

Оценка эволюционной значимости множественных эпистатических взаимодействий сайтов белков

Минкевич Мария НГУ ФЕН

Научный руководитель:

к.б.н. Афонников Д.А.

Рецензент:

к.б.н. Цепилов Я.А.

ВВЕДЕНИЕ



КЛАССИЧЕСКОЕ ОПРЕДЕЛЕНИЕ

Эпистаз — тип неаллельного взаимодействия генов, при котором проявление одного гена находится под влиянием другого гена (генов).

A	L	W	A	K	V	H	V	E	E	I	G	G	E	A
A	V	W	S	K	V	H	I	E	H	D	G	H	D	A
S	L	W	A	K	V	H	V	E	E	V	G	G	E	A

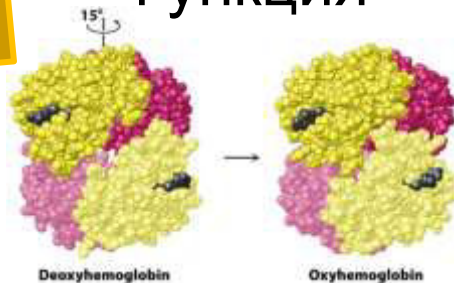
Последовательность



Структура



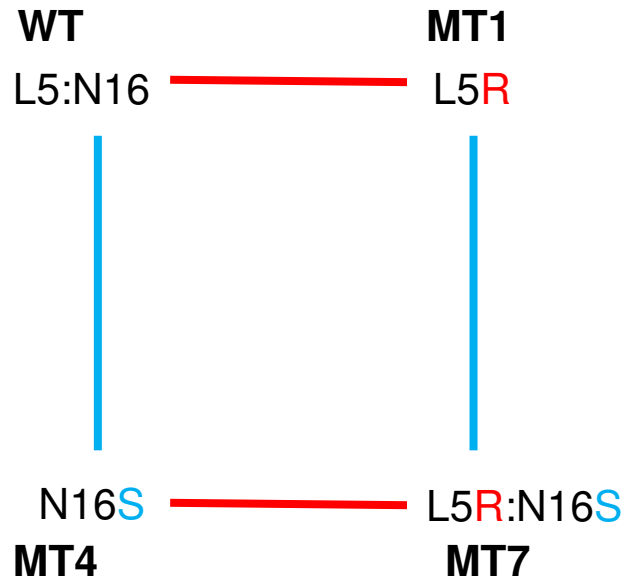
Функция



Эффект кооперативности – взаимодействие аминокислот в пределах одного белка, которое может оказывать влияние на его свойства.

АЛГОРИТМ HYPERCUBE

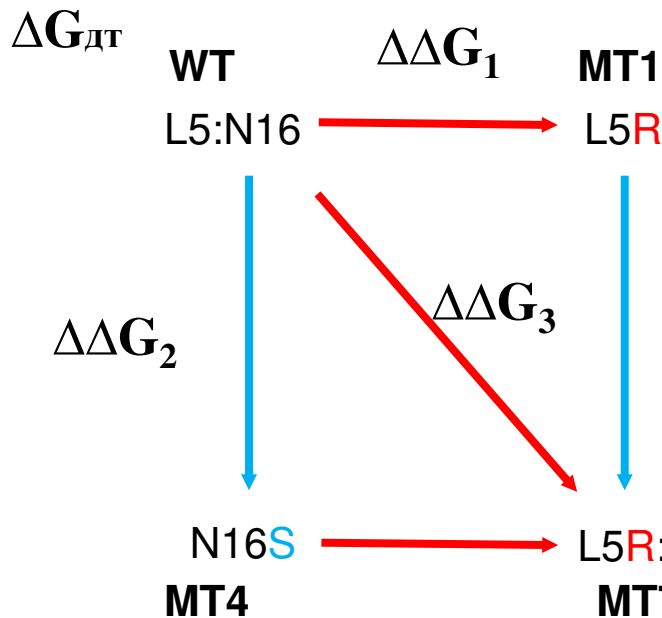
	5	16	
WT L5:N16	L	N	ADTLCIGYHANNST
MT1 L5R	R	N	ADTLCIGYHANNST
MT2 L5S	S	N	ADTLCIGYHANNST
MT3 L5D	D	N	ADTLCIGYHANNST
MT4 N16S	L	S	ADTLCIGYHANNST
MT5 N16L	L	L	ADTLCIGYHANNST
MT6 T19A	L	N	ALCIGYHANNST
MT7 L5R:N16S	R	S	ADTLCIGYHANNST



WT – референсная последовательность
 MT – последовательности с мутациями

Схема пути от дикого типа до варианта с двойной заменой: гиперкуб порядка 2

ОЦЕНКА КООПЕРАТИВНОГО ЭФФЕКТА



Если $\Delta\Delta G_3 = \Delta\Delta G_1 + \Delta\Delta G_2$, то кооперативного эффекта нет

$$\Delta\Delta G_1 = \Delta G_{\text{дт}} - \Delta G_{\text{MT1}}$$

$$\Delta\Delta G_2 = \Delta G_{\text{дт}} - \Delta G_{\text{MT4}}$$

$$\Delta\Delta G_3 = \Delta G_{\text{дт}} - \Delta G_{\text{MT7}}$$

Стабильность белка положительно коррелирует с приспособленностью организма (de Pristo et al., 2005).

WT – референсная последовательность
 MT – последовательности с мутациями

ОЦЕНКА СТАБИЛЬНОСТИ *in silico*

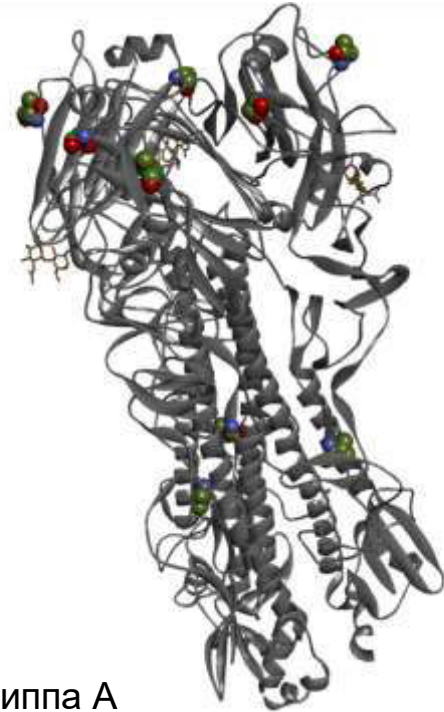


$$\Delta\Delta G = \Delta G_{\text{дт}} - \Delta G_{\text{мут}} \quad [\text{ккал}\backslash\text{моль}]$$

$\Delta\Delta G$ – разница в изменении свободной энергии фолдинга для вариантов белка

$\Delta G_{\text{дт}}$ – дикий тип

$\Delta G_{\text{мут}}$ – мутантный вариант белка



Гемагглютинин гриппа А
Мутации в позиции
A167T:P252S:A398V

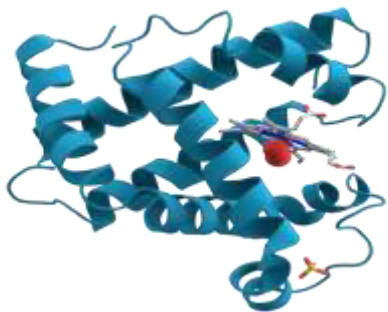
ЦЕЛЬ

Оценить влияние контекста на изменение стабильности белков в результате мутаций на основе анализа множественных аминокислотных замен, представленных в последовательностях гомологичных белков.

