

УДК 615.2/.3:547.022:541.2].001.24
©Погребняк, Глушко

КОМБИНИРОВАННЫЙ МЕТОД ПРОГНОЗИРОВАНИЯ БИОЛОГИЧЕСКОЙ АКТИВНОСТИ НА ОСНОВЕ СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНОГО ПОЛЯ И МЕТОДА ПОТЕНЦИАЛЬНЫХ ФУНКЦИЙ (MSpace)

Погребняк А.В., Глушко А.А.

Пятигорская государственная фармацевтическая академия, г.Пятигорск
эл. почта: pgfa@megalog.ru

Разработан новый метод прогнозирования биологической активности органических соединений на основе сравнительного анализа молекулярного поля с использованием метода потенциальных функций в пространстве распределения электронной плотности.

Ключевые слова: QSAR, CoMFA, ориентация молекул, молекулярное сходство, электронная плотность.

ВВЕДЕНИЕ. Предлагаемая нами модель оценки количественной связи структура-биологическая активность (MolecularSpace, MSpace) основана на сравнении трехмерного распределения электронной плотности в молекулах биологически активных веществ. Для расчета распределения электронной плотности могут использоваться как полуэмпирические, так и неэмпирические (*ab initio*) квантово-химические методы [1].

Исходными данными для прогнозирования служат координаты атомов в молекуле и набор собственных векторов молекулярных орбиталей либо матрица плотности. Для оптимизации параметров модели и для прогнозирования биологической активности используется база данных лекарственных препаратов и биологически активных веществ. База данных содержит трехмерные структуры веществ, их физико-химические дескрипторы, в том числе собственные векторы молекулярных орбиталей, используемые в данном методе, а также количественную информацию о фармакологической активности веществ [2].

МЕТОДИКА. Модель Mspace.

В основе данного алгоритма прогнозирования биологической активности лежит модифицированный нами метод потенциальных функций [2,3]. Результатом прогноза является значение логарифма обратной среднетерапевтической дозы данного вещества, необходимой для проявления прогнозируемой активности. Мера фармакологической активности аппроксимируется с помощью потенциальной функции:

$$P_i = \sum_{j=1}^m k_j f(r_{ij}) \quad (1)$$

где P_i - значение потенциальной функции для i -го препарата; k - вектор нормировочных множителей; m - количество активных препаратов; r - матрица расстояний; f - функция распределения; i - номер препарата в обучающей выборке; j - номер препарата в подгруппе активных веществ обучающей выборки.

В качестве функции распределения была использована экспоненциальная функция. Функция имеет параметр сглаживания, который может варьироваться для оптимизации модели:

$$f(r) = e^{(-\sigma r^2)} \quad (2),$$

где r - радиальная координата; σ - параметр сглаживания (характеризует относительную степень разброса структурных особенностей для данного вида активности).

Для оценки степени сходства молекулярных структур используется сравнительный анализ молекулярного поля (CoMFA). В классическом методе CoMFA для генерации дескрипторов молекулярной структуры используется трехмерная матрица значений электронной плотности, электростатического потенциала или других полей [4]. Используемый нами метод сравнения отличается тем, что вычисляются расстояния в функциональном пространстве функций распределения электронной плотности. В качестве метрики данного пространства используется величина (r):

$$r = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\rho_2 - \rho_1)^2 dx dy dz \quad (3),$$

где ρ_1 , ρ_2 - функции распределения электронной плотности в двух сравниваемых молекулах.

Данный интеграл зависит от взаимного расположения молекул. Для его расчета используется выравнивание молекул, соответствующее минимуму данной величины (r). Поиск минимума производится с использованием градиентного метода. При этом варьируются 6 параметров: смещения молекулы вдоль координатных осей x , y и z и поворот вокруг осей координат ориентируемой молекулы (α , β и γ).

Ускоренный поиск оптимального взаиморасположения молекул производится на основе метода Monte Carlo в сочетании с градиентным методом.

