КОМБИНИРОВАННАЯ ВЫЧИСЛИТЕЛЬНАЯ 3D-QSAR ТЕХНОЛОГИЯ BIS

Потемкин В.А., Гришина М.А., Барташевич Е.В.

Челябинский государственный университет, г. Челябинск

Создана комбинированная вычислительная 3D-QSAR технология BiS, включающая возможности количественного и качественного анализа и прогноза биологической активности с учетом стереоизомерии, конформационного и таутомерного состояния. Предложены подходы к оценке биологической акивности стереоизомеров, таутомеров и конформеров. Показаны возможности алгоритма на примере 43 видов активности. Ведется создание параллельной версии алгоритма на кластере СКИФ Союзного Государства России и Белоруссии.

Теоретическое исследование биологической активности соединений представляет собой сложную задачу, поскольку механизм биологического действия включает ряд стадий: адсорбцию, распределение, метаболизм, выделение, транспорт. Каждая стадия или несколько из них могут лимитировать процесс биологического действия. С этой точки зрения, процесс биологического действия является макрокинетическим, включающим ряд параллельных и последовательных стадий. Исследование таких процессов является весьма трудоемким даже в относительно несложных случаях классических химических реакций. Поэтому для изучения процессов биологического действия выделяют ряд наиболее важных стадий для рассматриваемого вида активности и производят исследование с учетом их параллельного или последовательного протекания.

Как правило, биологическое действие лекарственных средств включает в себя стадию взаимодействия с рецептором, образование комплекса “рецептор-лиганд”. Исследование данной стадии *in vitro* может проводиться достаточно надежно для многих видов биологической активности. Для характеристики эффективности данного процесса могут использоваться энергии образования или константы связывания, а также величины $IC_{50\%}$ или $EC_{50\%}$.

Для качественной и количественной оценки действия соединений на стадии образования комплекса “рецептор - лиганд” могут быть использованы различные методы: кластер-анализ, распознавание образов, нейронные сети, методы регрессионного, нелинейного, полиномиального описания, PLS-методы. В последние годы популярность завоевали методs 3D-QSAR и молекулярного докинга. Недостатком многих 3D-QSAR методов является необходимость априорного знания ориентации и расположения молекул в полости рецептора. Поэтому необходимо создание алгоритма, включающего в себя возможности ориентации молекул, определения функциональной зависимости биологической активности от характеристик взаимодействий в системе “рецептор - лиганд” и молекулярный докинг. Такие возможности предоставляет 3D-QSAR алгоритм BiS.

3D-QSAR Алгоритм BiS

Для решения задачи ориентации имеется ряд подходов. В основе большинства методов лежит предположение о геометрическом соответствии фармакофорной части молекулы сайту рецептора. Для “укладки” в полости рецептора, как и во многих традиционных подходах, используется пространство-заполняющая модель. В разных вариантах решения использовались жесткие сферы [1,2] и сферы с переменной жесткостью [3,4]. Данные попытки, как правило, позволяют установить верную ориентацию для рядов структурно аналогичных молекул. Попытки ориентации соединений разных классов приводят к отклонениям от экспериментально установленного расположения в полости рецептора. Причины отклонений очевидны: геометрическая форма молекулы является важной, но не единственной характеристикой, определяющей расположение молекулы в полости

рецептора. Ориентацию молекулы определяет весь комплекс ван-дер-ваальсовых, кулоновских и специфических взаимодействий с рецептором и сольватным окружением. В совокупности эти виды взаимодействий определяют молекулярное поле, которое обеспечивает комплементарность биологически активных соединений к рецептору. Следует иметь в виду, что молекула с данным видом активности содержит фрагменты, определяющие связывание с активными центрами рецептора. Эту часть молекулы можно назвать фармакофорной. Эффективность ее связывания с рецептором определяет величину активности. Остальные атомы, не взаимодействующие с рецептором, могут играть роль “балластной” части. Разные молекулы могут связываться с разными активными центрами рецептора. Поэтому генеральная совокупность активных молекул должна достоверно описывать поле рецептора. Тогда решение проблемы ориентации молекул в полости рецептора требует определения совокупного поля выборки молекул. Данная задача решается в рамках алгоритма BiS [5-12] следующим образом.

