

Дизайн стабильности белков

Дизайн белков, НТУ Сириус

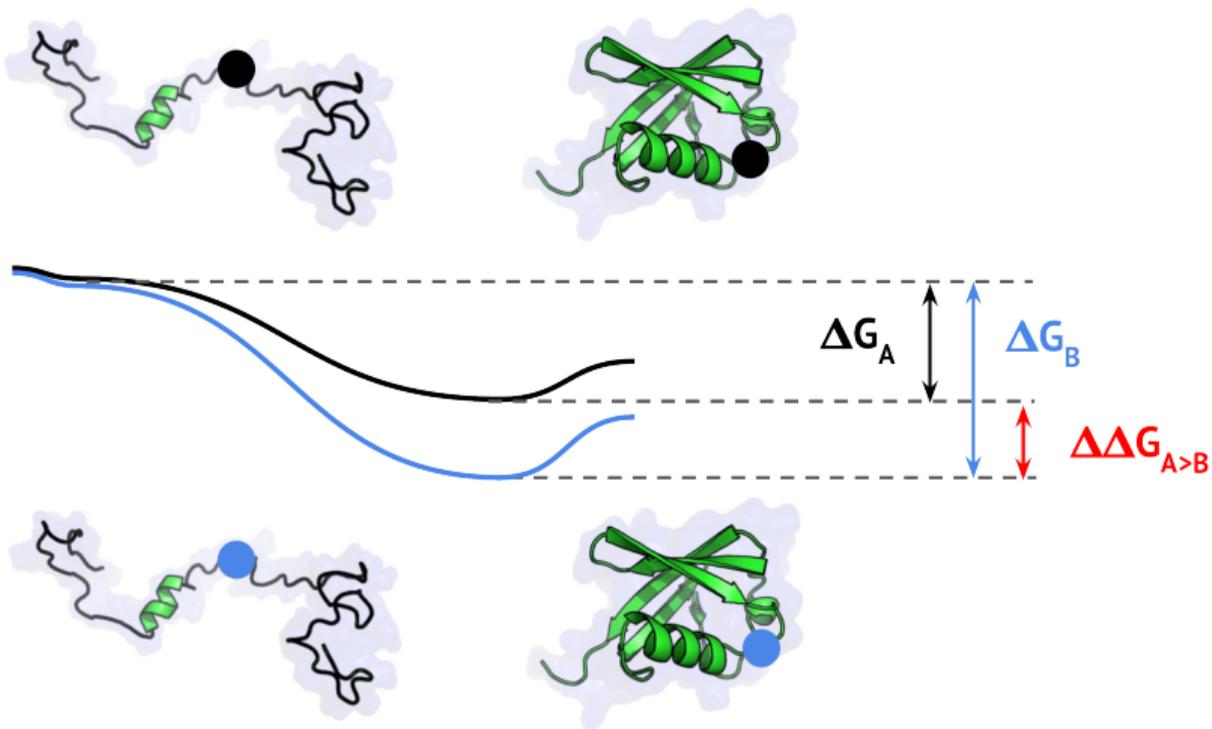
Головин А.В.^{1 2} Маслова В.Д.^{1 2}

¹НТУ Сириус, ЦИИиИТ

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

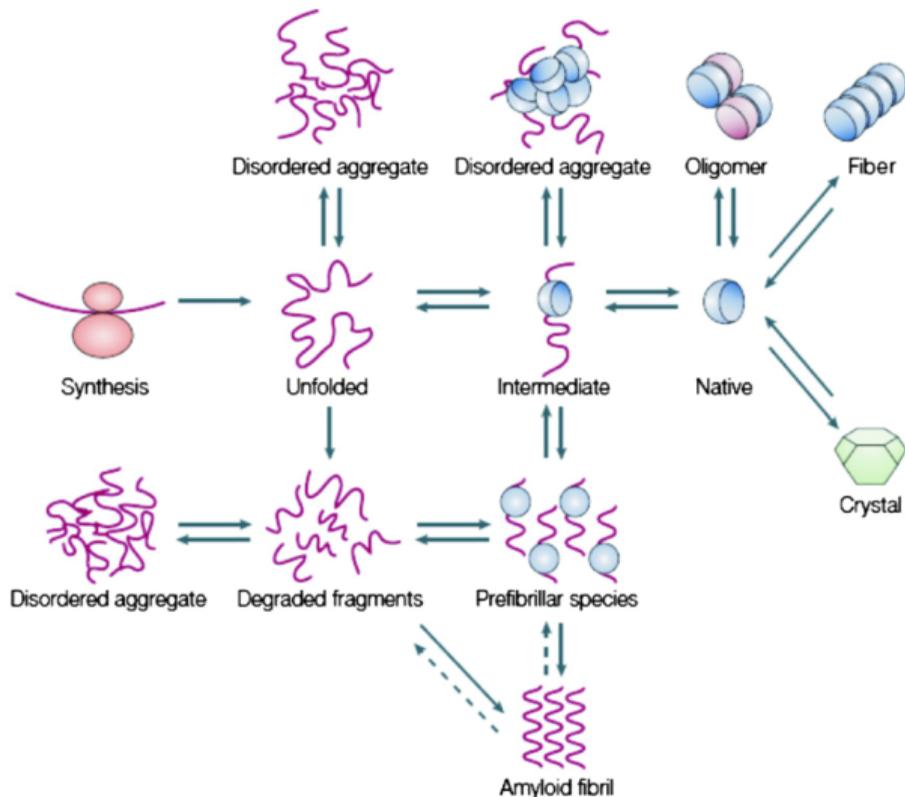
Сириус, 2022

Фолдинг и стабильность



Задача дизайна
стабильности: указать на
замену A>B, которая
приведет к
положительному
 $\Delta G_{A>B}$