Для упрощения расчетной схемы и ускорения расчета в процессе оптимизации вместо величины r используется интеграл произведения электронных плотностей (Q):

$$Q = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_2 \rho_1 dx dy dz \quad (5),$$

т.к.

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\rho_2 - \rho_1)^2 dx dy dz = \\ & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_2^2 dx dy dz - 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_2 \rho_1 dx dy dz + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_1^2 dx dy dz \end{aligned} \quad (6),$$

то при варьировании взаиморасположения молекул изменяется только интеграл произведения электронных плотностей. Минимум величины r соответствует максимуму величины Q . Поэтому процесс оптимизации интеграла произведения плотностей (Q) направлен на максимизацию. Оптимизация градиентным методом производится с аналитическим расчетом производных.

На рисунке 1 приведен пример выравнивания в пространстве молекул морфина и промедола (агонисты морфиновых рецепторов). На данном рисунке приведена карта распределения произведения электронных плотностей двух молекул, позволяющая визуализировать зоны сходства электронных структур. Выравнивание проводилось с использованием метода Monte Carlo в сочетании с градиентным методом путем максимизации величины Q . Процесс поиска глобального максимума функции Q для двух данных молекул занял 11 с на компьютере с процессором Intel 2 ГГц (1 Гб ОЗУ).

На основе расположения молекул с максимальной величиной Q рассчитывается расстояние r .

Таким образом производится попарное сравнение молекул и вычисление матрицы расстояний r_{ij} .

$$r_{i,j} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\rho_i - \rho_j)^2 dx dy dz \quad (7),$$

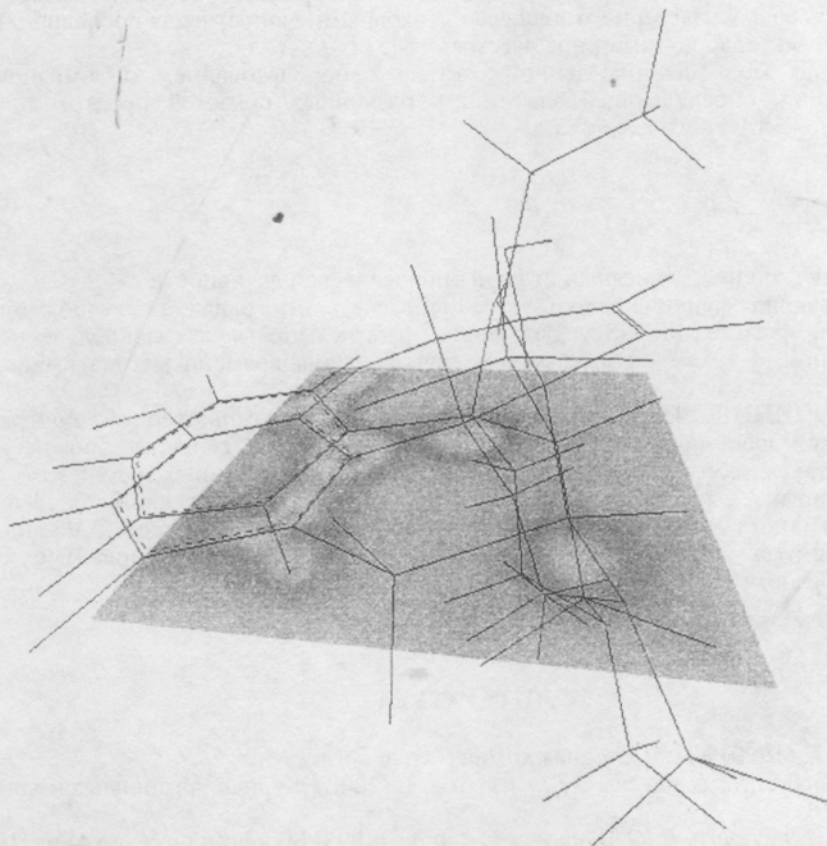


Рисунок 1.
Сравнение структур молекул морфина и промедола и карта распределения величины $\rho_1 \rho_2$ (произведение электронных плотностей) в плоскости.

