

Введение в дизайн белков

Дизайн белков, НТУ Сириус

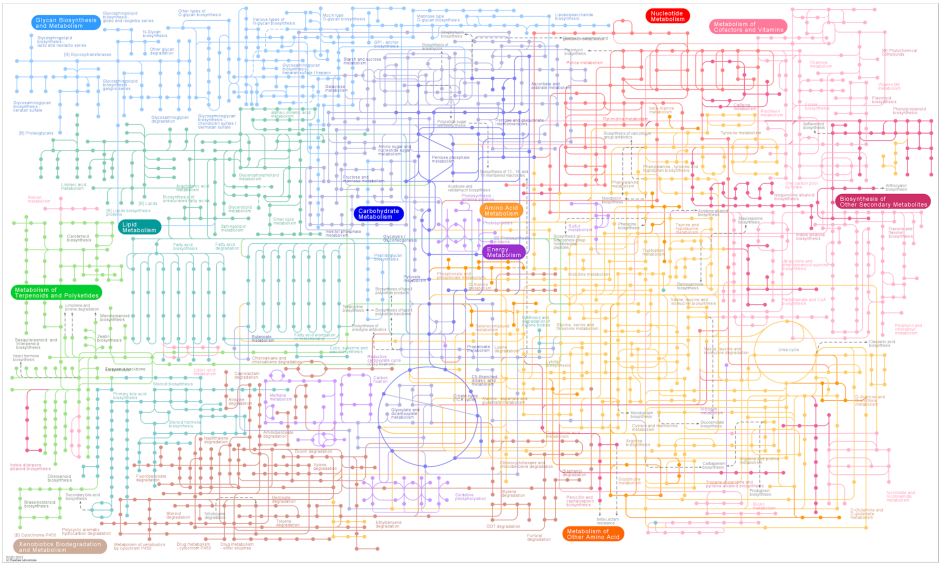
Головин А.В. ^{1 2}

¹НТУ Сириус, ЦИИиИТ

²МГУ им М.В. Ломоносова, Факультет Биоинженерии и Биоинформатики

Сириус, 2022

“Жизнь можно определить как активное, идущее с затратой полученной извне энергии, поддержание и самовоспроизведение молекулярной структуры” (Wikipedia, Biophysics way)



