

# Анализ поверхности взаимодействия белков и поиск наиболее значительных позиций методом *in silico* Ala-scan

Магистрант:

Татьяна Малыгина, СПбАУ

Научный руководитель:

Павел Яковлев, BIOCAD

СПбАУ, 2015

## Некоторые важные определения

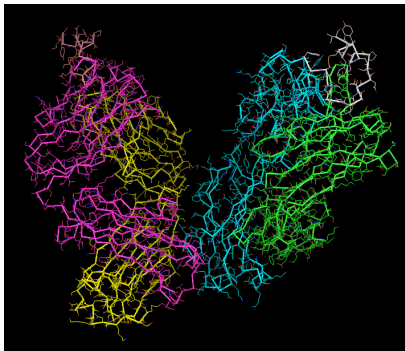
Когда белок состоит из нескольких цепей, говорят о его **четвертичной структуре**.

С точки зрения химии, разным видам структуры соответствуют разные виды химических связей и электростатических взаимодействий.

Когда мы рассматриваем несколько цепочек в составе одного белка или несколько белков, образующих комплекс, мы говорим о **белок-белковом взаимодействии**.

**Интерфейс** такого взаимодействия – это участки **поверхности** белков, непосредственно контактирующие между собой.

# Белок-белковое взаимодействие - I



Рассмотрим белок, имеющий четвертичную структуру.

**Вопрос:** можно ли изменить что-то в его структуре, чтобы образующие его цепочки были лучше сцеплены между собой?

# Белок-белковое взаимодействие - II

Пусть есть комплекс из двух белков (например, иммуноглобулин и эпитоп).

**Вопрос 1:** можно ли изменить что-то в его структуре, чтобы усилить связь между компонентами комплекса?

**Вопрос 2:** насколько специфична одна из компонент комплекса? Можно ли подобрать один из белков так, чтобы комплекс был более устойчивым? Насколько заменяема каждая из компонент?

Ответить нам поможет **аланиновое сканирование** (аланиновый мутагенез).

# Аланиновое сканирование

Аланиновое сканирование (ала-скан)<sup>1</sup> - метод для определения аминокислот в составе белка, играющих важную роль в сохранении его функций, стабильности или формы.

Проблемы ala-scan in vitro/in vivo:

- Большое пространство поиска.
- Сложность синтеза библиотек:  
необходим индивидуальный подход!
- Высокая стоимость.

	G	A	R	F	I	E	L	D
*	A	A	R	F	I	E	L	D
↓	G	A	R	F	I	E	L	D
Ala	G	A	A	F	I	E	L	D
↓	G	A	R	A	I	E	L	D
*	G	A	R	F	A	E	L	D
	G	A	R	F	I	A	L	D
	G	A	R	F	I	E	A	D
	G	A	R	F	I	E	L	A

<sup>1</sup>2001, "Combinatorial alanine-scanning" (Morrison K.L., Weiss G.A.).

# Энергетически горячие аминокислотные остатки


Энергетически горячий остаток (ЭГО)<sup>2</sup> – такая аминокислота в составе одного из компонент белок-белкового комплекса, мутагенез которой приводит существенному изменению свободной энергии комплекса  $\Delta\Delta G$ .

Существенным обычно считают изменение, превышающее по модулю 0.5-1 килокалорий на моль.

Цель аланинового сканирования – найти ЭГО.

Но для больших белков по всем аминокислотам его проводить долго, поэтому производится предварительный отбор аминокислот.

---

<sup>2</sup>в англоязычной литературе „energy hotspot residue” 

# Ala-scan in silico

Компьютерное моделирование аланинового сканирования. Постановка задачи

**На входе:** Пространственная структура белкового комплекса (в формате PDB).

**На выходе:** ЭГО.

**Как решить:** вычислить потери свободной энергии  $\Delta\Delta G$  при замене аминокислотного остатка на аланин для всех аминокислот, выбрать позиции с существенным значением потери (как правило, существенным считают изменение больше 1 килокалории на моль).



# Ala-scan in silico

Компьютерное моделирование аланинового сканирования. Методы

"Computational alanine scanning of protein-protein interfaces"(Kortemme, et al. - 2004)

"Computational Alanine Scanning Mutagenesis - An Improved Methodological Approach"(I.S. Moreira, et al. - 2006)

Готовые решения используют (на этапе мутагенеза):

- решения уравнения Пуассона-Больцмана (MM-PBSA)
- метод возмущения свободной энергии
- обобщенный метод Борна
- метод Монте-Карло

Посмотрим, как выбирают области для аланинового сканирования.