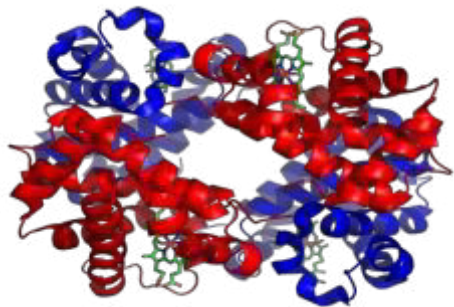
ЗАДАЧИ

1. Формирование выборок последовательностей из нескольких белковых семейств (глобинов, вирусных белков) для анализа, фильтрация данных по качеству последовательностей;
2. Построение множественных выравниваний для последовательностей белковых семейств;
3. Идентификация множественных замен и соответствующих им одиночных вариантов в выравненных аминокислотных последовательностях с помощью программы HupercubeME;
4. Оценка влияния единичных аминокислотных замен и их множественных комбинаций на стабильность белковых структур на основе компьютерного моделирования.

ОБЪЕКТЫ ИССЛЕДОВАНИЯ



Миоглобин 154 а-к остатка
Мономер

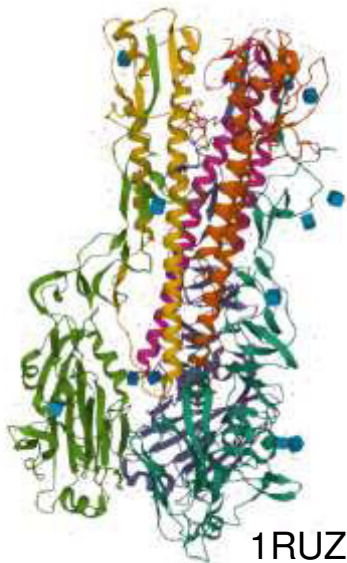


Гемоглобин 147 а-к остатка
Гетеротетрамер

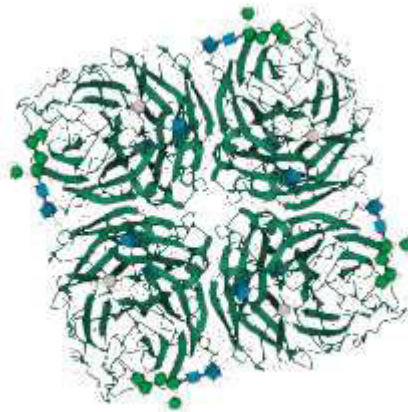
Общие параметры

- ✓ Полностью известна трехмерная структура
- ✓ Небольшой размер
- ✓ Относительно консервативен

ОБЪЕКТЫ ИССЛЕДОВАНИЯ



Гемагглютинин (HA)

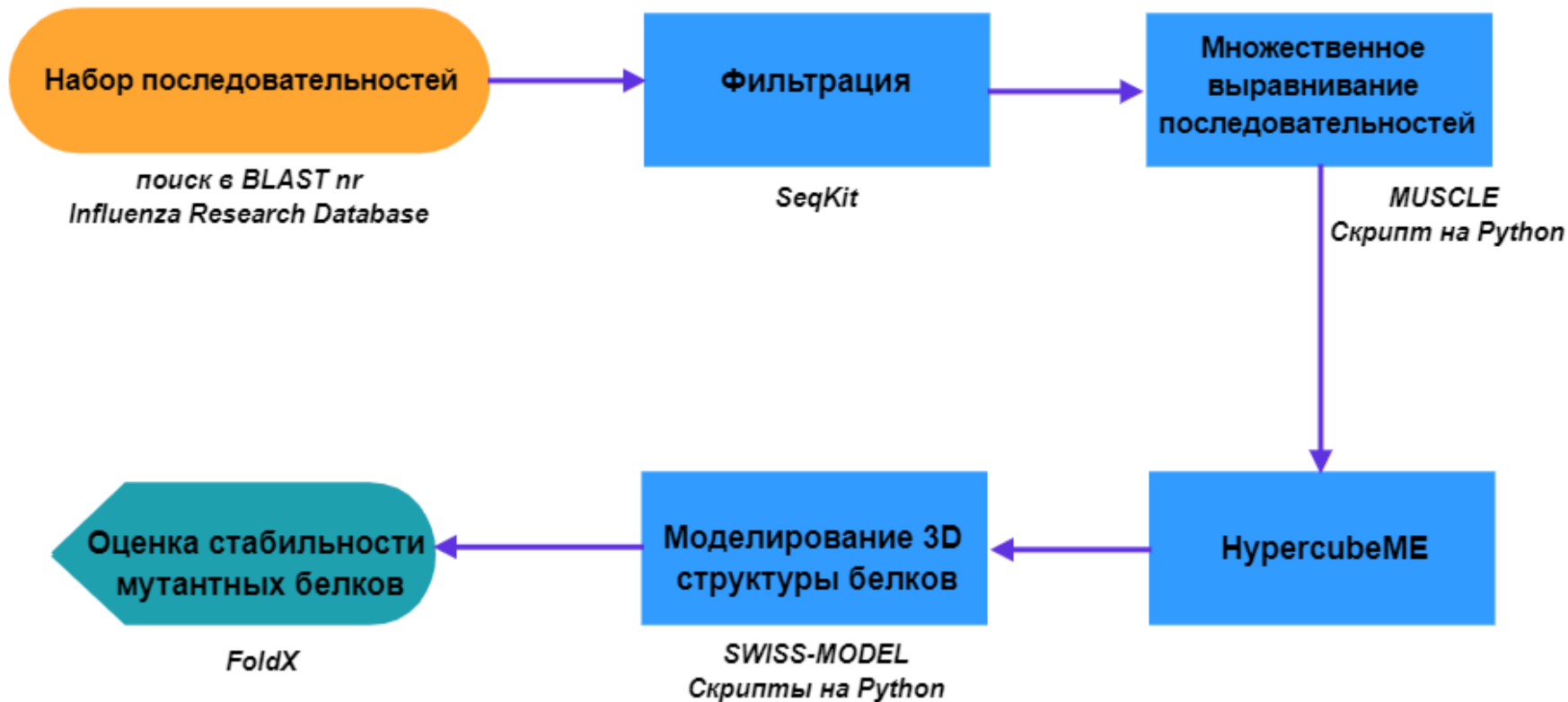


Нейраминидаза(NA)



Нуклеопротеин (NP)