Для определения молекулярного поля в алгоритме BiS используются потенциалы кулоновских и ван-дер-ваальсовых взаимодействий, наводимых на точку m поверхности молекулы. Кулоновский потенциал определяется из известного уравнения:

$$\varphi_m^q = \sum_{i=1}^{N} \frac{q_i}{R_{im}} k \quad (1)$$

где N- число атомов рассматриваемой структуры; q_i – заряд атома i; R_{im} – расстояние от точки m до атома i; k – коэффициент пересчета в единицы СИ.

Для расчета потенциала ван-дер-ваальсовых взаимодействий, наводимых на точку m, по аналогии с кулоновским предложено использование следующего уравнения [9]:

$$\varphi_m^{VDW} = -2\sum_{i=1}^{N} V_{im} \frac{2^3 r_i^3}{R_{im}^6} \quad (2)$$

где V_{im}- глубина минимума потенциальной энергии в соответствии с потенциалом Леннард-Джонса; r_i – ван-дер-ваальсов радиус атома i.

Расчет величин V_{im} и r_i производится в модели MERA [13-14]. В уравнении для упрощения расчетов использован только член притяжения ван-дер-ваальсовых взаимодействий.

Определение потенциалов поля производится для первой молекулы выборки. Полученные потенциалы позволяют определить характеристики каждого псевдоатома (пробной сферы) с центром в точке m (заряд q_m и радиус r_m), который обеспечивает комплементарность рассматриваемой молекуле в данной точке. Набор псевдоатомов является псевдоатомной моделью рецептора. Характеристики каждой пробной сферы можно вычислить по формулам:

$$q_m = -\frac{\varphi_m^q}{\sum_{i=1}^{N} \frac{k}{R_{im}}}; \quad r_m = \sqrt[3]{\frac{\varphi_m^{VDW}}{-2^3 \sum_{i=1}^{N} \frac{2V_{im} r_i^3}{R_{im}^6}}}$$

Затем в комплементарном поле в комбинации симплексного и квази-ньютоновского методов производится оптимизация ориентации второй молекулы выборки до достижения максимума вероятности контакта (P) ее атомов со всеми точками модельного рецептора.

$$P = 1 - \prod_{m=1}^{M} (1 - p_m)$$

где

$$p_m = \exp\left(-\frac{E_m}{RT}\right)$$

$$E_m = \sum_{i=1}^{N2} \left(\frac{k q_i q_m}{R_{im}} - 2V_{im} \frac{(r_m + r_i)^6}{R_{im}^6} + V_{im} \frac{(r_m + r_i)^{12}}{R_{im}^{12}} \right) \quad (3)$$

где $N2$ – число атомов во второй молекуле.

В найденной ориентации уточняется комплементарное поле рецептора путем добавки потенциалов поля второй молекулы к имеющимся:

$$\varphi_m^q = \varphi_m^q + \varphi_m^{q\,\prime}; \quad \varphi_m^{VDW} = \varphi_m^{VDW} + \varphi_m^{VDW\,\prime}$$

Потенциалы поля второй молекулы $\varphi_m^{q\,\prime}$ и $\varphi_m^{VDW\,\prime}$ вычисляются аналогично потенциалам первой. Данная процедура осуществляется для третьей, четвертой и всех последующих молекул выборки. По окончании процедуры производится рассмотрение расположения первой, второй и т.д. молекул выборки в уточненном поле. Процедура прекращается, если различие потенциала поля на новом и предыдущем шаге не превышает заданного предела. По окончании процедуры строится линейная зависимость биологической активности от параметров взаимодействия псевдоатомов модельного рецептора с молекулами:

$$BA = a_0 + \sum_{i=1}^{l1} a_i E_i + \sum_{i=1}^{l2} b_i F_i$$

В качестве параметров взаимодействий используются энергии, рассчитанные по уравнению (3), и силы F_m как первые производные энергии по расстоянию:

$$F_m = \sum_{i=1}^{N} \frac{dE_m}{dR_{im}}$$

Данное уравнение позволяет выделить псевдоатомы модельного рецептора, которые определяет величину биологической активности соединений. Эти псевдоатомы моделируют активные центры рецептора. Фрагменты лекарственного средства, расположенные вблизи активных центров, определяют фармакофорную часть молекулы. Анализ комплексов “модельный рецептор - лиганд” позволяет не только выявить фармакофорные части молекулы лекарственного средства, но также в случае различных механизмов действия, провести классификацию соединений по механизму действия [5,7-9,15,16]. Достаточно точное определение структуры комплексов позволяет устанавливать количественные взаимосвязи параметров взаимодействия в системе “рецептор - лиганд” с биологической активностью, что показано в [5-12,15,16] на примерах 28 видов активности. Определены внешние виды комплексов “модельный рецептор – лиганд”, найдены уравнения, позволяющие с высокой точностью описать величину активности. Для каждого вида активности выделены факторы, определяющие процесс биологического действия.

