ВВЕДЕНИЕ В ИНФОРАЦИОННУЮ БИОЛОГИЮ

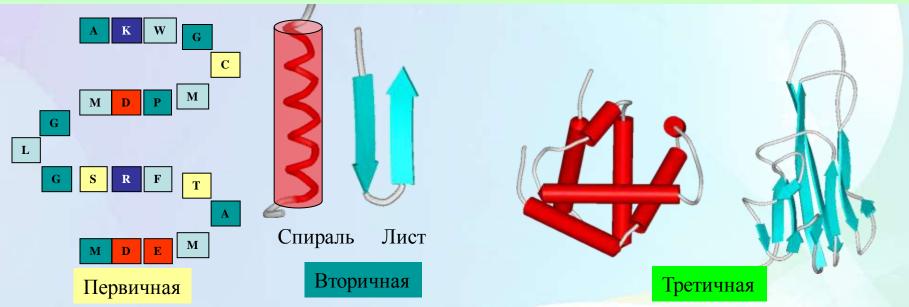
Н.А. Колчанов, С.А. Лашин

Электронно-лекционный курс разработан в рамках реализации Программы развития НИУ-НГУ 2012 год

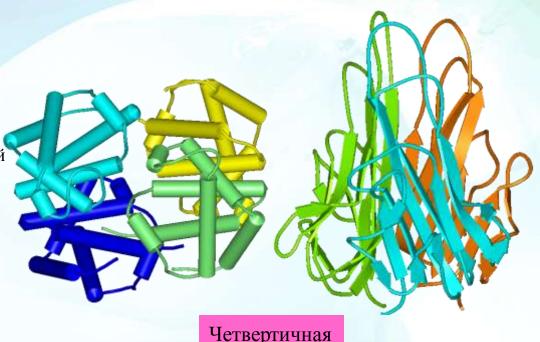
Лекция 3
Компьютерная протеомика:
молекулярный дизайн;
моделирование и анализ эволюции белков;
новая фармакология

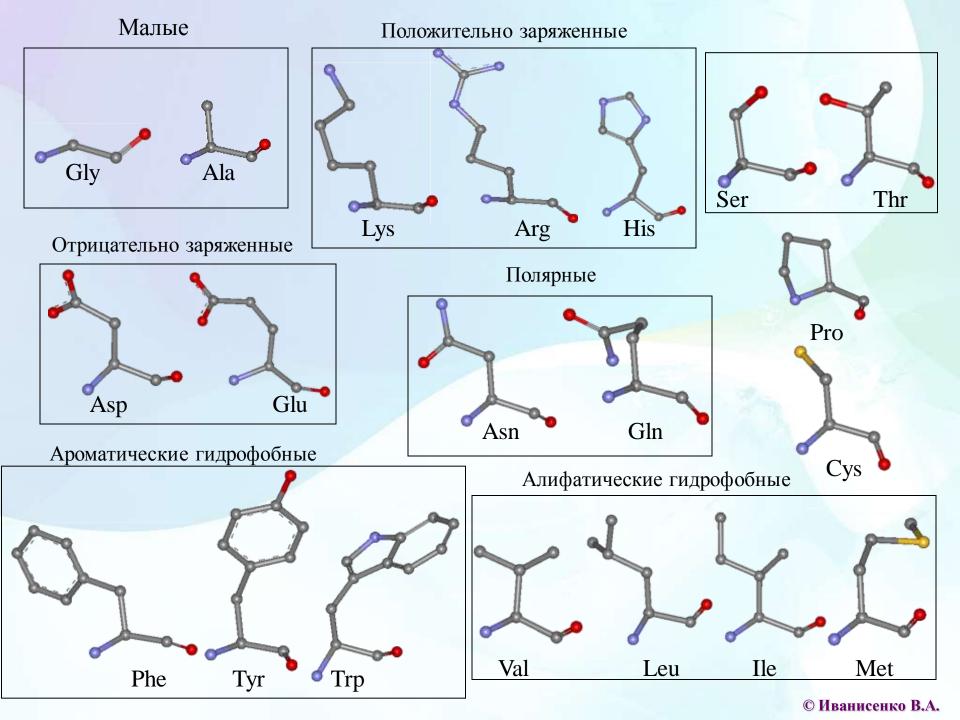
- 1. Основы структурной организации белков
- 2. Инструменты компьютерной структурной биологии
- 3. Функциональная аннотация белков
- 4. Планированию аминокислотных замен, направленно меняющих активность белка
- 5. Молекулярная эволюция функциональных сайтов
- 6. Мишени действия лекарств
- 7. Поиск потенциальных мишеней действия противовирусных лекарственных препаратов
- 8. Молекулярные механизмы побочных эффектов лекарственных препаратов

Различные уровни организации белковой структуры



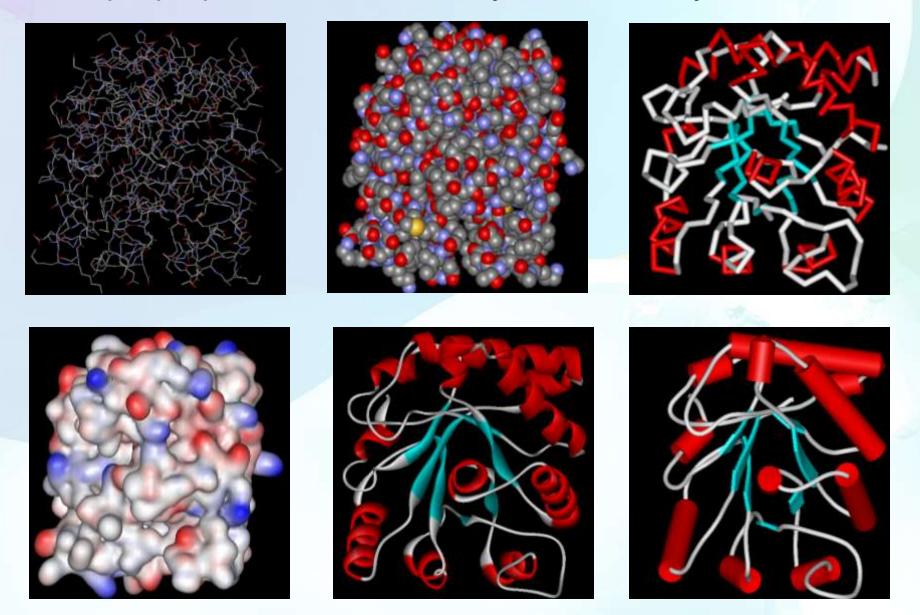
Первичная структура белковой цепи представляет собой аминокислотную последовательность. Вторичная структура состоит из регулярной организации спиралей и листов. Третичная структура является трехмерной нативной конформацией полипептидной цепи, как правило, включающей компактную упаковку элементов вторичной структуры. Четвертичная структура представляет собой комплекс нескольких полипептидных цепей (субъединиц), формирующих функцииональную биомолекулярную структуру. (Гемоглобин PDB ID 1HBR; ФНО-альфа человека PDB ID 5tsw).





Визуализация пространственных структур белков позволяет решать многие задачи по их структурно-функциональной организации.

Наиболее распространенными являются следующие модели визуализации.



В настоящее время секвенированы геномы более 230 организмов

В ближайшие годы ожидается секвенирование еще примерно 1000 геномов различных организмов

В базе данных SWISS-PROT содержится более 150 000 функционально аннотированных белковых последовательностей

В базе данных TREMBL содержится более 1 560 000 белковых последовательностей, требующих функциональную аннотацию

Задачи компьютерной протеомики и функциональной геномики

Функциональная аннотация белков

Биохимическая функция Клеточная функция

Молекулярный дизайн

Модифицированные белки с улучшенными медикобиологическими свойствами Модифицированные белки с приобретением новой функции Искусственные белки

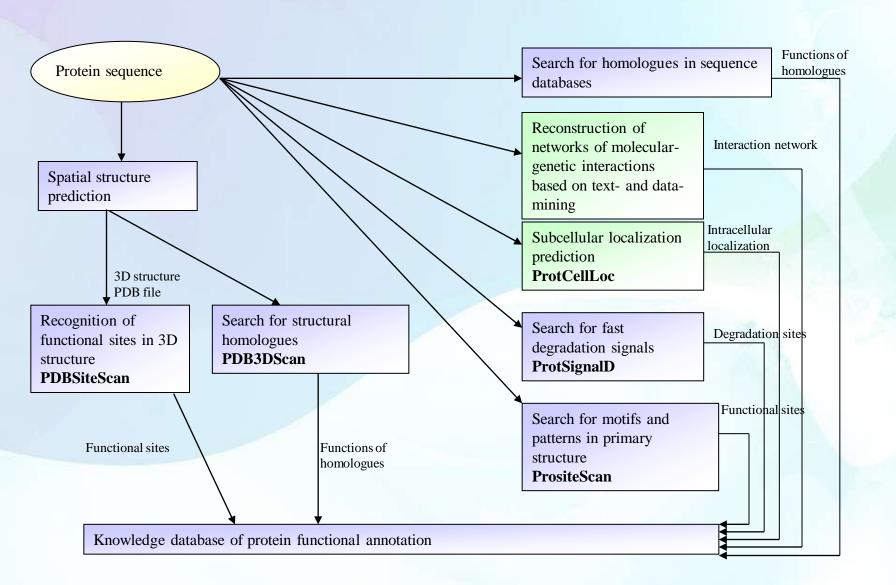
Молекулярная эволюция

Предковые последовательности Молекулярные механизмы эволюционного процесса

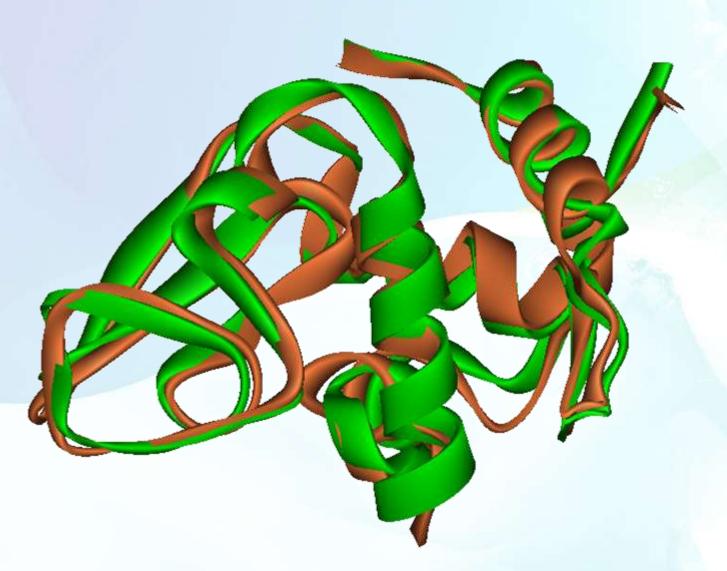
Поиск фармакологических мишеней

Блокирование функции определенных генов и белков
Возобновление поврежденной функции

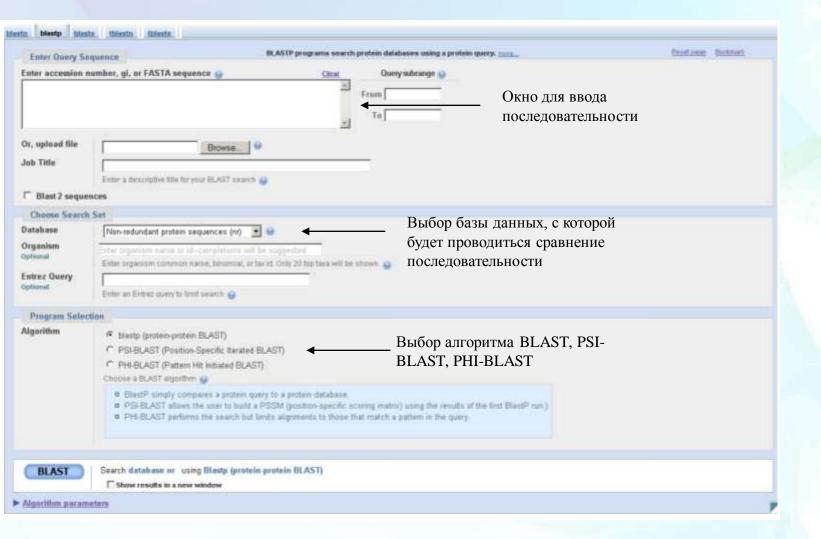
Module of protein function prediction



Структурное выравнивание



Поиск гомологии по базам данных белковых последовательностей

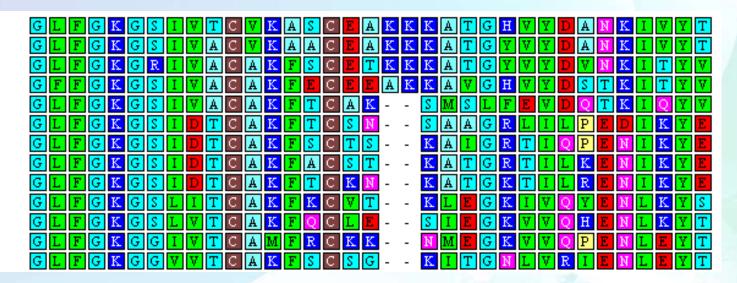


Поиск гомологии по базам данных белковых последовательностей

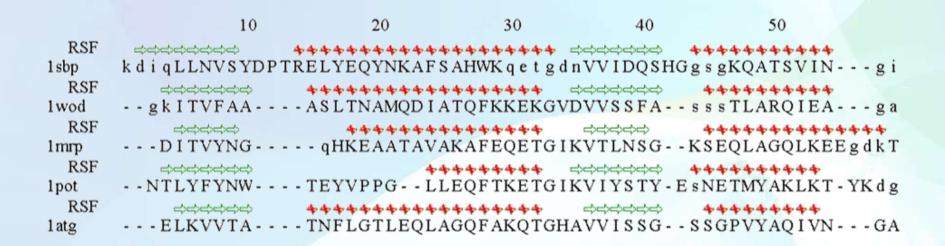
```
>prf||1402313A hemoglobin
   Length=162
    Score = 124 bits (310), Expect = 3e-27, Method: Compositional
matrix adjust.
    Identities = 68/153 (44%), Positives = 94/153 (61%), Gaps = 1/153
(0%)
   Query 3
VLTDVQVALVKSSFEEFNANIPKNTHRFFTLVLEIAPGAKDLFSFLKGSSEVPQNNPDLQ 62
              V T+ O ALV S+
                                        +FF + EIAP AK+LFS+LK S
                                                                  + NP
L+
   Sbict 9
VFTEEQEALVVKSWAVMKKNSAELGLKFFLKIFEIAPSAKNLFSYLKDSPIPLEQNPKLK 68
   Query 63
              AHAGKVFKLTYEAAIQLQVNGAVA-
SDATLKSLGSVHVSKGVVDAHFPVVKEAILKTIKE 121
               HA VF +T E+A+QL+ G V ++ LK LG++H GVV+ HF V +
A+L+TIKE
   Sbjct 69
PHAMTVFVMTCESAVQLRKAGKVTVRESNLKRLGAIHFKNGVVNEHFEVTRFALLETIKE 128
   Query 122 VVGDKWSEELNTAWTIAYDELAIIIKKEMKDAA 154
               V + WS E+ AW AYD+L
                                      IK EMK ++
   Sbict 129 AVPEMWSPEMKNAWGEAYDOLVAAIKSEMKPSS 161
```

Множественное выравнивание последовательностей

POLG_TBEVS 010383/281-581 Q82951/281-581 POLG_POWVL/279-579 POLG_YEFV1/286-579 POLG_MVEV/293-591 009754/1-298 POLG_KUNJM/291-589 POLG_STEVM/289-587 010246/1-296 POLG_DEN3/281-574 009234/281-576 POLG_DEN4/280-575



В случае правильно выравненных аминокислотных последовательностей совмещаются вторичные и третичные структуры белков.