МЕТОДЫ


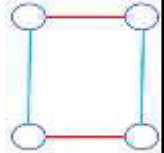
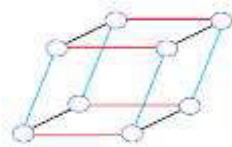


Таб 1 - Количество аминокислотных последовательностей до и после их обработки

Названия белков	Количество последовательностей	
	До фильтрации	После фильтрации
Гемоглобины и миоглобины	23 999	1825
Гемагглютинины	1 050 062	41 982
Нейраминидазы	757 599	31 765
Нуклеопротеины	668 727	12 212

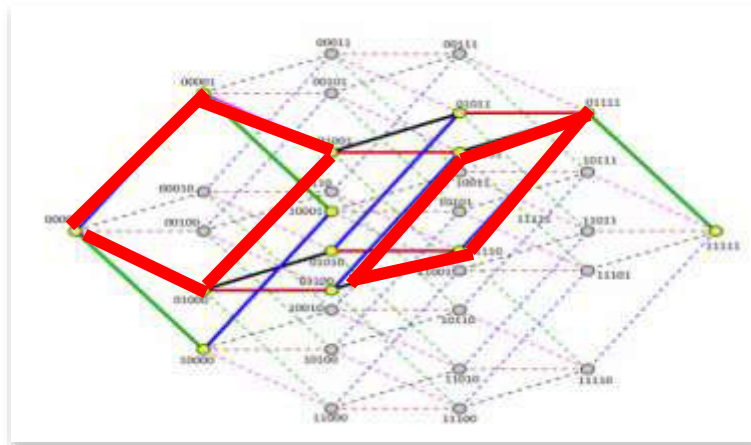
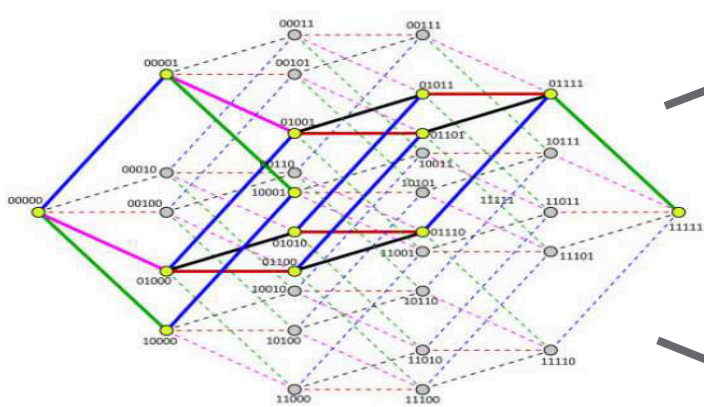
РЕЗУЛЬТАТЫ

Таб 2 - Количество гиперкубов для разных белков

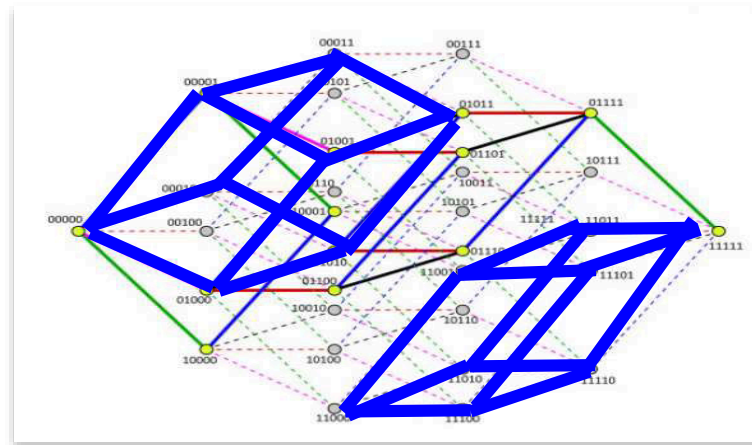
Кол-во гиперкубов Названия белков	одномерные 	двумерные 	трехмерные 
Глобины	593	13	-
Гемагглютинины	23 157	1934	10
Нейраминидазы	19 462	1565	6
Нуклеопротеины	11 992	2499	83

МАКСИМАЛЬНЫЙ ПОРЯДОК ГИПЕРКУБОВ

Для глобинов максимальный порядок составил 2



Максимальный порядок
гиперкуба для белков с
искусственной селекцией: 12
Количество гиперкубов **199 847
053** (Esteban et al., 2019)



