

3D QSAR МОДЕЛИРОВАНИЕ БИОЛОГИЧЕСКОЙ АКТИВНОСТИ С ИСПОЛЬЗОВАНИЕМ МЕТОДА НЕПРЕРЫВНЫХ МОЛЕКУЛЯРНЫХ ПОЛЕЙ

Жохова Н.И., Баскин И.И., Зефилов А.Н., Палюлин В.А.

Московский государственный университет им.М.В.Ломоносова, г.Москва

Жохова Н.И.,
Баскин И.И.,
Зефилов А.Н.,
Палюлин В.А.
Московский
государственный
университет
им.М.В.Ломоносова

Традиционные методы построения 3D QSAR моделей (CoMFA и CoMSIA), широко используемые для моделирования биологической активности химических соединений при создании новых лекарственных препаратов, обладают рядом ограничений. Эти методы крайне чувствительны к размеру, шагу и пространственной ориентации трехмерной гипотетической решетки, создаваемой вокруг молекулы и используемой для аппроксимации электростатического, стерического и гидрофобного молекулярных полей, потенциалы, которых рассчитываются в ее узлах в качестве дескрипторов молекулярной структуры [1]. Неоднозначность полученных таким образом моделей приводит к ненадежности прогноза на их основе.

Мы предлагаем новый метод для построения 3D QSAR моделей - Метод Непрерывных Молекулярных Полей (MCMF, Method of Continuous Molecular Fields). Его основная идея состоит в непосредственном анализе непрерывных молекулярных полей вместо дискретного массива их значений, рассчитанных на узлах решетки, что точнее соответствует физической природе этих полей. Идея реализована на базе одного из новейших методов машинного обучения - метода регрессии на опорных векторах (SVR, Support Vector Regression), который позволяет работать с бесконечным числом переменных [2]. Основным элементом метода MCMF является процедура расчета статистических ядер (kernels) с использованием предложенной формулы. Ядро, описывающее сходство между двумя молекулами, выражается через взвешенную сумму интегралов скалярных произведений значений потенциалов нескольких молекулярных полей по всему физическому пространству. Потенциалы полей молекул находятся путем суммирования потенциалов полей входящих в них атомов.

Возможности метода продемонстрированы на примере построения количественных корреляций "структура-ингибирующая активность" для 88 производных 3-амидинофенилаланина в отношении трех ферментов группы сериновых протеаз - тромбина, трипсина и фактора Ха [3]. Используются экспериментальные значения констант ингибирования, выраженные значением логарифма константы диссоциации pK_i ($-\log K_i$), которая характеризует степень связывания ингибитора с ферментом (K_i , моль/л).

Оценку прогнозирующей способности моделей проводили с использованием процедуры скользящего контроля с исключением по одному. Качество моделей оценивали по значениям коэффициента детерминации для скользящего контроля (q^2) и среднеквадратичной ошибки (RMSE). Как показано в таблице (, все модели, полученные с помощью метода MCMF, обладают лучшей прогнозирующей способностью по сравнению с моделями, полученными стандартными методами CoMFA и CoMSIA. Для методов CoMFA и CoMSIA приведены параметры моделей, построенных при оптимальных значениях шага решетки 1.5 Å, 1 Å, 1 Å для тромбина, трипсина и фактора Ха, соответственно.

Статистические характеристики моделей, полученных методами CoMFA, CoMSIA и MCMF для соединений обучающей выборки

Параметр	Тромбин			Трипсин			Фактор Ха		
	CoMFA	CoMSIA	MC MF	CoMFA	CoMSIA	MC MF	CoMFA	CoMSIA	MC MF
q^2	0.697	0.757	0.805	0.635	0.754	0.794	0.429	0.590	0.600
RMSE	0.378	0.244	0.210	0.204	0.159	0.137	0.294	0.194	0.189

Литература

- 3D QSAR in Drug Design. Theory Method and Applications. Ed. Kubinyi H., N.Y.: Kluwer, 1997, 515 p.
- Ivanciuc O. In: Reviews in Computational Chemistry. V. 23. Eds.: Lipkowitz K.B., Cundari T.R., Weinheim: Wiley-VCH. 2007. P. 291-400.
- Böhm M., Stürzebecher J., Klebe G. // J. Med. Chem. 1999. V. 42. P. 458-477.