Паттерны

Регулярные выражения, описывающие последовательность аминокислот в сайте.

KOKF HEKH FSHRGF KRTF HENH HSHRGY KVTF HYKH HSHIGM KRKF HEKH HSHSGY KRKM HEKH FSHAGF [H]-[EY]-[KN]-[H]

Профили

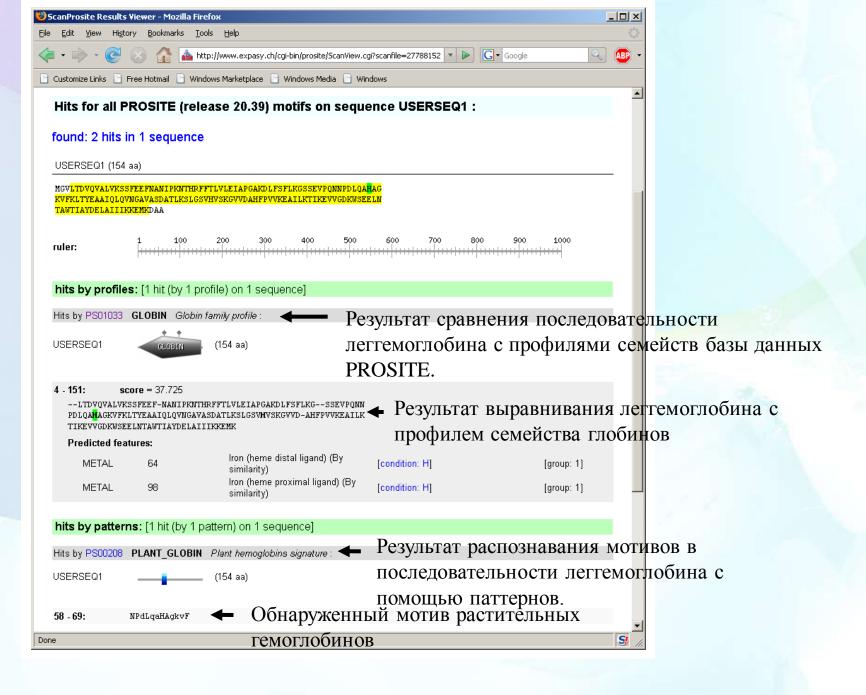
KQKFHEKHFSHRGF KRTFHENHHSHRGY KVTFHYKHHSHIGM KRKFHEKHHSHSGY KRKMHEKHFSHAGF

```
K 5.3...*******
Q .1....*******
R .3....
V .1....
T ..2...
F ...4..
M ...1..
H ....5.
E ....4
Y ....1
```

Позиционная весовая матрица или профиль выравнивания— это характеристика частот встречаемости каждой из 20-ти аминокислот в каждой из позиций множественного выравнивания анализируемого семейства белковых последовательностей

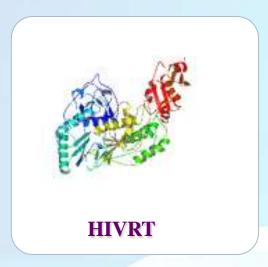


Веб-интерфейс базы данных PROSITE. В окно вводится последовательность анализируемого белка (в данном случае - леггемоглобина *Lupinus luteus*).

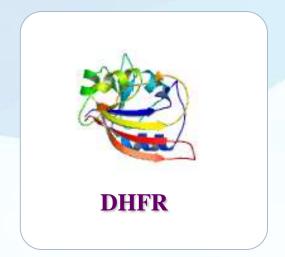


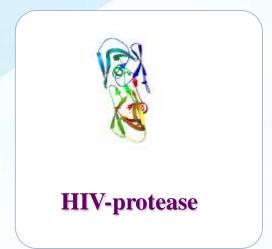
Молекулярная динамика





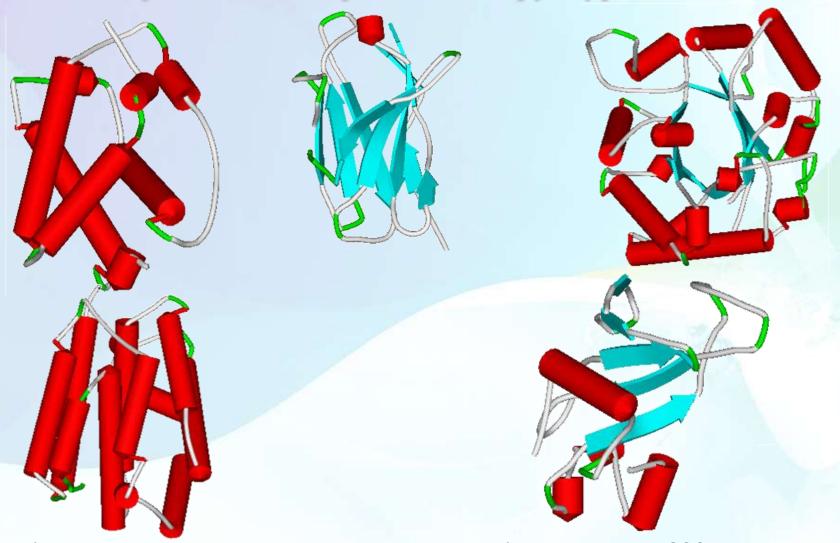








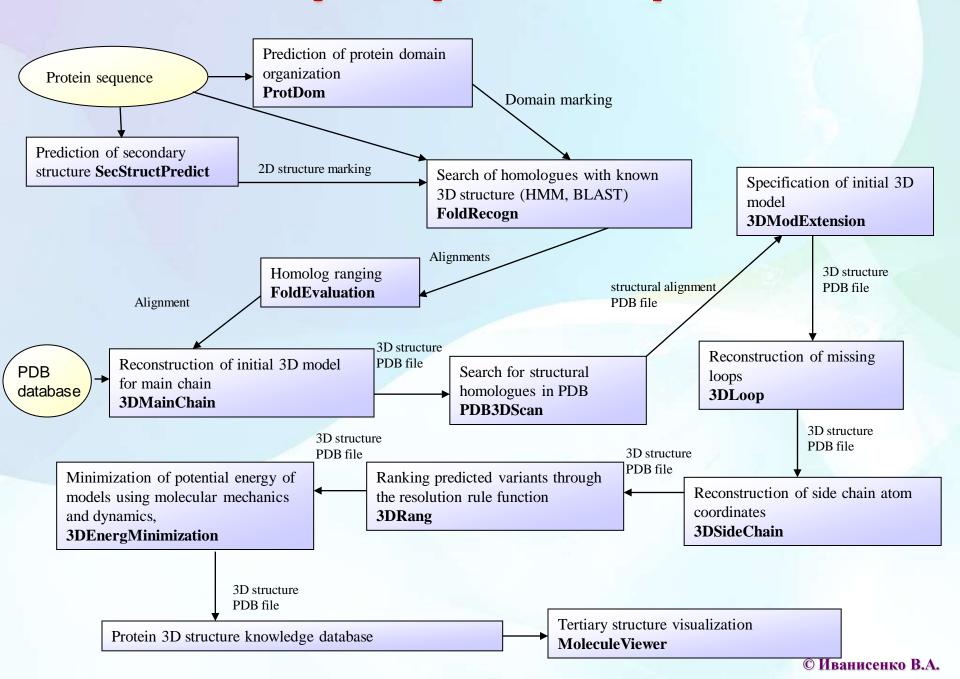
Предсказание третичной структуры белков



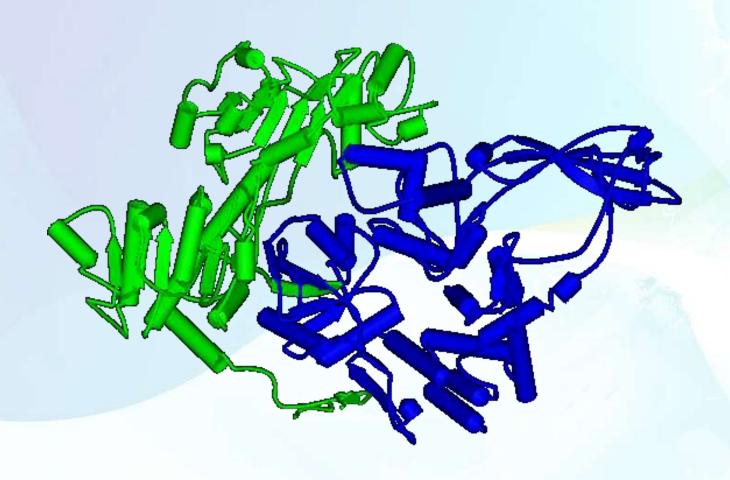
В базе данных структурной классификации 3D структур белковых доменов SCOP приведено 800 типов укладки (фолдов) из 54745 белковых доменов.

Разнообразие такое, что на 54745 доменов, представленных в 20619 белков из PDB приходится 800 типов укладок, указанных в SCOP.

Module of protein spatial structure prediction

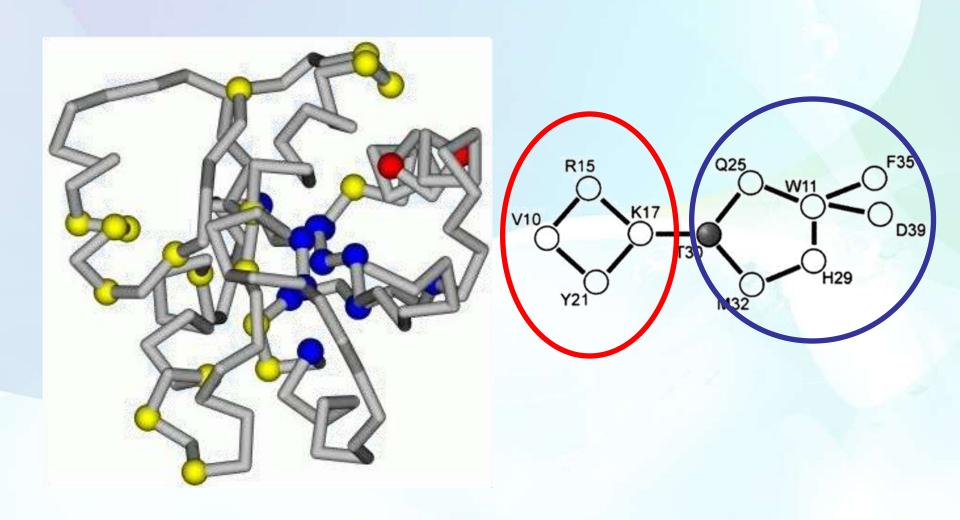


Определение доменной структуры белков



Обратная гираза (reverse gyrase) (PDB ld 1GL9). Зеленым выделен домен, выполняющий функцию геликазы, синим – топоизомеразы. Организм: ARCHAEOGLOBUS FULGIDUS

Поиск кластеров аминокислотных остатков, обладающих определенными физико-химическими или эволюционными свойствами



Инструменты компьютерной структурной биологии



База данных функциональных сайтов



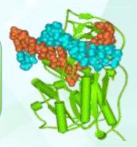
Программа распознавания функциональных сайтов



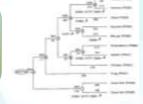
Программа реконструкции молекулярных комплексов



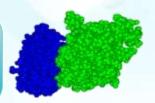
Программа анализа взаимосвязи структураактивность Функциональная аннотация белков



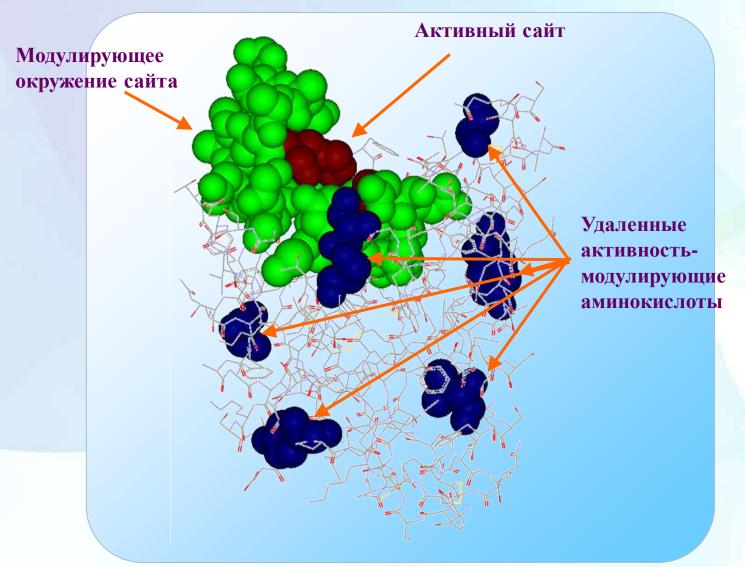
Молекулярная эволюция



Поиск фармакологических мишеней

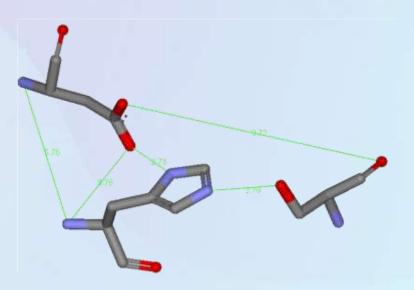


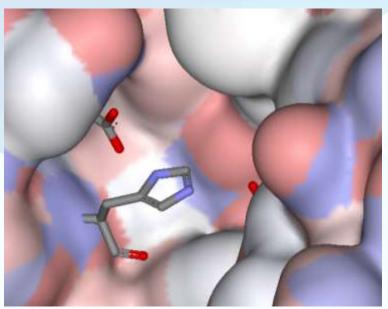
Концепция структурно-функциональной организации белков