# Выбор регионов для сканирования I

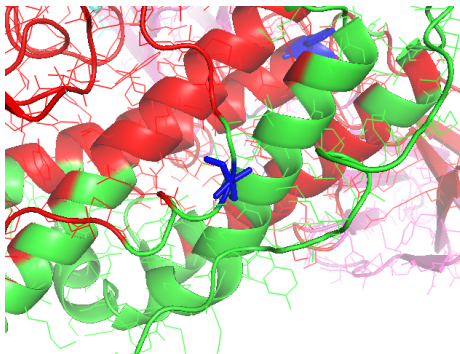
## Использование отсечки по расстоянию

- Аланиновому сканированию подвергаются аминокислоты цепочки, образующей комплекс совместно с другой цепочкой, содержащие атомы, удаленные от каких-либо атомов цепочки, также участвующей в образовании комплекса, на расстояние, не превышающей некоторой фиксированной величины
- В качестве порогового значения расстояния используются, например, величины 4, 5, 8 Å
- в Rosetta Alascan Protocol используется усложнение: дополнительно рассматриваются аминокислоты,  $\beta$ -углерод которых после формирования комплекса в шаре определенного фиксированного радиуса содержит существенно больше атомов  $\beta$ -углерода, чем содержал до этого.

- Рассмотрим базу данных с информацией о результатах экспериментов по аланиновому сканированию ASEdb.
- Найдем объекты со ссылкой на Protein Data Bank.
- Среди всех таких объектов, найдем те, в которых есть аминокислоты, мутация которых приводит к существенному изменению свободной энергии комплекса ( $\geq 1$  ккал/моль)
- Посмотрим, всегда ли они удалены от интерфейса в пределах стандартно используемой отсечки (в качестве примера возьмем расстояние, не превышающее  $8 \text{ \AA}$ ).

# Результаты эксперимента

Комплекс человеческого гормона роста и рецептора  
человеческого гормона роста<sup>3</sup>



# Выбор регионов для сканирования II

Поиск по гомологии

ASEdb: 76/101 корректных записей о белках, из них много 3hr.

Еще одна замечательная база данных:

tsailab.chem.pacific.edu/wikiBID/index.php/Main\_Page

page discussion view source history

## Main Page

### Binding Interface Database Wiki

An overflow of information on the characterization of protein-protein interactions at the amino-acid level understanding protein interfaces. For this reason it is necessary to acquire a protein-protein interaction data that can be easily accessed. The Binding Interface Database (BID) is structured to organize vast amounts of protein maps, and descriptive functional profiles. Detailed protein descriptions, interaction descriptions, bond formation to binding were systematically obtained by mining the primary literature containing alanine scanning analysis as well as more detailed can be found in our publication: [Fischer, et al., Bioinformatics 2003](#).

There are currently more than 467 interacting protein pairs with over 7000 hot spots documented. More than 1000 searched and entered include: Calmodulin (CAM) interacting proteins, Mitogen activated protein kinase (MAPK), Jak/STAT, Chaperones and chaperonins, Antigen/Antibody interactions, Cell Adhesion and Junction molecule pathway, tumor suppressors and various other proteins.

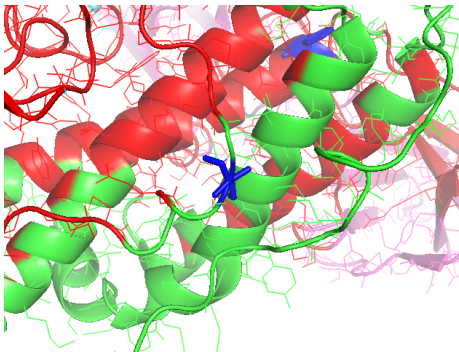
By extracting and centralizing information on protein-protein interactions from the primary literature, the experimental data in a readily searchable and accessible format. In this format, more relevance can be derived from correlations. We have develop some preliminary associations in the "browse by" box on the left. If you wish to join the BID wiki community. Also, [general comments/letters of support](#) are also greatly appreciated.

# Выбор регионов для сканирования II

Выводы. Основная задача.

Выводы: эффективного и универсального метода способа найти ЭГО – не придумали.