Кроме того, возможность ориентации молекул, имеющаяся в алгоритме BiS, позволяет выяснить расположение молекул в реальном рецепторе. Сопоставление результатов теоретического расположения с данными РСА показывает хорошие результаты, что позволяет выяснять детали механизма биологического действия [9,15,16].

Мультиконформационная оценка биологической активности в алгоритме BiS

Многие молекулы являются конформационно подвижными системами, способными в существовать в нескольких (а то и нескольких десятках) конформаций. Данные формы существуют в динамическом равновесии. При этом неочевидно, что максимальной активностью обладает наиболее выгодный конформер. Более активным может являться конформер, присутствующий в меньшей концентрации. Кроме того, в комплексе с рецептором может быть более выгодной конформация, минорная в свободном состоянии. То есть минорная конформация будет являться более комплементарной к рецептору. Данный конформер определяет действие соединения. При образовании комплекса “рецептор - лиганд” он выводится из раствора, а в растворе вновь устанавливается динамическое равновесие с прежней долей активного конформера. Таким образом, процесс биологического действия определяется долей активного конформера, а также скоростью установления конформационного равновесия. Тогда активность соединения (A) может быть определена как суперпозиция активностей конформеров (a_i), то есть:

$$A = \sum_{i=1}^{K} a_i p_i \qquad (4)$$

где p_i – вероятность нахождения i-го конформера; K – число конформеров.

Вероятность нахождения i-го конформера может быть рассчитана по формуле:

$$p_i = \frac{\exp(-E_i / kT)}{\sum_{j=1}^{K} \exp(-E_j / kT)}$$

где E_i, E_j – энергии i-го и j–го конформеров, соответственно; k – постоянная Больцмана; T – абсолютная температура.

В качестве исходного допущения ряда методов 3D-QSAR [17–24] принимается, что биологическая активность соединения пропорциональна энергии взаимодействия “рецептор – лиганд”. При этом рассматривается наиболее выгодный конформер. Если считать, что подобное допущение справедливо для всех прочих конформеров, то

$$a_i = b_0 + b_1 E_i^R$$

где E_i^R – энергия взаимодействия рецептора с i-ым конформером; b_0 и b_1 – константы.

Подставив данное выражение в формулу (4), получаем

$$BA = \sum_{i=1}^{K} p_i \left(a + bE_i^R\right) = a\sum_{i=1}^{K} p_i + b\sum_{i=1}^{K} p_i E_i^R \qquad (5)$$

поскольку $\sum_{i=1}^{K} p_i = 1$, то

$$BA = a + b\sum_{i=1}^{K} p_i E_i^R \qquad (6)$$

Для использования последнего уравнения необходимо определить число конформеров, долю каждого из них в веществе, рассчитать их энергию взаимодействия с рецептором. В этом случае, метод анализа биологической активности может быть создан в виде комбинации метода поиска конформеров с определением их доли в веществе и 3D-QSAR алгоритма, обеспечивающего построение комплекса “модельный рецептор – лиганд” и определение энергий взаимодействия рецептора с конформерами. Коэффициенты a и b могут быть определены методами регрессионного анализа. Подробное изложение методики поиска конформеров и расчета их доли в веществе приведено в [10,11].

В рамках алгоритма BiS формируется псевдоатомная модель рецептора и рассчитывается энергия взаимодействия каждого конформера с модельным рецептором. Модельный рецептор, единый для всех конформеров всех молекул, формируется в виде псевдоатомов (сфер с радиусами r_m и зарядами q_m), расположенных вокруг лигандов. Ориентация каждого конформера в поле модельного рецептора оптимизируется до достижения минимума потенциальной энергии взаимодействия. Вычисляется энергия взаимодействия данной молекулы (**E**) во всем множестве ее конформаций с модельным рецептором

$$\mathbf{E} = \sum_{i=1}^{K} p_i E_i^R \qquad (7)$$

Аналогичные операции проводятся со всеми прочими молекулами выборки и устанавливается зависимость вида $BA = a + b\mathbf{E}$, что при подстановке выражения (7) приводит к искомому уравнению (6).