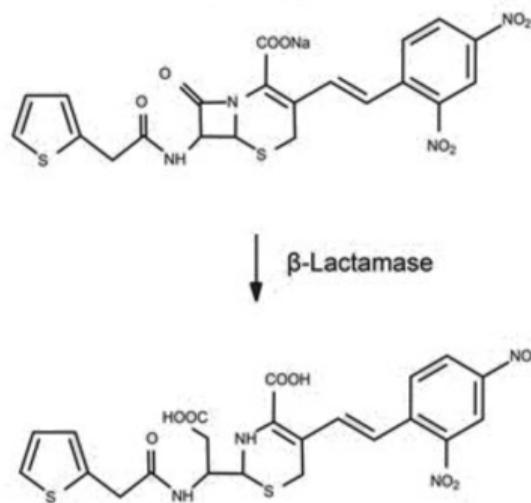
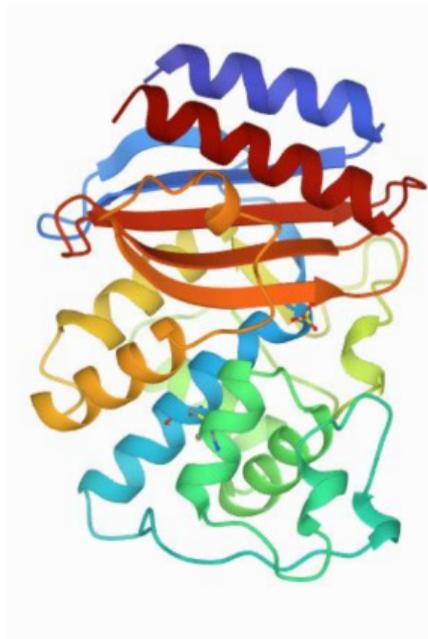
Денатурация и агрегация



$$35^{\circ}\text{C} < T_m \leq 65^{\circ}\text{C}$$

$$3 < \text{pH} < 10$$

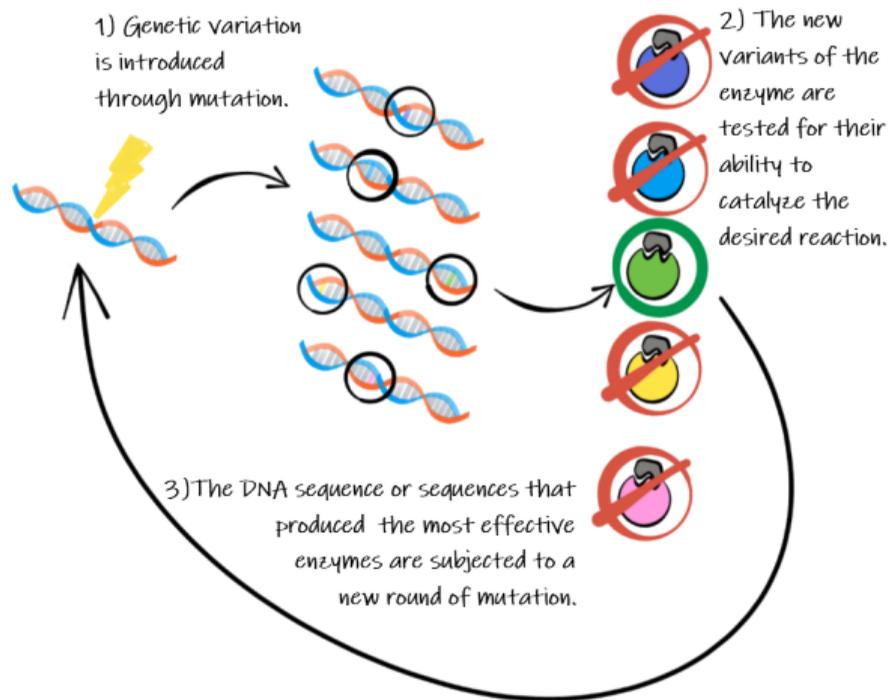
Области применения



Получение коммерчески-применимых ферментов

Стабилизация терапевтических пептидов и белков

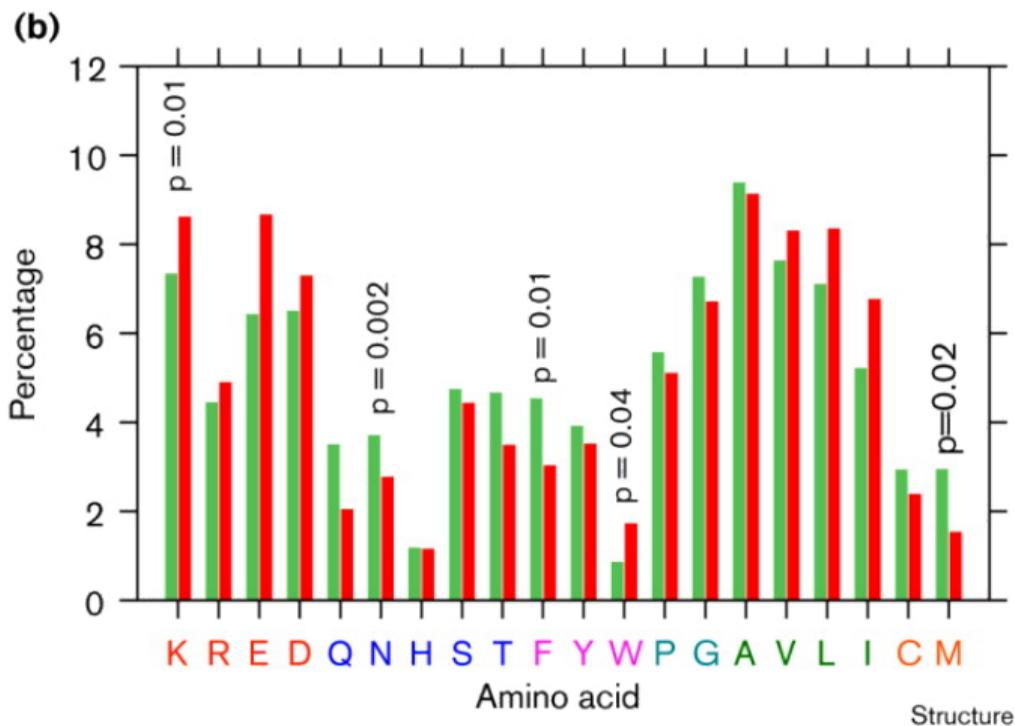
Направленная эволюция



Особенности термостабильных белков экстремофилов

В последовательностях:

- Ile, Val, Leu, Trp ↑
- Gly ↓, Pro ↑
- Ser, Gln, Cys, Met ↓
- Больше полярных АК



Красные это термофилы

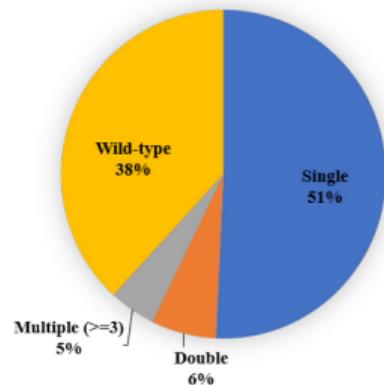
Особенности термостабильных белков экстремофилов

В структурах:

- Короткие петли
- Меньше полостей
- Больше ионных пар на поверхности белка
- Плотно упакованные концы
- Более протяженные участки вторичной структуры

