

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

Магистрант:

Татьяна Малыгина, СПбАУ

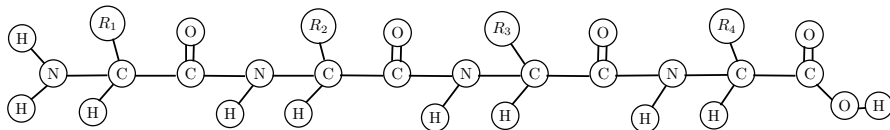
Научный руководитель:

Павел Яковлев, BIOCAD

СПбАУ, 2015

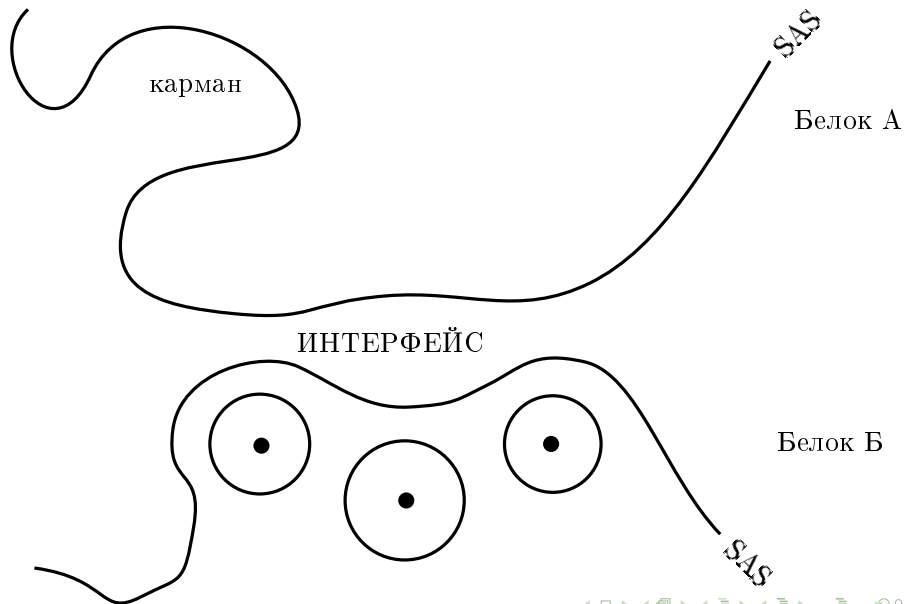
Белок-белковое взаимодействие I

Первичная структура белка задается последовательностью (цепочкой) аминокислот:



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

Белок-белковое взаимодействие II



Дизайн лекарств: проблемы

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

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

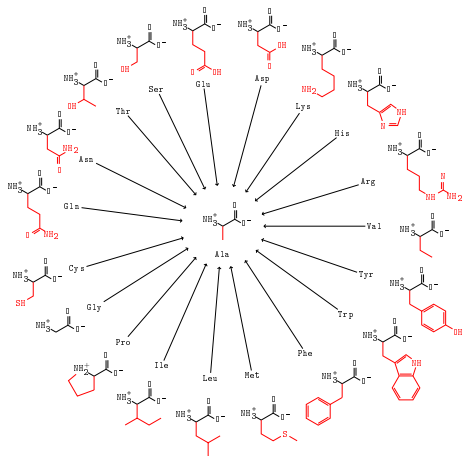
Вопрос 2: насколько специфична одна из компонент комплекса? Можно ли подобрать один из белков так, чтобы комплекс был более устойчивым? Насколько заменяема каждая из компонент?

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

Аланиновое сканирование

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

	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



Методы выбора аминокислот для сканирования

1. Отсечка по расстоянию

- выбираются все аминокислоты, удаленные от атомов парной цепочки не далее, чем на фиксированную величину.

Плюсы: быстро работает, легко реализовать.

Минусы: вблизи интерфейса расположены не все энергетически значимые аминокислоты.

Kortemme T., Baker D. A simple physical model for binding energy hot spots in protein-protein complexes / The National Academy of Sciences. Vol. 99(22). 2002. P. 14116–14121.

Kortemme T., Kim D. E., Baker D. Computational Alanine Scanning of Protein-Protein Interfaces // Sci. STKE. 2004. Vol. 18, no. 219.

Методы выбора аминокислот для сканирования

2. Предсказание

- проводится предсказание позиций, в которых расположены энергетически важные аминокислоты.

Плюсы: широкий и гибкий класс методов.

Минусы: зависит от выбора базы, не работает при дизайне de novo.

Liu B., Wang X., Lin L. et al. Prediction of protein binding sites in protein structures using hidden Markov support vector machine // BMC Bioinformatics. 2009.

Gavenonis J., Sheneman B. A., Siegert T. R. et al. Comprehensive Analysis of Loops at Protein-Protein Interfaces for Macrocycle Design // Nat Chem Biol. 2014 September. Vol. 10(9). P. 716–722.

Цели и задачи

Цель работы - построить алгоритм выбора аминокислот для компьютерного аланинового сканирования.

Задачи:

1. Определить признаки, которые определяют энергетическую значимость аминокислот, выбрать среди признаков биологически важные и обоснованные.
2. Предложить алгоритм выбора аминокислот, формирующих протяженные регионы, определяющие специфичность по отношению ко второй цепочке комплекса. Написать программную реализацию.
3. Провести анализ полученного алгоритма, проинтерпретировать полученные результаты.

Энергетически значимые аминокислоты:

- часто находятся вблизи интерфейса взаимодействия белков и в этом случае консервативны с точки зрения эволюции белков;
- могут находиться в глубине карманов;
- могут находиться на подвижных участках вторичной структуры белка - β -поворотах и петлях.

