

БИОИНЖЕНЕРИЯ РЕКОМБИНАНТНЫХ БЕЛКОВ: ОТ СТРУКТУРЫ К ФУНКЦИИ И БИОЛОГИЧЕСКОЙ АКТИВНОСТИ

© 2025 г. Д.А. Долгих^{a,b,*}

^aИнститут биоорганической химии имени М.М. Шемякина и Ю.А. Овчинникова РАН, Москва, Россия

^bМосковский государственный университет имени М.В. Ломоносова, Москва, Россия

*E-mail: dolgikh@nmr.ru

Поступила в редакцию 13.12.2024 г.

После доработки 27.12.2024 г.

Принята к публикации 12.05.2025 г.

Статья посвящена развитию инженерии рекомбинантных белков, прежде всего искусственных белков, или белков *de novo*, — от создания первых белков с заданной пространственной структурой и биологической активностью до современных работ в этой области, в которых широко используются методы машинного обучения и искусственного интеллекта. Применение этих методов, в частности компьютерных платформ Rozetta и AlphaFold, привело к огромному прогрессу в данной области, о чём свидетельствует Нобелевская премия по химии прошлого года. В настоящее время эти методы должны быть рекомендованы для использования в любой современной лаборатории, проводящей работы по физико-химии белков и белковой инженерии.

Статья подготовлена на основе доклада на научной сессии Отделения биологических наук РАН 10 декабря 2024 г.

Ключевые слова: белки, белковая инженерия, структура и функция белков, инженерная биология, искусственный интеллект, AlphaFold.

DOI: 10.31857/S0869587325070063, **EDN:** FIIRKS

Становление и развитие белковой инженерии — относительно новой отрасли науки и технологии — происходило у нас на глазах. Число работ, в которых применяются подходы белковой инженерии, стремительно растёт. Если ввести термин “protein engineering” в поисковую строку базы данных NCBI, получим свыше 30 000 результатов. Очевидно, в коротком обзоре невозможно охватить всё разнообразие этих работ.

В нашей стране очень много лабораторий, в которых ведутся исследования в области белковой ин-

женерии, — они есть практически в каждом институте направления физико-химической биологии. Это и ИБХ РАН, и Институт молекулярной биологии им. В.А. Энгельгардта РАН, и Институт биологии гена РАН, Институт белка РАН и многие-многие другие. Я остановлюсь на одном направлении, которое мне особенно близко и, на мой взгляд, наиболее интересно — получении новых белков или “белков *de novo*”, о важности которого свидетельствует Нобелевская премия 2024 г. по химии — я имею в виду работы Д. Бейкера.

Белковая инженерия включает в себя два возможных подхода к получению новых белков — рациональный дизайн и направленную эволюцию. В первом случае конструирование белка опирается на знание его структурно-функциональных свойств, во втором — идёт случайный мутагенез с последующим отбором нужных вариантов. Сочетание этих подходов позволяет эффективно получать новые белки с заданными свойствами.

Современная белковая инженерия может очень многое: это и изменение каталитических свойств ферментов, и изменение структурных свойств бел-



ДОЛГИХ Дмитрий Александрович — доктор биологических наук, профессор, руководитель лаборатории инженерии белка ИБХ РАН, профессор кафедры биоинженерии биофака МГУ.

ков и их функциональных свойств, в том числе, улучшение эффективности белковых лекарственных препаратов. Наконец, это создание совершенно новых белков, или “белков *de novo*”, — направление, которое прямо связано, если говорить в современных терминах, с синтетической биологией биомолекулярных систем. Именно о нём здесь и пойдёт речь.

В этих исследованиях особенно ярко проявляется уровень наших знаний о принципах формирования структуры белковых молекул: если правильно понимать эти принципы, можно создавать белки, обладающие заданной структурой. Задача конструирования таких белков была поставлена на самых ранних этапах развития белковой инженерии. Вначале исследователи пробовали повторять небольшие простые структуры, имеющиеся в природных белках. То есть задачей было придумать такую аминокислотную последовательность белка (первичную структуру), которая приобретёт требуемую вторичную структуру и свернётся в заданную пространственную. Первая успешная работа в этом направлении была выполнена Л. Риган и В. ДеГрадо, которые получили белок *de novo* — тетраспиральный комплекс, состоящий из 4 спиралей, образующих гидрофобное ядро. Получить заданную вторичную структуру было достаточно легко, расположив в соответствующих позициях альфа-образующие аминокислотные остатки и задав расположение перетяжек, а вот обеспечить правильное пространственное расположение спиралей удалось не сразу. Пришлось разбить аминокислотную последовательность на две части, состоящие из двух спиралей (своего рода “шпилька”, одна сторона которой гидрофобная, а другая гидрофильная), закристаллизовать их, посмотреть, как они укладываются в пространстве, и на основе этого подобрать правильное расположе-

ние перетяжки между шпильками. Подобные конструкции на основе природных белков были позднее получены в нескольких лабораториях. Однако более амбициозной задачей было попытаться получить пространственную структуру с такой топологией, которая не наблюдается в природных белках, но не противоречит известным принципам формирования пространственных структур белков.