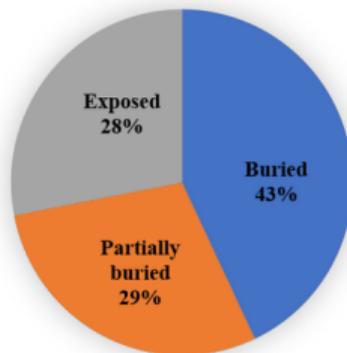
| | | T_{opt} (°C) | 45 | | 50 | | 52.5 | | | 55 | | 60 | | 72.5 | | 75 | | 80 | | 100 | | | | | | | | | | |
|----------------------|-------------|----------------|-------|--------|--------|-------|-------|------|------|------|-------------|--------|------|------|------------|-------|------|------|------|-------|------|------|-------|------|------------|---------|------|------------|--------|--------|
| Protein | | TAGAH | Xyl-1 | Phyc-a | Phyc-b | TAGAH | GAPDH | PGK | PFK | NPR | G/T reduct. | CGTase | ADK | CP | Subtilisin | Xyl-2 | CDGT | SRP | MDH | PPase | SOD | CheY | GAPDH | PFK | Ferredoxin | TATA-BP | OCT | Rubredoxin | Glu-DH | TIF-2B |
| Cavities | number | Blue | Red | Red | Red | Red | Blue | Red | Red | Red | Blue | Blue | Red | Red | Red | Red | Red | Red | Red | Red | Blue | Red | Blue | Red | Red | Red | Red | Red | Red | Red |
| | volume | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue |
| | area | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue |
| Hydrogen bonds | number | Blue | Red | Red | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | |
| | unsatisfied | Red | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | |
| Ion pairs | < 4.0 Å | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | |
| | < 6.0 Å | Red | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | |
| | < 8.0 Å | Red | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | |
| Secondary structure | α | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | |
| | β | Red | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | |
| | irregular | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | |
| Polarity of surfaces | exposed | Blue | Red | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | |
| | buried | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | Blue | |

Mutation count based type



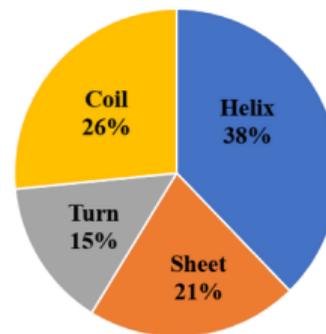
■ Single ■ Double ■ Multiple (≥ 3) ■ Wild-type

Data distribution based on ASA



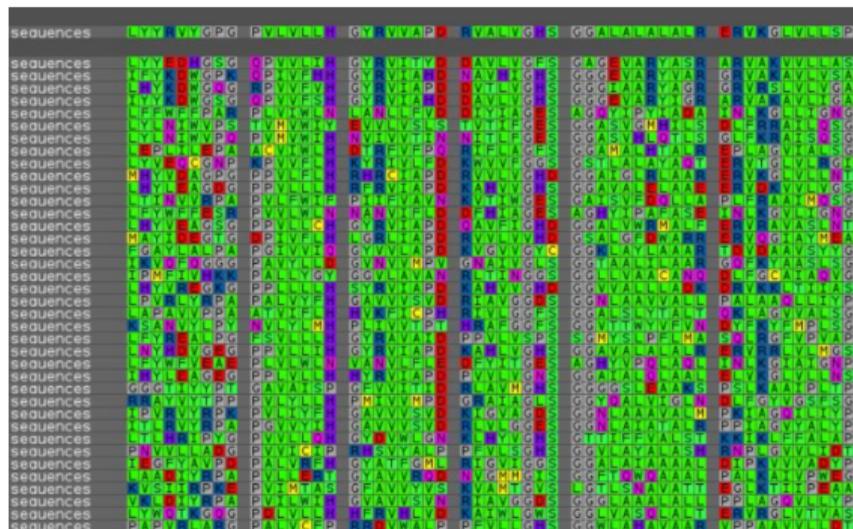
■ Buried ■ Partially buried ■ Exposed

Mutation distribution based on Secondary structure

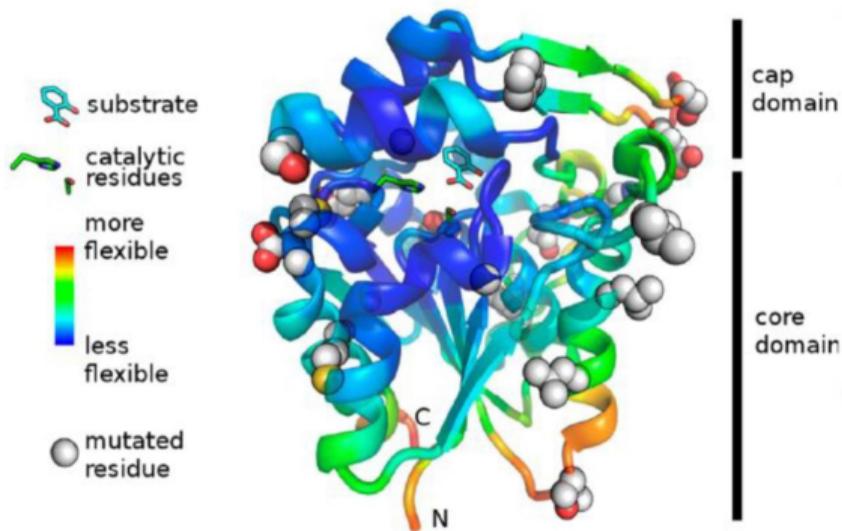


■ Helix ■ Sheet ■ Turn ■ Coil

Вычислительный дизайн стабильности



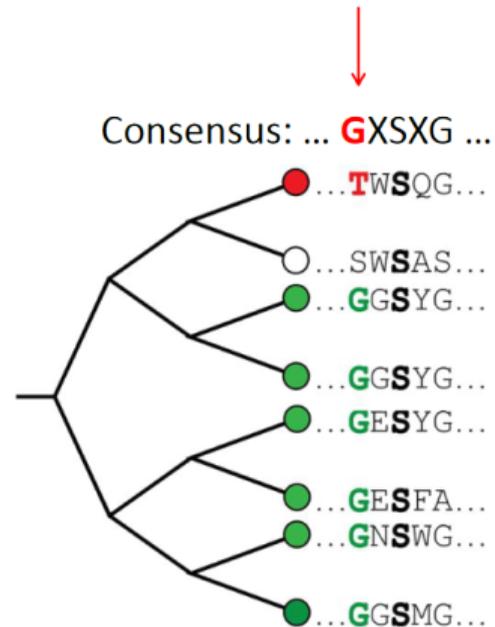
последовательность



структура

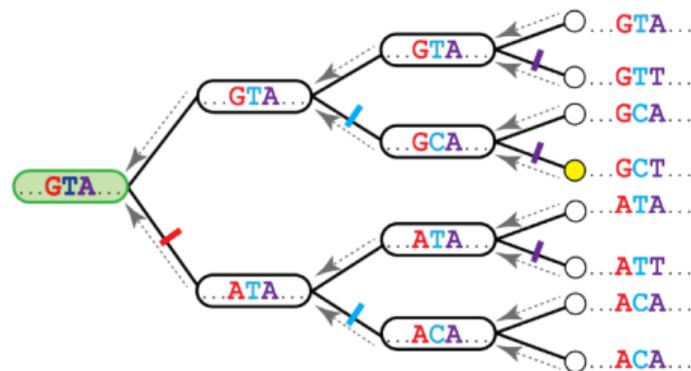
Back-to-consensus

- Предполагается, что консенсусные остатки в функционально разнообразных гомологах отвечают за стабильность, а отличия отражают случайные дестабилизирующие мутации, которые оказались нейтральными при отборе
- Замена в позиции, отличающейся от гомологов, на консенсусную АК приведет к увеличению стабильности