Функциональные центры белков ключевой объект изучения в компьютерной протеомике

Физико-химические и структурные характеристики, определяющие специфичность функциональных сайтов

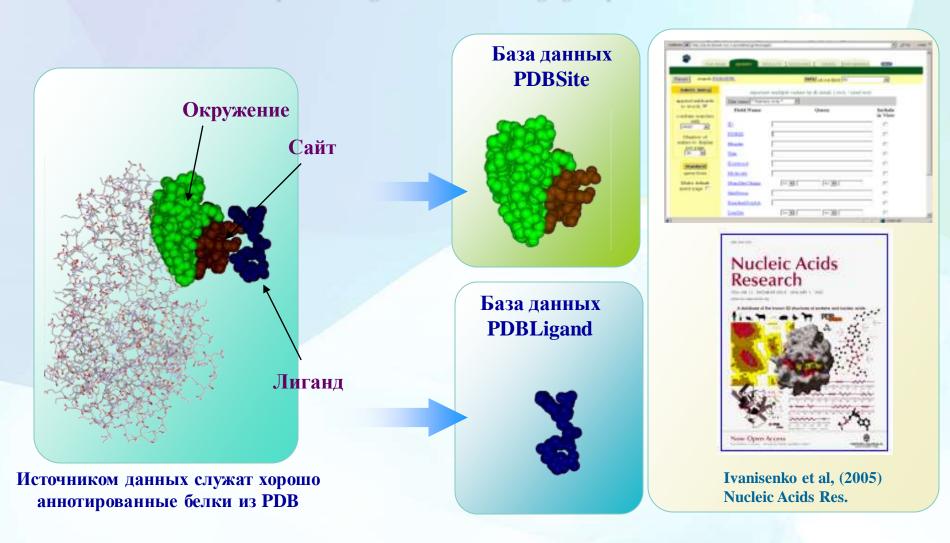




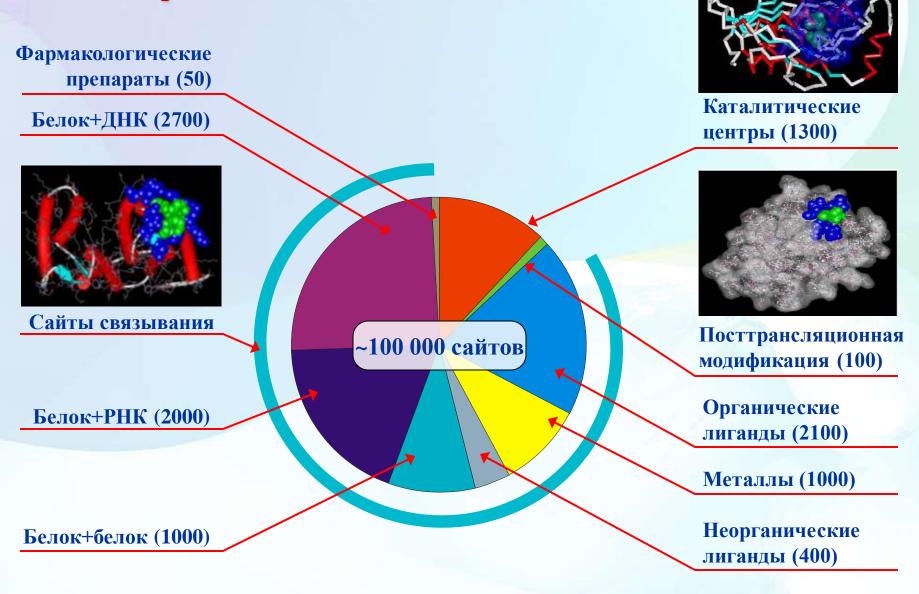
- Физико-химические свойства аминокислот сайта
- Особенности расположения аминокислот в первичной структуре
- Особенности конформации
- Вторичная структура
- Доступность для растворителя
- Полярность и заряд окружения

В базах данных PDBSite и PDBLigand накапливаются пространственные структуры и основные характеристики функциональных сайтов белков и их лигандов

http://wwwmgs.bionet.nsc.ru/mgs/gnw/pdbsite/



Содержание базы данных PDBSite



Gene Ontology классификация белков, функциональные сайты которых представлены в базе PDBSite

Молекулярные функции 951 Биологические процессы 598 Клеточные компоненты 165

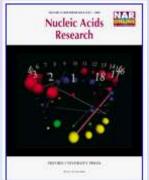
Распределение сайтов по организмам

Eukaryota – 13385

Bacteria - 7824

Archaea -1743

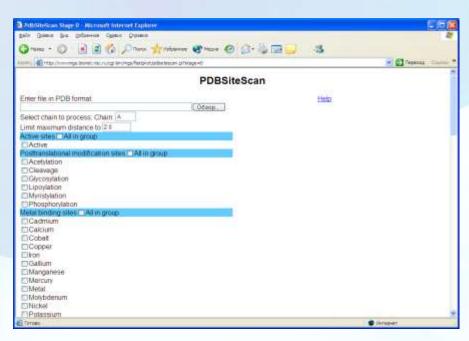
Viruses − 1041



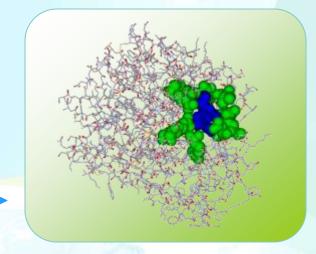
Ivanisenko et al, (2004) Nucleic Acids Res.

PDBSiteScan: программа распознавания функциональных сайтов

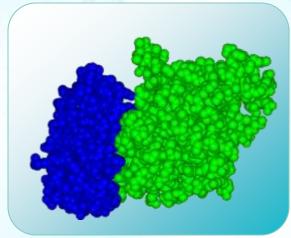
http://wwwmgs.bionet.nsc.ru/mgs/systems/fastprot/pdbsitescan.html



Распознавания функциональных сайтов

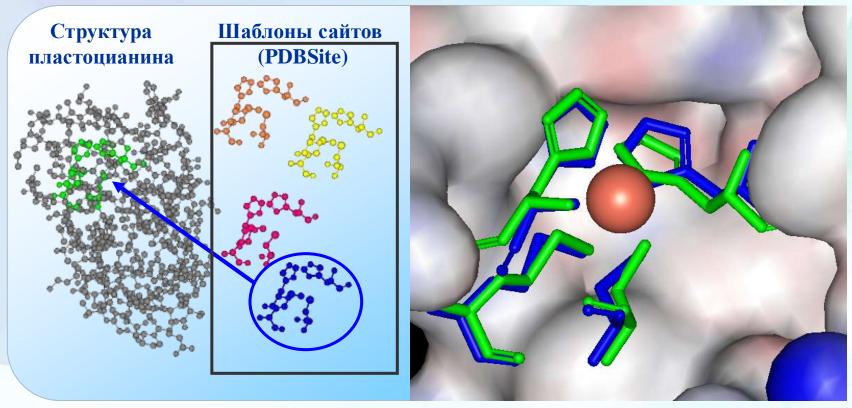


Реконструкция молекулярных комплексов сайт-лиганд



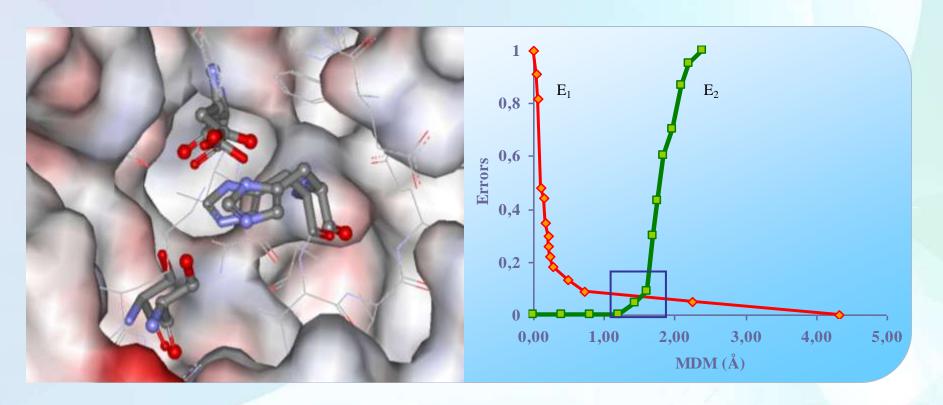
Программа поиска активных сайтов в структурах белков: PDBSiteScan

Поиск сайта связывания меди в молекуле пластоцианина (PDB ID 1BXU).



Зеленым цветом изображены остатки распознанного сайта в пластоцианине, синим цветом изображены остатки сайта-шаблона из базы PDBSite (ID 1B3ICU). Ион меди показан оранжевым шариком.

Анализ точности распознавания каталитических центров в суперсемействе гидролаз

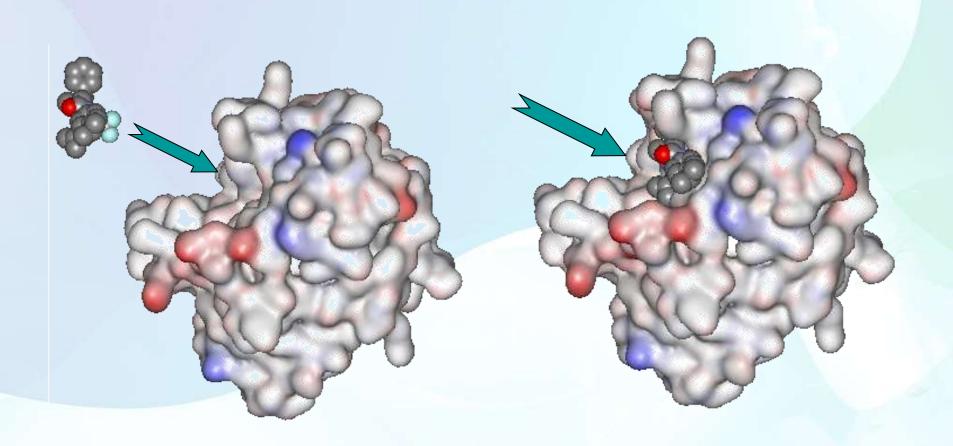


MDM – мера качества совмещения в пространстве шаблона активного сайта из базы PDBSite и фрагмента белка.

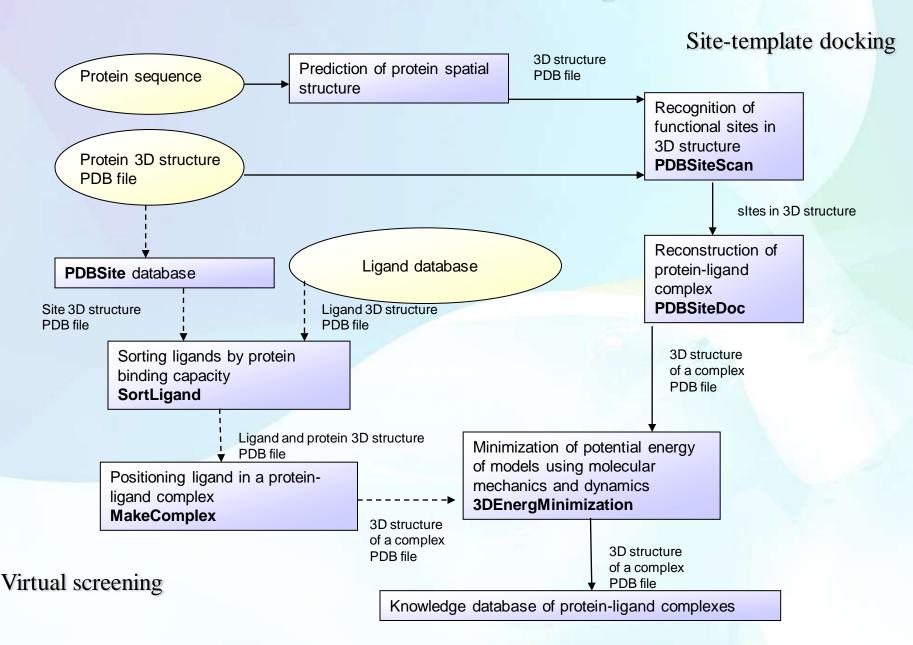
Е₁ – ошибка І-рода (недопредсказания)

Е₂ – ошибка ІІ-рода (перепредсказания)

