

Белок представляет собой линейный полимер. Первичная структура белка - это последовательность аминокислот, составляющих белок. Пространственная структура - это расположение всех атомов белка в пространстве. Всего имеется 20 аминокислот, отличающихся друг от друга по физико-химическим свойствам. Аминокислоты можно разделить на группы с похожими свойствами.

**Описание проблемы:**

В настоящее время известно несколько тысяч первичных структур ДНК- и РНК-связывающих белков, для многих из которых установлены и депонированы в базе данных Protein Data Bank (PDB, [www.rcsb.org](http://www.rcsb.org)) пространственные структуры белков отдельно либо в комплексе с ДНК или РНК. В подавляющем большинстве случаев ДНК и РНК связываются с белками в положительно заряженной бороздке в структуре белка. Эта часть структуры может быть сформирована аминокислотными остатками, расположенными в разных местах первичной структуры. Разумеется, не все положительно заряженные остатки белка участвуют в формировании ДНК/РНК-связывающей бороздки. Поскольку для формирования бороздки полипептидная цепь сворачивается в пространстве, можно ожидать, что положительно и отрицательно заряженные остатки будут образовывать некоторый скрытый паттерн в первичной структуре белка.

Примеры ДНК- и РНК-связывающих белков будут предоставлены.

**Описание задачи:**

С помощью методов машинного обучения сделать классификатор, который по первичной структуре белка определяет, является ли он ДНК/РНК-связывающим.