

# Строение и функции белков

## 1. Уровни организации структуры белков

Существует несколько уровней организации структуры белков: первичная, вторичная, третичная и четвертичная. Установление первичной структуры в простом случае происходит с помощью секвенирования, но большинство исследований делается с применением масс-спектрометра. Этот метод в какой-то мере предсказательный, так как включает в себя этап сравнения теоретических и экспериментальных данных. Ограничение этого метода в том, что с его помощью нельзя установить посттрансляционные модификации белка.

## 2. Генетический код

Первичная структура белка определяется последовательностью нуклеотидов в ДНК и мРНК. Впервые расшифровка генетического кода была получена в эксперименте 1962 года (Marshall, Matthaei) с помощью трансляции *in vitro*. Бактериальному экстракту добавлялась матричная РНК трех разных типов: содержащая только уридиновые нуклеотиды, только адениновые нуклеотиды или только цитозиновые нуклеотиды. В итоге в пробирке транслировались пептиды либо только из фенилаланина, либо из лизина, либо из пролина. По результатам эксперимента открыли такое важное свойство генетического кода как триплетность.

Итак, генетический код обладает следующими свойствами:

- триплетность — одна аминокислота кодируется тремя нуклеотидами;
- вырожденность — одна аминокислота кодируется несколькими триплетами;
- универсальность для всех живых организмов.

## 3. Вторичная структура белка

Основные элементы вторичной структуры белка — это **альфа-спирали** и **бета-слои**. Различают право- и лево- закрученную альфа-спираль (см. рис. ??). Правозакрученная (альфаR) встречается в белке чаще всего. Альфа-спираль стабилизирована водородными связями между NH-группой первого аминокислотного остатка и CO-группой четвертого аминокислотного остатка. Альфа-спираль также имеет такие характеристики как шаг и количество аминокислот на виток. В левой спирали невозможны водородные связи, шаг ее больше, молекула не устойчива. Ее стабилизация происходит при взаимодействии с такими же левыми спиралями, как в случае с коллагеном (тримеризация).

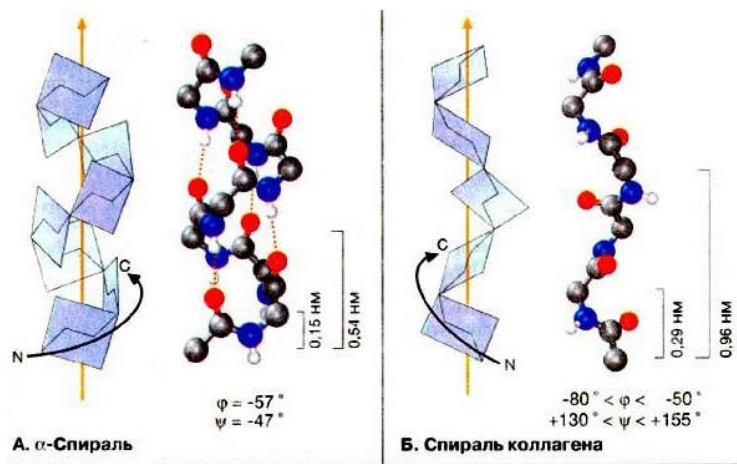


Рис. 1: Альфа-спирали

**Бета-слои (Складчатые структуры, бета-листы)** Бета-слои формируются за счет водородных связей между двумя пептидными цепями, при чем цепи могут быть направлены как параллельно, так и антипараллельно (см. рис. ??).

