

Секция «Биоинженерия и биоинформатика»

Создание *in silico* метода предсказания сайтов регуляторного протеолиза на основе полипептидной последовательности субстратов.

Белушкин Александр Александрович

Студент

Московский государственный университет имени М.В. Ломоносова, Факультет
биоинженерии и биоинформатики, Москва, Россия

E-mail: alexb177@rambler.ru

Протеолизом называется процесс каталитического гидролиза белков, выполняемый специализированными ферментами - протеазами. Являясь необратимой посттрансляционной модификацией белков, протеолиз играет важную роль в апоптозе, иммунных процессах, пищеварении, нервной и эндокринной системах. Развитие раковых опухолей часто связано с нарушениями протеолитической регуляции [3]. По этим причинам исследование механизмов протеолиза крайне важно, а одной из наиболее актуальных проблем, связанных с протеолизом, является идентификация субстратов протеаз. Высокая стоимость, большие временные затраты и низкая процессивность экспериментальных методов исследования протеолитических взаимодействий делают задачу биоинформатического поиска субстратов протеаз особенно актуальной. <?xml:namespace prefix = o ns = "urn:schemas-microsoft-com:office:office" />

На сегодняшний день существуют несколько специализированных методов предсказания протеолитических сайтов (участков полипептидной цепи субстратов, с которыми непосредственно взаимодействуют протеазы), однако большинство из них основано на использовании только первичной специфичности протеаз – профиля частот аминокислот в окрестности сайта разрезания. Однако известно, что структурные характеристики сайтов протеолиза также влияют на эффективность протеолитических взаимодействий. По этой причине определение структурных параметров субстратов, влияющих на каталитическую эффективность протеолиза, и дальнейшая разработка метода по предсказанию протеолитических сайтов с учетом влияния определенных структурных параметров может стать мощным средством поиска субстратов протеаз *in silico*.

В данном исследовании был разработан биоинформатический метод предсказания сайтов регуляторного протеолиза на основе полипептидной последовательности субстратов. Ранее было оценено влияние на процесс протеолиза семи различных структурных характеристик, вычисленных при помощи 11-ти ведущих биоинформатических методов. При построении предсказательного алгоритма были отобраны характеристики и методы, имеющие наибольшую предсказательную способность. Также были использованы дополнительные структурные параметры, которые ранее не применялись ни в одном из существующих методов предсказания протеолитических сайтов. В частности, в недавнем исследовании разрешенных структур протеолитических субстратов [2] была показана подверженность длинных петель субстрата протеолитической дегградации. Также был использован параметр стабильности вторичной структуры, влияние которого было показано ранее. Тестирование метода было проведено на выборке, включающей более 2000 белков из базы данных CutDB [1], которая на данный момент содержит максимальное количество экспериментально зарегистрированных фактов протеолиза. Для выбора оптимального алгоритма, используемого для предсказания позиций сай-

тов протеолиза, было протестировано пять различных методов машинного обучения. В качестве используемого метода был выбран алгоритм классификации с использованием деревьев решений. Веса первичной специфичности протеолитических сайтов были рассчитаны при помощи позиционно-весовых матриц из базы данных Merops. Для построения результирующего алгоритма структурные предсказания были дополнены данными о первичной специфичности протеаз. Результаты работы полученного метода сравнивались с результатами работы двух современными методами предсказания протеолитических сайтов. По предварительным оценкам, выполненным с использованием таких метрик машинного обучения, как true positive rate, accuracy и specificity, разработанный метод превосходит рассмотренные методы. Представленный алгоритм является универсальным методом, позволяющим выполнять предсказания сайтов различных протеаз, и может быть успешно использован для поиска неизвестных субстратов протеаз с последующей экспериментальной проверкой.

Литература

1. Igarashi Y. et al., CutDB: a proteolytic event database. *Nucleic Acids Res.* 2007, 35:D546-9.
2. Kazanov M.D. et al., Structural determinants of limited proteolysis. *J Proteome Res.* 2011;10:3642-51.
3. Lopez-Otin, C.; Bond, J. S. Proteases: multifunctional enzymes in life and disease. *J. Biol. Chem.* 2008, 283, 30433–37.