Для вирусов максимальный порядок составил 3

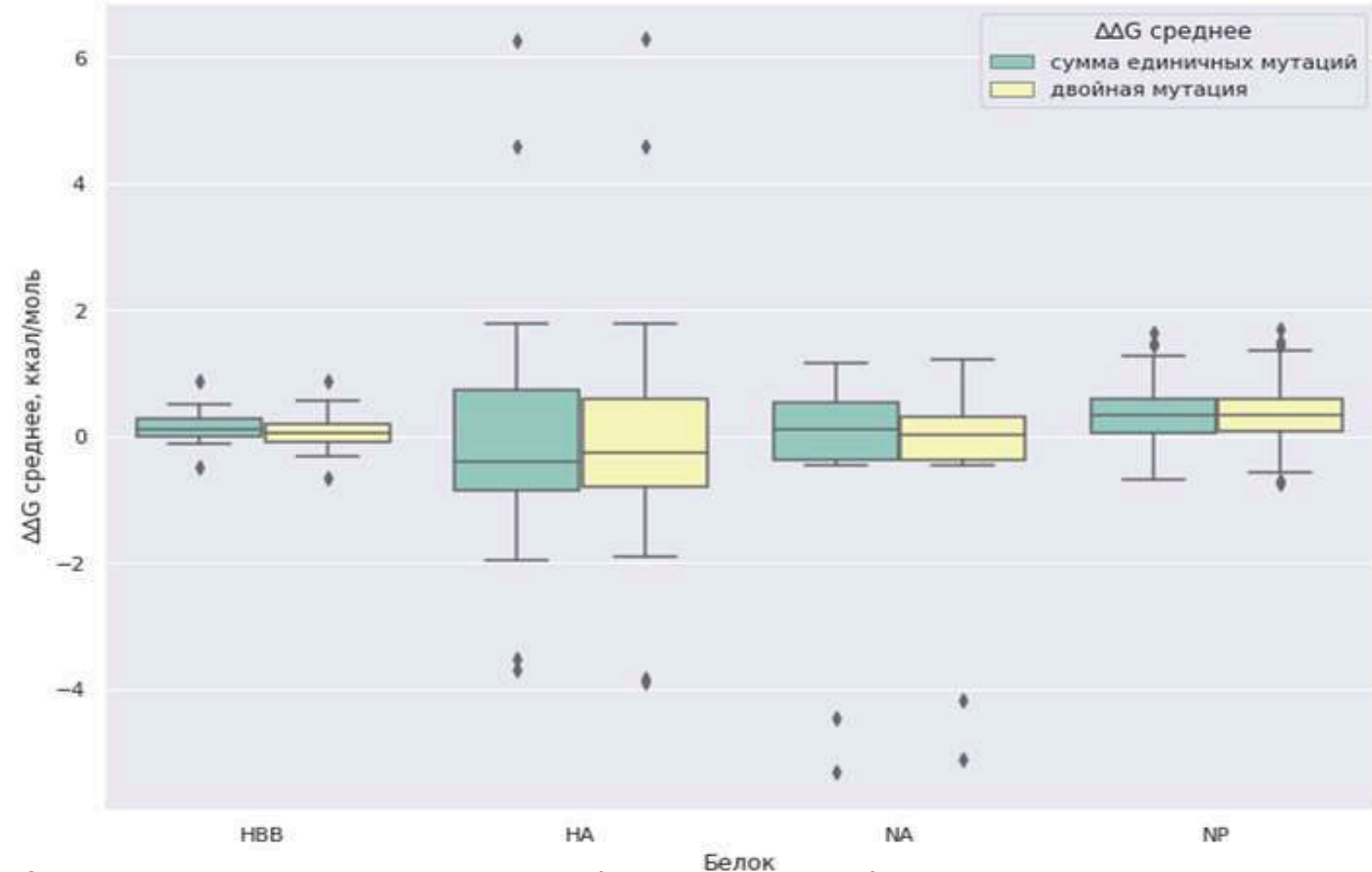
ОЦЕНКА СТАБИЛЬНОСТИ

Таб 3 - Изменение стабильности вариантов глобинов с двойными заменами

Позиции мутаций	$\Delta\Delta G_{\Sigma(ВА, АВ)}$, ккал/моль	$\Delta\Delta G_{(ВВ)}$, ккал/моль
A102G:A182S	1.0257	1.1506
A102G:G112S	0.0480	-0.0303
A182S:A189S	0.1967	0.1996
A83S:A172T	1.7544	1.7562
C92S:A98V	-0.0214	-0.0155
D113G:K144R	0.5655	-0.6502
G112S:A182S	-0.9776	-1.3003
G123S:K189R	-0.2276	-0.2261
H15Q:S31T	0.2874	0.2887
I33V:G39S	0.4901	0.4017

AA – дикий тип
AB, BA – единичные
замены
BB – двойные замены

$\Delta\Delta G_{(ВВ)}$ – изменение
энергии для двойных
замен
 $\Delta\Delta G_{\Sigma(ВА, АВ)}$ - изменение
энергии для суммы
единичных замен



HBV – бета-глобины;
 HA – гемагглютинины;
 NA – нейраминидазы;
 NP – нуклеопротеины

Сравнение средних значений свободной энергии фолдинга в пересчете на 1 замену для двойных замен белков.

Таб 4 - U-критерий Манна-Уитни для оценки значимости для разности средних $\Delta\Delta G$

$\Delta\Delta G_{(11)}$ ср - $\Delta\Delta G_{(10,01)}$ ср, ккал/моль	р-значение	Результат для $p < 0.05$
НВВ:НА	0.441	Нет значимости
НВВ:NA	0.522	Нет значимости
НВВ:NP	0.327	Нет значимости
НА:NA	0.960	Нет значимости
НА:NP	0.757	Нет значимости
NA:NP	0.682	Нет значимости

ВЫВОДЫ

1. Программа HypercubeME может быстро локализовать множественные замены в большом числе белковых последовательностей из баз данных, однако не способна оценить эволюционную значимость (приспособленность) обнаруженных мутантных белков. Для этого требуется отдельный метод, и в данной работе была использована программа FoldX, где фактором кооперативности мутаций служила разница свободной энергии фолдинга между диким типом белка и мутантным вариантом.
2. Ввиду небольшого размера выборки и малой скорости накопления замен в аминокислотной последовательности, миоглобины и гемоглобины имеют небольшое число двойных замен, что описывается программой HypercubeME в виде модели гиперкубов второго порядка.

ВЫВОДЫ

3. Белки вируса гриппа А, имеющие размер выборки на порядок превышающий таковой у глобинов, а также подчиняющиеся другому режиму эволюции, накапливают большое число двойных и тройных замен, что описывается программой HypercubeME в виде модели гиперкубов второго и третьего порядков;
4. Оценка кооперативного вклада мутаций в приспособленность белка с помощью оценки стабильности не выявила эффекта, похожего на эпистаз, в исследуемых белках.

АЛГОРИТМ HYPERCUBE

$n = 0$

0000

1000

1001

0001

$n = 1$

0000 — 0001

1000 — 1001

$n = 2$

0000 — 0001

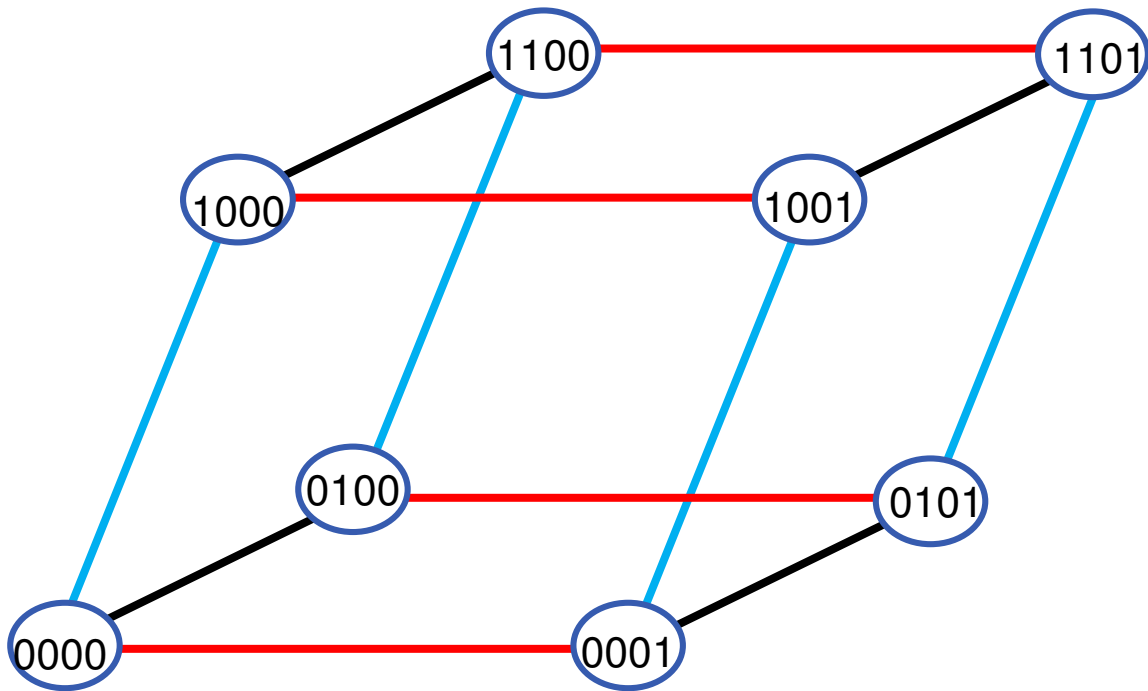
1000 — 1001

Красные линии соответствуют заменам в 4 белковом сайте

Голубые линии соответствуют заменам в 1 белковом сайте

На каждом шаге все возможные n -мерные гиперкубы генерируются из множества $(n-1)$ -мерных гиперкубов.