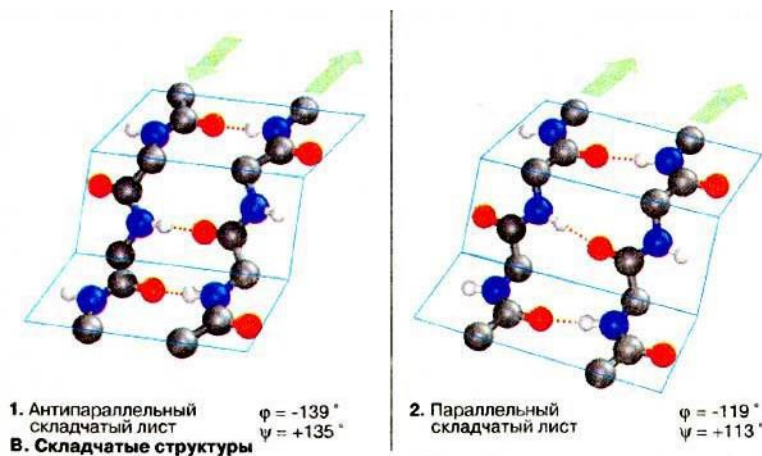


Рис. 2: Бета-слои

В тех участках, где пептидная цепь изгибается достаточно круто, часто находятся **бета-петли**, образованные 4 аминокислотами. Петли стабилизированы водородными связями между 1 и 4 аминокислотными остатками. Это очень распространенный элемент белков.

Описанные выше структуры составляют примерно 95% архитектуры любого белка и они являются основными структурными единицами для построения более сложных третичных комплексов.

Существуют биоинформатические методы определения вторичной структуры белка по ее первичной структуре. Точность этих методов порядка 80%.

## 4. Третичная структура белка

Третичная структура формируется элементами вторичной структуры. Джеймс Бетчеллер Самнер впервые выделил фермент (уреазу) и доказал белковую природу ферментов, за что получил Нобелевскую премию в 1946 году. Позже Дж.Кендрью определил третичную структуру миоглобина кашалота на основе рентгеноструктурного анализа.

Итак, третичная структура белка образована за счет:

- ковалентных связей (пептидных и дисульфидных);
- водородных связей;
- электростатических взаимодействий;
- Ван-дер-Ваальсовых сил;
- гидрофобных взаимодействий боковых радикалов аминокислот.

**Супервторичная структура** — это функциональная группа белка, составленная из элементов вторичной структуры, которая определяет свойства третичной структуры. Основные супервторичные структуры: спираль-петля-спираль, цинковые пальцы (распространенные у транскрипционных факторов), coiled coil motif.

Также третичная структура белка может иметь доменную организацию. **Домен** — это элемент третичной структуры белка, представляющий собой достаточно стабильную и независимую подструктуру белка, чей фолдинг происходит независимо от остальных частей. В состав домена обычно входит несколько элементов вторичной структуры. На сегодняшний день в базах данных зарегистрировано 173536 различных доменов, и большое разнообразие белков (не всех) образовано комбинациями этих доменов. Существует гипотеза о том, что на ранних этапах формирования жизни на Земле у эукариот каждый домен происходил как минимум из одного экзона, и в результате перетасовок и мутаций на сегодняшний день имеется такое большое разнообразие белков.

Зная доменную организацию белка, можно предположить какие функции он выполняет.

Различные методы быстрого, дешевого и точного предсказания третичной структуры белков сейчас активно развиваются, так как у такого огромного количества белков, которое известно науке, экспериментально ее определить нереально. Эти методы структурной протеомики основаны на следующих постулатах:

- сворачивание белка (фолдинг) — процесс укладки полипептидной цепи в компактную пространственную структуру;
- аминокислотная последовательность белка однозначно определяет его пространственную структуру;
- пространственная структура белка определяет его функцию.

Подходы к предсказанию третичной структуры белка можно разделить на три группы:

- Ab initio — моделирование укладки «из первых принципов» без использования дополнительной информации о структурах сложных белков.
- Предсказание на основе гомологии (homology modeling) — моделирование на основе известных структур схожих белков.
- Тридинг (threading) — моделирование на основе слабой гомологии.

**CASP (Critical Assessment of protein Structure Prediction)** — конкурс методов предсказания структуры белков. Лаборатории, которые участвуют в этом конкурсе, имея на руках только первичную последовательность аминокислот, разрабатывают более точные методы предсказания третичной структуры белка. К концу проведения конкурса организаторы высылают лабораториям образцы белков, первичную структуру которых они получили ранее. Этот белок биоинформатики могут кристаллизовать и проверить достоверность своего метода.

## 5. Четвертичная структура белка

Четвертичная структура — взаимное расположение нескольких полипептидных цепей в составе единого белкового комплекса. Впервые четвертичную структуру гемоглобина определили Кендрью и Перуц, за что получили Нобелевскую премию в 1962 году. Перуц придумал вводить в белок атомы тяжелых металлов для стабилизации комплексов, которые в итоге давали четкие сигналы рентгеноструктурного анализа.