Первая такая структура, показанная на рисунке 1, была предложена, получена и исследована в нашей работе в “Докладах Академии наук” [1]. Этот белок мы называли альбебетин, так как он состоит из двух повторяющихся элементов альфа-бета-бета. Дизайн этого белка был разработан под руководством О.Б. Птицына, выдающегося учёного, к сожалению, рано ушедшего из жизни.

Дизайн альбебетина включал следующее: использование α - и β -образующих аминокислотных остатков — Leu для α -спиралей (в положениях i , $i+3$, $i+4$, $i+7$) и Val для β -участков (в положениях i , $i+2$, $i+4$); введение разноимённых зарядов на концы α -спиралей (Glu и Asp на N- и Arg на C-конец) для их стабилизации; введение гидрофильных заряженных остатков в крайние β -участки для правильного взаимного расположения β -участков; использование перетяжек минимальной длины между α - и β -участками для предотвращения образования других пространственных структур; использование в перетяжках Pro и Gly, разрушающих α - и β -структуру. Благодаря такому дизайну заданная структура альбебетина, включающая взаимное расположение регулярных элементов и петельных участков (топология), стала энергетически более выгодной, чем другие возможные топологии. Был синтезирован соответствующий ген, после чего альбебетин удалось наработать в микрограммовых количествах в бесклеточной системе экспрессии, а затем уже в препаративных количествах в *E.coli*.

Исследование альбебетина различными физико-химическими методами показало, что он компактен, стабилен к разворачиванию мочевиной, обладает заданной вторичной структурой и динамикой расщепления трипсином, соответствующей заданной пространственной структуре. В то же время его структура обладает значительной подвижностью, она лабильна, поэтому альбебетин находится в состоянии, которое в природных белках называется состоянием расплавленной глобулы.

Здесь я немного остановлюсь на свойствах этого состояния, тем более что оно было впервые в мире обнаружено и описано в наших работах по альфа-лактальбумину и карбоангидразе. Это состояние наблюдается в присутствии слабоденатурирующих агентов или даже в нативных или близких к нативным условиях, а также во многих белках как промежуточная фаза при их сворачивании. Иначе говоря, это состояние обычно для многих белков и, как мы показали в своей недавней работе в журнале

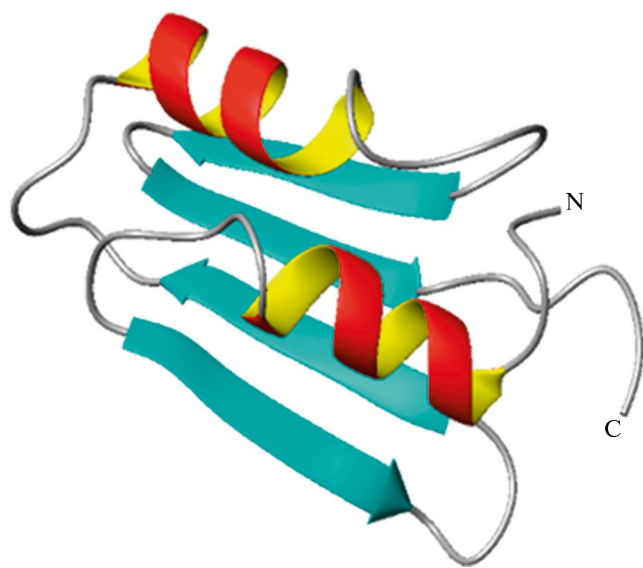


Рис. 1. Заданная пространственная структура искусственного белка альбебетина

“Molecules” [2], является, скорее, правилом, чем исключением. Сейчас для белков (или их фрагментов), находящихся в этом состоянии, широко используется термин “natively unfolded” или “intrinsically disordered”. Такие белки могут приобретать жёсткую структуру при взаимодействии с другими макромолекулами. Поэтому тот факт, что альбумин находится в состоянии расплавленной глобулы, не помешал нам попробовать получить на основе этого белка с заданной пространственной структурой искусственные белки с заданной биологической активностью.

