

УДК 577.21; 577.29; 621.37

**М. И. Богачев**, канд. техн. наук,

**О. А. Маркелов**, аспирант,

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»

**А. Р. Каюмов**, канд. биол. наук.,

Казанский (Приволжский) федеральный университет

# Алгоритм предварительного отбора белковых последовательностей для множественного выравнивания в условиях априорной неопределенности на основе теории графов<sup>1</sup>

*Ключевые слова:* белковые последовательности, множественное выравнивание, априорная неопределенность, теория графов, перколяция.

*Key words:* protein sequences, multiple alignment, a priori uncertainty, graph theory, percolation

*Предложен подход к предварительному отбору последовательностей при поиске консенсусных фрагментов в первичной структуре белковых молекул методом множественного выравнивания. Подход основывается на положениях теории случайных графов и теории перколяции. Применение показано на примерах множественного выравнивания группы аминокислотных последовательностей белков, относящихся к различным функциональным классам, специфично распознаваемых внутриклеточным регуляторным белком.*

зы, такие как ClpAP, ClpCP, ClpXP [3,4]. Поскольку протеолитическое расщепление — процесс неотрашимый, распознавание белка мишени должно быть высокоспецифичным. Некоторые белки напрямую распознаются AAA+ протеиназами, другие же распознаются и доставляются к месту протеолиза специальными адаптерными белками [4]. Несмотря на то что интерес к направленному протеолизу не иссякает, во многих случаях механизм узнавания белка-мишени остается неизвестным [1,2,5].

Протеиназа ClpP в клетках бактерий находится в олигомерном комплексе с АТФазами ClpA, ClpC или ClpX и осуществляет направленный протеолиз регуляторных и дефектных белков [3]. В бактериях сигналом для распознавания, как правило, является последовательность гидрофобных ароматических аминокислот (тирозина, триптофана, фенилаланина и лейцина) [6]. На сегодняшний день неизвестна консенсусная аминокислотная последовательность, необходимая для распознавания АТФазами ClpA, ClpC и ClpX.

Для решения задачи поиска консенсусных последовательностей биомолекул используется группа так называемых алгоритмов выравнивания. Среди классических работ в этой области следует в первую очередь отметить алгоритмы *Needleman Wunsch* [7] и *Smith-Waterman* [8]. Если первый подход ориентирован на выравнивание по всей длине последовательности, то второй направлен на поиск локальных консенсусных фрагментов произвольной длины. На базе данных двух подходов, получивших дальнейшее развитие в работах [9,10], были синтезированы алгоритмы множественного выравнивания для групп последовательностей [11,12].

## Введение

Протеолитические ферменты, широко распространенные в природе и выполняющие гидролиз белков, участвуют в различных процессах, в частности, в адаптационных и деструктивных катаболических процессах, функционируют на этапе посттрансляционного расщепления белков, осуществляют избирательное расщепление химически устойчивой пептидной связи в белках. Направленный высокоспецифичный протеолиз необходим для поддержания баланса клеточного протеома, регуляторных процессов и контроля качества белков в клетке [1–3]. В клетках *B.subtilis* эту функцию выполняют внутриклеточные АТФ-зависимые AAA+ протеина-

<sup>1</sup> Работа выполнена при поддержке Совета по грантам Президиума Российской Федерации (грант для государственной поддержки молодых российских ученых, шифр МК-556.2011.8, дог. № 16.120.11.556-МК от 18.02.2011).