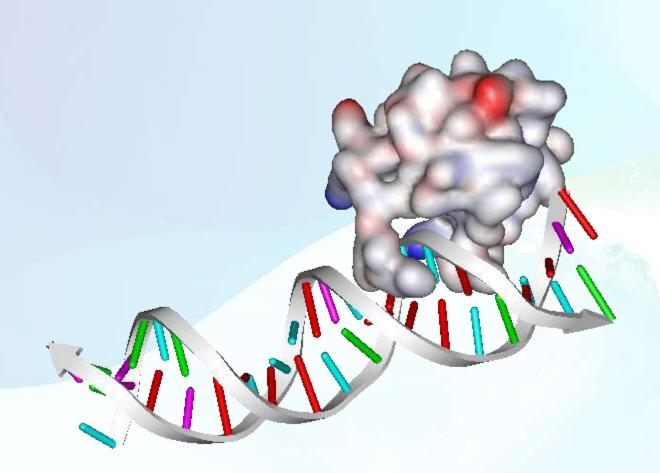
Докинг белков с малыми молекулами



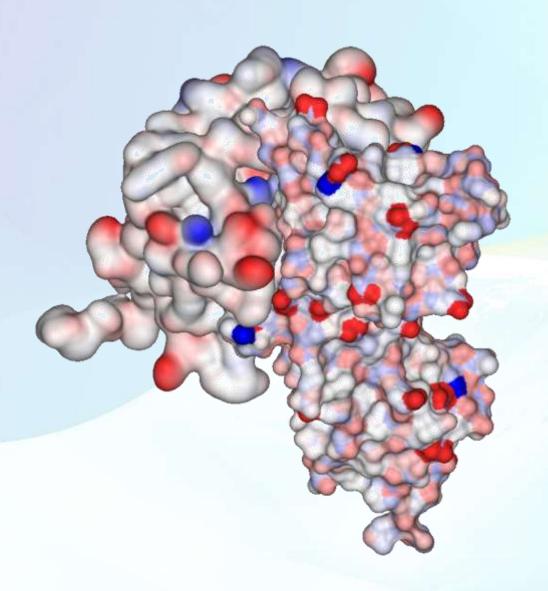
Module molecular docking



Докинг белков с ДНК

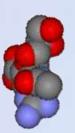


Белок-белковые взаимодействия



Site-template docking

PDBLigand

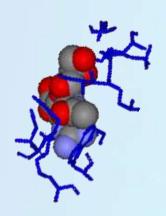


PDBSite

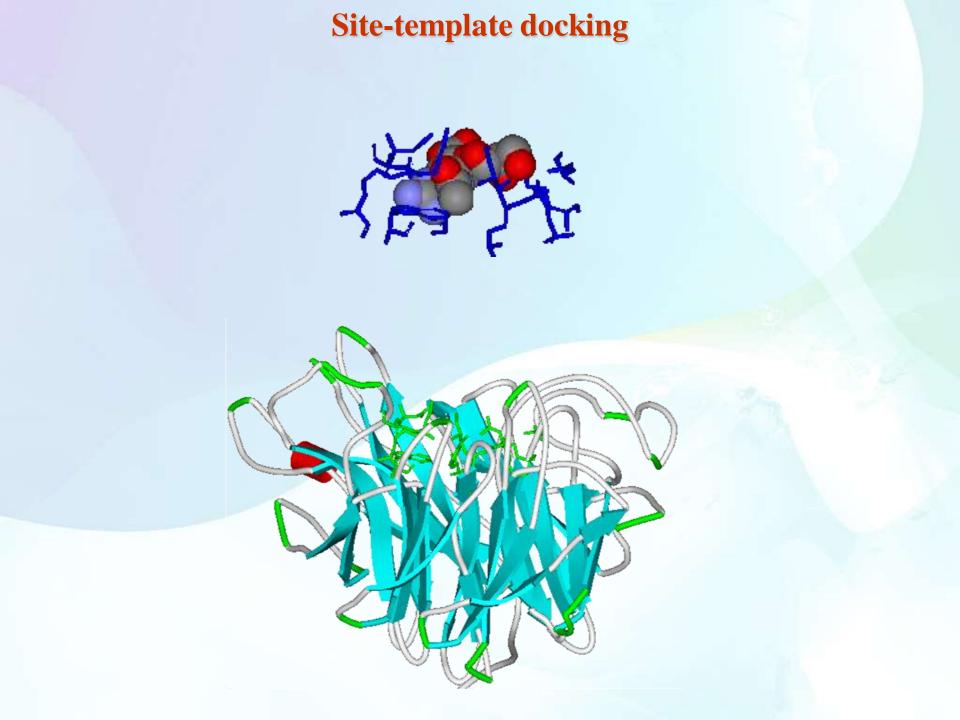




Site-ligand complex



Site-template docking



Site-template docking

Компьютерно-экспериментальное исследование структурнофункциональной организации белка PII Mycobacterium tuberculosis

Белок PII Mycobacterium tuberculosis перспективная мишень для действия противотуберкулезных лекарственных препаратов

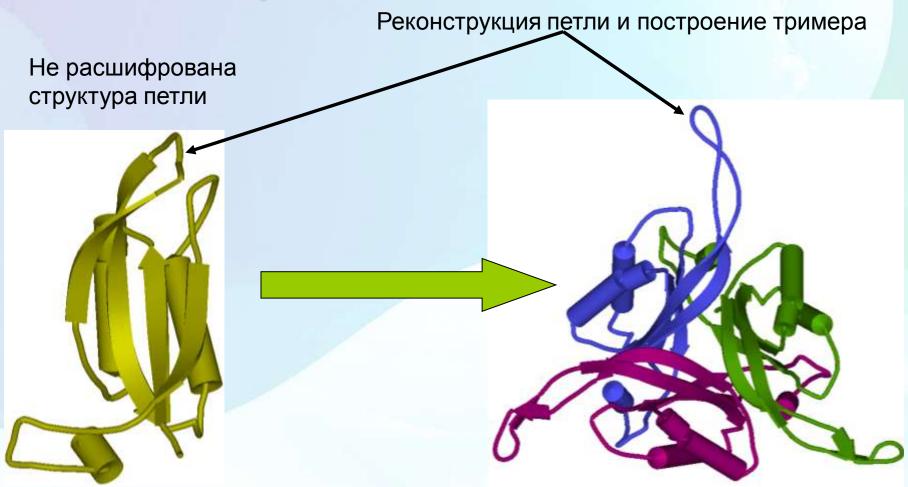
Белок РП осуществляет транспорт веществ через мембрану, участвует в регуляции метаболизма азота и углерода.

Однако механизмы его молекулярных функций в Mycobacterium tuberculosis до сих пор остаются плохо изученными.

Регуляция белком РІІ глутамин-синтетазы в *E. coli*



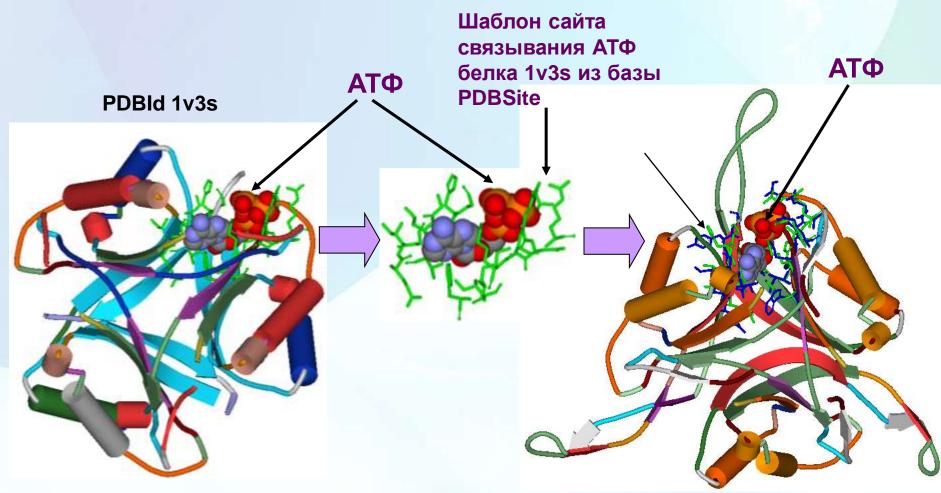
Моделирование пространственной структуры PII Mycobacterium tuberculosis



3d структура PII *Mycobacterium tuberculosis*, представленная в базе данных PDB (3bzq).

Реконструированная модель полной структуры тримера белка PII, с восстановленными петлями © Иванисенко В.А.

Распознавание сайтов связывания нуклеотидов (АТФ, АДФ, ГТФ, ЦТФ, УТФ) в пространственной структуре белка PII *Mycobacterium tuberculosis* и построение молекулярных комплексов с помощью PDBSite-шаблонов и программы PDBSiteScan



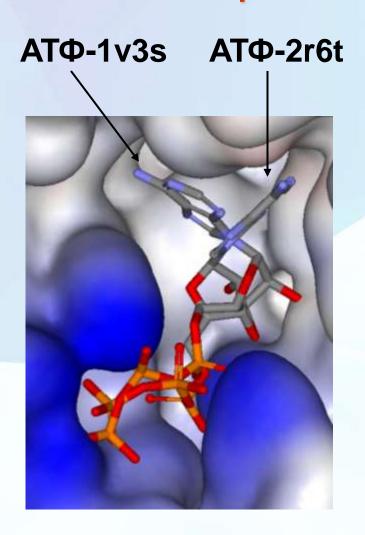
P-II
Thermus thermophilus

P-II

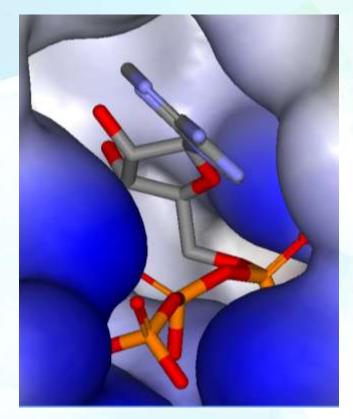
Mycobacterium tuberculosis

© Иванисенко В.А.

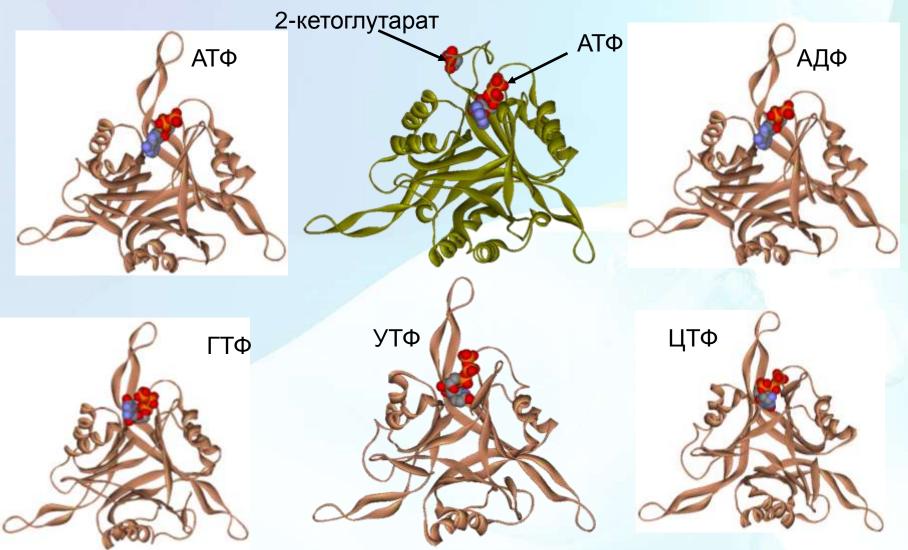
Расположение АТФ в кармане связывания белка PII *Mycobacterium tuberculosis*, полученное с использованием шаблонов АТФ-связывающих сайтов из различных белков



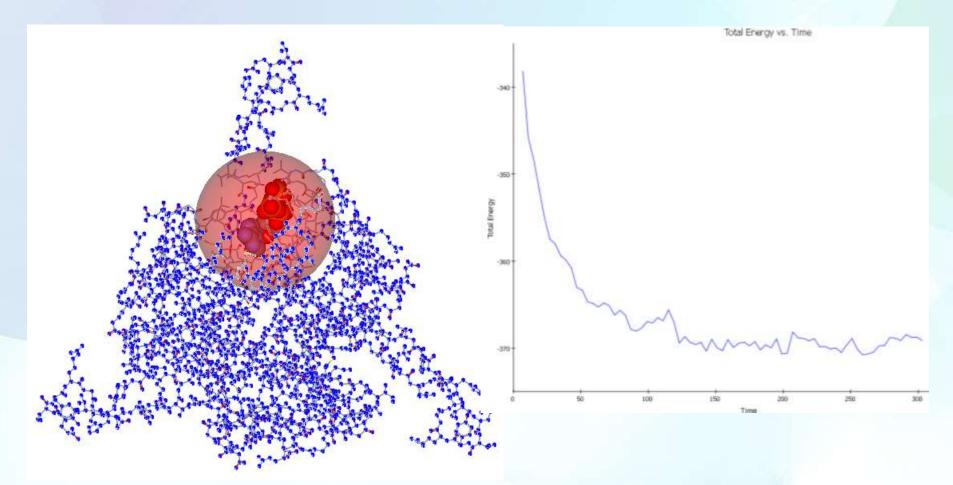
ATΦ-1r10



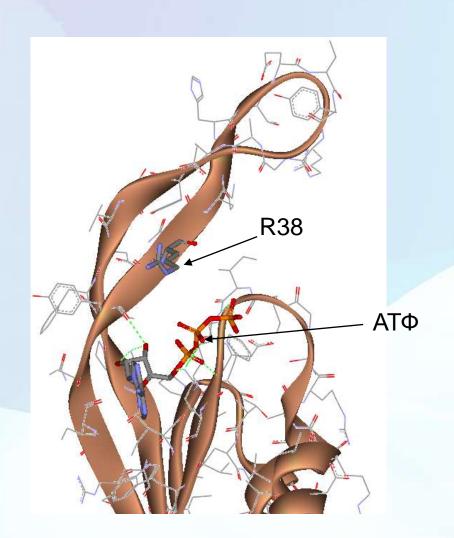
Предсказанные молекулярные взаимодействия белка PII *Mycobacterium tuberculosis* с ATФ, ATФ+2-кетоглутарат, АДФ, ГТФ, УТФ, ЦТФ



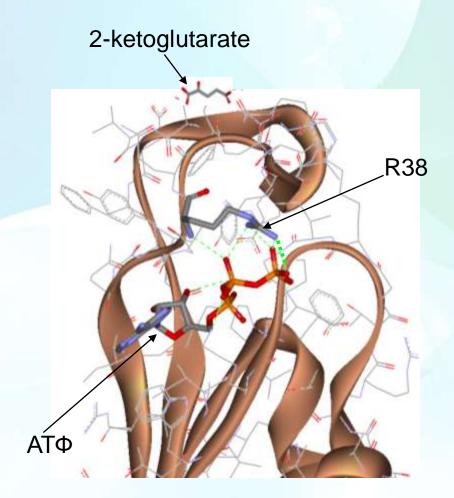
Оптимизация пространственной структуры построенных комплексов PII/лиганд с использованием молекулярной динамики с помощью пакета программ Accelrys discovery studio



Моделирование эффекта аллостерической регуляции 2кетоглутаратом взаимодействий PII с АТФ