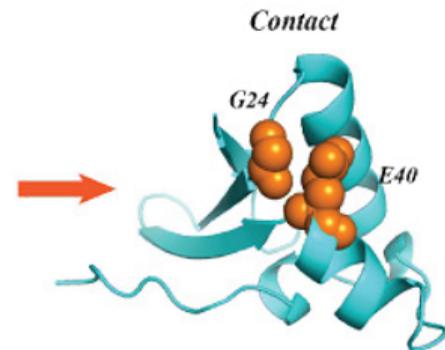
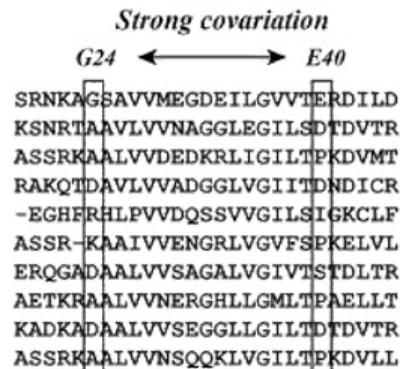
Восстановление предшественника

- Жизнь произошла от термофильных организмов. Предковые гомологи всех белков - термостабильные.
- Построение филогенетических деревьев и поиск предковой последовательности приведет к созданию термостабильного белка.

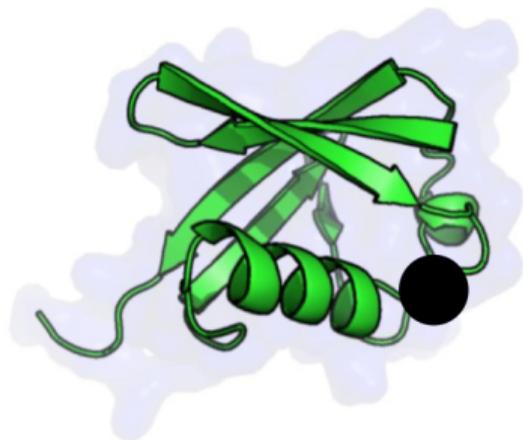


Сочетание структурных данных и выравниваний

- Рациональный анализ структур
- Поиск коварирующих позиций
- Сохранение остатков вторичной структуры
- Сохранение функционально важных позиций

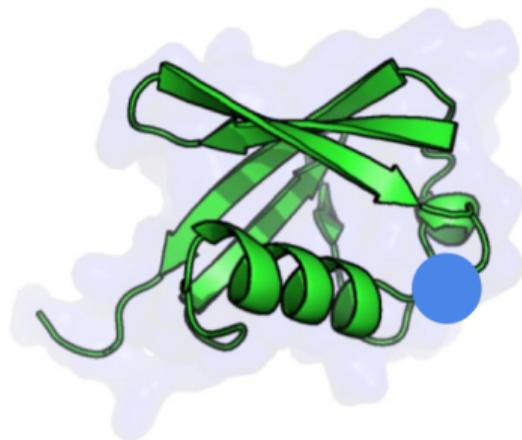


Оценка ddG единичных замен



Physics-based

Оценка межатомных взаимодействий



Descriptors-based

Воспроизведение известных структур с применением ML

Knowledge-based

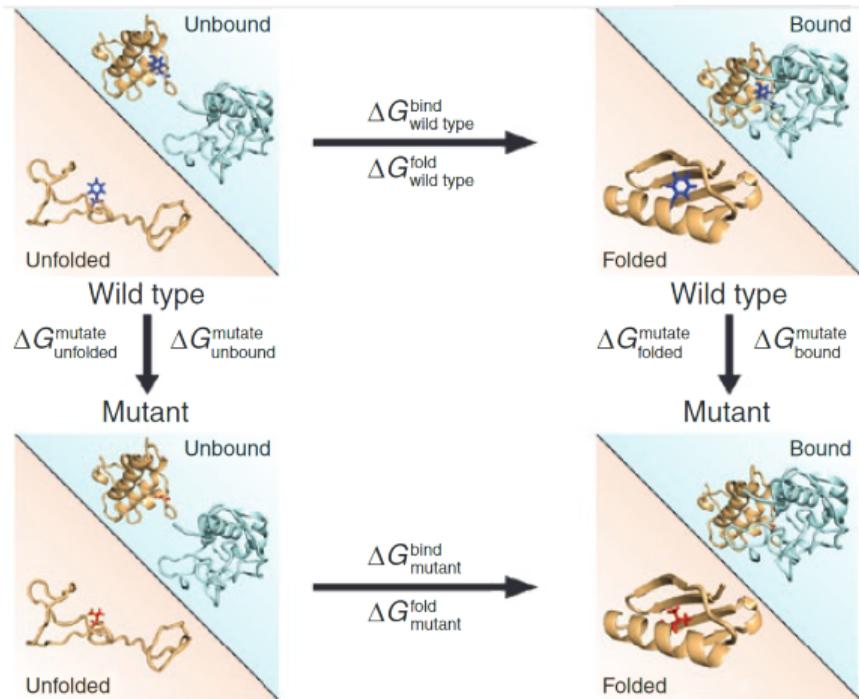
Воспроизведение статистических данных

Оценка ddG единичных замен: CC/PBSA

- Physics-based

$$\Delta G_{CC/PBSA} = \Delta G_{EL} + \Delta G_{VdW} + \Delta G_S$$

- Вычислительно затратный



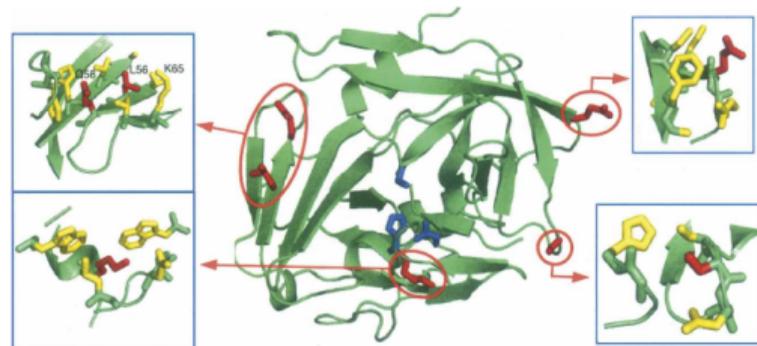
Оценка ddG единичных замен: PopMusic

- Physics-based

$$\Delta\Delta G_P = \sum_{i=1}^{13} a_i(A)\Delta\Delta W_i + a_{14}(A)\Delta V_+ +$$