Вначале мы получили альбумин — искусственный белок с некоторой модельной активностью, добавив на N-конец альбумина фрагмент интерферона альфа-2 (рис. 2а) и показали, что такой белок в полной мере обладает этой заданной модельной активностью — эффективно направляет blast-трансформацию тимоцитов мыши. Таким образом, альбумин стал первым в мире искусственным белком, на который была привита биологическая активность. Логическим развитием этих работ стало получение подобным образом ряда искусственных белков, обладающих заданной биологической активностью, представляющей медицинский интерес, — противораковой, инсулин-подобной и противовирусной (рис. 2, б–д). При этом в одном случае активность была опосредована не линейным фрагментом, а пространственно-организованным центром интерферона, который моделировался двумя разнесёнными по первичной структуре участками белковой цепи (рис. 2е).

Перейдём от этих первых работ по белкам *de novo* к современным исследованиям в рамках этого направления, за которые была присуждена Нобелев-

ская премия по химии 2024 г. Д. Бейкеру. Единственная работа, прямо упомянутая в пресс-релизе Нобелевского комитета, — получение искусственного белка Top7 с заданной структурой, не встречающейся в природных белках (рис. 3). Структура отличается от нашего альбумина наличием ещё одного бета-стрэнда. Бейкер и его сотрудники использовали при конструировании Top7 разработанную ими программу Rosetta, которая позволяет с высокой точностью рассчитывать энергетические параметры белков, анализировать различные варианты возможных мутаций и таким образом находить оптимальные варианты последовательностей. Они смогли закристаллизовать Top7 и получить точную структуру этого белка, которая, как было показано, соответствует заданной [3]. Но, конечно, Нобелевская премия была получена не только за Top7, а за создание компьютерной платформы Rosetta, позволившей учёным проводить многие подобные работы. Только за последние два года в ведущих журналах опубликованы работы по белкам *de novo* с различными пространственными структурами — это и альфа-структурные белки, и бета-белки, и альфа/бета белки, и бета-бочонки, которые не обнаружены в природных белках.

Чтобы наглядно показать, насколько возрастают возможности белковой инженерии с появлением компьютерных методов, я хочу привести пример антител, точнее, антителомиметиков. Инженерия антител — это наиболее коммерциализированная область белковой инженерии. Очень многие фирмы и лаборатории активно ведут работы в данной области: только в нашей небольшой лаборатории за последние годы получены десятки антител против самых различных антигенов. Но если мы говорим

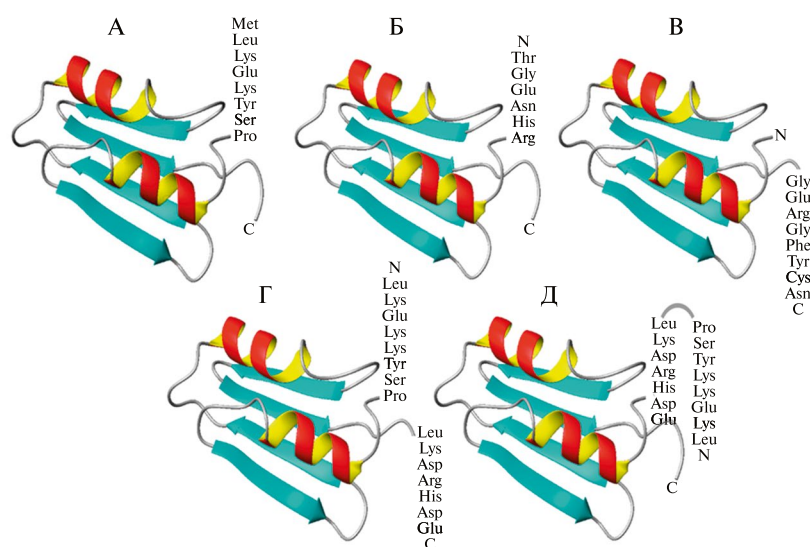


Рис. 2. Искусственные белки на основе альбумина с заданными биологическими активностями: А — blast-трансформирующая активность (альбумин), Б — противораковая активность, В — инсулин-подобная активность, Г и Д — противовирусная активность



Нобелевская премия по химии 2024



David Baker – "for computational protein design"