$n = 3$



1. Порядок гиперкуба = число белковых сайтов с множественными мутациями

2. Число гиперкубов = число послед-тей с множественными мутациями

АЛГОРИТМ ПОЛУЧЕНИЯ ДИКОГО ТИПА БЕЛКА

L5R:N16S

МКАI**R**VVLLYTFTTA**S**ADTLCIGYHANNST

Замена на
оригинальную
аминокислоту
Скрипт на Python

L5:N16

МКАI**L**VVLLYTFTTA**N**ADTLCIGYHANNST

SWISS-MODEL

Получение
трехмерной
структуры

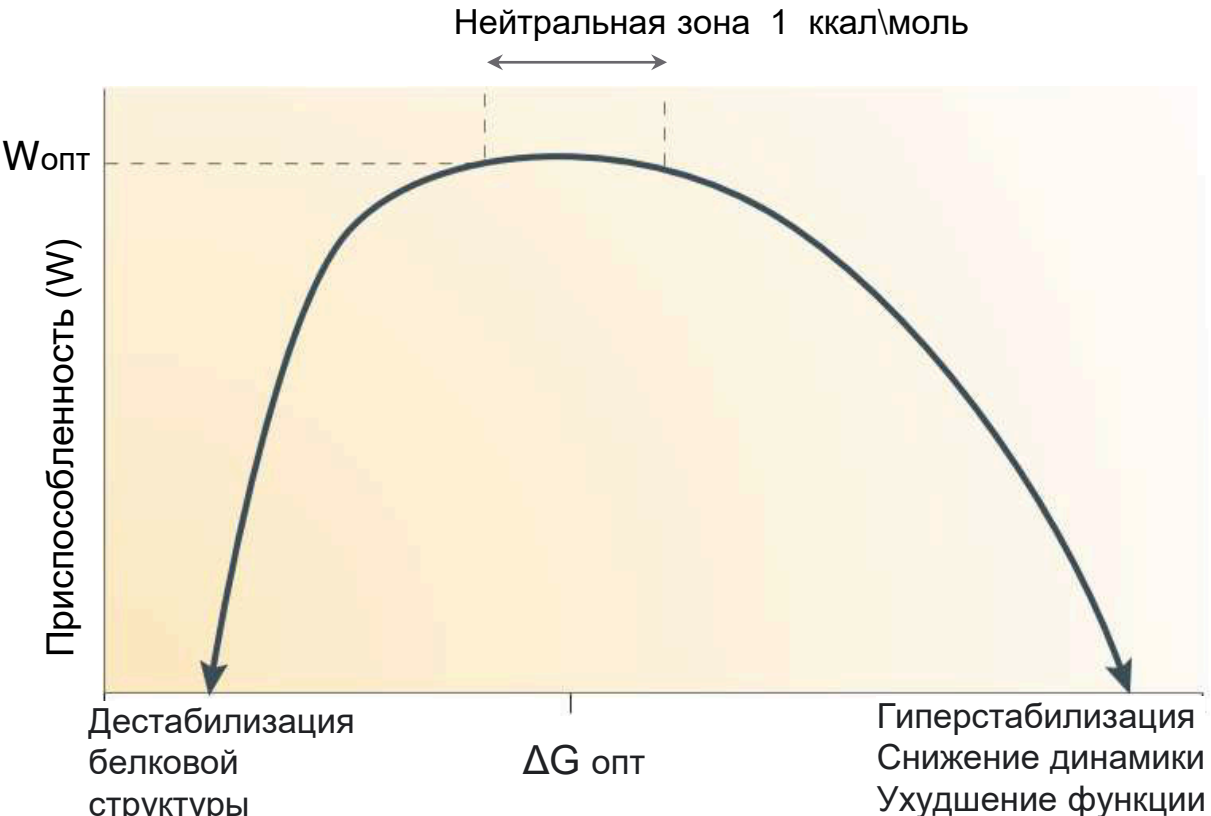


Список мутаций
из HypercubeME

$\Delta\Delta G$

FoldX

ОЦЕНКА СТАБИЛЬНОСТИ



Экспериментально была обнаружена положительная корреляция между стабильностью белка и его приспособленностью

