

# Краткий обзор протеома бактерии *Saccharophagus degradans*

Елизавета Минина<sup>1</sup>

<sup>1</sup>Факультет биоинженерии и биоинформатики, МГУ им. М. В. Ломоносова, Москва, Россия

## РЕЗЮМЕ

В данной работе приводится краткий обзор протеома бактерии *Saccharophagus degradans* на основании данных, полученных из базы данных NCBI и обработанных при помощи программы Microsoft Excel 2010. Данная работа была проведена в рамках прохождения курса биоинформатики на Факультете биоинженерии и биоинформатики МГУ имени М. В. Ломоносова.

## 1 ВВЕДЕНИЕ

К настоящему моменту накоплен колоссальный объём информации о геномах, протеомах и других аспектах генетики отдельных видов. Секвенирование геномов более не является трудоёмким научным экспериментом, а стало рутинной процедурой, благодаря которой к настоящему моменту полностью прочитаны геномы многих прокариотических и эукариотических организмов, а также вирусов. Получить информацию о геноме интересующего организма может любой желающий, имеющий доступ к сети Интернет: эти данные размещены на открытом сайте Национального центра биологической информации (<http://www.ncbi.nlm.nih.gov/>). Более того, элементарную обработку геномной информации из базы данных NCBI с целью получения базовых представлений о геноме интересующего организма можно даже с помощью самых доступных программ, в частности, Microsoft Excel.

В данной работе излагается пример такой обработки. С помощью Microsoft Excel 2010 был проведён анализ протеома бактерии *Saccharophagus degradans*, геном которой полностью секвенирован [2]. Было оценено количество РНК- и белкокодирующих генов на прямой и комплементарной цепях ДНК, а также составлена гистограмма, отображающая распределение белков бактерии по длинам (длина выражена в количестве аминокислотных остатков).

Изучение протеома *Saccharophagus degradans* представляется в высшей степени интересным. Как известно, многие природные полисахариды, такие как целлюлоза и хитин, не расщепляются большинством организмов (в том числе и человеком) по причине отсутствия соответствующих ферментов. *Saccharophagus degradans* способна расщеплять такие соединения, поскольку обладает необходимыми ферментами [1]. Анализ протеома с целью выявления и дальнейшего углублённого изучения структуры и механизмов работы этих ферментов не только представляет научный интерес, но и может быть полезен с практической точки зрения. Он может позволить оценить целесообразность применения этих ферментов на практике,

например, для удаления загрязнений или в пищевой промышленности для повышения усвояемости целлюлозы и хитина. Эти ферменты можно будет либо выделять из бактерий *S. degradans*, выращивая их в промышленных масштабах, или клонировать гены необходимых ферментов и внедрить их в более подходящие для промышленных целей организмы, например, кишечную палочку *Echerichia coli*. Таким образом, изучение протеома *S. degradans* может быть перспективной научной задачей.

## 2 МАТЕРИАЛЫ И МЕТОДЫ

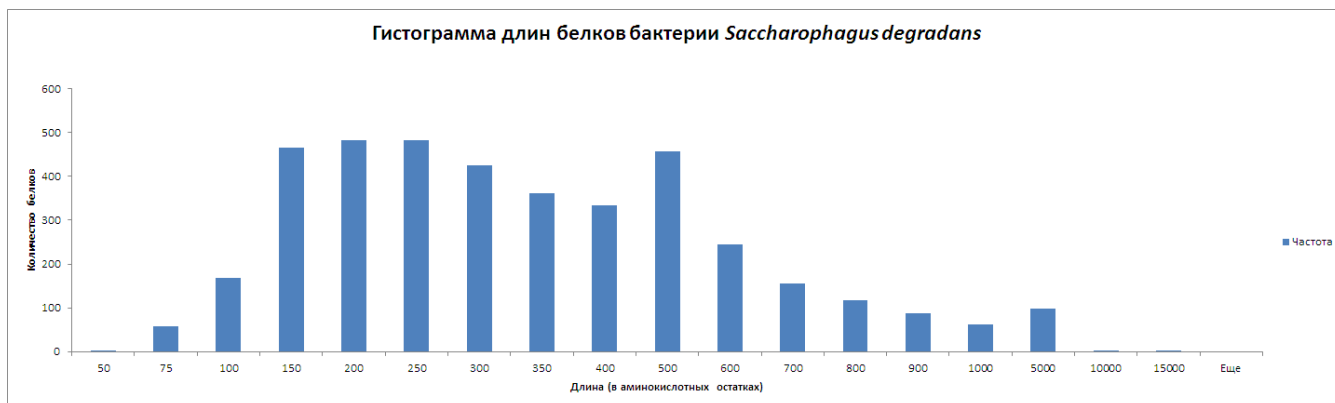
Все данные, использованные в данной работе, были взяты из базы данных NCBI (<http://www.ncbi.nlm.nih.gov/genome>). Данные по *S. degradans* находятся по адресу [http://www.ncbi.nlm.nih.gov/genome/1079?genome\\_assembly\\_id=170430](http://www.ncbi.nlm.nih.gov/genome/1079?genome_assembly_id=170430). Обработка данных осуществлялась с помощью программы Microsoft Excel 2010.