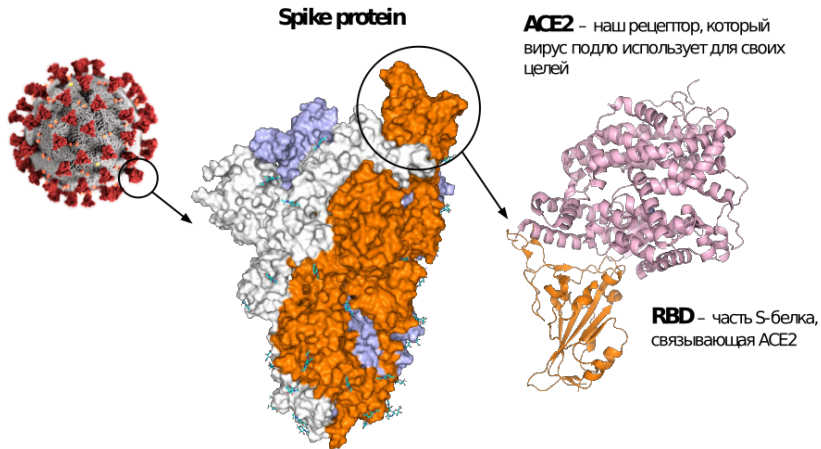
Metabolic Pathways

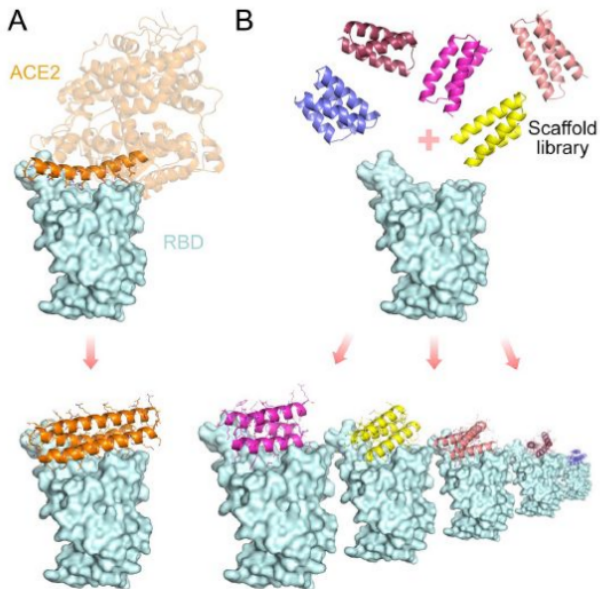
This diagram illustrates the complex network of metabolic pathways, showing the breakdown and synthesis of various biomolecules. Key pathways include:

- Glycolysis:** The breakdown of glucose into pyruvate, leading to the production of ATP and NADH.
- Citric Acid Cycle:** The central energy-producing cycle that releases CO₂ and generates high-energy electron carriers.
- Lipid Metabolism:** Pathways for the synthesis and breakdown of fatty acids and triglycerides.
- Carbohydrate Metabolism:** Pathways for the synthesis and breakdown of various sugars and polysaccharides.
- Amino Acid Metabolism:** Pathways for the breakdown of amino acids into intermediates that enter the Citric Acid Cycle or other metabolic pathways.
- Energy Metabolism:** Pathways for the synthesis of ATP and other energy carriers.
- Nucleotide Metabolism:** Pathways for the synthesis and breakdown of nucleotides, essential for DNA and RNA.
- Metabolism of Other Amino Acids:** Pathways for the synthesis and breakdown of various amino acids, including those involved in neurotransmitter synthesis.

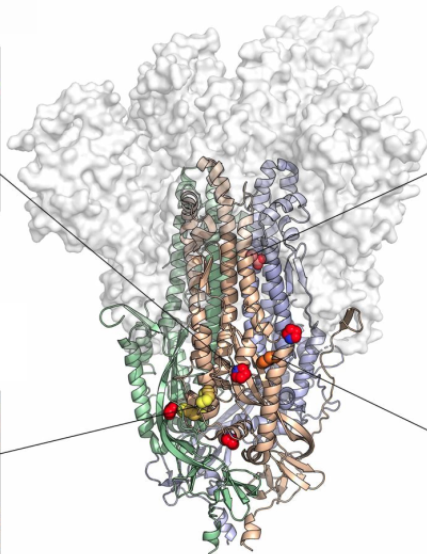
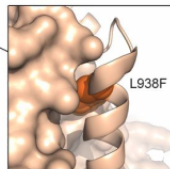
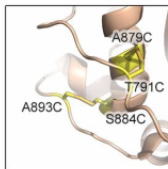
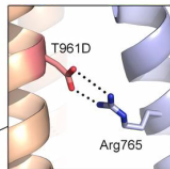
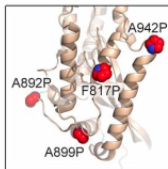
The diagram also shows the synthesis of various biomolecules, including lipids, carbohydrates, and nucleotides, and the breakdown of these molecules into their constituent parts. The pathways are interconnected, allowing for the regulation of metabolism in response to the cell's needs.