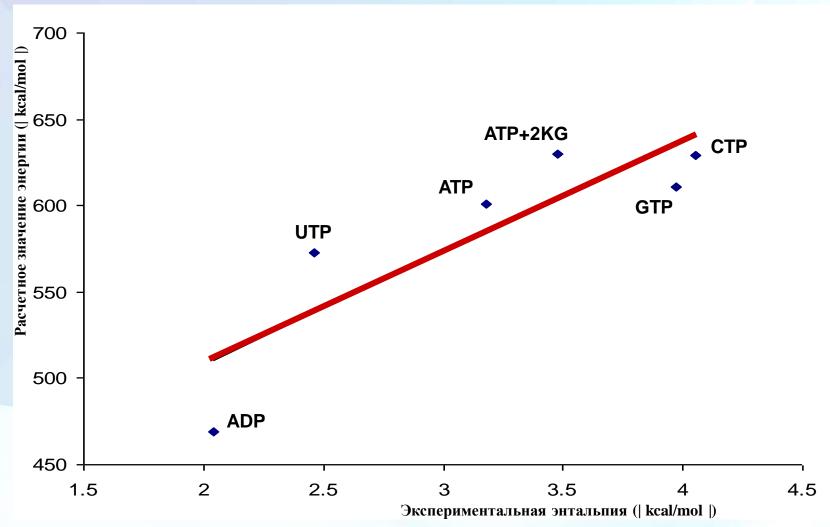


Комплекс РІІ-АТФ



Двойной комплекс PII-2-кетоглутарат-АТФ

Сравнение экспериментально измеренных и теоретически рассчитанных энергий связывания

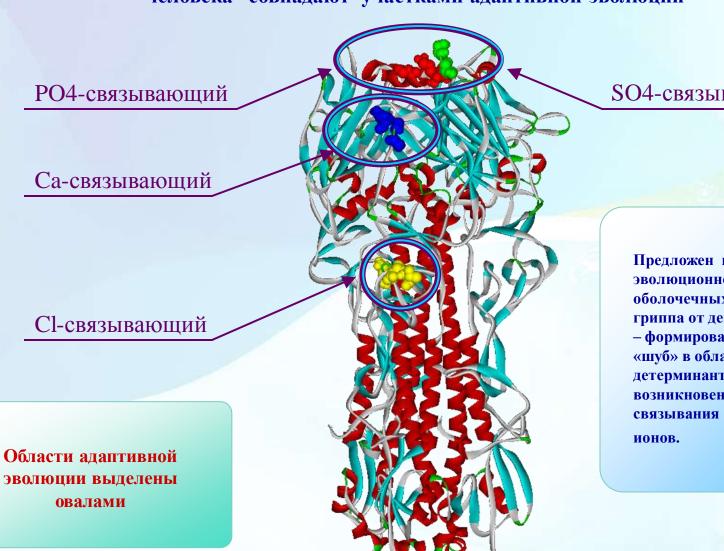


Экспериментальная проверка проводилась в IGIB Дели, Индия

Bandyopadhyay et al (2009) BMC Micr. In press.

Адаптивная эволюция: мутационное ускользание вируса от иммунной системы

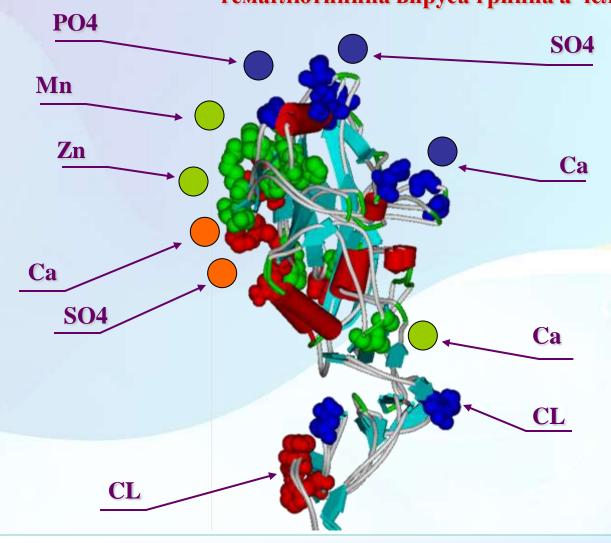
Сайты связывания ионов в пространственной структуре гемагглютинина вируса гриппа А человека совпадают участками адаптивной эволюции



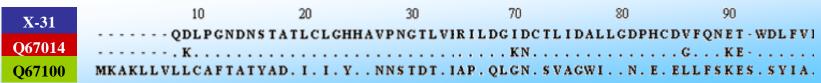
SO4-связывающий

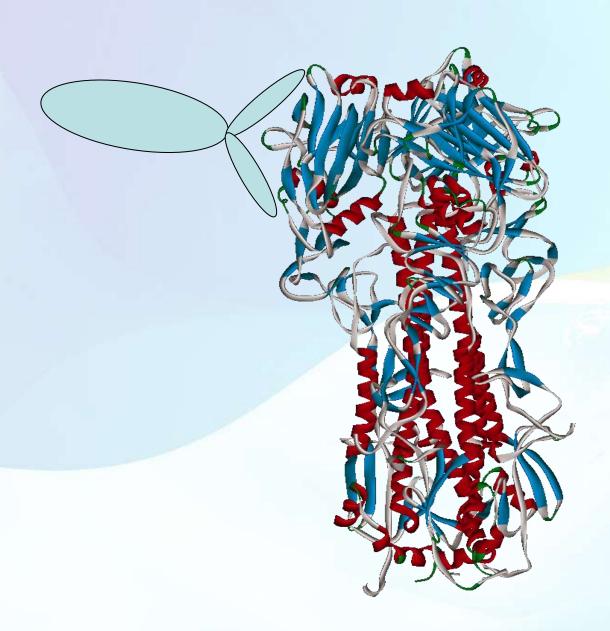
Предложен новый механизм эволюционной защиты оболочечных белков вируса гриппа от действия антител - формирование ионных «шуб» в области антигенных детерминант, обусловленное возникновением сайтов связывания различных

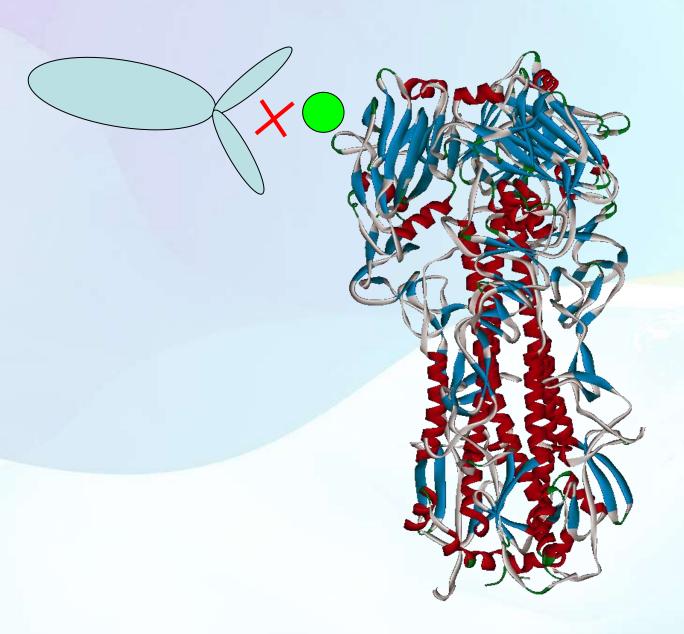
Адаптивная эволюция: мутационное ускользание вируса от иммунной системы. Полиморфизм сайтов связывания ионов в пространственной структуре гемаглютинина вируса гриппа а человека

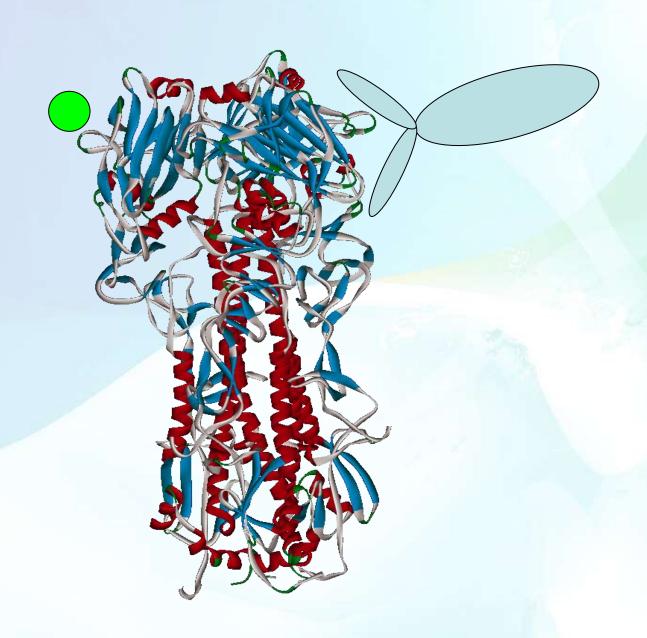


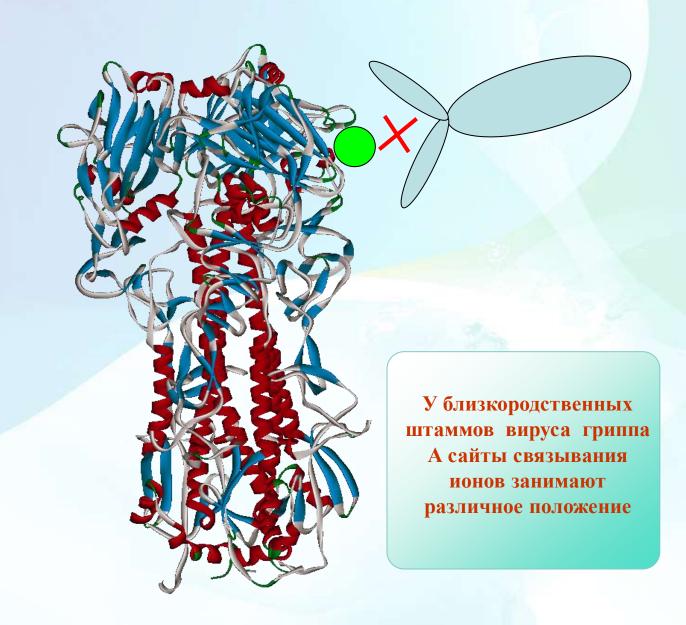
В пространственной структуре гемагглютининов близкородственных штаммов вируса гриппа А сайты связывания ионов занимают различное положение, что может быть связано с изменением направления атаки имунной системы

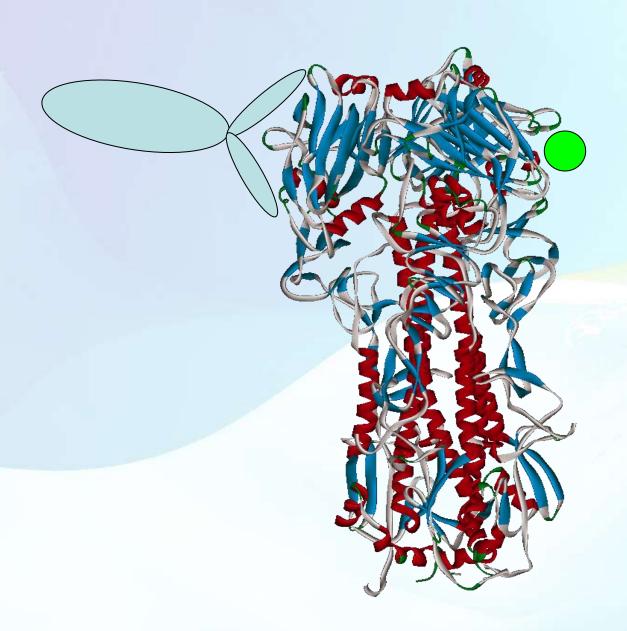


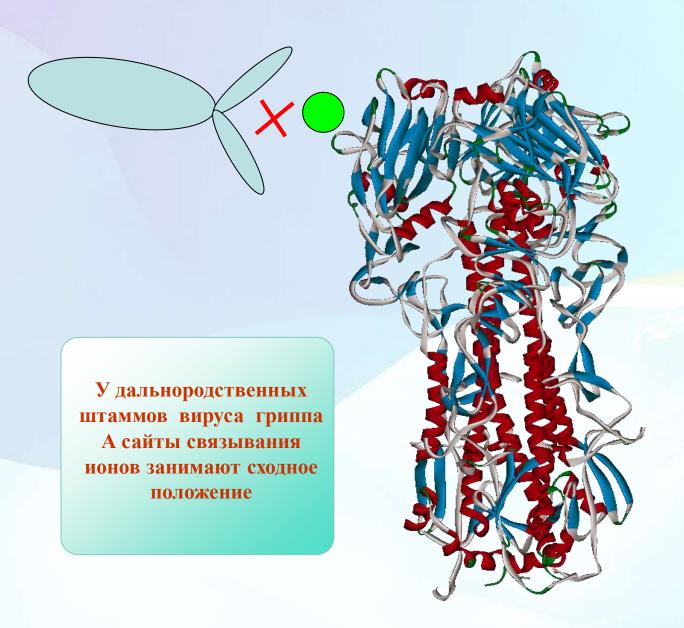






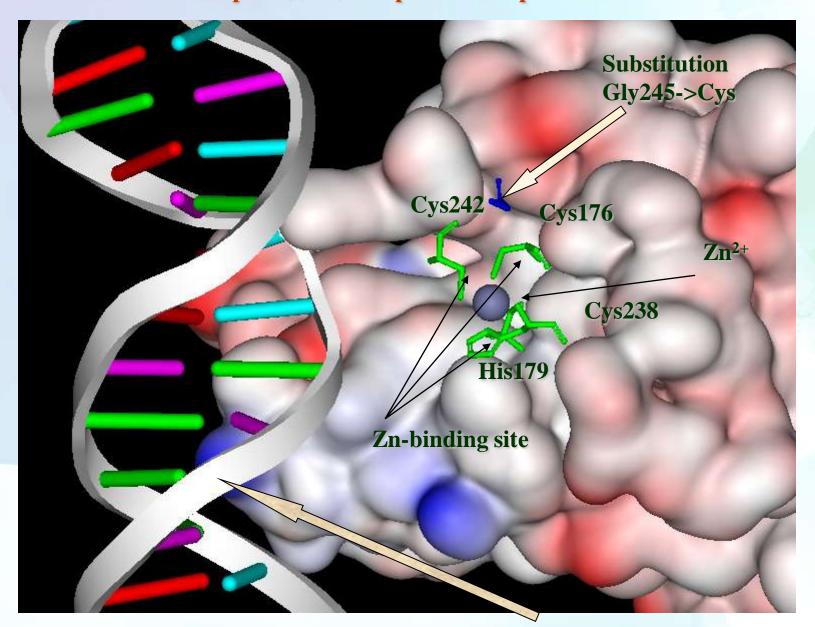




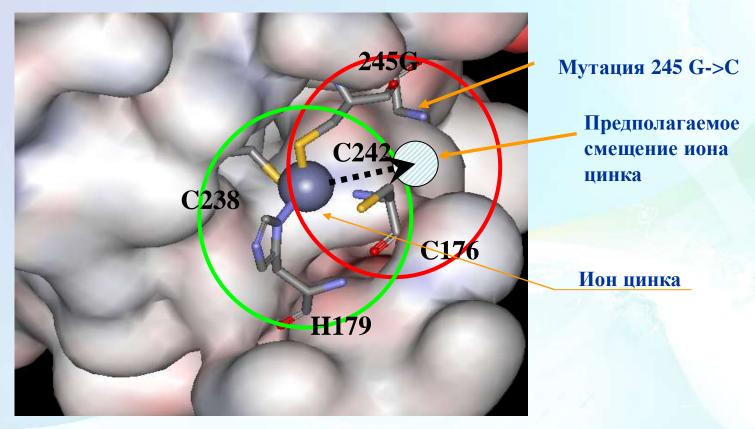


Молекулярные механизмы нарушения функции мутантных белков

Молекулярные механизмы нарушения функции мутантного белка р53, приводящие к развитию рака



