

СТАТИСТИЧЕСКИЙ АНАЛИЗ И МОДЕЛИРОВАНИЕ КОНФОРМАЦИОННОЙ ДИНАМИКИ ПЕПТИДОВ

Бабкина А. Е., Ефимов П. В.

Харьковский национальный университет имени В.Н.Каразина
pavel.v.efimov@univer.kharkov.ua

Белки играют важнейшую роль в природе и жизни человека. Они представляют большой интерес для изучения, т.к. выполняют много биологических функций, зависящих от их трехмерной структуры и динамики сворачивания. Поэтому, изучение конформационной динамики пептидов дает возможность получать белки с нужными свойствами, задавая лишь первичную структуру.

Для исследования конформационной динамики пептидов, может быть использовано молекулярное моделирование, с последующим анализом полученных данных. Компьютерное моделирование пептидов показывает, что большинство перестановок в молекуле происходит через быстрые изменения значений двугранных углов. В результате наблюдаются длинные периоды флуктуации значений углов в пределах одного состояния. Следовательно, необходимо разделение конформационных состояний на кластеры. Но при этом возникают определенные проблемы, связанные с многомерностью и цикличностью данных. Классические методы кластерного анализа требуют либо знания количества предполагаемых кластеров либо могут работать с небольшими выборками. Нейронные сети Кохонена также не приводят к нужным результатам. Возникает необходимость разработки новых методов и подходов к анализу результатов молекулярно-динамического моделирования конформационной динамики пептидов.

В данной работе предлагается разведочный метод многомерных угловых данных. В качестве первого этапа предлагается проведение предварительного анализа, целью которого является выявление количества кластеров в имеющемся наборе данных. N -мерное пространство разбивается на $2N+1$ равных $(N-1)$ -мерных симплексов. Для этого вычисляются косинусы и синусы всех углов, и по их знаку имеющийся набор данных последовательно разделяется на группы. Вычисляется заселенность каждого симплекса. Отбрасываются симплексы с числом точек ниже установленного порога. Таким образом, многомерное пространство разбивается на множество подпространств, внутри которых становится возможным применение классических методов кластерного анализа в пределах каждой группы. Для проверки данного метода были смоделированы различные наборы данных, подобные реальным системам, с заведомо известным количеством кластеров. Метод показал хорошие результаты, позволяя обрабатывать большие наборы данных. Предложенный подход позволяет сократить расчетное время и свести задачу к методам традиционного кластерного анализа.