"Proteins generally consist of 20 different amino acids, which can be described as life's building blocks... David Baker succeeded in using these blocks to design a new protein that was unlike any other protein. Since then, his research group has produced one imaginative protein creation after another, including proteins that can be used as pharmaceuticals, vaccines, nanomaterials and tiny sensors". (Press release)

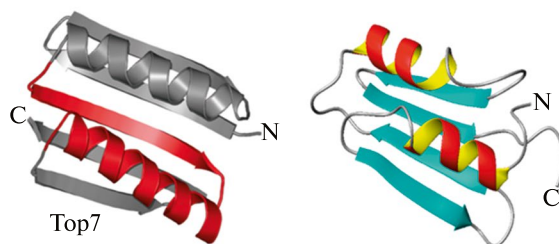


Рис. 3. Из пресс-релиза Нобелевского комитета о присуждении Нобелевской премии по химии 2024 года Дэвиду Бейкеру. Внизу – структура искусственного белка Top7 (слева) и альбумина (справа)

о возможностях направленной доставки эффекторных молекул, то вместо антител лучше использовать малые белки/полипептиды – скаффолды, обладающие свойством связываться с антигенами. В ИБХ РАН подобные работы ведут в лабораториях С.М. Деева, А.Г. Габибова, нашей лаборатории и других. Скаффолды, как правило, основаны на природных белках, структура которых может быть похожа на иммуноглобулиновую. Однако использование других, неприродных, скаффолдов, открывает совершенно новые возможности для белковых инженеров. Так, Д. Бейкер и его коллеги с помощью разработанной ими платформы Rosetta рассмотрели сотни подобных мини-белковых конформаций, рассчитали более 20 000 последовательностей и показали, что лучшие из них – а лучших десятки – связываются с двумя антигенами с наномолярными константами. Были синтезированы и исследованы 14 самых лучших вариантов и показано, что они стабильны, не теряют активности после нагревания до 70°C или выше, в экспериментах на мышах обладают протективной активностью и неиммуногенны [4]. Можно упомянуть также работу этого года В. ДеГрадо, где с помощью Rosetta проанализированы 75 000 вариантов пептидов, которые, по замыслу, должны связывать амилоиды альфа-синуклеина; в результате найден вариант, который делает это очень эффективно, различая даже конформеры амилоидов [5].

Нельзя не упомянуть ещё одно событие, имеющее прямое отношение к белковой инженерии. Это успешное применение искусственного интеллекта

для определения пространственной структуры белков по первичной структуре – я имею в виду платформу AlphaFold [6]. Вторая половина Нобелевской премии 2024 г. была присуждена сотрудникам компании Google DeepMind Д. Хассабису и Д. Джамперу за использование искусственного интеллекта в данной области. Как сказано в пресс-релизе Нобелевского комитета, они “взломали код белков”. Решена задача, поставленная 50 лет назад: как из первичной структуры белка получить его пространственную структуру. Это действительно достижение нобелевского масштаба: исследователи, занимающиеся белками, получили мощный инструмент, который позволяет вывести работы по физико-химии белков на качественно новый уровень.

Однако означает ли это, что определение структурно-функциональных отношений в белках и белковая инженерия перешли в статус чисто технологических задач? И да, и нет. Очевидно, что эффективное применение искусственного интеллекта стало возможным только после того, как в базах данных было накоплено огромное количество экспериментальных данных о белковых структурах, на основе которых работают методы машинного обучения – до этого такие подходы были если не невозможны, то существенно менее эффективны. В недавней работе А.В. Финкельштейна и Д.Н. Иванкова в “Journal of Molecular Biology” показано, что при меньшем количестве известных белковых структур точность предсказания падает, и это неудивительно, так как фактически искусственный интеллект предсказывает пространственную структуру белков

на основе гомологии с уже имеющимися структурами [7]. Если задать вопрос о том, насколько эти подходы учитывают физические законы формирования белковых структур, ответ очевиден: искусственному интеллекту не нужно знать физические законы — он решает задачу распознавания образов по известным шаблонам. И если в какой-то необычной группе белков число её элементов пока относительно невелико, задача отнесения к этой группе новых белков вряд ли может быть решена такими методами. В работе Д. Чакраварти и Л.Л. Портера [8] показано, что если взять три группы белков, а именно, белки, обладающие единственной нативной структурой, белки, могущие менять структуру в зависимости от условий, и белки неструктурированные, то есть белки/фрагменты белков в состоянии расплавленной глобулы, то, как и следовало ожидать, хуже всего предсказываются структуры последних. Но это несколько не умаляет того вклада, который обеспечило применение искусственного интеллекта, вклада, который важен для всех исследователей, занимающихся белками.

