

Построение метрики согласованности между петлевыми и каркасными регионами переменного участка антитела

Кочкова С.В., Университет ИТМО, Санкт-Петербург

Научный руководитель – Сергушичев А.А., кандидат технических наук, научный сотрудник факультета информационных технологий и программирования Университета ИТМО, Санкт-Петербург

Введение

На данный момент остается актуальной задача построения трехмерной структуры белка, в том числе антитела. Существует несколько компаний, специализирующихся на биоинформатике, которые предлагают несколько вариантов решения этой задачи, но оптимального варианта нет. В том числе интересно такое направление этой задачи, как построение антитела. Если константные участки антитела не представляют интереса, то переменные участки открывают целое многообразие для создания новых антител для лечения заболеваний.

Для того, чтобы построение переменных участков антител происходило более оптимально, в рамках этой работы предлагается разработать метрику согласованности фрагментов переменного участка.

Ранее в этом направлении было выпущено несколько статей, результаты по которым суммируются в статье 1997 года [1]. С тех пор прошло много времени и было открыто еще очень много антител, в связи с чем результаты, представленные в этой статье, потеряли свою актуальность.

В связи с вышеизложенной информацией, в рамках данной работы планировалось составить более полную базу открытых на данный момент антител, а также при помощи различных подходов разработать новую метрику согласованности между фрагментами переменного участка антитела.

Цель работы

Целью данной работы является построение метрики согласованности между фрагментами переменного участка антитела на основе составленной базы данных антител.

Базовые положения исследования

Антитело состоит из легкой и тяжелой цепей – последовательностей аминокислот. Тяжелая цепь состоит из четырех сегментов – сначала идет переменный регион (V), потом сегмент разнообразия (D) и потом сегмент соединения (J), и в конце – константный регион (C). Соответственно VDJ регионы составляют переменный регион антитела, состоящий из трех регионов петель и четырех регионов каркасов. D и J регионы состоят из петли и фрагмента каркаса, и конкретно про них никакой закономерности выявлено не было. При этом про первые две петли и три каркасных региона было выпущено множество статей [2][3], показывающих, что такие V регионы похожи и их можно разделять на семейства. Также есть статьи [1], показывающие, что для каких-то из этих изученных семейств существуют канонические структуры петлевых регионов. Но так как результаты в этих статьях уже устарели, в рамках данной работы была в том числе собрана актуальная база открытых антител.

База открытых антител была собрана из PDB [4], а также из дополнительных источников, которые закрыты коммерческой тайной компании ЗАО «Биокад».

Так как V регионы антител могут быть кластеризованы на семейства, идея построения метрики была выведена такой: метрика согласованности представляет из себя таблицу, в которой ячейки означают вероятность петли присутствовать в V регионе. Таблица вероятностей должна быть построена для трех возможных типов цепей (тяжелая цепь и два вида легкой цепи) и для двух петель.

Для построения семейств по последовательностям аминокислот V региона использовался иерархический алгоритм кластеризации[5]. Метрикой в данном случае выступает метрика схожести, которая строится на основе расстояния Хэмминга по последовательности. Кластеризация для петель проводилась по трехмерным структурам на основе метрики среднеквадратичного отклонения.

В итоге были получены семейства для последовательностей для трех типов цепей, а также кластера по структурам петель, так же для трех типов цепей. Тогда, если в таблице вероятностей для метрики согласованности по вертикали располагаются семейства последовательностей, а по горизонтали располагаются кластера структур, то значение в ячейке и будет показывать, что из себя представляет каноническая структура петли для определенного семейства последовательностей.

Результаты

В результате выполнения данной работы была собрана актуальная база антител, а также получена метрика согласованности между петлевыми и каркасными регионами варибельного участка антитела.

Список литературы

1. Bissan Al-Lazikani, Arthur M. Lesk, Cyrus Chothia. Standard Conformations for the Canonical Structures of Immunoglobulins. /1997 – J. Mol. Biol. (1997) 273, 927-948.
2. VBase 1997 MRC Centre for Protein Engineering [Режим доступа: <https://www2.mrc-lmb.cam.ac.uk/vbase/alignments2.php>].
3. Harry W. Schroeder, Jan L. Hillson. Structure and evolution of mammalian VH families. / 1989 – International Immunology, Vol. 2, No. 1.
4. Protein Data Bank [Режим доступа: <http://www.rcsb.org/pdb/>].
5. Stephen C. Johnson. Hierarchical clustering schemes. / 1967 – Psychometrika, 1967, Vol. 32, I. 3, 241-254