Будем решать эту задачу. Для этого попробуем по полученной картинке понять, что еще необходимо учесть:



# Как будем выбирать регионы для сканирования

- разумно начинать с области интерфейса белок-белкового взаимодействия, затем расширять ее.
- при мутации не гидрофобные аминокислоты могут стать гидрофобными и оказаться значимыми, поэтому в первую очередь можно расширить область на такие аминокислоты, находящиеся на границе интерфейса.
- поскольку при выборе интерфейса с отсечкой по расстоянию в интерфейсе могут появляться дыры, будем их заполнять. Для этого будем искать поверхностные карманы (углубления в поверхности белка в границах интерфейса).
- Из всех вторичных структур на поведение белка больше всего влияют петли. Добавим в интерфейс все аминокислоты, содержащиеся в петлях, которые частично уже попали в область рассмотрения.

# Алгоритм поиска протяженных регионов, потенциально содержащих „энергетически горячие точки”

Включим в состав множества протяженных регионов,  
содержащих „энергетически горячие аминокислотные остатки”,  
следующее:

- аминокислоты, образующие „интерфейс” взаимодействия с парной цепочкой или белком (с использованием отсечки по расстоянию от второй цепочки)
- аминокислоты, образующие поверхность „карманов”, находящихся в области взаимодействия пары белков
- не-гидрофобные аминокислоты, являющиеся соседними по отношению к аминокислотам, образующим интерфейс
- если интерфейс взаимодействия образован петлями, то добавим все аминокислоты, образующие петли



# „Интерфейс”

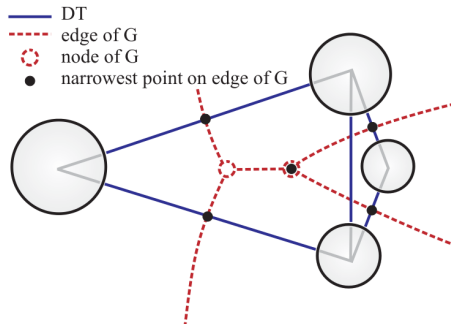
- определяем множество треугольников выпуклой оболочки, для которых хотя бы одна вершина удалена от центров атомов второй цепочки не больше, чем на выбранное значение отсечки
- Далее расширяем интерфейс
  - шаг 1: добавляем к интерфейсу все треугольники выпуклой оболочки, содержащие атомы аминокислот, которые уже туда попали
  - шаг 2: продлеваем регион до границы гидрофобности
  - шаг 3: продлеваем регион за границы гидрофобности на 1 аминокислоту.

В результате у нас есть одна или нескольких протяженных связанных областей выпуклой оболочки, по которым можно восстановить аминокислоты.

# Триангуляция Делоне

- Рассматриваем одновременно 2 цепочки, образующие белковый комплекс.
- Начнем с построения выпуклой оболочки и триангуляции Делоне для каждой из них, будем искать протяженные регионы с энергетически горячими аминокислотными остатками на одной из них. Строить будем по центрам атомов, формирующих аминокислоты цепочки.
- Выберем все треугольники выпуклой оболочки, в которых хотя бы одна вершина удалена от некоторых атомов второй цепочки не далее, чем на выбранное (фиксированное) значение отсечки.

# Построение графа по триангуляции Делоне



[Computation of tunnels in protein molecules using Delaunay triangulation, P.Medek, et al., 2007]

Используем модифицированный алгоритм Дейкстры, аналогично упомянутому в оригинальной работе.

Перед добавлением петель треугольники триангуляции преобразуются в фрагменты последовательности аминокислот, продлеваем их, используя информацию о вторичной структуре.

# Ala-scan in silico с фильтрацией данных

## Реализация

**Полученный алгоритм аланинового сканирования основан на Rosetta alascan protocol и образован следующей последовательностью действий:**

- читаем 2 цепочки атомов из PDB,
- выбираем аминокислоты одной из цепочек с помощью приведенного выше алгоритма поиска,
- для этих аминокислот пробуем провести мутагенез с использованием метода Монте-Карло выводим те, изменение которых привело к существенным изменениям свободной энергии системы.

# Результаты

- Получен скрипт, который итеративно формирует протяженные регионы, содержащие ЭГО (в процессе тестирования), и который может использоваться в модифицированном rosetta ala-scan protocol для поиска аминокислотных последовательностей.
- Планируется: переделать в плагин к PyMol, добавить промежуточный вывод аминокислот для визуального воспроизведения (в виде mesh-объекта)
- Предположительно, в алгоритм можно добавить поиск по гомологии (на стадии идеи)

Вопросы?