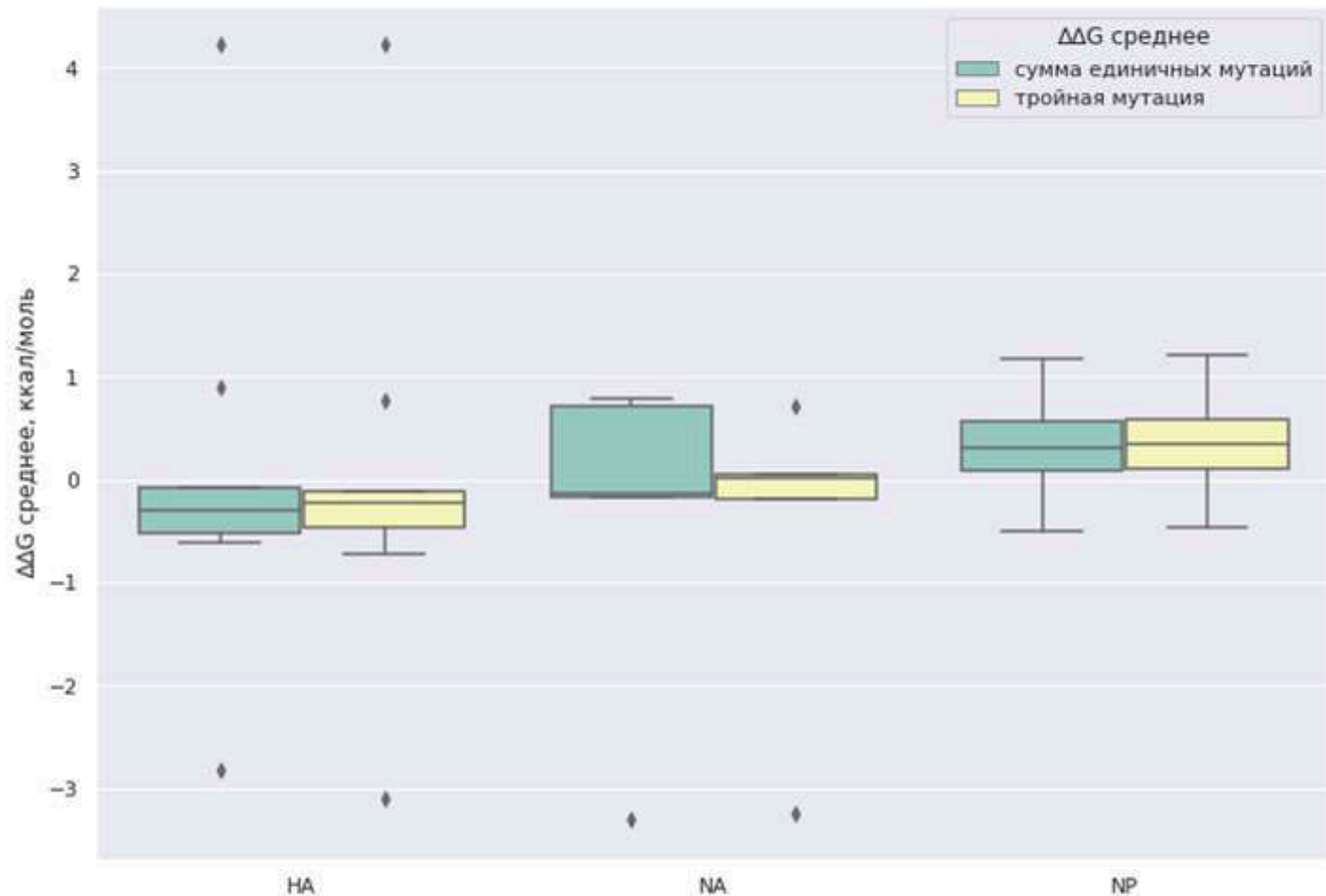
Источник: dePristo et al, 2005

ОГРАНИЧЕНИЯ

1. Размер выборки: в больших выборках выше вероятность идентифицировать множественные замены
2. Режим отбора: разные режимы отбора на белки способствуют разной скорости накопления замен
3. Погрешность в алгоритме FoldX
4. Сложность оценки филогенетического сигнала

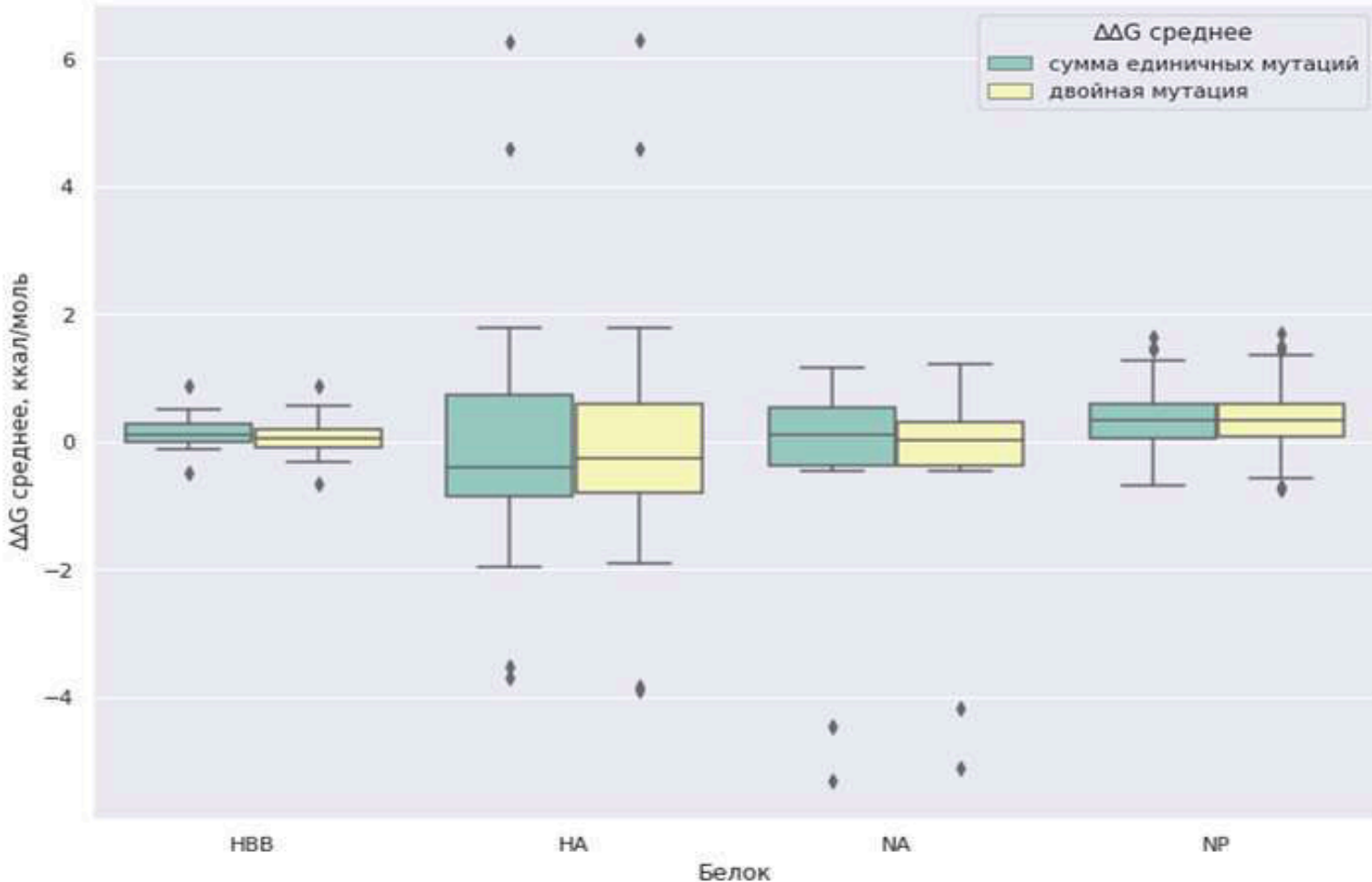
АЛЬТЕРНАТИВНОЕ ОПРЕДЕЛЕНИЕ

Кооперативный эффект – физические взаимодействия аминокислот в пределах одного белка, которые могут оказывать влияние на его биологические и физико-химические свойства.



HA –
гемагглютинины;
NA –
нейраминидазы;
NP –
нуклеопротеины.