$$+ a_{15}(A)\Delta V_- + a_{16}(A)$$

- WWW available

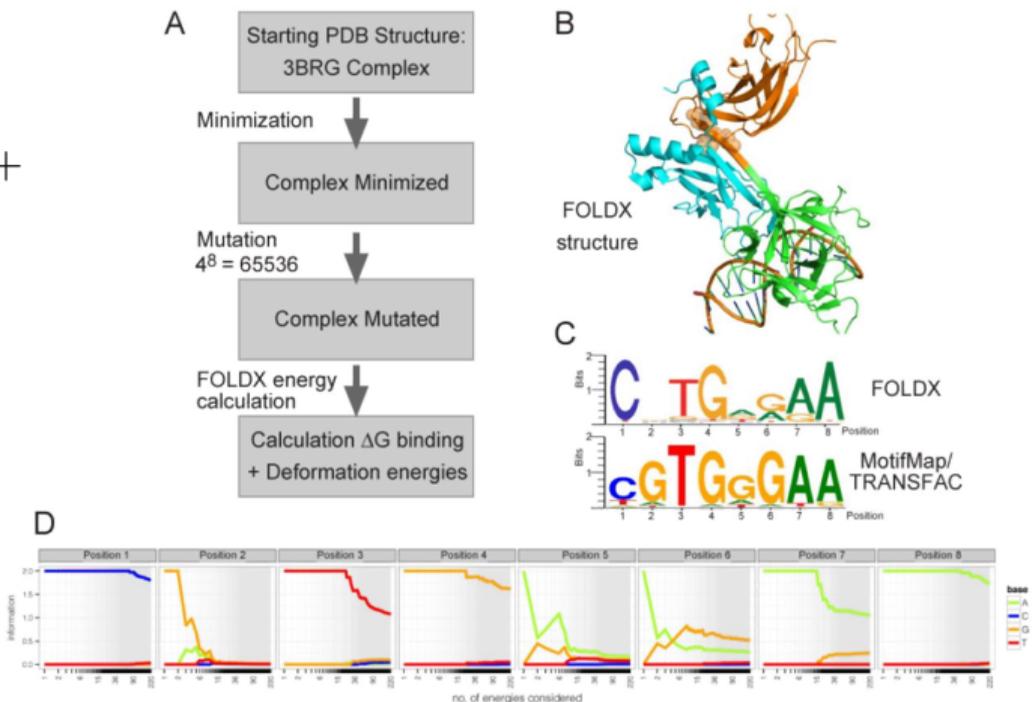


Оценка ddG единичных замен: FoldX

- Physics-based + Knowledge-based

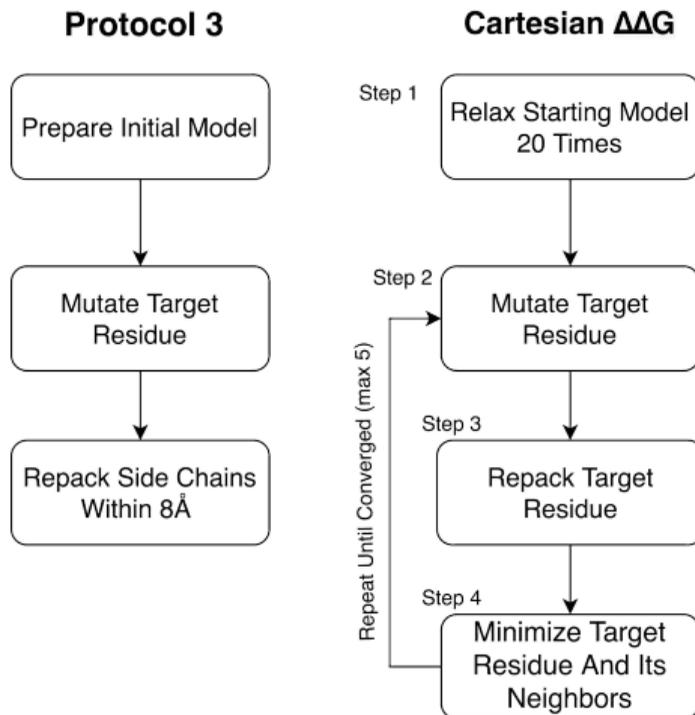
$$\Delta G = \Delta G_{vdw} + \Delta G_{solH} + \Delta G_{solP} + \Delta G_{hbond} + \Delta G_{wb} + \Delta G_{el} + \Delta S_{mc} + \Delta S_{sc}$$

- Есть в виде веб-сервера
- Умеет работать с ДНК
- Умеет работать с димерами
- Умеет учитывать стабильность при смене pH или ионной силы раствора



Оценка ddG единичных замен: Rosetta

- Knowledge-based в сочетании с Монте-Карло динамикой для минимизации структур
- До 2016 года для подсчета ddG применялись кастомные протоколы
- После - cartesian_ddG
- Вычислительно затратен



Оценка ddG единичных замен: I-Mutant

- Predictor-based
- Support vector machine regression
- Есть в виде веб-сервера
- Умеет работать с SNP
- Умеет работать на последовательности без структуры

I-Mutant Suite

Predictor of effects of single point protein mutation

Prediction of protein stability changes upon single point mutation from:

- Protein Structure (if available)
- Protein Sequence**

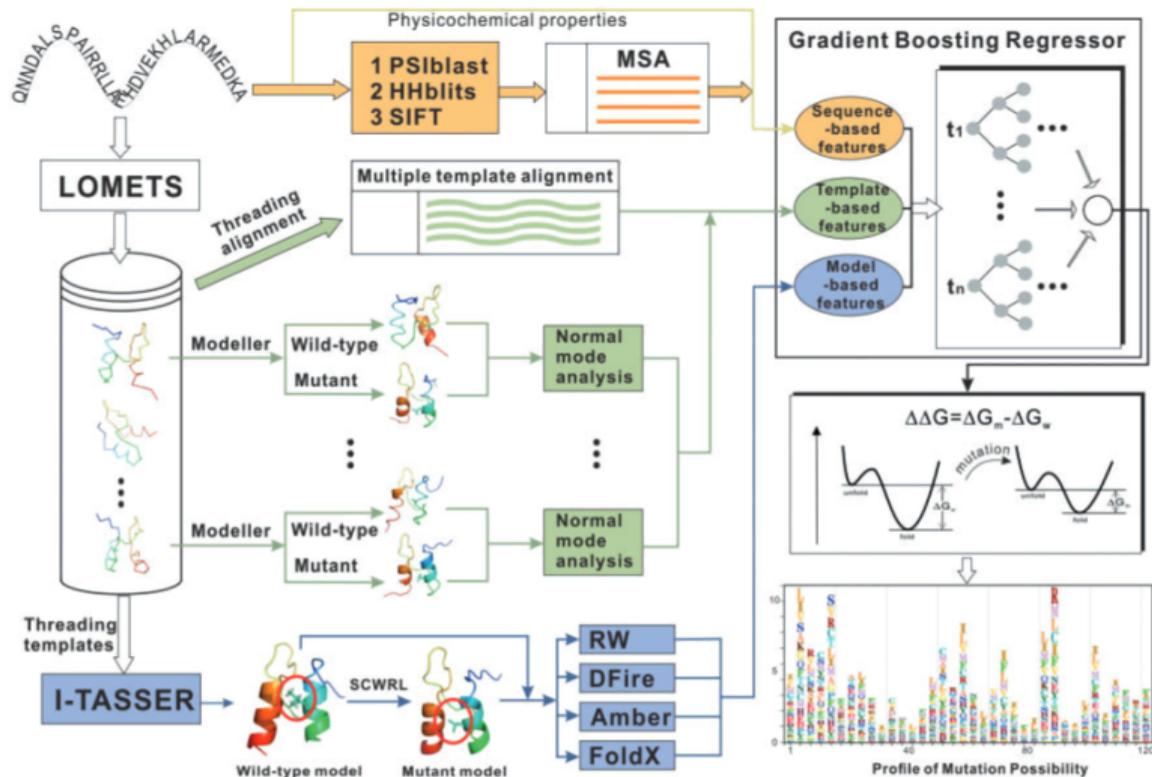
Prediction of Disease associated single point mutation from

- Protein Sequence

Enter

Оценка ddG единичных замен: STRUM

- Predictor-based
- Gradient boosting regression
- Был в виде веб-сервера - сейчас сервер временно мёртв



Оценка ddG единичных замен: Критика

FoldX as Protein Engineering Tool: Better Than Random Based Approaches?

Oliver Buß  , Jens Rudat , Katrin Ochsenreither 

A critical review of five machine learning-based algorithms for predicting protein stability changes upon mutation 

Jianwen Fang 

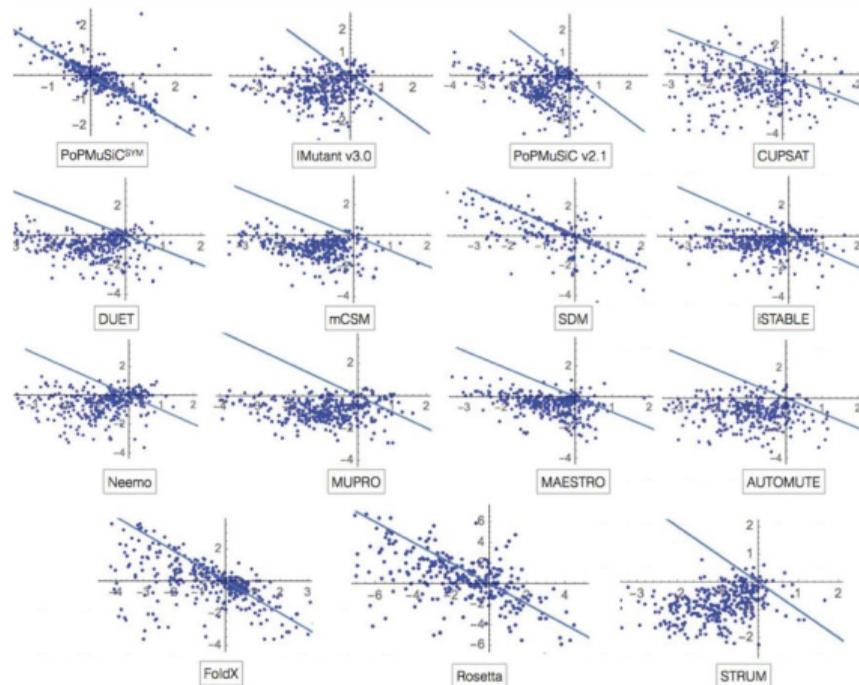
Quantification of biases in predictions of protein stability changes upon mutations 