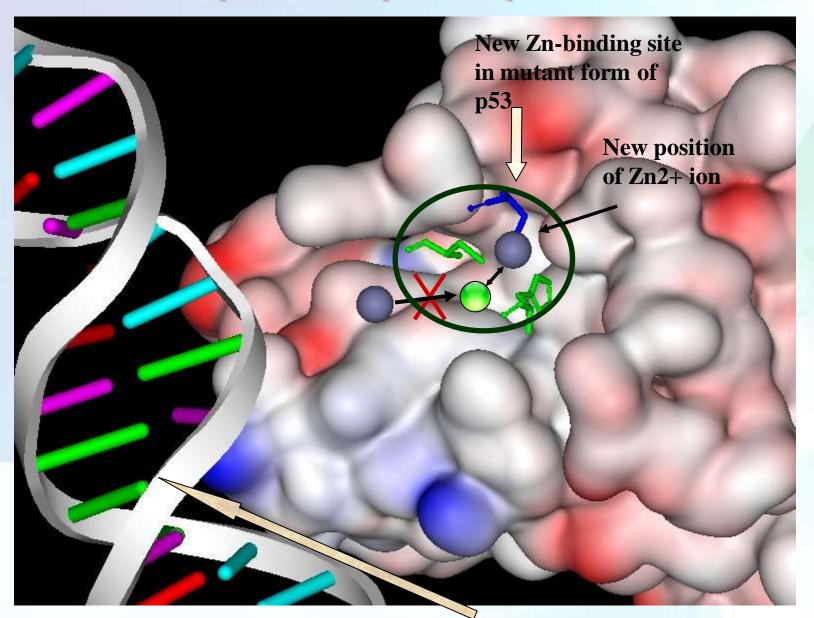
Молекулярные механизмы нарушения функции мутантного белка р53, приводящие к развитию рака



- Зеленым обведены аминокислоты реального сайта связывания Zn²⁺
- **Красным обведены аминокислоты потенциального сайта связывания Zn**²⁺

Структурный анализ мутации Gly245→Cys в ДНК-связывающем домене белка p53, вызывающую наследственную предрасположенность к раку (синдром Ли-Фраумени).

Молекулярные механизмы нарушения функции мутантного белка p53, приводящие к развитию рака



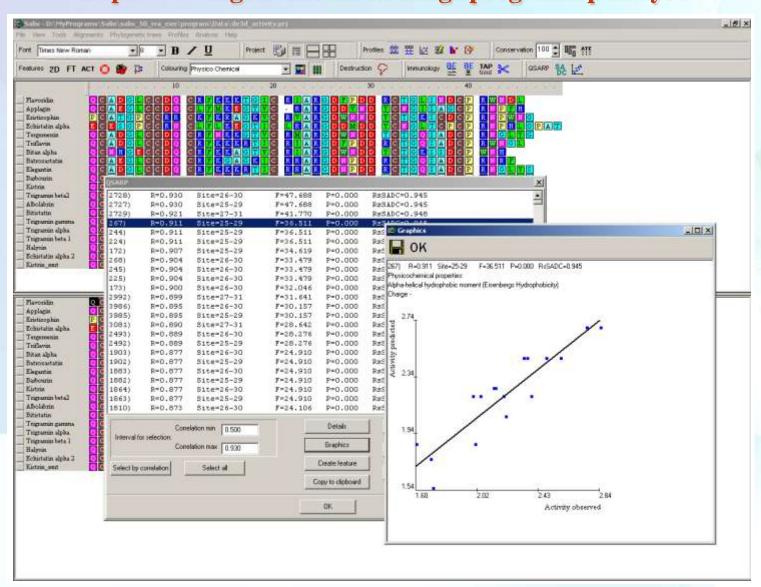
Количественный анализ взаимосвязи структура-активность в семействах гомологичных белков

- Выявление аминокислотных остатков, влияющих на активность белков — активность модулирующих центров.
- Установление зависимости между физико-химическими свойствами активность модулирующих центров и белковой активностью.

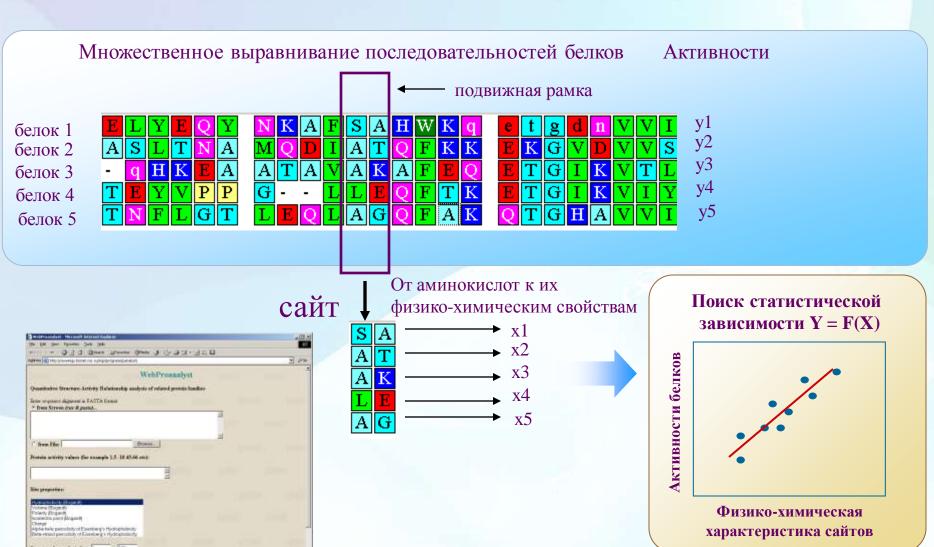
Задачи:

Планирование белково-инженерных экспериментов. Характеризация активностей белков в базах данных белковых последовательностей. Характеризация активностей искуственных белков.

WebProAnalyst, Salix: программы для количественного анализа взаимосвязи структура-активность в семействах гомологичных белков http://wwwmgs.bionet.nsc.ru/mgs/programs/panalyst/



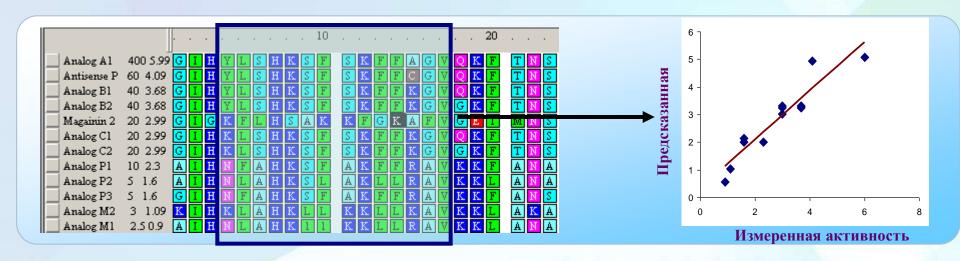
WebProAnalyst: программа для количественного анализа взаимосвязи структура-активность в семействах гомологичных белков http://wwwmgs.bionet.nsc.ru/mgs/programs/panalyst/



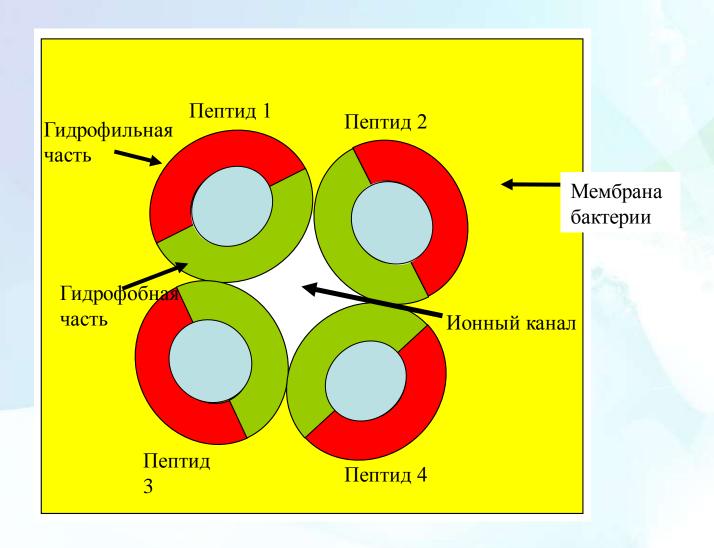
Milde window house

Количественные зависимости структура-активность в белковых семействах

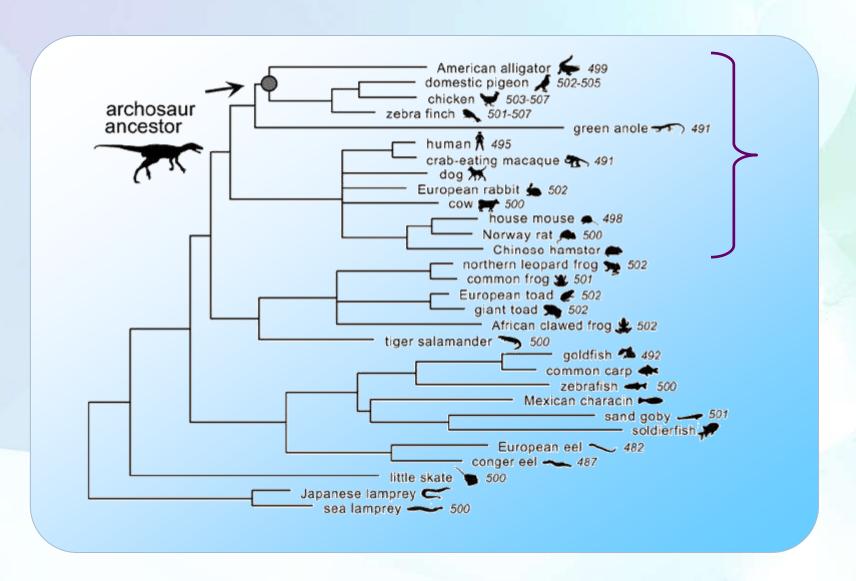
Зависимость между антимикробной активностью пептидов и альфаспиральным моментом гидрофобности сайта 3-18



Ионный канал в мембране бактерии, образованный 4 пептидами



Предсказание видимого спектра для зрения археозавра (\(\text{\chinax}\) предкового родопсина археозавра)



Предсказание \(\lambda \) тах для родопсина археозавра

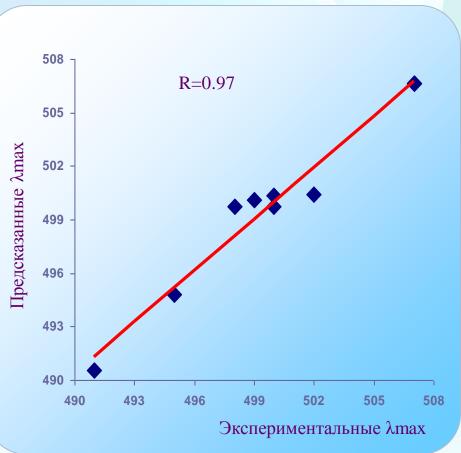
Остатки влияющие на сдвиг длины волны, результаты согласуются с данными других авторов Briscoe, Mol. Biol. Evol. (2001).

