

Сборка аминокислотных подпоследовательностей антител на основе top-down данных

Виталий Демьянюк

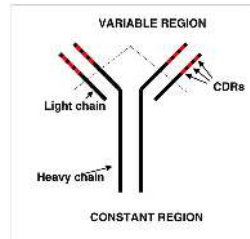
Институт Биоинформатики

Научный руководитель: Кира Вяткина

21 декабря 2013 года

Антитела

- Антитела — протеины производимые в В-клетках. Связываются с антигеном и ускоряют его уничтожение
- Секвенирование антител нужно для изучения их биологических и медицинских свойств
- Состоят из двух идентичных тяжелых и идентичных лёгких цепей, каждая из которых состоит из константного и переменного региона
- Наиболее интересными, являются CDR-ы — участки в переменном регионе, определяющие антигенную специфичность антитела



Масс-спектры

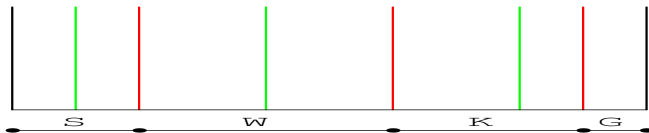


Рис.: Теоретический спектр для пептида SWKG

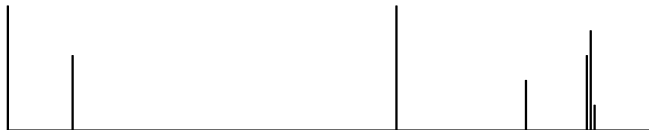


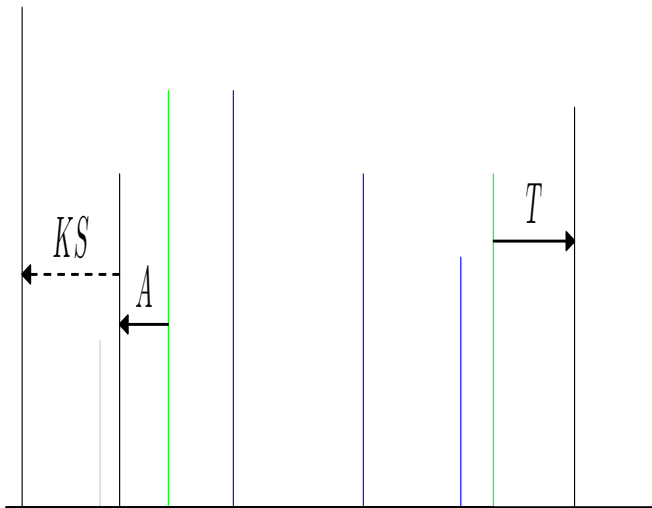
Рис.: Пример экспериментального спектра для пептида SWKG

- Top-down спектры — масс-спектры всего протеина
- Bottom-up спектры — спектры небольших аминокислотных подпоследовательностей (~20 аминокислот)

Этап 1: Построение каркаса

- Получить надежный пептид (например, часто встречающийся шести- или восьмимер)
- Выровнять все спектры относительно теоретического спектра этого пептида (выбрав в теоретическом спектре только массы префиксов)
- Спектр был выровнен успешно если количество совпавших пиков оказалось не менее заданного порога
- Построить суперспектр, состоящий из пиков успешно выровненных спектров, склеив при этом пики, позиции которых отличаются не более чем на 0.005 Da
- Интенсивность склеенного пика равна количеству склеенных пиков
- Удалить все *недостижимые* пики

Недостижимые пики



Этап 2: Пересчет интенсивностей

- Выровнять относительно суперспектра, полученного на этапе 1, все исходные спектры
- Построить новый суперспектр для успешно выровненных спектров
- Оставить в новом суперспектре только те пики, для которых имеются соответствующие пики в прежнем суперспектре
- Выдать новый суперспектр

Этап 3: Заполнение пробелов и поиск аминокислотной последовательности

- Выровнять относительно суперспектра, полученного на этапе 2, все исходные спектры
- Построить новый суперспектр для успешно выровненных спектров
- В прежний суперспектр добавить из нового суперспектра пики, заполняющие пробелы. Пик заполняет пробел, если:
 - Он отсутствует в прежнем суперспектре
 - В прежнем суперспектре имеется пик, отстоящий от рассматриваемого на массу одной аминокислоты влево, а также есть пик, отстоящий на массу одной аминокислоты вправо
- В графе, соответствующем полученному суперспектру, найти наилучший путь

Результаты

- Набор данных : 24519 top-down спектров для антитела MabCampath

Пептид	Результат
ASLSSPSQ	<i>LSLLKLLHKQQYWNLYKDLNQSAKCTLTVRDGVSASLSSPSQTMQLD</i>
LSSPSQ	<i>GLAQKQQYWNLYKDLNQSAKCTLTVRDGVSASLSSPSQTMQLN</i>
SASLSS	<i>DGVSASLSSPSQTM, LHKQQYWNLYKDLNQSAKCT</i>

- ASLSSPSQ встречается в 20 спектрах, LSSPSQ в 25 спектрах, SASLSS в 12 спектрах
- Рассматриваемые пептиды находятся в варибельном регионе легкой цепи антитела MabCampath
- На втором этапе пороговое значение — 8, на третьем — 6, на первом $\min(8, peplen + 1)$, где *peplen* — длина надежного пептида, используемого на первом этапе

Результаты

- KQYWNLYKDLNQSAKCTLTVRDGVASLSLSPSQTMQ
MGSFRPSVGTQLNNTNYLLLKPAKGPKQYWNLYKDLNQSAKCTLTVRDGVASLSLSPSQTMQLD
- VGTQLNNTNYLLLKPAKGPKQYWNLYKDLNQSAKCTLTVRDGVASLSLSPSQTMQLD
NDKQSTFDTGADKTM~~TS~~QSVGTQLNN...ASLSLSPSQTMQLDS
- VSASLSLSPSQTM
DLQMTLYLSLLKLHLKQYWNLYKDLNQSAKCTLTVRGDV~~ASLSLSPSQTMQLN~~

Bottom-up результаты

Перед корректировкой	После корректировки
YNTNNLQTVPSFRST	YNTNNLQTVPSF
ANLFGFVGTQLNNTNY	FVGTQLNNTNY
YNTNNLQTVFcENT	YNTNNLQTVF
YNTNNLQTVPS	YNTNNLQTVPSF
GSRSPYFNLLcVVSATGSK	PYFNLLcVVSATGSK
AASVA YFNLLcVVSATGSK	AYFNLLcVVSATGSK
AYFNLLcVVS	AYFNLLcVVS

Выводы

- Предложенный метод позволяет восстановить более 50% переменного региона
- В дальнейшем планируется осуществить сборку аминокислотной последовательности антитела целиком