Sars-Cov-2





DOI: 10.1126/science.abd0826



Дизайн de novo

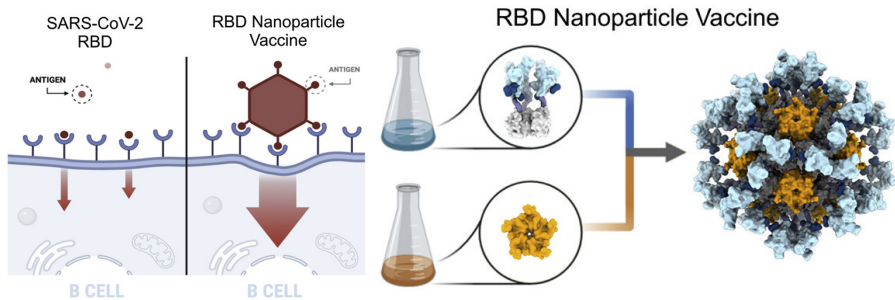
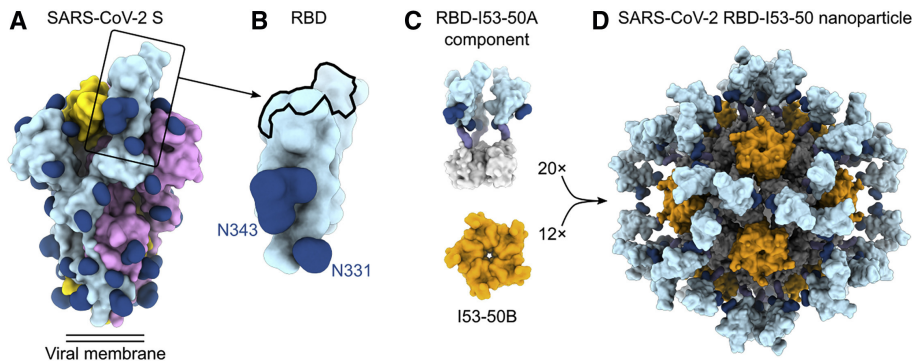


Схема дизайна



Дизайн белков

- Дизайн белков - это рациональный дизайн новых белковых молекул для разработки новых вариантов активности, а также для углубления базового понимания функции белка.
- Подходы к рациональному дизайну белков позволяют делать предсказания последовательности белков, которые складываются в конкретные структуры.

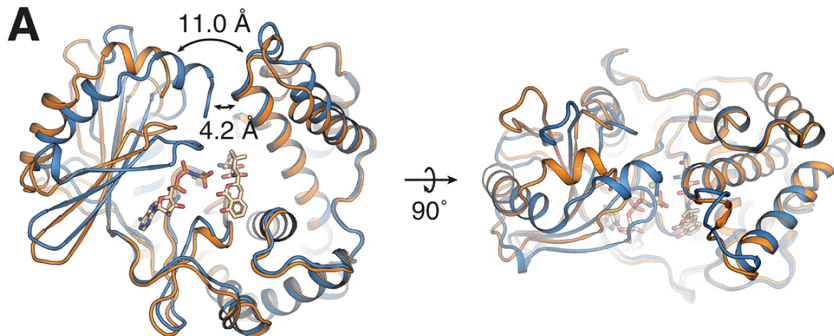
Структура-Функция

- Функция белка в значительной степени зависит от структуры белка, и рациональный дизайн белка использует эту взаимосвязь для разработки белков.
- В рациональном дизайне белка целевая структура или ансамбль структур должны быть известны заранее.
- В белковой инженерии, где используются различные методы для поиска белков из библиотек, которые выполняют определенную функцию.