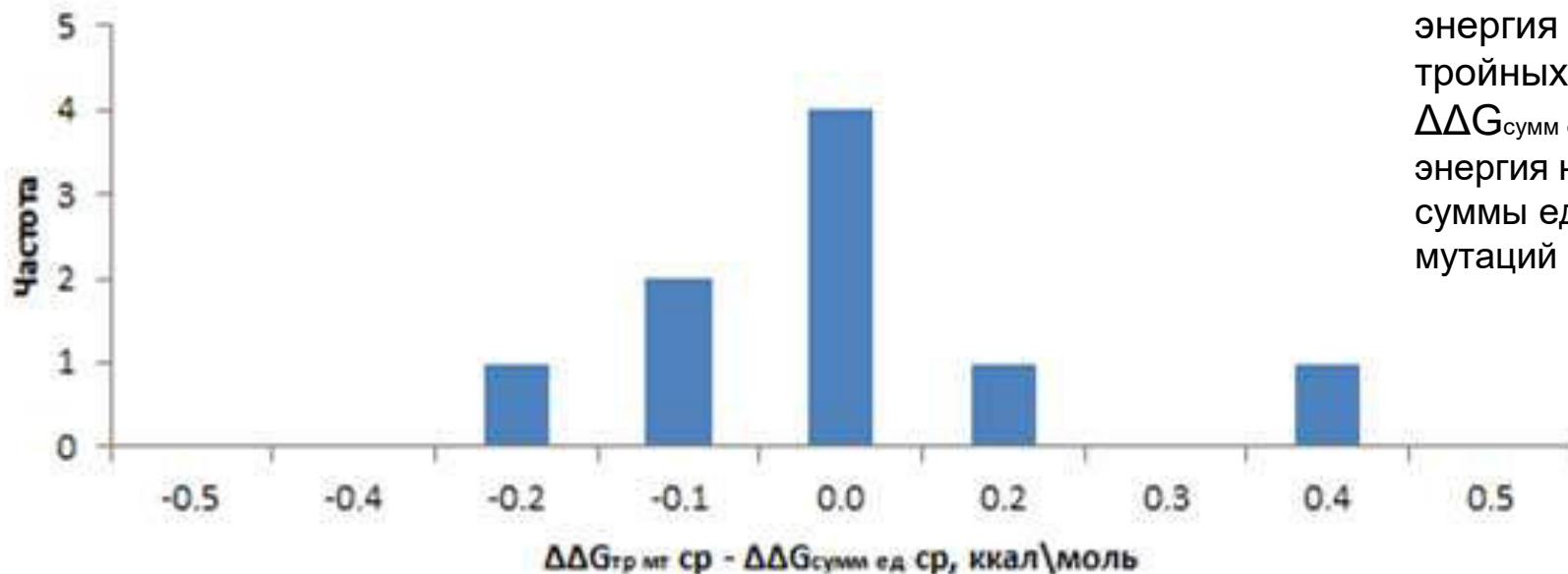
Сравнение средних значений свободной энергии фолдинга в пересчете на 1 замену для тройных замен белков.



HBV – бета-глобины;
 HA – гемагглютинины;
 NA – нейраминидазы;
 NP – нуклеопротеины

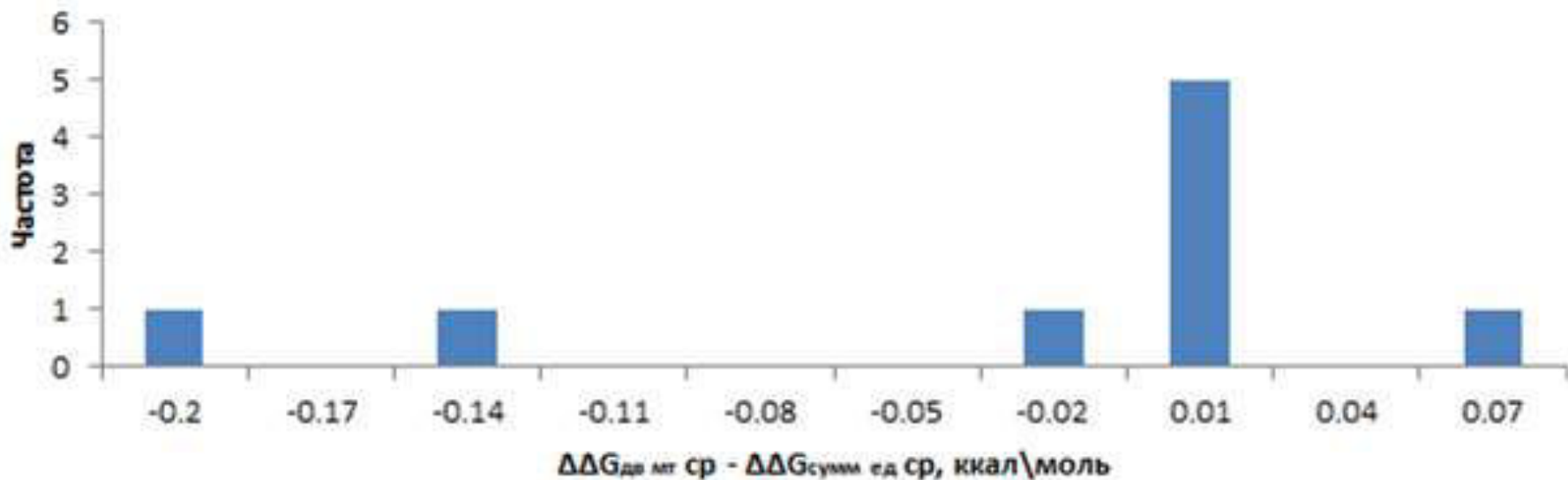
Сравнение средних значений свободной энергии фолдинга в пересчете на 1 замену для двойных замен белков.

Разница между средней энергией на одну замену для $\Delta\Delta G_{\text{тр мт}}$ и $\Delta\Delta G_{\text{сумм ед}}$ для тройных замен в гемагглютинаинах

