Предсказание спектров возбуждения флуоресценции на основе данных КМ/ММ молекулярной динамики флуоресцентных белков с помощью моделей машинного обучения

Степанюк Р.А.

В работе предлагается использование значений изменения дипольного момента при возбуждении (DMV) хромофора для предсказания спектров возбуждения флуоресценции GFP-подобных белков с учетом квадратичного эффекта Штарка мы предлагаемый оригинальный подход, совмещающий применение методов КМ/ММ молекулярного моделирования и машинного обучения. Предлагаемый нами алгоритм предполагает совмещение двух моделей машинного обучения (ММО1 и ММО2) с промежуточной стадией градиентной оптимизации коэффициентов Штарка на обучающей выборке по метрике Вассерштайна и перехода к расширенной обучающей выборке (ОВ2) с заменой целевого признака, DMV, на значения E возбуждения с учетом оптимизированных коэффициентов уравнения Штарка для каждого белка обучающей выборки. Отработка инструмента на данной задаче позволит затем расширить этот подход на предсказание спектров флуоресценции, при условии получения МД траекторий систем в возбужденном состоянии, например, методом LR-TDDFT, и последующим расчетом дескрипторов и значений DMV. Таким образом, наша работа, направлена на создание инструмента быстрой оценки спектральных свойств, ускоряющего пайплайн рациональной разработки и модификации флуоресцентных белков. Предлагаемый инструмент своим функционалом расширит имеющиеся возможности исследователей при использовании других моделей (ESM3, AlphaFold3 и др.). Работа выполнена в рамках проекта НОШ МГУ №23-Ш03-04.



Рисунок 1. Принципиальная схема алгоритма предсказания спектров флуоресцентных белков