## 6. Структуры белков

Основоположник структурной протеомики — Кристиан Бемер Анфинсен — ввел понятие **фолдинга белка** (укладки аминокислотной последовательности, которая уникальным образом определяет трехмерное строение белковой молекулы). Анфинсен вместе с Муром и Стайном в 1972 году получили Нобелевскую премию по химии за работу по «установлению связи между аминокислотной последовательностью рибонуклеазы А и ее биологически активной конформацией». На основе этой работы возникла такая область как структурная протеомика, основной мировой задачей которой является получение как можно большего числа трехмерных структур белков.

**Гомологичные последовательности белков** Зарегистрированных организмов и их секвенированных генов очень много и новые описания происходят чуть ли не каждый день. Понятно, что такой колоссальный массив информации невозможно обработать. Поэтому в научном мире приняты некие допущения, основанные на гомологичности последовательностей.

Основной инструмент для определения гомологичности — выравнивание — заключается в том, что аминокислотные последовательности белков из разных организмов выравниваются относительно друг друга (см. рис. ??) и по этому выравниванию можно понять какие аминокислоты схожи и различны у разных организмов.

По степени схожести аминокислотной последовательности можно предполагать третичную структуру у неизвестных белков.

**Гомологичные последовательности** — последовательности, имеющие общее происхождение (общего предка). Признаки гомологичности белков: сходная трехмерная структура и аминокислотная последовательность, аналогичная функция (для ортологов) и пр.

Различают несколько типов гомологичных белков:

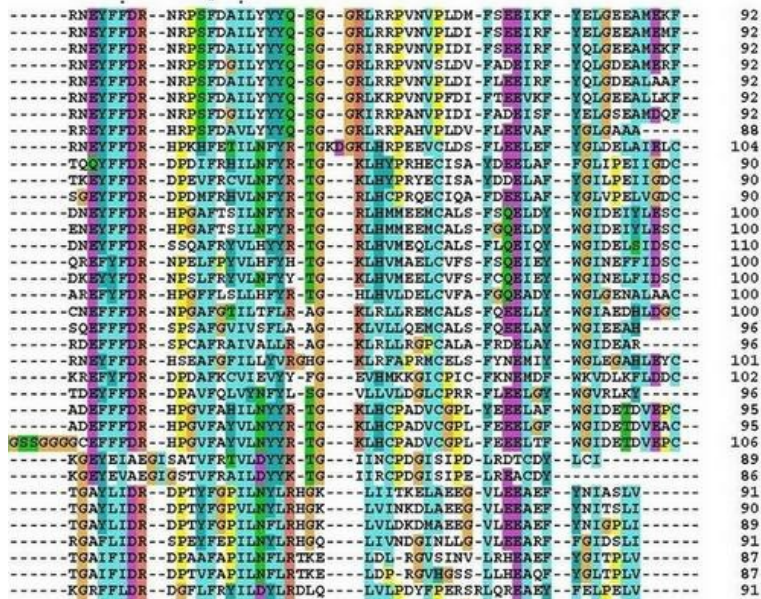


Рис. 3: Множественное выравнивание

- **Ортологи** — последовательности, возникшие из одного общего предшественника в процессе видообразования у разных организмов. Ортологи, как правило, имеют одну и ту же функцию.
- **Паралоги** — последовательности, возникшие из одного общего предшественника в результате дупликации одного гена в одном организме. Паралоги, как правило, имеют разные функции.

## 7. Изчение структуры белков

Структурная протеомика занимается определением структуры белка. Было рассчитано, что для изучения большей части многообразия белков человека необходимо установить структуру только 16000 его белков. А дальнейшее разнообразие определить на основе схожести и гомологичности. К 2013 году в базе данных NCBI есть информация о структуре 22000 белков. Но это не уникальные белки из всего разнообразия, а имеющие прикладное значение. У таких модельных белков может быть по несколько структур, соответствующих разным мутациям. Также есть данные о структуре взаимодействий белков с ДНК, РНК и простейшими субстратами.