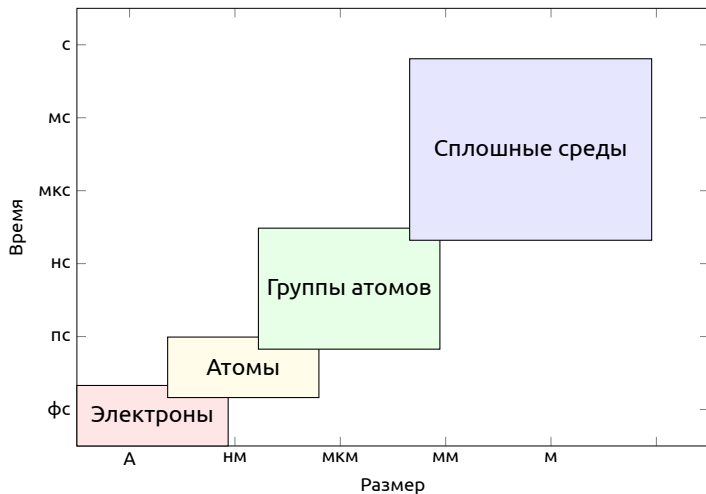
Пространство последовательностей

- Белки могут быть изменены на основе последовательности и структуры известного белка
- Последовательность может быть создана полностью с нуля при разработке белков de novo.
- И дизайн de novo, и редизайн белка могут устанавливать правила в пространстве последовательностей: конкретные аминокислоты, которые разрешены в каждом положении изменяемого остатка.
- Пример: Состав поверхности зонда RSC3 для отбора нейтрализующих ВИЧ антител был ограничен на основании данных об эволюции и балансировки заряда.
- Появление производительных вычислительных методов позволяет конструировать белки без вмешательства человека в выбор последовательности.

Подвижность белка



Масштабы в моделировании



Дизайн ферментов

- Разработка структуры белка может отличаться от разработки фермента, надо учитывать многие состояния, участвующие в каталитическом механизме.
- Однако дизайн белка является предпосылкой дизайна ферментов de novo.
- Примеры: de novo разработали ферменты для ретроальдольной реакции, реакции элиминирования Кемпа и для реакции Дильса-Альдера

Дизайн связывания

- Белковые взаимодействия участвуют в большинстве биотических процессов.
- Многие формы рака и вирусные инфекции, связаны с межбелковыми взаимодействиями.
- Связывание одного из партнеров взаимодействия нарушает вызывающее заболевание взаимодействие.
- Взаимодействия белок-белок могут быть разработаны с использованием алгоритмов дизайна белков
- Интерфейсы между белками более полярны, чем ядра белков, и связывание включает компромисс между десольватацией и образованием водородных связей. [40]
- Брюс Тидор и его коллеги разработали метод повышения аффинности антител, сосредоточив внимание на электростатических факторах.

Дизайн связывания - особенности

Функции энергии при дизайне белков должны быть адаптированы для связывания, поскольку связывание включает компромисс между конформациями с наименьшей энергией свободных белков (E_P и E_L) и конформацией с наименьшей энергией связанного комплекса.

$$\Delta_G = E_{PL} - E_P - E_L \Delta_G = E_{PL} - E_P - E_L.$$

Дизайн специфичности

- Дизайн белок-белковых взаимодействий должен быть высокоспецифичным, потому что белки могут взаимодействовать с большим количеством белков
- Одним из наиболее ярких примеров дизайна для специфичности является дизайн специфических bZIP-связывающих пептидов
- Недавняя вычислительная модернизация позволила экспериментально переключить кофакторную специфичность ксилоредуктазы *Candida boidinii* с NADPH на NADH

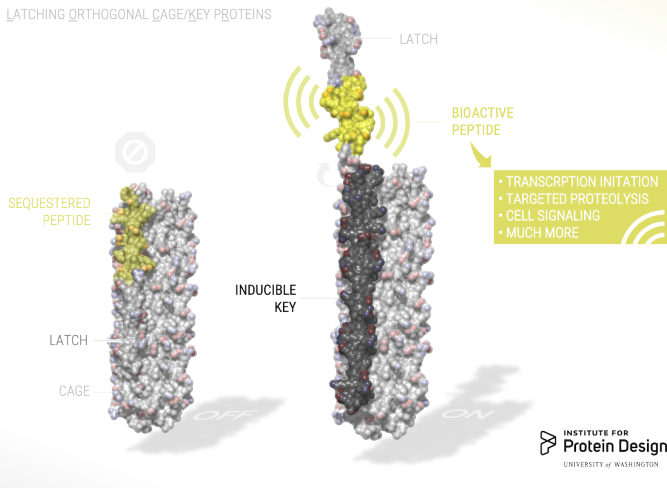
Изменение поверхности белка

- Идея заключается в создании поверхности протеина с сохранением целостности укладки
- "Белковая шлифовка" особенно полезна для изменения связывания белка с другими белками. Одним из наиболее важных применений этого подхода было создание зонда RSC3 для отбора широко нейтрализующих антител к ВИЧ.