Fabrizio Pucci , Katrien V Bernaerts, Jean Marc Kwasigroch, Marianne Rومان

Оценка ddG единичных замен: Критика

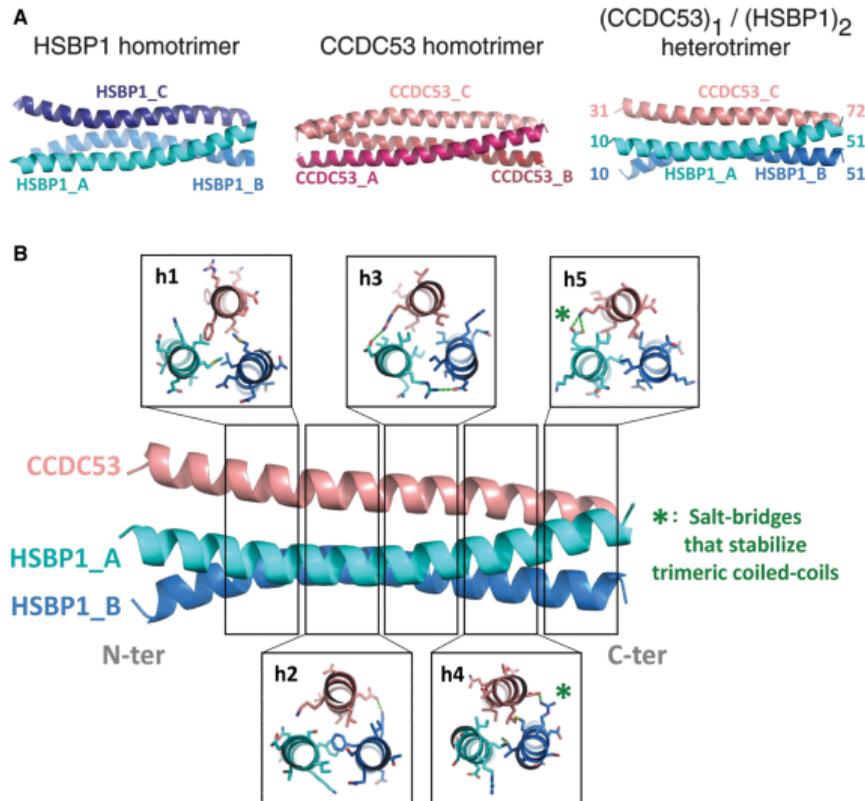
- Критерий асимметричности ddG

$$\Delta\Delta G_{WT \rightarrow mut} = -\Delta\Delta G_{mut \rightarrow WT}$$

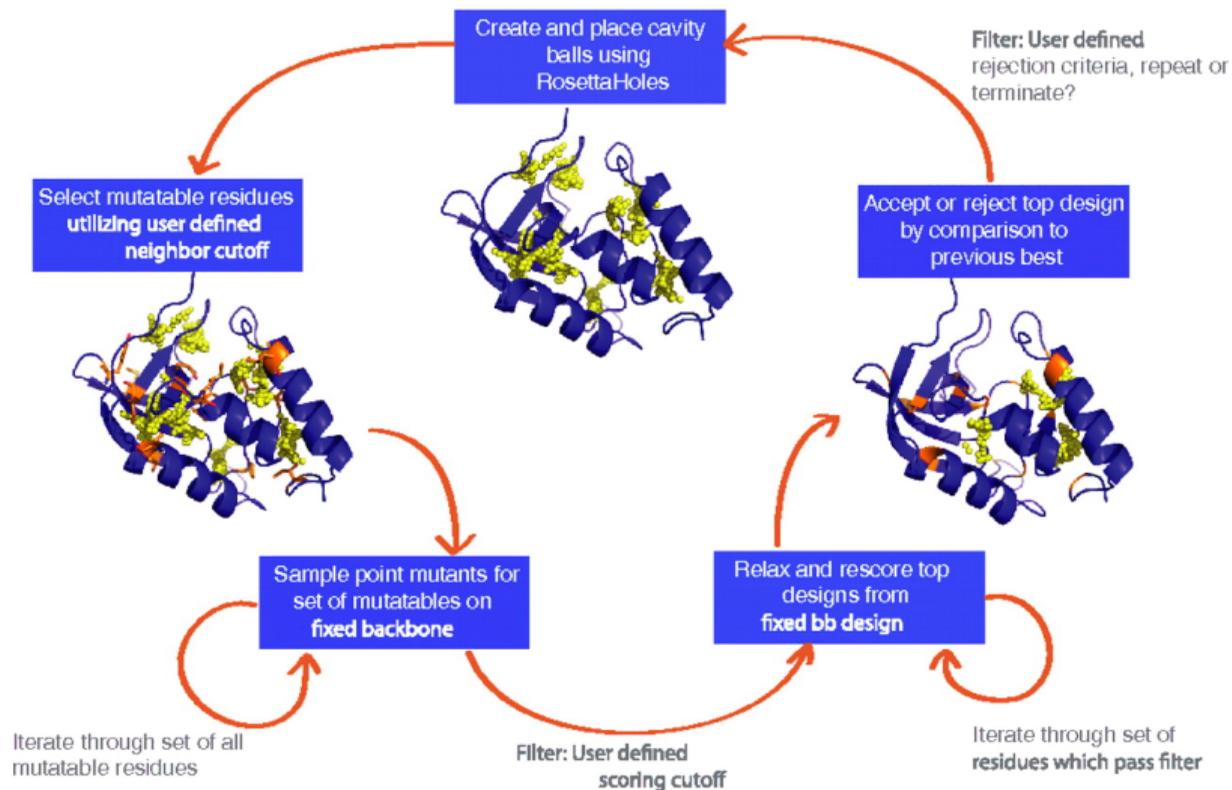


Дизайн гидрофобных ядер

- Наибольшие успехи - с помощью алгоритмов Rosetta
- Получены и стабилизированы неприродные фолды