Помогают в установлении структур белков различные проекты. Например, проект **PSI (Protein Structure Initiative)**. Для него были созданы специальные машины, в которых автоматизированы процессы клонирования, выделения белка, его кристаллизация. Выбрано для участия в проекте 154143 белка, из них наработано в клетке с помощью клонирования 106932 белка, выделить и закристаллизовать удалось 26868, а определить структуру в итоге только у 5321 белков.

Считается, что у идентичных аминокислотных последовательностей одинаковые структуры и функции. Но по мере накопления данных стало понятно, что есть исключения из этого правила. Оказалось, что около 15% белков эукариот не имеют четкой организации, более того, некоторые белки могут изменяться в процессе своего существования в клетке. К примеру, белок **лимфотактин** из семейства хемокинов, стимулирующий хемотаксис у Т-лимфоцитов, имеет две различные структуры. Одна из них содержит 3 бета-листа и одну концевую альфа-спираль и может переходить во вторую, в составе которой 4 бета-листа. Вторая структура образует димеры, а первая мономерна.

Фермент **хоризмутаза**, катализирующий превращение хоризмовой кислоты в префеновую кислоту, не образует третичной структуры. В несвязанном с лигандом состоянии определить структуру фермента не удается из-за высокой подвижности.

Очень интересное семейство белков — **прионы** — обладает аномальной трехмерной структурой. Прионы способны катализировать структурное превращение гомологичного им нормального клеточного белка в себе



подобный (прионный), присоединяясь к белку-мишени и изменяя его конформацию. В результате развиваются прионные болезни.

## 8. Функции белков

Оказалось удобным связывать функции белка с его доменной организацией. То есть, зная белковые домены, можно примерно предсказать в каких клеточных процессах белок участвует.

Белки можно классифицировать по их функциям. Различают структурные, регуляторные, транспортные, защитные, каталитические, сигнальные, рецепторные, резервные белки.

Для примерно половины известных белок-кодирующих генов известны функции, и каждый год открывают функцию примерно 100 новых генов. Около 23-24% белков с совершенно неизвестной функцией. Транскрипционные факторы — это белки, регулирующие транскрипцию, составляют примерно 12% (2000 белков) от общего числа белков человека. 8,5% белков имеют домен, ответственный за связывание с нуклеиновыми кислотами, но они еще не охарактеризованы как транскрипционные факторы или белки, выполняющие какую-нибудь другую функцию, связанную с нуклеиновыми кислотами. То есть для части генов функция точно не определена, а для части вообще пока не известна.

Наиболее престижный журнал для молекулярного биолога — журнал «Cell». В этом журнале в 2012 году была опубликована статья (Castello et al.), суть которой описана ниже. На культуре клеток фибробластов человека специальными методами были зафиксированы белок-, РНК- и ДНК-белковые взаимодействия и с помощью олиго-dt (олигонуклеотиды, состоящие из тиминов, прим. автора) были выделены за полиА-хвосты мРНК-белковые комплексы. Далее эти мРНК секвенировали, а белки проходили процедуру масс-спектрометрии. В результате было выявлено 860 РНК-связывающих белков. Из них 144 РНК-связывающих белка было известно по литературным данным, 222 белка определялись как таковые за счет гомологии с первыми, 179 белков имеют РНК-связывающий домен и 315 белков были впервые идентифицированы как РНК-связывающие. Считается, что эти белки осуществляют компарментализацию РНК, регуляцию трансляции или организуют комплексы из других белков и РНК для каких-то определенных функций.

## 9. Структурные белки

Это самые распространенные белки, выполняющие опорную функцию. Они формируют цитоскелет, мембраны и многие другие структуры. Очень важны фибриллярные белки: коллаген в соединительной ткани, кератин в волосах, ногтях и коже, эластин в сосудистой стенке и др.

**Коллагеновые фибриллы** состоят из 3 левозакрученных альфа-спиралей. У человека имеется 28 типов коллагена.

**Эластин** определяет эластичность тканей. Он распространен в коже, легких и кровеносных сосудах. Молекула характеризуется наличием большого количества неполярных радикалов, в результате чего эластин нерастворим в воде. В растворе образует полимерные сети.

Другое семейство фибриллярных белков — **кератины** — наиболее прочные среди структурных белков.