Сенсоры и преключатели

INTRODUCING LOCKR

LATCHING ORTHOGONAL CAGE/KEY PROTEINS



Основные проблемы:

- Монте-Карло: 100 а.к. $3N$ степеней свободы, получаем 10^{48} конформаций.
- **Парадокс Левинталя:** "Промежуток времени, за который полипептид приходит к своему скрученному состоянию, на много порядков меньше, чем если бы полипептид просто перебирал все возможные конфигурации".
- Для решения разумно использовать накопленные знания для моделирования.

Последовательность-структура

Причины парадокса Левинталя:

- Теоретические модели, не соответствуют тому, что природа старается оптимизировать;
- В ходе эволюции были отобраны только те белки, которые легко сворачиваются;
- белки могут сворачиваться разными путями, не обязательно следуя глобально оптимальному пути.
- Считается, что структура определяется последовательностью, но иногда нужны другие факторы.
- Структура более консервативна чем последовательность

Сравнительное моделирование

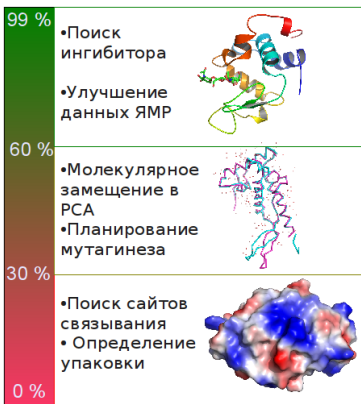
- Зачем искать конформации если можно представить, что при подобии последовательностей подобны и структуры.
- Надо оценить насколько вероятно, что отличие в последовательности может привести изменению способа укладки цепи.
- Надо отфильтровать ошибки полученные при определении структуры.

Известные структуры и последовательности

- Сейчас известно порядка 10^5 структур. Примерно 10% это уникальные белки.
- Только 30% из первого пункта имеют разрешение лучше 3.0 ангстрем.
- Примерно 25% известных последовательностей можно использовать для сравнительного моделирования.
- Для 50% последовательностей можно предсказать способ укладки.

Степень идентичности и сравнительное моделирование

Sali, A. & Kuriyan, J.
Trends Biochem. Sci. 22,
M20–M24 (1999)



Как это реализовать?

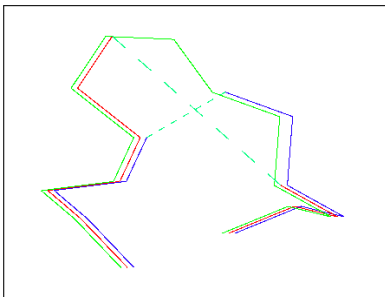
- Надо найти белок заготовку с известной структурой.
- Построить первичное выравнивание.
- Улучшить выравнивание.
- Построить ход основной цепи.
- Моделирование петель
- Достроить/моделировать положение боковых радикалов
- Проверка модели

Поиск белка заготовки

- Поиск по PDB с помощью:
 - Blast
 - Psi-Blast
 - Методов распознавания упаковки
- Используя биологическую информацию.
- Функциональное аннотирование в базах данных.
- Используя информацию об активных сайтах, или мотивы.