Оптимизация параметров модели *Mspace*.

В качестве параметра оценки качества модели была использована стандартная ошибка прогноза (SPRESS), определяемая методом скользящего контроля. Модель *Mspace* имеет ряд параметров: параметр сглаживания потенциальной функции (σ) и коэффициенты k для потенциальной функции (1). Поскольку SPRESS возрастает и убывает вместе с суммой квадратов невязок прогноза ($PRESS_{cv}$), то проводилась оптимизация параметров, направленная на минимизацию $PRESS_{cv}$:

$$PRESS_{cv} = \sum_{i=1}^n \left(\left(\sum_{j=1}^m k_j f(r_{ij}) \right) - A_i \right)^2, \quad j \neq i \quad (8)$$

где: k - вектор нормировочных множителей;
 r - матрица расстояний;
 f - функция распределения;
 m - количество активных препаратов;
 n - количество препаратов в обучающей выборке;
 i - номер препарата в обучающей выборке;
 j - номер препарата в подгруппе активных веществ из обучающей выборки;
 A_i - фармакологическая активность i -го препарата из обучающей выборки.

Прогнозирование биологической активности.

Для прогнозирования биологической активности исследуемого вещества производится попарное сравнение структуры его молекулы со структурами молекул подгруппы активных веществ (расчет расстояний по вышеприведенному алгоритму) и расчет логарифма обратной среднетерапевтической дозы на основе оптимизированной потенциальной функции (1).

Примечание. Данный подход к сравнению молекулярных структур может быть использован также для кластерного анализа, т.к. входными данными для последнего могут быть не только векторы, но и матрица расстояний [5].

Возможная модификация данного метода с использованием оптимизируемой весовой функции, определяющей значимость различных областей пространства для определения молекулярного сходства:

$$r_{i,j} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W^2 (\rho_i - \rho_j)^2 dx dy dz \quad (9),$$

где W - весовая функция, зависящая от вида биологической активности.

Взвешивающая функция используется при расчете интеграла r , а целевой функцией ее оптимизации является PRESScv. Дальнейшая работа направлена на определение вида весовой функции, а также оптимального с точки зрения времени метода вычисления интеграла (9).

ЗАКЛЮЧЕНИЕ. Нами проводится разработка компьютерной программы, предназначенной для управления базой данных химических веществ, биологических активностей, других сопутствующих объектов, а также качественных и количественных связей между ними. Для выявления характера данных связей и построения моделей связи структура активность планируется использовать методику MSpace. Особенностью данного алгоритма является его высокая производительность. Последнее делает возможным обработку достаточно больших массивов молекул.

ЛИТЕРАТУРА

1. Кларк Т. (1990) Компьютерная химия (пер. с англ.), Мир, М.
2. Погребняк А.В., Оганесян Э.Т., Глушко А.А. (2002) Журнал органической химии, **38**, 1618.
3. Тетко И.В., Танчук В.Ю., Васильев С.А. и др. (1995) Биоорганическая химия, **10**, 809.
4. Kubinyi H. in: The Encyclopedia of Computational Chemistry. Vol. 1, pp. 448-460.
5. Tryon R.C. (1939) Cluster Analysis. Ann Arbor, MI: Edwards Brothers.

PREDICTION OF BIOLOGICAL ACTIVITY ON THE BASIS OF COMPARATIVE MOLECULAR FIELD ANALYSIS AND POTENTIAL FUNCTIONS METHOD (MSpace)

Pogrebnyak A.V., Glushko A.A.

Pyatigorsk State Pharmaceutical Academy;
357500 Pyatigorsk, Kuchury 2, r. 73, e-mail: pgfa@megalog.ru

Method for prediction of biological activity of organical substances on the basis of comparative molecular field analysis and potential functions method has been developed. It is based on the measure of molecular similarity, according to distances between molecular structures in functional space of the electronics density distribution.

Key words: QSAR, CoMFA, alignment, data base, molecular similarity, electronic density.