

Информационная система, поддерживающая процесс построения моделей прогнозирования свойств химических соединений

*Сенкова Т. Н., Кумсков М. И., Миловидов А. Н.,
Свитанько И. В.*

qsar_msu@mail.ru, kumskov@mail.ru

Москва, Кафедра Вычислительной математики, Мехмат МГУ

Накоплены значительные массивы экспериментальных данных по физико-химическим свойствам, биологической активности, а также большое число QSPR- и QSAR-моделей (Quantitative Structure-Property Relationship и Quantitative Structure-Activity Relationship). Актуальной является задача унификации представления таких моделей в рамках информационной системы (ИС), способной единообразно хранить как данные о дескрипторах молекул, так и параметры этапов построения моделей прогнозирования, использующих разнообразные методы классификации и распознавания образов.

Задача системы — информационная поддержка процесса построения и сравнения QSPR/QSAR-зависимостей (различающихся как по качеству, так и по сложности моделей прогнозирования) за счет единообразного их хранения на основе реляционной СУБД.

Система ориентирована на представление ранее найденных QSAR-зависимостей, основанных на различных (по детализации) формах хранения молекул, которые могут иметь вид графа или молекулярной поверхности, «раскрашенной» локальным физико-химическим свойством (ЛФХС), например, с использованием потенциала или способности отдавать или принимать электрон (донорно-акцепторные факторы) и т. п. При формировании признаков на молекулярной поверхности вычисляются экстремумы ЛФХС — особые точки. В результате структурным объектом, подлежащим анализу и классификации, становится маркированный граф, вершины которого располагаются в особых точках на молекулярной поверхности.

Для анализа конкретного свойства неизвестно, на каком уровне следует описывать молекулы — на топологическом, на планарном или на пространственном. Выбор уровня представления молекул и адаптация признаков для конкретного свойства проводится математиком-исследователем динамически в процессе построения QSAR-моделей.

Пользователь-химик должен суметь запустить прогноз новых молекул без присутствия математика спустя месяцы и годы после построения модели. Проблема состоит в том, что последовательность преобразования молекул может быть построена на основе модулей и программ, имеющих

различные алгоритмы и различные значения входных параметров. Так, существует десяток программ, позволяющих построить пространственную укладку атомов молекулы, т. е. вычислить ее «3D-конформацию». Эти программы используют различные принципы, начиная с методов молекулярной механики и кончая квантово-химическими расчетами. При проведении прогнозирования свойств новых молекул химиком, эти новые структуры должны пройти тот же путь преобразований теми же самыми программными модулями, что прошли молекулы обучающей выборки при построении модели математиком. Таким образом, все версии программ, участвующих в цепочке преобразований молекул, и их параметры должны единообразно сохраняться в репозитории, чтобы была возможность полной повторяемости результатов модели при прогнозировании.

Молекулярные структуры после построения QSAR модели не должны «стираться», поскольку, во-первых, они определяют «область допустимых значений» (ОДЗ) модели и, во-вторых, могут быть повторно использованы (при появлении новых данных о механизмах действия) для QSAR моделирования.

ИС должна поддерживать три типа пользователей:

1. *Пользователь, ответственный за контент ИС*, выполняет загрузку структурных массивов в базу данных, отвечает за проверку правильности данных о свойствах, за формирование фрагментного представления молекул и его описания и за загрузку пространственного представления молекул и описания его характеристик; пополняет каталог выборок, представленных в системе, и их описаний.
2. *Пользователь-исследователь* выполняет построение и верификацию моделей — отвечает за создание и тестирование программ в среде MATLAB, реализующих модели, выбранные для поддержки в ИС, за сохранение параметров моделей с лучшими показателями по качеству прогноза, за сохранение протоколов тестирования моделей и их аналитического сравнения, а также за определение «области действия» модели.
3. *Конечный пользователь* проводит расчеты по прогнозированию свойств новых соединений, используя модели, поддерживаемые в ИС; сохраняет протоколы расчетов, делает запросы на изменение, включая необходимую корректировку моделей или области их действия.

В моделях системы будут реализованы ранее разработанные методы [1–3], основанные на использовании структурных фрагментов молекул, а также:

- 1) метод идентификации и классификации информативных особых точек (ОТ) на пространственной молекулярной структуре, описы-

- вающих физико-химические экстремумы локального свойства на молекулярной поверхности;
- 2) метод формирования матрицы «молекула–признак» на основе символьных дескрипторов (ЭО — элементов описания), представляющих 3D конфигурации двоек, троек и четверок ОТ на молекулярных поверхностях;
 - 3) метод идентификации ЭО — определение отношения эквивалентности ЭО с использованием нечеткого (fuzzy) кодирования взаимных расстояний между ОТ в ЭО;
 - 4) методы расчета молекулярного подобия по матрице «молекула–элемент описания».

Работоспособный прототип системы предполагается создать на основе реляционной СУБД в двухзвенной архитектуре клиент-сервер. На первом этапе в репозиторий планируется загрузить пять выборок со структурами молекул и их свойствами, а для вычисления «топологического сходства» молекул на фрагментах будут построены индексные файлы, описывающие состав молекул в виде цепочек из 2-х, 3-х и четырех атомов. Для вычисления «пространственного сходства» планируется загрузить 3D конформации молекул, рассчитанные потенциалы на молекулярных поверхностях и сформировать соответствующие индексные файлы по дескрипторам, описывающим 3D отношения между активными центрами.

Работа поддержана грантом РФФИ № 07-07-00282.

Литература

- [1] *Svitanko I. V., Kumskov M. I., Tcheboukov D. E., Dolmat M. S., Zakharov A. M., Ponomareva L. A., Grigor'eva S. S., Chichua V. T.* QSAR modeling on the base of electrostatic molecular surface (amber fragrances) // 16-th Eur. Symp. on Quantum Structure-Activity Relationships and Molecular Modelling, Italy: EuroQSAR-2007. —
- [2] *Захаров А. М., Кумсков М. И., Пономарева Л. А.* Эволюционный алгоритм построения моделей «структура-свойство» для молекулярных поверхностей с использованием аппарата нечеткой логики // ММРО-12, Москва, 2005. — С. 109–112.
- [3] *Захаров А. М., Свитанько И. В., Григорьева С. С., Чичуа В. Т.* Поиск особых точек на молекулярных поверхностях с использованием нечеткого кластер-анализа // ММРО-12, Москва, 2005. — С. 333–335.