Улучшение выравнивания

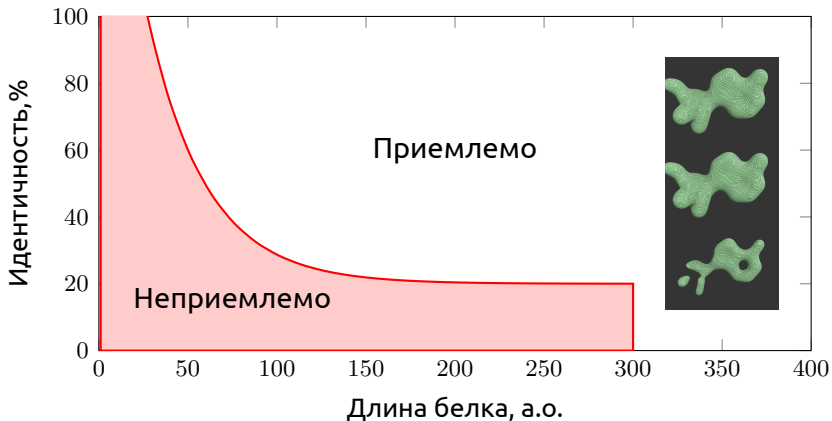
1	2	3	4	5	6	7	8	9	10	11	12	13	14
PHE	ASP	ILE	CYS	ARG	LEU	PRO	GLY	SER	ALA	GLU	ALA	VAL	CYS
PHE	ASN	VAL	CYS	ARG	THR	PRO	---	---	---	GLU	ALA	ILE	CYS
PHE	ASN	VAL	CYS	ARG	---	---	---	THR	PRO	GLU	ALA	ILE	CYS



Из книги "Professional Gambling" от Gert Vriend

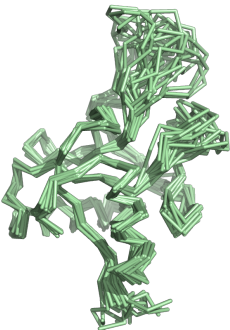
Качество белка заготовки

- Выбор качественного белка заготовки очень важен.
- Лучший вариант не обязательно обладает лучшей степенью идентичности.
 - Белок 1: ID 93%, 3.5 ангстрема разрешение. Хуже.
 - Белок 2: ID 90%, 1.5 ангстрема разрешение. Лучше!

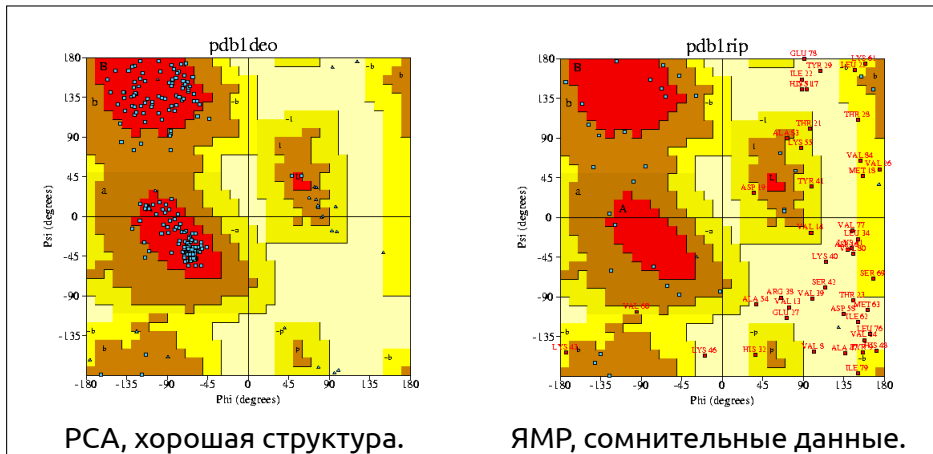


Если структура белка заготовки получена ЯМР

- Определимся какие области определены лучше.
- Соотнесём с выравниванием.
- Если низкая гомология выпадает на “подвижные” области, то структура подходит.



