-Понти бы сказал, что живопись, о которой он пишет (Сезанн, Клее, да Винчи и др.), не имеет ничего общего с производством красивых картинок из наборов данных на компьютерах биоинформатиков. Но сами биоинформатики думают иначе: на обложке учебника Певзнера трехмерные изображения спирали ДНК и белка альбумина соседствуют с рисунками Леонардо (ил. 10), а эпиграфом к главе о структуре белка служат еще более красноречивые слова, принадлежащие Эдгару Мейеру, одному из основателей *Protein Data Bank*: «Посетитель Академии во Флоренции может видеть величественные образы, созданные руками Микеланджело из мраморных блоков. По аналогии с этим не-кристаллограф может увидеть то, что видит кристаллограф, восхищающийся строгой формой кристалла и собирающийся изучить изумительную структуру, скрытую внутри него. *Protein Data Bank* – это наш музей, в котором модели молекул отражают чудеса природы и сложные формы, возможно, такие же древние, как сама жизнь. С помощью интерактивной графики и сетевых технологий PDB делает эти изображения легкодоступными. Какие чудеса все еще остаются сокрытыми, пока мы строим, сравниваем и расширяем нашу базу данных?» [Pevsner 2015, 589].



Ил. 10. Фрагмент обложки третьего издания учебника Певзнера. [Pevsner 2015, cover]. Обложка второго издания (2009) также содержала рисунок из тетрадей Леонардо, а первого (2003) – рисунок человека из «Анатомических таблиц» Пьетро Берреттини

Построение компьютерных визуализаций, таким образом, ассоциируется с практикой художественной, но при этом познавательной. Интуиция того, что, чтобы быть *познанным*, нечто должно быть *изображено*, в явном виде восходит к ренессансной эпистемологии, к декларациям о необходимости союза между людьми учености и мастерами изобразительного искусства (на таком союзе настаивали, например, Альбрехт Дюрер, Андреас Везалий, Джон Ди) [Лисович 2015, 35–46; Йейтс 2018, 53–97]. Речь идет не только о том, что изображение должно помочь в познании чего-то нового или в лучшем освоении учебного материала (как, например, чертеж помогает понять доказательство геометрической теоремы). В случае с биоинформатикой само получение детализированного трехмерного изображения расценивается как самодостаточный результат, подводящий итог научному поиску ответа на вопрос, что есть тот или

иной белок. Несколько выше мы назвали это «невыносимой жаждой визуального»; В. Л. Глазычев в «Гемме Коперника» пишет о «неутолимой жадности взора человека Ренессанса» [Глазычев 1989, 158].

С эпохи Возрождения и вплоть до середины XIX века научная репрезентация биологических объектов во многом подчинялась регулятиву, концептуализированному Дастон и Галисоном как «истина-по-природе»¹³. Желаемой визуализацией научного объекта является художественный образ, выхватывающий из многообразия случайных проявлений «идеальный тип» объекта (например, рисунок растения, который не отсылает ни к какому реально существующему растению, но содержит в себе характеристические черты биологического вида). Для достижения этой цели биолог и художник буквально должны работать вместе, и высококачественные иллюстрации в ботаническом атласе, подобно 3D-моделям белков в базах данных, являются значимым и самодостаточным научным достижением¹⁴. Но когда речь идет о размерах, измеряемых в ангстремах, художник вынужден уступить свое место техническому специалисту по микрофотографии или рентгенокристаллографии; из режима «истины-по-природе» практика визуализации переходит в режим «механической объективности»¹⁵. Здесь ценностью становятся регистрация явлений природы «как они есть сами по себе», минимизация всего, что связано со свойствами наблюдателя, привносимых им интерпретаций, погрешностей его органов чувств, максимизация точности и достоверности отображения. Таковы были рентгенограммы миоглобина, полученные Кендрю и Перутцем (см. ил. 3), а несколько ранее – известные работы Розалинд Франклин и Мориса Уилкинса.