## 3 РЕЗУЛЬТАТЫ

Геном *Saccharophagus degradans* представлен единственной хромосомой. Эта бактерия не имеет плазмид. Геном состоит из 5057531 пар оснований, в нём закодировано 4007 белков и 50 некодирующих РНК. Ниже рассматривается распределение белков по длинам, а также распределение генов по прямой и комплементарной цепям ДНК.

### 3.1 Распределение белков по длинам

Для получения представления о том, белки каких длин имеются в протеоме *S. degradans*, была построена гистограмма, отображающая распределение белков по длинам (рис. 1). Как можно наблюдать, разброс длин белков протеома этой бактерии довольно значителен: самые маленькие белки имеют длину менее 50 аминокислотных остатков (а. о.), а длина самого крупного белка – клеточного поверхностного рецептора IPT/TIG - составляет 14609 а. о. Наибольшее количество белков имеет длину в пределах 150 – 500 а. о., а наиболее многочисленны белки с длиной от 150 до 200 а. о. и от 200 до 250 а. о.: в каждой из этих групп насчитывается 482 белка, так что белки длиной от 150 до 250 а. о. составляют, в общей сложности, 24 % от всех белков. 62 белка имеют длину менее 100 а. о., а длина двух белков превосходит 10000 а. о.

Рис 1. Гистограмма длин белков *S. degradans*

### 3.2 Распределение генов по цепям

Помимо анализа белков *S. degradans*, в данной работе также рассматривалось распределение генов по двум цепям ДНК. Как известно, молекула ДНК состоит из двух цепей: прямой цепи и цепи, комплементарной ей. Некоторые гены локализованы на первой цепи, другие – на второй.

В случае *S. degradans* имеет следующее распределение по цепям генов, кодирующих белки и не кодирующие РНК:

Таблица 1. Распределение генов *S. degradans* по двум цепям ДНК

	Гены РНК	Гены ДНК
Прямая цепь	29	1962
Комплементарная цепь	21	2045

## 4 ОБСУЖДЕНИЕ

### 4.1. Распределение белков по длинам

Исходя из сравнительных данных, полученных по протеому *S. degradans*, наибольшую часть протеома составляют белки со средним числом аминокислотных остатков (от 150 до 500). Эти белки функционально столь разнообразны, что связать длину с функцией в их случае вряд ли удастся: среди них есть транскрипционные факторы, метаболические ферменты, мембранные переносчики и белки со многими другими функциями. В целом, можно сказать, что для клетки наиболее выгодно синтезировать белки длиной от 150 до 500 а. о.: этого количества аминокислотных остатков достаточно для выполнения необходимых функций, при этом синтез их не столь затратен, как синтез очень крупных белков (напомним, что в процессе синтеза белка напрямую расходуется энергия). Однако чем

можно объяснить существование в протеоме таких белков-гигантов, как вышеупомянутый рецептор IPT/TIG? Можно предположить, что это обусловлено функцией белка. Возможно, лиганд этого рецептора имеет такую пространственную структуру и атомную конфигурацию, что для его специфического связывания необходим определённый большой набор аминокислотных остатков. Кроме того, возможно, что этот рецептор может распознавать несколько разных лигандов, соответственно, для связывания каждого из них ему необходим отдельный домен с определёнными аминокислотными остатками в сайте связывания. Наконец, многие трансмембранные белки (то есть белки, расположенные в плазматической мембране клетки и пронизывающие её насквозь; именно такими белками является большинство рецепторов бактериальной клетки) имеют несколько длинных трансмембранных альфа-спиралей, состоящих из большого количества аминокислотных остатков. Стоит отметить, что приведённые выше рассуждения в значительной мере гипотетичны и умозрительны: мы ещё знаем слишком мало о протеоме *S. degradans*, многие белки этой бактерии, внесённые в базу данных и включённые в статистические данные в настоящей работе, ещё не были выделены, и их существование пока ещё только теоретически предсказано.

### 4.2. Распределение генов по цепям

Существует гипотеза, что гены распределены между двумя цепями случайно с вероятностью 0,5 нахождения на каждой из цепей. Как можно видеть, у *S. degradans* и для генов, кодирующих белки, и для генов, кодирующих РНК, соотношение генов на прямой и комплементарной цепях примерно равно 1, что подтверждает гипотезу о случайном распределении генов по цепям с вероятностью 0,5.

## 5 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Файл Excel, в котором проводились все расчёты и который содержит результаты работы, можно найти по адресу <http://kodomofbb.msu.ru/~lizaminina/term1/genome.xls>.

## БЛАГОДАРНОСТИ

Я выражаю большую благодарность своим преподавателям информатики за интересные и познавательные задания, дав-

шие хорошую пищу для размышлений. Особенно хочется поблагодарить преподавателей за то, что они показали, что даже с помощью самых распространённых программ можно обрабатывать геномные данные и получать интересные и важные результаты.

## СПИСОК ЛИТЕРАТУРЫ

- 1 Ekborg NA, Gonzalez JM, Howard MB, Taylor LE, Hutcheson SW, Weiner RM. (2005) *Saccharophagus degradans* gen. nov., sp. nov., a versatile marine degrader of complex polysaccharides. *Int J Syst Evol Microbiol.* **55**, 1545-9
- 2 Weiner RM et al. (2008) Complete genome sequence of the complex carbohydrate-degrading marine bacterium, *Saccharophagus degradans* strain 2-40 T. *PLoS Genet.* **4**(5), e1000087