Уравнение регрессии:

Y=15.784*X₁-467.266*X₂-37.661

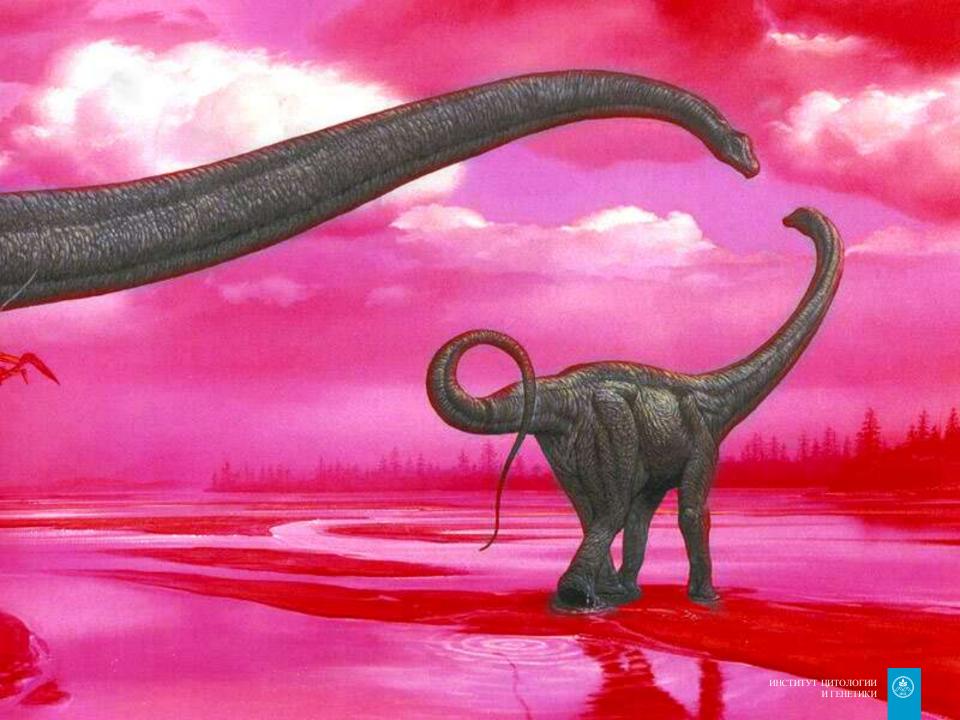
X₁ - Среднее значение для изоэлектрической точки (Bogard)

X₂ – Момент гидрофобности (Eisenberg)



 λ max предсказанное : **Y**= **506.7**

 λ max измеренное: Y = 508



Коэффициент корреляции структура-активность для идеальной регрессионной модели

Анализируемый сайт			Активности			
Группа 1	ADERKSP ADERKSP	A A	ADERKSP ADERKSP			$\sum_{i=1}^{m} \sum_{j=1}^{ni} \left(Y_{i,j} - \overline{Y}_{i,j} \right)^{2}$
Группа 2	ADERKSP ADERKSP ADERKSP	R	ADERKSP ADERKSP ADERKSP	\mathbf{Y}_{22}	$R^2 = 1 -$	$\frac{\sum_{i=1}^{N} (x_{ij} - x_{i})}{\sum_{i=1}^{N} (x_{ij} - x_{i})^{2}}$
Группа 3	ADERKSP ADERKSP	H H	ADERKSP ADERKSP	\mathbf{Y}_{31} \mathbf{Y}_{32}		$\sum_{k=1}^{\infty} (I_k - I)$

 n_i - число белков в i-й группе,

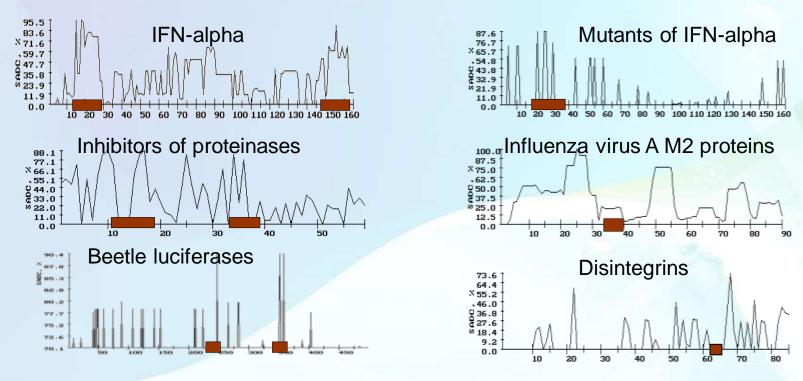
N - общее число белков,

 Y_{ij} - активность j-го белка из i-й группы,

 Y_i - среднее значение активности белков в i-й группе,

Y - общее среднее активности по всем белкам.

Профили корреляций структура-активность для различных семейств белков



обозначены известные функциональные районы

Для шести белковых семейств показано, что пики на профилях корреляций расположены вблизи функциональных районов

Молекулярная эволюция функциональных сайтов

Проблемы изучения функции белков

- 1. «Неразборчивость» при взаимодействии с другими молекулярными соединениями
- 2. Высокая полифункциональность
- 3. Часто проявляют не свойственные данному классу белков функции

Nobeli et al, (2009) Protein promiscuity and its implications for biotechnology. Nature

Структурные аналоги функциональных сайтов белков

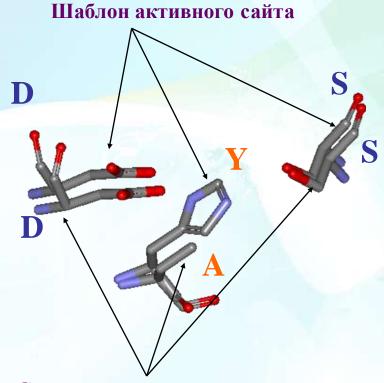
Структурный аналог функционального сайта — участок пространственной структуры белка, сходный с функциональным сайтом по конформации основной цепи, но различающийся по структуре боковых радикалов.

пост-сайт - сайт утративший функцию

пре-сайт – сайт, способный приобрести функцию в результате мутаций

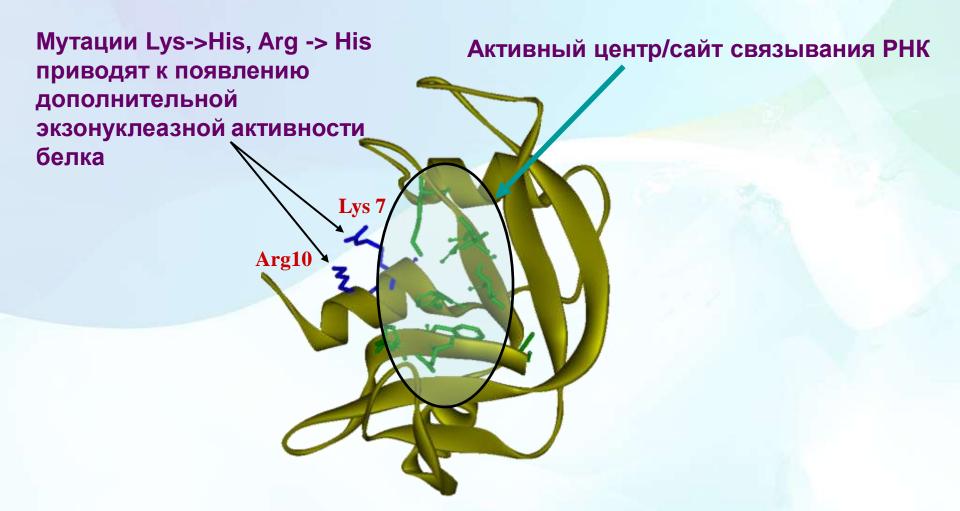
псевдо-сайт — обладает полным структурным сходством с функциональным сайтом, но не проявляет активность в полной мере

Структурные аналоги функциональных сайтов могут быть идентифицированы с помощью пространственных шаблонов

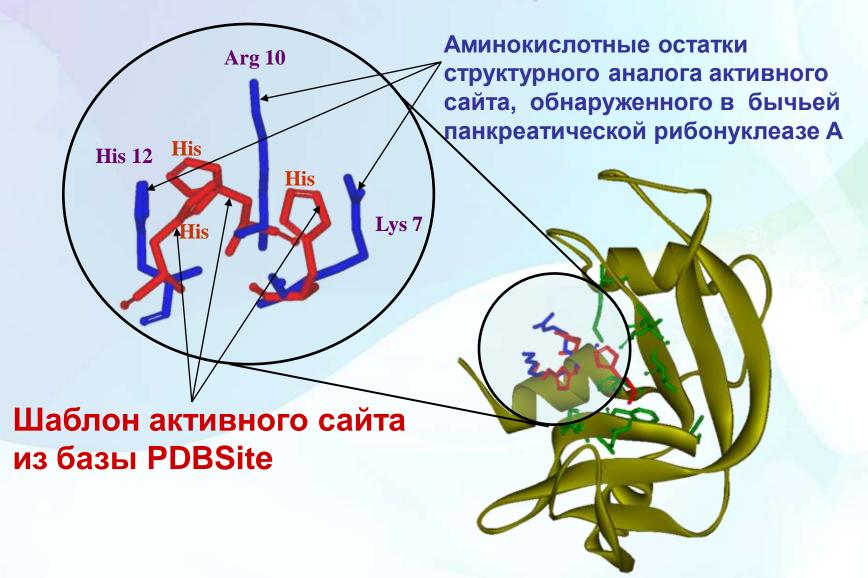


Структурный аналог активного сайта

Пример конструирования нового дополнительного активного сайта в бычьей панкреатической рибонуклеазе А путем внесения двух аминокислотных замен в (Toscano et al, 2007)



Распознавание структурного аналога рибонуклеазного активного сайта с помощью PDBSiteScan



Анализ распределения структурных аналогов сайтов в пространственной структуре белков суперсемейства гидролаз

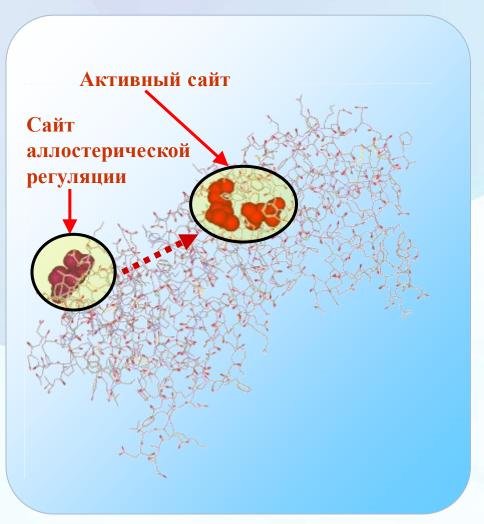


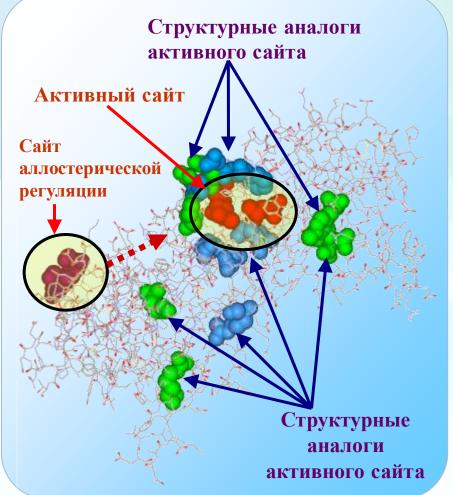
Показано, что концентрация структурных аналогов активных сайтов увеличивается при приближении к действующему активному сайту

ПРЕДЛАГАЕМАЯ КОНЦЕПЦИЯ СТРУКТУРНО-ФУНКЦИОНАЛЬНОЙ ОРГАНИЗАЦИИ БЕЛКОВ

Существующее представление

Новый взгляд





Мишени действия лекарств

РЕПЛИКАЦИЯ/ТРАНСКРИПЦИЯ ДНК

ДНК – алкилирующие агенты, minor groove binders, intercalating agents, wrong substrates (trojan horses)

ФУНКЦИИ РНК

РНК – интерференция (подавление трансляции и ускорение деградации РНК)

СТРУКТУРА И ФУНКЦИЯ БЕЛКОВ

Процесс фолдинга после трансляции - ингибиторы белкового фолдинга Процесс создания биологически активных модулей (units) для полисубъединичных белков - ингибиторы белкового фолдинга Активные центры ферментов - обратимые и необратимые ингибиторы Транспортеры - uptake inhibitors

УРОВЕНЬ ГЕННОЙ СЕТИ И ПУТЕЙ ПЕРЕДАЧИ СИГНАЛА

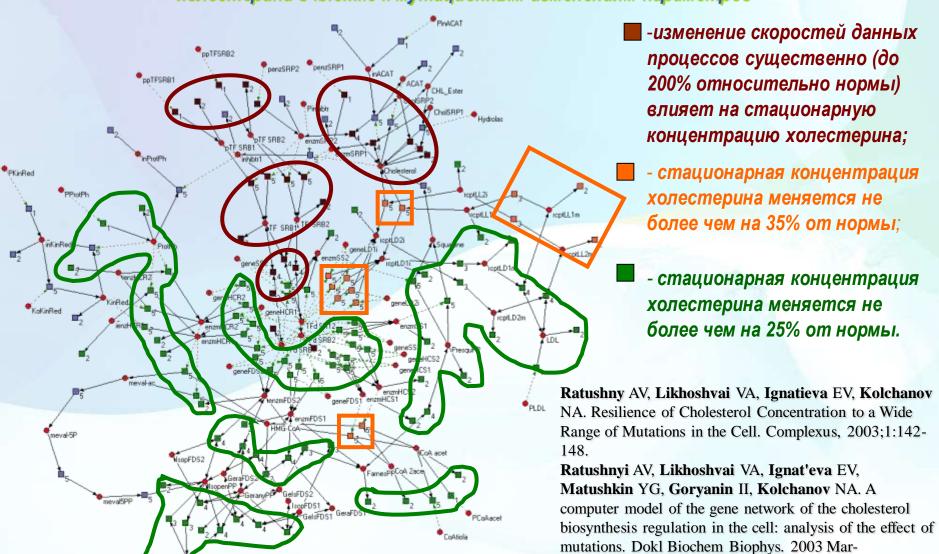
Рецепторы – агонисты и антагонисты

СУПРАМОЛЕКУЛЯРНЫЕ СТРУКТУРЫ

Ионные каналы – блокаторы и активаторы

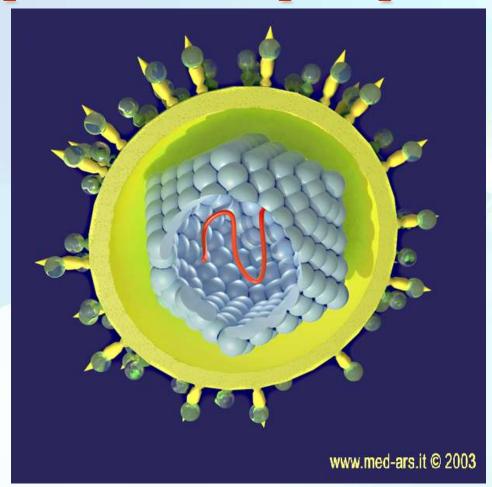
ЛИМИТИРУЮЩИЕ ПРОЦЕССЫ В ГЕННОЙ СЕТИ

регуляции биосинтеза холестерина в клетке Генная сеть с указанием чувствительности стационарного содержания свободного холестерина в клетке к мутационным изменениям параметров



Apr;389:90-3. [in Russian].