Качество заготовки, Рамачандран



Построение остова

- Генерируем координаты остова моделируемого белка для остатков из выравненных областей.
- Не обязательно использовать координаты, могут подойти дистанционные ограничения.
- Большинство исследователей предпочитают Modeller. Modeller использует дистанционные ограничения.

Моделирование петель

- Эмпирическое моделирование:
 - Поиск подходящего фрагмента по PDB
 - Использовать базы данных (LIP, etc..)
- Молекулярная механика.
- Монте-Карло.
- Rosseta:
 - Поиск фрагментов близких по последовательности.
 - Комбинирование результатов поиска с помощью Монте-Карло.

Комбинации выше перечисленных.

Моделирование боковых радикалов

- Если идентичность последовательностей высока то можно ожидать высокую консервативность третичных контактов.
- Если анализ показывает, что важные контакты консервативны то:

Лучше оставить конформацию боковых радикалов из заготовки чем моделировать.

Моделирование боковых радикалов

- Конформация боковых радикалов зависит от конформации основной цепи.
- Существуют базы данных ротамеров.
- Некоторые исследователи считают, что SCWRL метод самый удачный.

Это эмпирический метод на основе теории графов.
<http://dunbrack.fccc.edu/SCWRL3.php>

Точность моделирования боковых радикалов

- Высокая точность моделирования достигается для боковых радикалов внутри глобулы.
 - Причина: в экспериментах остатки на поверхности более подвижны.
 - Вычислительное проще упаковать гидрофобные остатки, чем учесть полярные контакты и водородные связи с водой или с участием воды.

Улучшение модели

- Методы минимизации энергии.
- Моделирование молекулярной динамики (оптимизация гидрофобики)
- Моделирование Монте-Карло.
- Любой известный подход для оптимизации структуры.

Ошибки

- Обычно ошибки не исправляются на последующих этапах моделирования.
 - Хорошее выравнивание не исправит плохой выбор белка заготовки.
 - Хорошее моделирование петель не исправит плохое выравнивание.
- При обнаружении ошибки необходимо повторять некоторые этапы.

Проверка

- Большинство программ для моделирования по гомологии выдают правильные значения для связей и валентных углов.
- Карта Рамачандрана в большинстве случаев для модели выглядит также, как для белка заготовки
- Проверка на ориентацию или положение заряженных остатков может быть полезна.
- Использование любых экспериментальных данных:
 - Остатки активного центра.
 - Места модификаций.
 - Места контактов.

ProQ сервер оптимизирован на поиск правильной модели а не нативной структуры.

Ресурсы для гомологичного моделирования

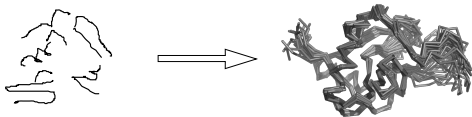
- Modeller
- SwissModel
- Eva-CM
- Nest И т.д.

Предсказание структуры белка *Ab initio*

- Теоретически можно использовать молекулярную динамику.
- Моделирование отжига, как в МД так и в Монте-Карло.
- На основе фрагментов, Rosseta

Ab initio, Rosseta

- Метод использует информацию о предсказании вторичной структуры
- Сравниваем фрагменты от 3 до 9 остатков с библиотекой известных структур. Строим эти фрагменты.
- Соединяем эти фрагменты и используем Монте-Карло для оптимизации третичной структуры.