Биоинформатическую научную деятельность, в концепции Дастон и Галисона, логичнее всего связывать с другими типами эпистемических добродетелей: «тренированным суждением» (экспертно интерпретированные репрезентации) или «видением как производством» (манипулируемые интерактивные образы)¹⁶. Это оправданно исторически и проще концептуально. Но есть и менее тривиальная возможность: показать, что в высокотехнологичной научной прак-

¹³ См.: [Дастон, Галисон 2018], глава 2.

¹⁴ Так, художник Георг Эрет, иллюстратор Линнея, за свои заслуги на этом поприще был принят в Лондонское Королевское общество [Глазычев 1989, 200].

¹⁵ См.: [Дастон, Галисон 2018], глава 3.

¹⁶ Особенно, конечно, с последней: «Эти цифровые образы предназначены для использования, вырезания, корреляции, вращения и цветовой обработки при помощи кликов мыши и нажатия клавиш» [Дастон, Галисон 2018, 542].

тике XXI века продолжают просвечивать и «истина-по-природе», и «механическая объективность». Амбивалентность визуализации, которую мы отмечали – *удобное* представление данных vs *точное* представление объекта – становится более понятной, если отследить эти регулятивы.

Свобода выбора визуальных эффектов, имеющаяся у любого пользователя PDB, конечно, отличается от свободы художника, сопровождавшего натуралиста: во втором случае художник – визуальный проводник интеллектуальной воли ученого и не вправе самостоятельно решать, в какой форме и каким цветом изобразить объект. Но нужно помнить, что это ограничение художественного произвола осуществляется во имя выявления идеальной сути исследуемого, отображения его существенных свойств. В визуализациях белковых структур «существенность» свойств релятивизируется: можно быть «существенным» для определенных задач и в определенном контексте. Но общая ориентация выбора средств сохраняется. В принципе, каждый из около 2 500 атомов молекулы миоглобина кашалота можно окрасить в свой уникальный цвет, но этого не делают, так как нет задачи, для решения которой это существенно. Изображения молекул, хранящиеся в PDB, принципиально нереалистичны, но выражают «смысл» этих молекул – как того требует регулятив «истины-по-природе».

Однако аннотация каждой структуры в PDB содержит указание на ее разрешение, то есть степень точности. Жажда детализации, столь явно заметная у Кендрю, – типичная характеристика «механической объективности». Максимальное устранение субъекта из процесса производства модели обеспечивается просто тем, что модель строится вообще не человеком: приемлемое разрешение структуры достижимо лишь использованием компьютера¹⁷. Непосредственным результатом, однако, является визуально неоформленный массив данных, подлежащий интерпретациям и манипуляциям, опять же с помощью компьютерных программ. Общая «машинность» происходящего является, по сути, реализацией идеала механической объективности¹⁸.

¹⁷ Усиливающееся отстранение человека от процесса производства научного знания в дисциплинах, активно использующих компьютерные технологии, дало повод для дискуссии о необходимости разработки «нечеловеческой эпистемологии» [Humphreys 2009, 625; Lenhard 2019, 120–126].

¹⁸ «Машина, как представлялось, давала изображения, не зараженные интерпретациями... Притязания ученых на такого рода свободную от суждения репрезентацию являются доказательством интенсивности их желания получить совершенное, 'чистое' изображение. В этом контексте машина олицетворяла собой подлинность» [Дастон, Галисон 2018, 216].

Само стремление визуализировать результаты машинной работы – как акт удостоверения в том, что процесс познания доведен до некоторого логического конца, – имеет, как мы указывали, ренессансное происхождение. Изображение при этом строится компьютером, а не художником, но ученый определяет то существенное, что должно быть тем или иным способом представлено; и, возможно, неявно предполагается, что это «существенное», как и идеальные свойства организмов у ботаников XVIII века, было перед тем обнаружено некоторым интеллигентным образом – если считать, что для «созерцания идей» в XX веке обнаружился суррогат в виде компьютерной обработки данных.

