

РАЗРАБОТКА ПРОГРАММНОЙ БИБЛИОТЕКИ ДЛЯ РАСЧЕТА ГЕОМЕТРИЧЕСКИХ ПАРАМЕТРОВ ДНК НА ЯЗЫКЕ PYTHON

Васильев В. А., Армеев Г. А.

Биологический факультет Московского государственного университета
имени М.В.Ломоносова, Москва, Ленинские горы, д. 1, стр. 12,
info@mail.bio.msu.ru, +7(495)939-10-00

Геометрия двуцепочечной ДНК в А и Б формах была определена в 1953 году и на данный момент в банке данных белковых структур содержится более 6000 структур, содержащих 2 и более цепей ДНК. Известно, что геометрия ДНК зависит от ионного окружения, последовательности ДНК и взаимодействия с белками. Для того чтобы изучать закономерности изменения геометрии необходимы методы описания конформации ДНК. Существует несколько подходов к описанию локальной геометрии двуцепочечной ДНК. Один из наиболее распространенных подходов описывает ДНК в виде набора динуклеотидов. При таком подходе положение каждой следующей пары описывается 6 параметрами относительно предыдущей пары: 3 описывают пространственное смещение вдоль взаимно перпендикулярных осей и 3 – повороты вокруг этих осей.

Существующие программы для расчета геометрических параметров такие как DSSR(3DNA) и CURVES+ имеют ряд недостатков. Данные программы не содержат интерфейса для интеграции в другие программные библиотеки, их исходный код отсутствует в свободном доступе. Также эти программы не оптимизированы для обработки траекторий молекулярных динамики. В данной работе мы разработали аналог программ 3DNA find_pair и analyze на языке программирования Python.

В нашей программной библиотеке расчёт параметров ДНК осуществляется в несколько этапов. Первый шаг - поиск нуклеотидов в PDB структуре и определение их типа (независимо от имени остатка и имен атомов). На втором шаге алгоритм определяет пары нуклеотидов. Последний этап - расчет внешних и внутренних параметров пар на основе определенных нуклеотидов и их пар. Работа данной программной библиотеки сравнивалась с работой 3DNA. Рассчитываемые значения параметров в среднем отличаются на 0.00274. Наша программная библиотека работает несколько медленнее 3DNA на отдельных структурах, но вычисления для траекторий производятся более чем в 10 раз быстрее.

Программная библиотека используется в работе программы для моделирования нуклеосомных фибрилл PyNAMod (<https://github.com/intbio/PyNAMod>).

Работа была поддержана Российским Научным Фондом, грант №21-74-00033