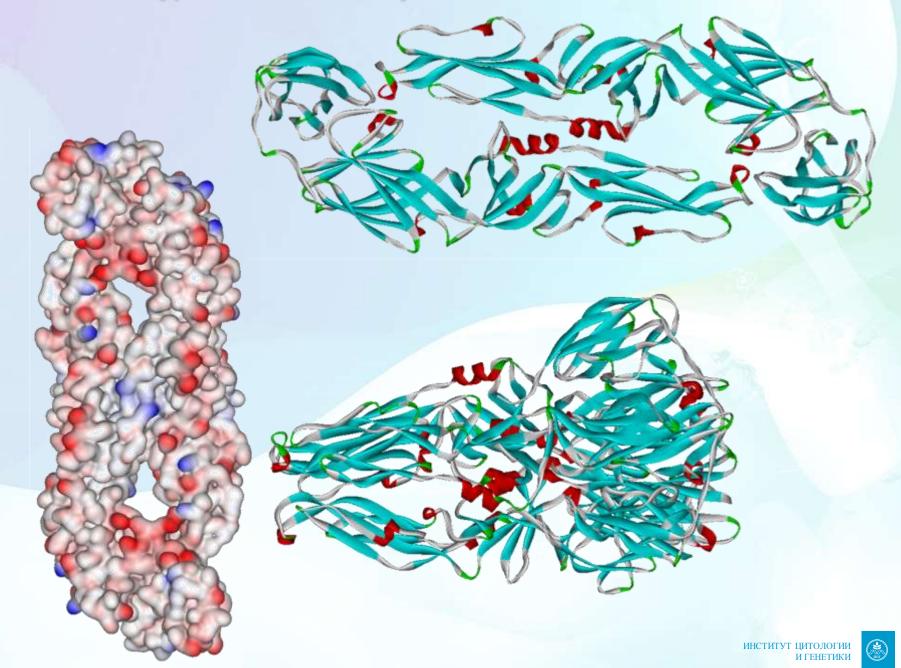
Поиск потенциальных мишеней действия противовирусных лекарственных препаратов



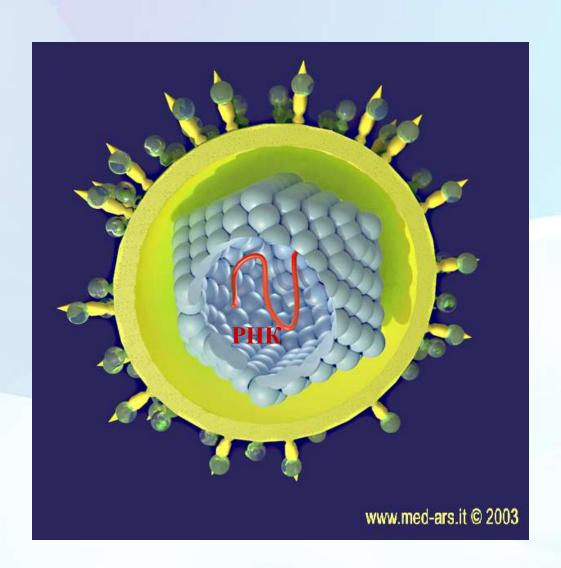
Бактериофаг Т4. Проникновение в клетку бактерии.



Вирус клещевого энцефалита. Оболочечный белок Е.

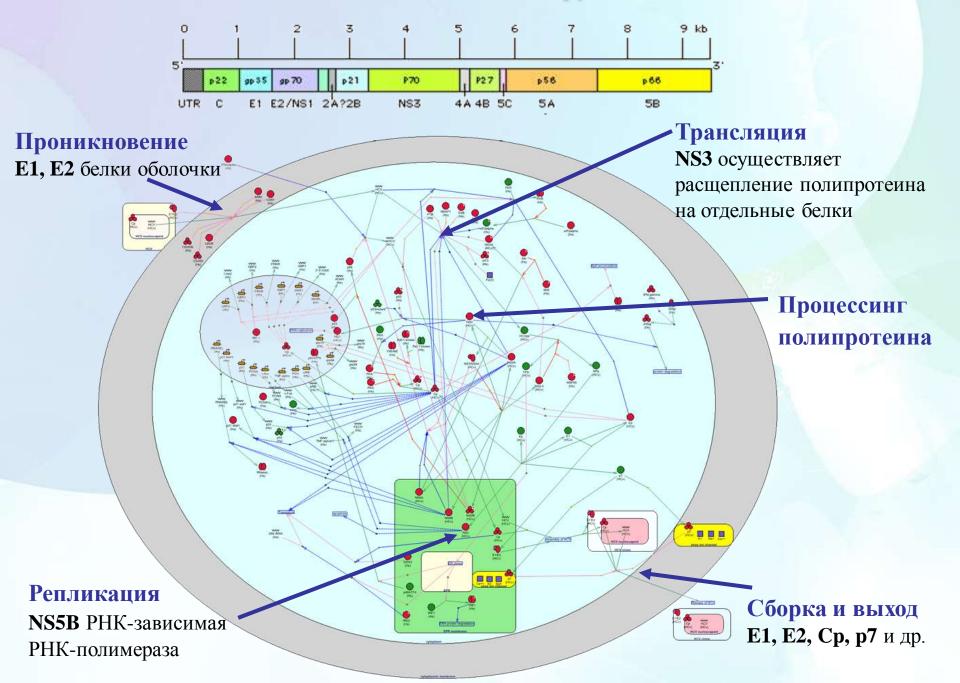


Вирус гепатита С



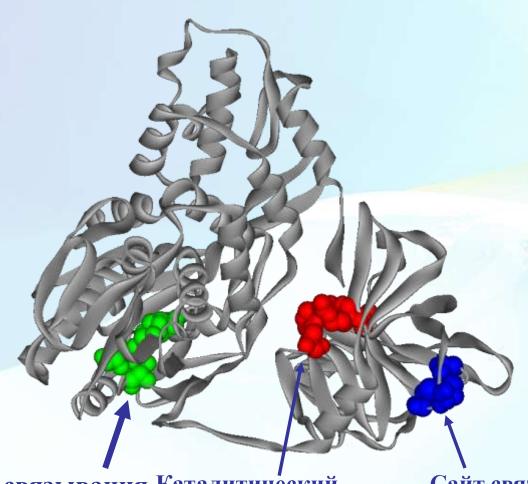
Вирус гепатита С относится к семейству Флавивирусов, род Гепацивирусов. Он попадает в организм путем, сходным с ВИЧ, и поражает печень. Хроническая форма инфекции длится 10-15 лет и вызывает цирроз печени, провоцирует рак, угнетает иммунную систему организма. В России число инфицированных оценивается в 2 млн. человек.

Генная сеть жизненного цикла вируса гепатита С



ПОТЕНЦИАЛЬНЫЕ МИШЕНИ ДЛЯ ДЕЙСТВИЯ ПРОТИВОВИРУСНЫХ ПРЕПАРАТОВ

NS3 гидролаза ВГС



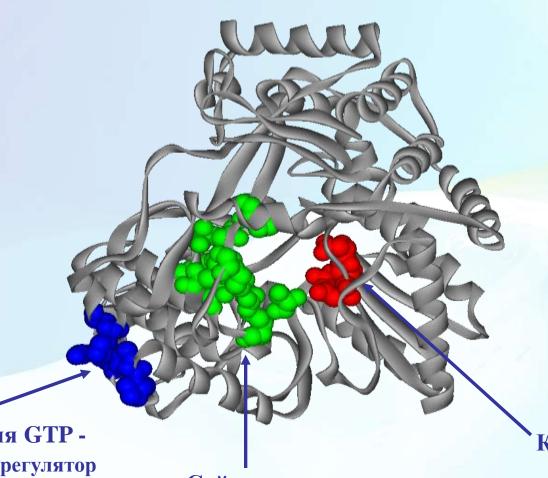
Сайт связывания Каталитический МО4 протеазный

центр

Сайт связывания цинка

ПОТЕНЦИАЛЬНЫЕ МИШЕНИ ДЛЯ ДЕЙСТВИЯ ПРОТИВОВИРУСНЫХ ПРЕПАРАТОВ

NS5В полимераза ВГС



Сайт связывания **GTP** -

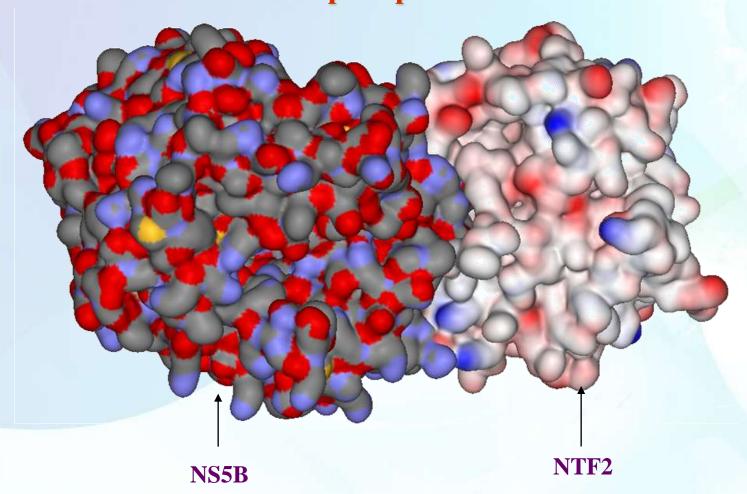
аллостерический регулятор

активности

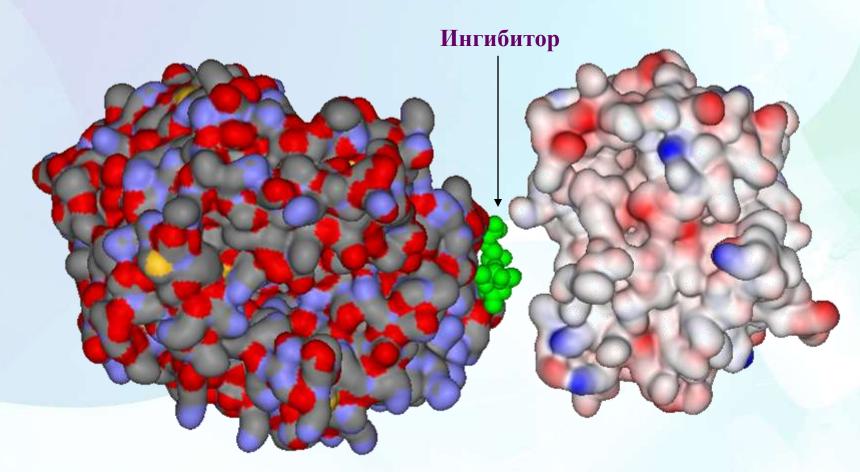
Cайт связывания rNTP (UTP, AAT, GTP, CTP)

Каталитический центр

© Иванисенко В.А.

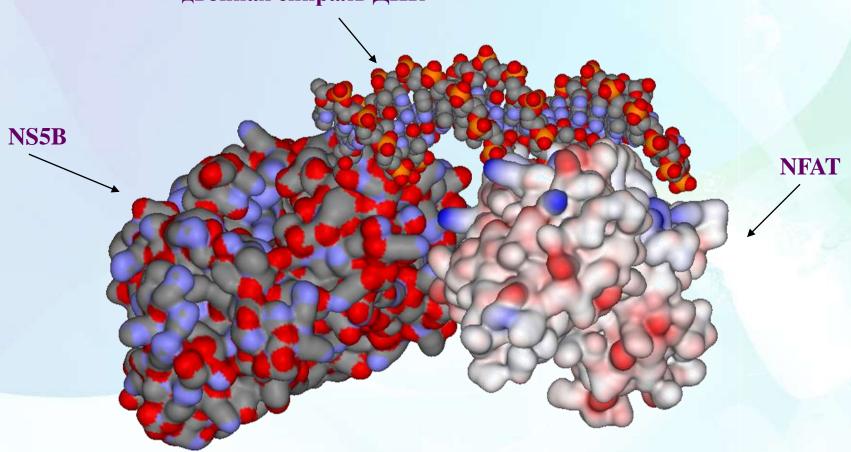


Теоретически предсказанный комплекс, образуемый белком NS5B вируса гепатита С с белком человека NTF2, осуществляющим транспорт белков в ядро клетки.

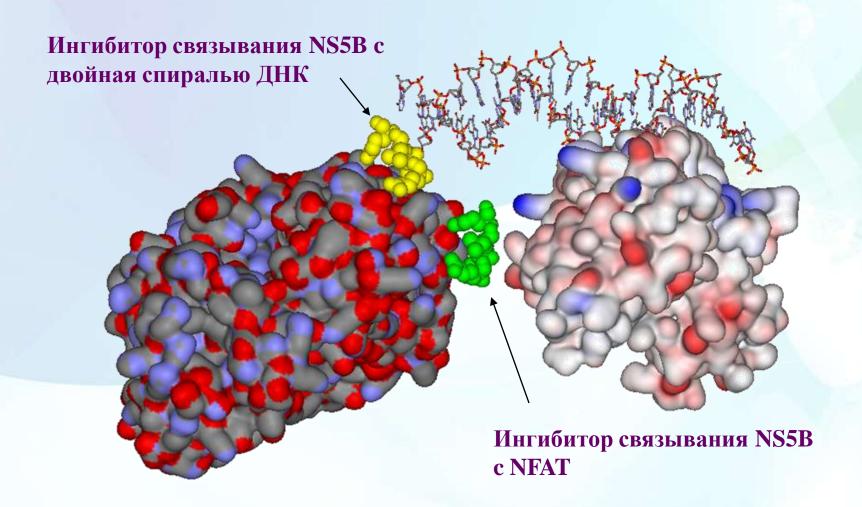


Ингибитор образования комплекса NS5B-NTF2 будет препятствовать транспорт NS5B в ядро клетки.

двойная спираль ДНК



Теоретически предсказанный комплекс, образуемый белком NS5B вируса гепатита С, с транскрипционным фактором человека NFAT и двойной спиралью ДНК.

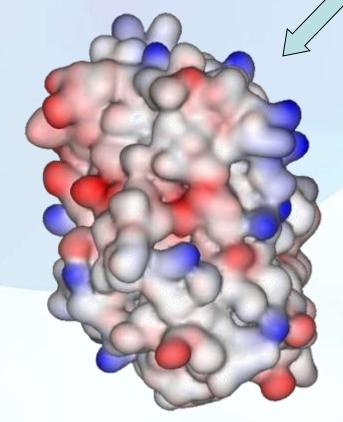


Изучение молекулярных механизмов побочных эффектов лекарственных препаратов

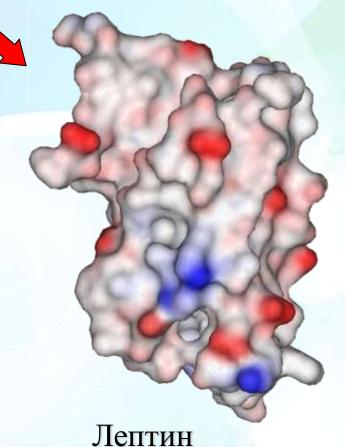
Пептидомиметик разработан для ингибирования функции циклина и используется как модель для конструирования противоопухолевых препаратов

Целевое взаимодействие с циклином человека

Побочные взаимодействия с лептином, рассчитанные компьютерными методами



Циклин А2



Побочные эффекты действия таких лекарств могут быть сопряжены с их способностью ингибировать функцию лептина





Циклин А2

Лептин