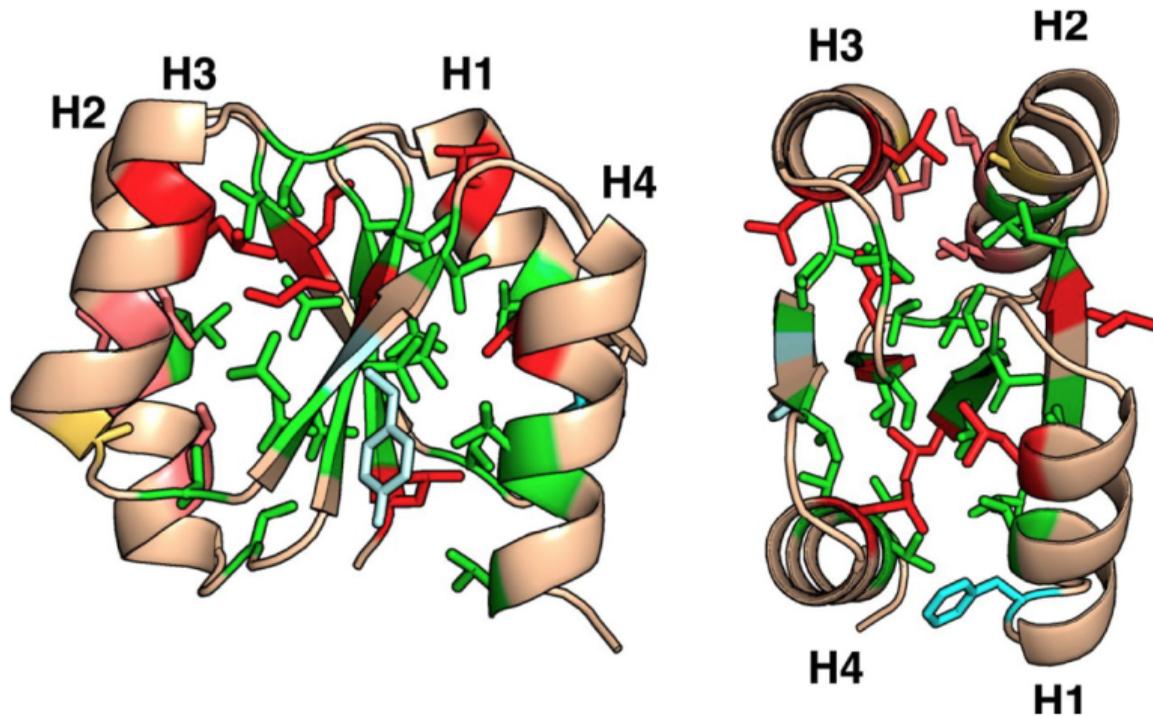


Rosetta VIP



10.1073/pnas.1115172109

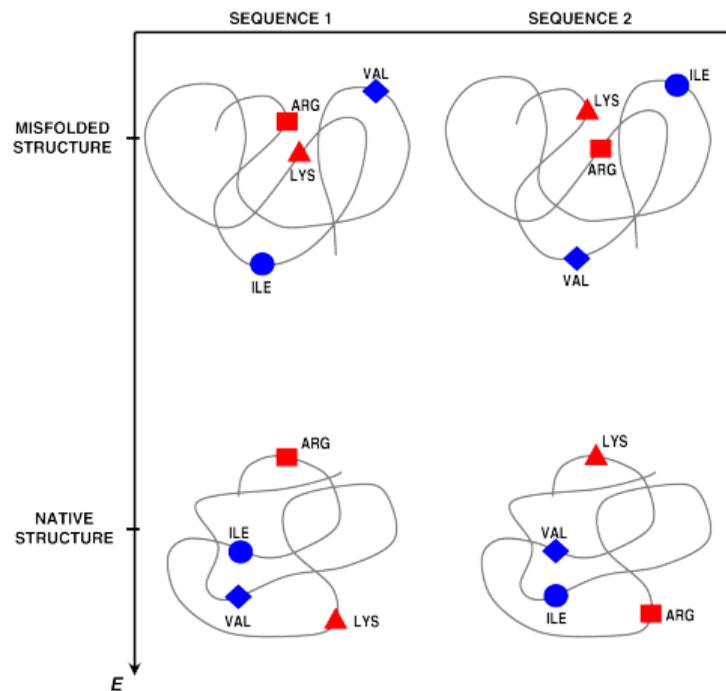
Недоупакованное гидрофобное ядро



Robust folding of a de novo designed ideal protein even with most of the core mutated to valine. <https://doi.org/10.1073/pnas.2002120117>

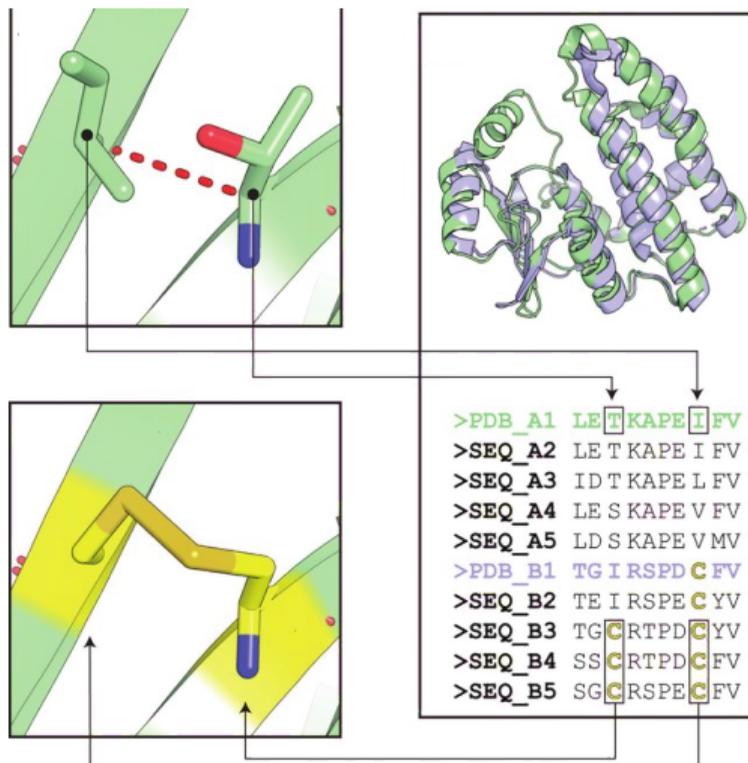
Внесение заряженных АК

- Формирование солевых мостиков
- Негативный дизайн - дестабилизация несвернутых состояний



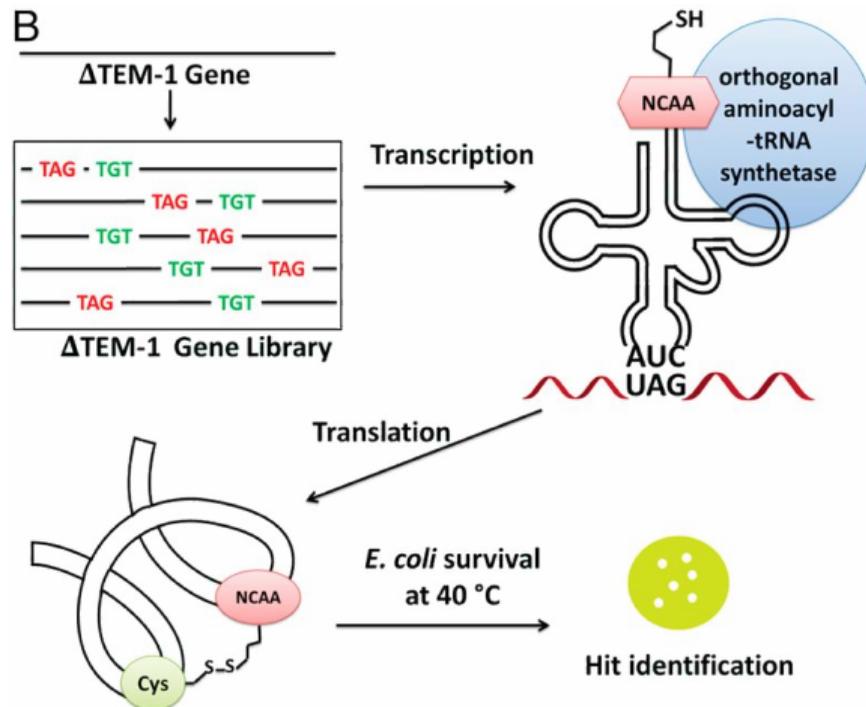
Дизайн дисульфидных связей

- MODIP, DbD - поиск на основе статистической скоринг-функции
- SSbondPre - предиктор на основе нейросети



Дизайн дисульфидных связей с нестандартными АК

- Дисульфидные мостики формируют тиол-содержащие АК
- Используются методы биоинформатики для поиска возможных позиций для мутации на тиол-содержащие АК и методы биоинженерии для вставки таких АК в процессе трансляции



Заключение

- В процессе дизайна стабильности белка приходится иметь дело со сложным процессом фолдинга белка, предсказать который пока невозможно - приходится аппроксимировать
- Выборки для обучения методов предсказания ограничены
- Несмотря на локальные успехи - мало системных универсальных методов