

ОПРЕДЕЛЕНИЕ ЗНАЧИМОСТИ ГЕОМЕТРИЧЕСКИХ ПАРАМЕТРОВ ХРОМОФОРА ДЛЯ ПРЕДСКАЗАНИЯ СПЕКТРОВ ПОГЛОЩЕНИЯ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Кулакова А.М.¹, Захарова Т.М.¹, Хренова М.Г.^{1,2}

¹Московский государственный университет имени М.В. Ломоносова, Химический факультет, кафедра физической химии,
Россия, 119991, г. Москва, Ленинские горы, д. 1, стр. 3
Тел.: (495)939-22-86, E-mail: kulakova@lcc.chem.msu.ru

²ФИЦ Биотехнологии РАН, Россия, 119071, Москва, Ленинский проспект, д. 33, стр. 2

Для визуализации клеточных процессов часто используют зеленый флуоресцентный белок и его аналоги. Они представляют собой бочонки из β -листов, внутри которых расположен хромофор, сформированный из трех аминокислотных остатков. Изменение спектральных свойств таких белков возможно не только за счет изменения самого хромофора, но и за счет внесения изменений в хромофорсодержащую область белка. Известно, что для таких белковых систем, соответствующих минимумам на энергетической поверхности, имеется взаимосвязь между спектральными свойствами и геометрическими характеристиками, наиболее важными из которых являются длины мостиковых связей между фенильным и имидазолидоновым фрагментами хромофора.

В данной работе, для белка EYFP семейства зеленого флуоресцентного белка проведены молекулярно-динамические расчеты с комбинированными потенциалами квантовой механики/молекулярной механики (PBE0-D3/cc-pvdz) в NPT ($p = 1$ атм, $T = 300$ К) ансамбле. Для набора кадров из МД траектории были рассчитаны изменения дипольного момента при возбуждении ($\Delta\mu$), для которого известна квадратичная взаимосвязь с изменением разности энергий между основным и возбужденным электронным состояниями. С помощью методов регрессионного анализа и метода случайного леса были построены модели расчета значений $\Delta\mu$, исходя из геометрических характеристик хромофора. Полученные модели хорошо предсказывают значения изменения дипольного момента при возбуждении. При этом показано, что для расчета значений $\Delta\mu$ по результатам молекулярной динамики недостаточно использовать только значения длин мостиковых связей в хромофоре. В минимальный набор данных для рассматриваемого белка также необходимо включать связи, соседние с мостиковыми.

Работа выполнена с использованием оборудования Центра коллективного пользования сверхвысокопроизводительными вычислительными ресурсами МГУ имени М.В. Ломоносова при финансовой поддержке научно-образовательной школы МГУ «Мозг, когнитивные системы, искусственный интеллект» (проект 23-Ш03-04).