Заключение

Настойчивое стремление к визуальной репрезентации в биоинформатике остается непонятным, если оставаться внутри этой дисциплины и руководствоваться логикой рациональной экономии ресурсов – как концептуальных, так и вычислительных. Мы предложили более широкий эпистемологический контекст, в котором компьютерная визуализация белковых молекул предстает хоть и очень далеким, но все же потомком ранних союзов ученых и художников, а также более поздних союзов ученых и техников. Она несет на себе следы их ценностных ориентиров, порой радикально несовместимых; но техника XX – начала XXI века позволяет этим следам сосуществовать в причудливом переплетении – в компьютерах, базах данных, иллюстрациях к учебникам по биоинформатике.

БИБЛИОГРАФИЯ

- Глазычев 1989 – *Глазычев В. Л.* Гемма Коперника. Мир науки в изобразительном искусстве. М.: Советский художник, 1989.
- Дастон, Галисон 2018 – *Дастон Л., Галисон П.* Объективность. М.: Новое литературное обозрение, 2018.
- Йейтс 2019 – *Йейтс Ф.* Театр Мира. М.: Циолковский, 2019.
- Леск 2013 – *Леск А.* Введение в биоинформатику. М.: Бином, Лаборатория знаний, 2013.
- Лисович 2015 – *Лисович И. И.* Скальпель разума и крылья воображения: научные дискурсы в английской культуре раннего Нового времени. М.: Изд. дом ВШЭ, 2015.

- Мерло-Понти 1992 – *Мерло-Понти М.* Око и дух. М.: Искусство, 1992.
- Chadarevian 2018 – *Chadarevian S.* John Kendrew and myoglobin: protein structure determination in 1950s // *Protein Science*. 2018. Vol. 27. P. 1136–1143.
- He, Petoukhov 2011 – *He M., Petoukhov S.* Mathematics of bioinformatics. Theory, practice and applications. Hoboken: John Wiley & Sons Inc., 2011.
- Humphreys 2009 – *Humphreys P.* The philosophical novelty of computer simulation methods // *Synthese*. 2009. Vol. 169 (3). P. 615–626.
- Kendrew 1964 – *Kendrew J. C.* Myoglobin and the structure of proteins : Nobel lecture, December 11, 1962 // *Nobel Lectures, Chemistry, 1942–1962*. Amsterdam, 1964. P. 676–698.
- Lenhard 2007 – *Lenhard J.* Computer simulation: the cooperation between experimenting and modelling // *Philosophy of science*. 2007. Vol. 74 (2). P. 176–194.
- Lenhard 2019 – *Lenhard J.* Calculated surprises: a philosophy of computer simulation. New York: Oxford University Press, 2019.
- Morrison, Morgan 1999 – *Morrison M., Morgan M.* Models as mediating instruments // *Models as mediators: perspectives on natural and social science* / eds. M. Morrison, M. Morgan. New York: Cambridge University Press, 1999. P. 10–37.
- Pevsner 2015 – *Pevsner J.* Bioinformatics and functional genomics. 3rd ed. Singapore: John Wiley & Sons Inc., 2015.
- Rheinberger 2000 – *Rheinberger H.-J.* Cytoplasmic particles. The trajectory of a scientific object // *Biographies of scientific objects* / ed. by L. Daston. Chicago, 2000. P. 270–294.
- Stevens 2013 – *Stevens H.* Life out of sequence: a data-driven history of bioinformatics. Chicago: University of Chicago Press, 2013.
- Tjio, Levan 1956 – *Tjio J.H, Levan A.* The chromosome number of man // *Hereditas*. 1956. Vol. 42. P. 1–6.

Материал поступил в редакцию 05.06.2021

Материал поступил в редакцию после рецензирования 10.09.2021