Ab initio, Rosseta

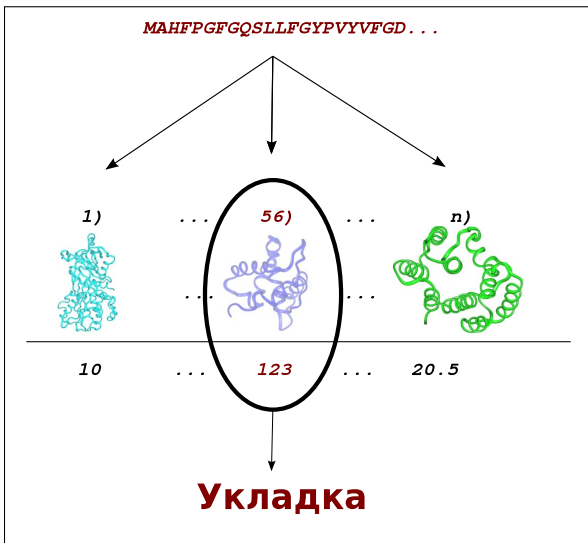
- Для определения хорошей конформации использую специальные потенциалы, которые делают модель похожей на нативную
- Что можно использовать:
 - Потенциалы для третичных контактов
 - Гидрофобные потенциалы
 - Потенциал для уменьшения радиуса вращения молекулы
 - Водородные связи и т.д.

Можно добавить знание об дисульфидных мостиках, местах связывания катионов металлов и т.д.

Threading — протягивание нити

- Сравниваем последовательность со всеми известными способами укладки.
- Используем потенциалы для определения тенденций в известных способах укладки.
 - Каждую аминокислоту из модели помещаем в позиции белков разных укладок
 - Определяем как хорошо эта аминокислота подходит белку заготовке на основе парных взаимодействий
 - На основе суммарного результата определяем белок заготовку.

Threading — протягивание нити

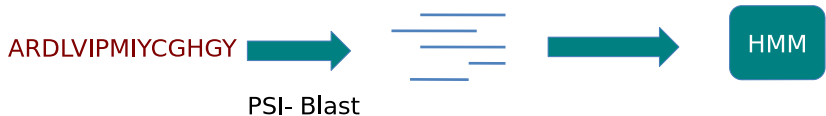


Threading — недостатки

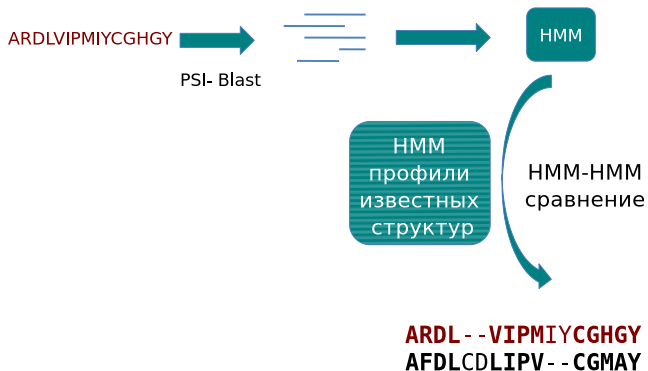
- Взаимодействия в белке не всегда описываются парными контактами.
- Потенциалы часто основываются на профилях последовательностей.

Есть гибридные методы Rosseta/Threading: I-Tasser

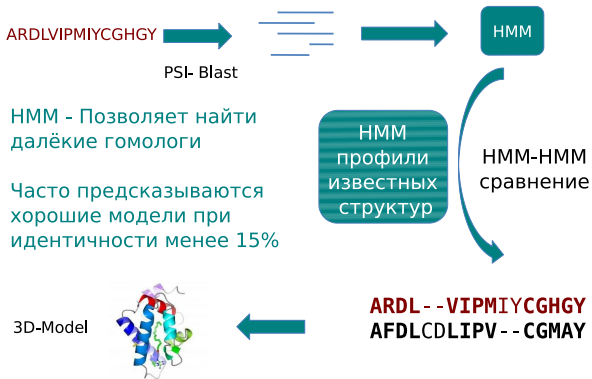
Распознавание укладки, Phyge2



Phyre2



Phyre2



Мета серверы

- Сравнение разных методов.
- Большинство методов предсказывают правильную укладку в первых 10-20 результатах.
- Удаление структур с высоким значением параметров модели, но с единственной укладкой.
- Суперпозиция результатов, взвешивание.
- Часто выдают только позиции атомов остова.