Хотелось бы вспомнить один забавный эпизод, связанный с программой AlphaFold. Программа сейчас очень популярна, и года полтора назад к нам обратился молодой исследователь, использующий AlphaFold2, который прочитал нашу первую статью по альбегетину, взял его аминокислотную последовательность, ввёл в программу и с удивлением увидел, что AlphaFold выдал пространственную структуру искусственного белка, соответствующую заданной. Он не знал, что мы уже получили и исследовали альбегетин и экспериментально подтвердили заданную структуру, поэтому предложил экспрессировать и исследовать наш белок. Мы рассказали ему, что это уже сделано и предложили какие-то другие варианты сотрудничества. К сожалению, его интерес заметно угас, но нам всё равно было приятно, что наши первые работы находят подтверждение в современных методах — значит, мы выбрали правильное направление.

В заключение хочу ещё раз подчеркнуть, что белковая инженерия, зародившаяся сорок лет тому назад, прошла большой путь от первых работ по мутагенезу природных и созданию простых искусственных белков до использования современных

компьютерных программ, позволяющих решать комплексные задачи по созданию сложных белков с заданными полезными свойствами для нужд человека. Мощные компьютерные платформы, в которых используются методы машинного обучения и искусственный интеллект, в последнее время находят самое широкое применение, поднимают белковую инженерию на совершенно новый уровень, и Нобелевская премия по химии 2024 г. свидетельствует о признании выдающегося прогресса в этой научной области.

ЛИТЕРАТУРА

1. Долгих Д.А., Федоров А.Н., Чемерис В.В. и др. Получение и исследование альбегетина, искусственного белка с заданной пространственной структурой // Доклады АН. 1991. Т. 320. С. 1266—1269. Dolgikh D.A., Fedorov A.N., Chemeris V.V. et al. Production and study of albebetin, an artificial protein with a given spatial structure // Reports of the Academy of Sciences. 1991, vol. 320, pp. 1266—1269.
2. Bychkova V.E., Dolgikh D.A., Balobanov V.A. et al. The Molten Globule State of a Globular Protein in a Cell Is More or Less Frequent Case Rather than an Exception // Molecules. 2022, vol. 27, no. 4361.
3. Kuhlman B., Dantas G., Ireton G.C. et al. Design of a novel globular protein fold with atomic-level accuracy // Science, 2003, v. 302, pp. 1364—1368.
4. Chevalier A., Silva D.-A., Rocklin G.J. et al. Massively parallel de novo protein design for targeted therapeutics // Nature, 2017, vol. 550, pp. 74—79.
5. Wallace H.M., Yang H., Tan S. et al. De novo design of peptides that bind specific conformers of α -synuclein // Chemical Science, 2024, vol. 15, pp. 8414—8421.
6. Jumper J., Evans R., Pritzel A. et al. Highly accurate protein structure prediction with AlphaFold // Nature, 2021, vol. 596, pp. 583—589.
7. Finkelstein A.V., Ivankov D.N. Protein 3D Structure Identification by AlphaFold: a Physics-Based Prediction or Recognition Using Huge Databases? // Journal of Molecular Biology, 2024, vol. 6, pp. 1—10.
8. Chakravarty D., Porter L.L. AlphaFold2 fails to predict protein fold switching // Protein Science, 2022, vol. 31, e4353.

ENGINEERING RECOMBINANT PROTEINS: FROM STRUCTURE TO FUNCTION AND BIOLOGICAL ACTIVITY

D.A. Dolgikh^{a,b,*}

^a*Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, Moscow, Russia*

^b*Lomonosov Moscow State University, Moscow, Russia*

**E-mail: dolgikh@nmr.ru*

The article discusses the development of recombinant protein engineering, primarily artificial proteins or *de novo* proteins, from the creation of the first proteins with a given spatial structure and biological activity to modern work in this area, which widely uses machine learning and artificial intelligence methods. The use of these methods, in particular the Rozetta and AlphaFold computer platforms, has led to tremendous progress in this area, as evidenced by last year's Nobel Prize in Chemistry. Currently, these methods should be recommended for use in any modern laboratory conducting work on the physical chemistry of proteins and protein engineering.

The article is based on the author's report at a scientific session of the Division of Biological Sciences of the Russian Academy of Sciences on December 10, 2024.

Keywords: proteins, protein engineering, protein structure and function, engineering biology, artificial intelligence, AlphaFold.