

ПЕРИОДИЧНОСТЬ С ДЛИНОЙ ПЕРИОДА 2 В АМИНОКИСЛОТНЫХ ПОСЛЕДОВАТЕЛЬНОСТЯХ БЕЛКОВ

Руденко В.М., Коротков Е.В.¹

Институт Биоинженерии в составе ФИЦ Биотехнологии, Россия, 117312, Москва, пр-т 60-летия Октября, 7, корп.1, +7 926-701-81-08, v.m.rudenko@gmail.com
¹НИЯУ МИФИ, Россия, 115409, Москва, Каширское ш., 31, +7 926-724-82-71, katrin2@biengi.ac.ru

Широко известен тот факт, что аминокислотные последовательности могут содержать в своем составе тандемные повторы, которые связаны с пространственной структурой белка и его биологической функцией [1]. Белки с тандемными повторами вовлечены в такие жизненно-важные процессы как дупликация генов, рекомбинация, кроссинговер. Это подчеркивает их значимость и эволюционную роль.

С точки зрения математики тандемные повторы можно интерпретировать как периодичность, хотя и размытую, содержащую замены, вставки и делеции. В данной работе мы предлагаем производить поиск периодичности с учетом корреляций соседних символов последовательности. Для этих целей сначала мы группировали аминокислоты на 5 групп в соответствии с полярностью их радикалов и перекодировали исходную последовательность, определенную на алфавите в 20 символов к новому 5-символьному алфавиту. Затем рассматривали все парные комбинации соседних символов последовательности, строили для них позиционно-весовую матрицу и циклическое выравнивание [2]. В качестве меры периодичности выступала величина Z , которая определялась методом Монте-Карло. Пороговое значение величины было взято $Z_0=6$.

Метод был применен к анализу последовательностей банка данных uniprot. Было обнаружено достаточно много белков, с периодичностью 2. Так используемый нами подход выявил 2115 белков, в то время как популярная программа T-REKS обнаружила на том же множестве всего 1090 периодических фрагментов.

Анализ белков с периодичностью 2 показал, что среди них встречается много уже аннотированных ранее как белки, содержащие повторы: collagen-like protein, serine-aspartate repeat-containing protein и т.д. Их последовательности имеют экстремальное большое значение Z .

Но наибольший интерес представляют последовательности, где определяется периодичность с небольшими значениями Z . Например, последовательность Q3SZ45 из генома *Bos taurus* (Stromal cell-derived factor 2), длина 211. Периодичность с $Z=8.2$ наблюдается на участке 25-193 aa, что составляет около 80% ее длины.

Литература

1. *Kajava A.V.* Tandem repeats in protein: From sequence to structure // *Journal of Structural Biology* vol 179, N 3, 2012. p. 279-288
2. *Turutina V., Laskin A. et.al* Identification of amino acid latent periodicity within 94 protein families // *Journal of Computational Biology* vol 13, N 4, 2006. p. 946-964