Алгоритм выбора регионов

На входе PDB-файл: получаем координаты атомов в пространстве и информацию об их типе.

Этапы алгоритма

- 1 **Предварительная обработка:** строим триангуляцию Делоне по центрам атомов и двойственный граф.
- 2 **Выбор начального множества треугольников.**
- 3 **Поиск карманов.**
- 4 **Переходим от элементов триангуляции к позициям аминокислот**
- 5 **Добавление петель и β -поворотов:** используем информацию из PDB-файла или алгоритм DSSP.

На выходе: набор позиций аминокислот для аланинового сканирования.

- построен скрипт на python, реализующий выбор протяженных регионов для аланинового сканирования с возможностью предварительной маскировки и совместимый с протоколом аланинового сканирования пакета Rosetta;
- плагин для PyMol;
- проведено тестирование работы алгоритма на наборе данных экспериментального аланинового сканирования Kortemme&Baker.

https://github.com/latticetower/MS_thesis

Результаты II

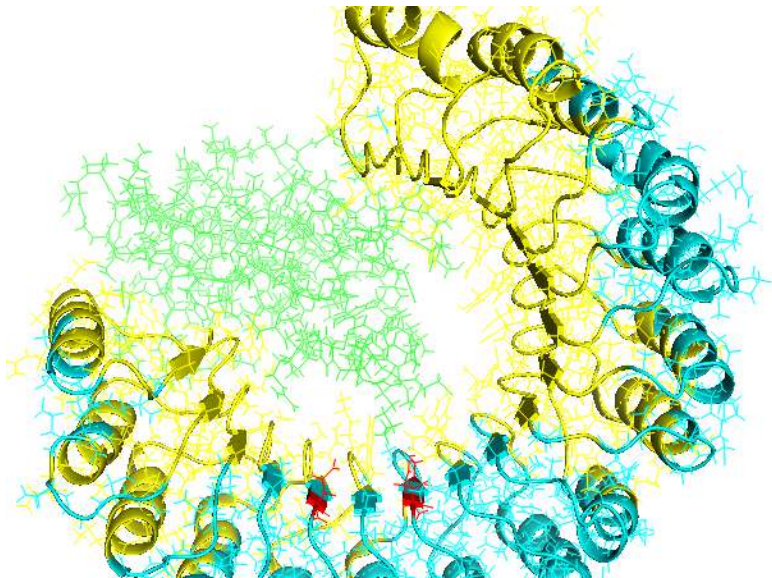
PDB	chain	P+N	TP+FP	TP	TP+FN	FN	TPR	ACC
1	2	3	4	5	6	7	8	9
1A22	A	180	85	29	29	0	1.0000	0.6889
1A22	B	192	166	36	36	0	1.0000	0.3229
1BXI	A	83	67	29	30	1	0.9667	0.5301
1DFJ	I	456	215	11	14	3	0.7857	0.5461
1F47	A	17	17	9	9	0	1.0000	0.5294
1FC2	C	45	41	3	3	0	1.0000	0.1556
2PTC	I	58	52	1	1	0	1.0000	0.1207

Результаты III

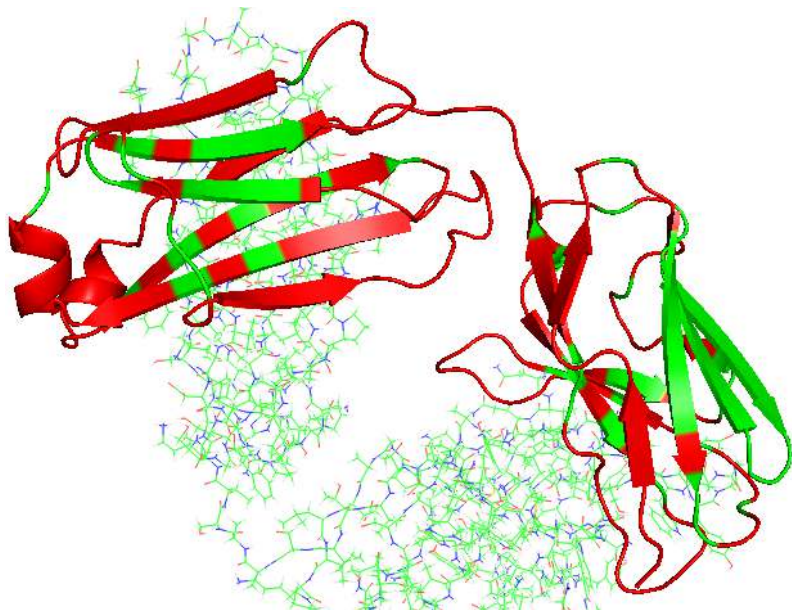
Выбор аминокислот вблизи интерфейса взаимодействия с отсечкой по расстоянию 5Å

PDB	chain	P+N	TP+FP	TP	TP+FN	FN	TPR	ACC
1	2	3	4	5	6	7	8	9
1A22	A	180	33	24	29	5	0.8276	0.9222
1A22	B	192	33	27	36	9	0.7500	0.9219
1BXI	A	83	23	16	30	14	0.5333	0.7470
1DFJ	I	456	39	13	14	1	0.9286	0.9408
1F47	A	17	9	8	9	1	0.8889	0.8824
1FC2	C	45	16	3	3	0	1.0000	0.7111
2PTC	I	58	14	1	1	0	1.0000	0.7759

Пример работы алгоритма: структура 1DFJ



Пример работы алгоритма: структура 2OSL



Спасибо за внимание!