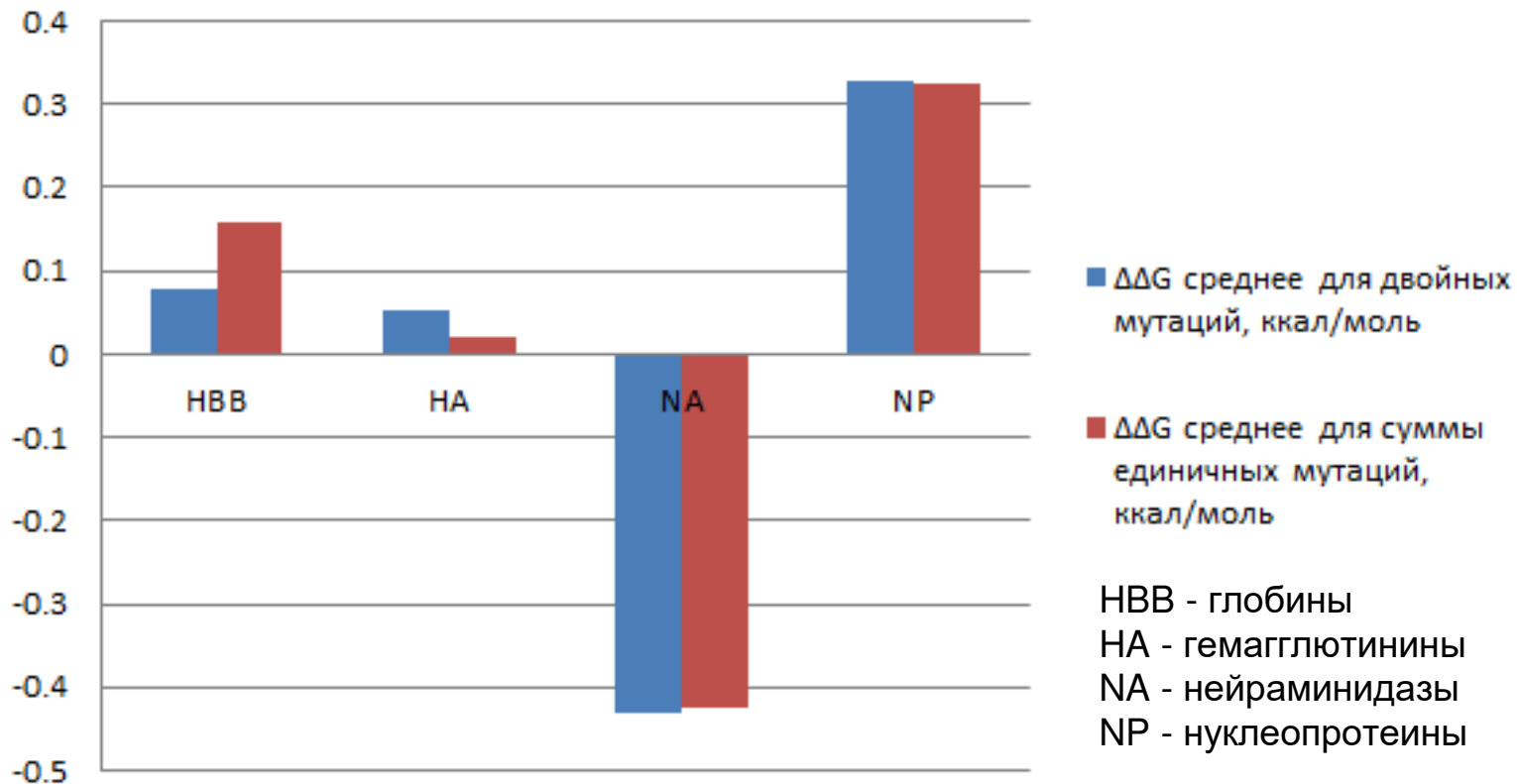


$\Delta\Delta G_{\text{тр мт ср}}$ - средняя энергия на замену для тройных мутаций
 $\Delta\Delta G_{\text{сумм ед ср}}$ - средняя энергия на замену для суммы единичных мутаций

Разница между средней энергией на одну замену для $\Delta\Delta G_{\text{дв мт}}$ и $\Delta\Delta G_{\text{сумм ед}}$ для двойных замен в глобинах



Сравнение средних значений $\Delta\Delta G$ для разных белков

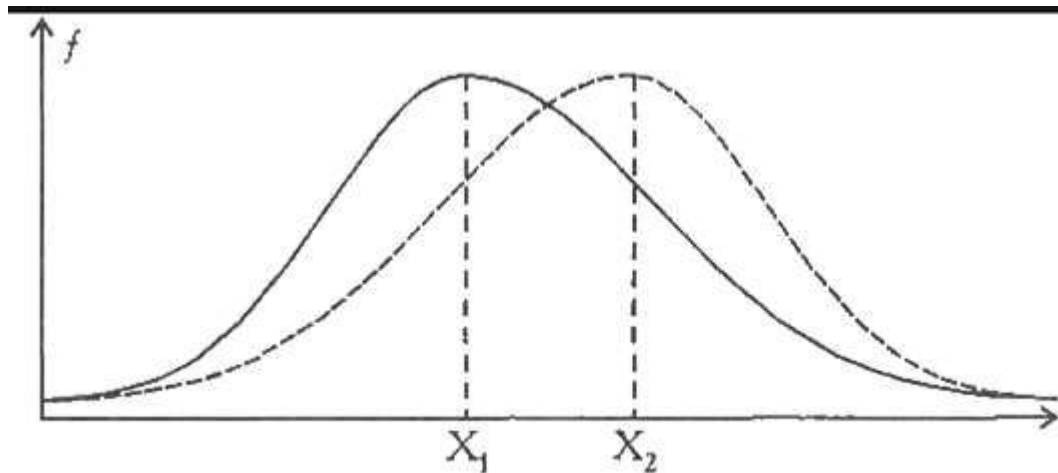


ОЦЕНКА ДОСТОВЕРНОСТИ РЕЗУЛЬТАТОВ

Непараметрический U-критерий Манна-Уитни

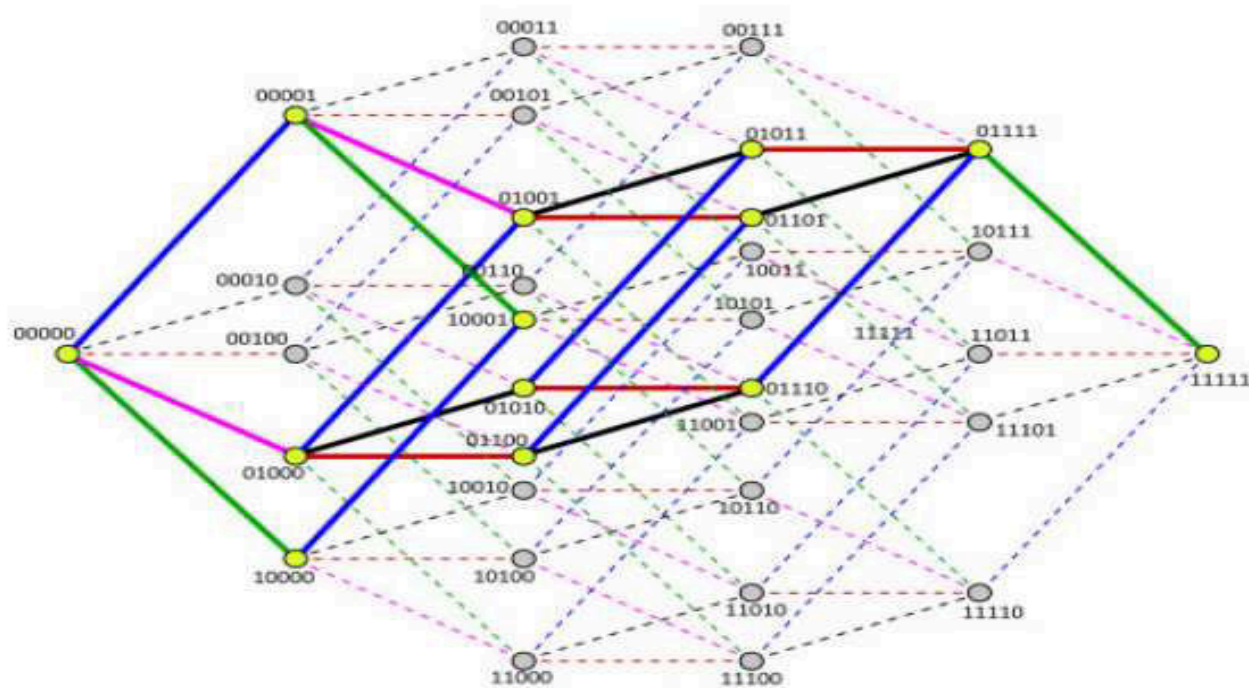
$$U = R - \frac{n(n+1)}{2}$$

где n – размер выборки, а R –
сумма рангов.



Сравнение двух независимых выборок

ОЖИДАЕМЫЙ РЕЗУЛЬТАТ



199 847 053 гиперкубов
размерности 1-12
для белка HIS3

Источник: Laura A.
Esteban et al..
HypercubeME: two
hundred million
combinatorially complete
datasets from a single
experiment

- 1) Можно ли в белках, возникших в процессе эволюции, найти комбинации мутаций, образующих гиперкубы?
- 2) Если да, то какой размерности будут найденные гиперкубы?
- 3) Можно ли оценить фактор кооперативности?