Проводится варьирование параметров пробных сфер (r_m и q_m) до достижения минимальной погрешности уравнения (4) по оценке “скользящий контроль”. Изменение параметров изменяет силовое поле взаимодействия. В новом поле молекулы могут переориентироваться, поэтому после каждого изменения параметров проводится оптимизация расположения

каждого из конформеров в поле измененного модельного рецептора. Варьирование r_m и q_m производится методом градиентного спуска.

Мультитаутомерная оценка биологической активности в рамках алгоритма BiS

Очевидной является зависимость биологической активности от таутомерного состояния соединений. Достаточно вспомнить различия молекулярной и цвиттер-ионной форм аминокислот. Поэтому учет таутомерных форм лекарственных средств является весьма важным. Таутомерные формы соединения существуют в некотором динамическом равновесии. Аналогично конформерам, максимальной активностью может обладать не самая выгодная таутомерная форма. Более активной может являться таутомерная форма, присутствующая в меньшей концентрации. Кроме того, в комплексе с рецептором может быть таутомер, минорный в свободном состоянии. То есть минорная таутомерная форма, в этом случае будет являться более комплементарной к рецептору, нежели доминирующая. Данный таутомер будет определять биологическое действие соединения в целом. При образовании комплекса “рецептор - лиганд” этот таутомер выводится из раствора, а в растворе вновь устанавливается динамическое равновесие с прежней долей активного таутомера. Таким образом, процесс биологического действия определяется долей активной таутомерной формы и скоростью установления таутомерного равновесия. Тогда активность соединения (A) может быть определена как суперпозиция активностей таутомеров (a_i):

$$A = \sum_{i=1}^{L} a_i p_i^t$$

где p_i^t – вероятность нахождения i-го таутомера; L – число возможных таутомеров.

Вероятность нахождения i-го таутомера может быть рассчитана по формуле:

$$p_i^t = \frac{\exp(-E_i / kT)}{\sum_{j=1}^{L} \exp(-E_j / kT)}$$

где E_i, E_j – энергии i-го и j–го таутомеров, соответственно.

Если, считать, что биологическая активность каждого таутомера пропорциональна энергии взаимодействия “рецептор – лиганд”, то

$$a_i = b_0 + b_1 E_i^R \qquad (8)$$

где E_i^R – энергия взаимодействия рецептора с i-ым таутомером; b_0 и b_1 – константы.

Считая, что активность вещества есть суперпозиция активностей таутомеров с учетом их доли в веществе, формула (8) может быть переписана в виде

$$BA = \sum_{i=1}^{L} p_i^t \left(a + bE_i^R\right) = a\sum_{i=1}^{L} p_i^t + b\sum_{i=1}^{L} p_i^t E_i^R \quad (9)$$

поскольку $\sum_{i=1}^{L} p_i^t = 1$, то

$$BA = a + b\sum_{i=1}^{L} p_i^t E_i^R \qquad (10)$$

Для использования уравнения (10) необходимо определить число таутомеров, долю каждого из них в веществе, рассчитать энергию взаимодействия с рецептором. Тогда метод анализа биологической активности может быть создан в виде комбинации метода поиска таутомеров с определением их доли в веществе и 3D-QSAR алгоритма, обеспечивающего построение комплекса “модельный рецептор – лиганд” и определение энергий взаимодействия рецептора с таутомерами. Коэффициенты a и b могут быть определены методами регрессионного анализа. В отличие от мультиконформационного подхода, для таутомерного анализа сложно предложить автоматическую схему генерации возможных

таутомеров, поскольку наличие возможных таутомерных форм определяется набором эмпирических правил. Кроме того, изменение таутомерной формы влечет за собой изменение молекулярно-топологической формы, что усложняет быстрый анализ энергий таутомеров методами молекулярной механики. Поэтому в работе [25] были предположены возможные таутомеры, а расчет их энергий проводился *ab initio* DFT B3LYP/6-31*.

Каждая таутомерная форма соединения может иметь ряд конформаций, поэтому более полным будет описание, объединяющее мультитаутомерный и мультиконформационный подходы. Тогда уравнение (10) с учетом уравнения (5) будет иметь следующий вид:

$$BA = \sum_{i=1}^{L} p_i^t \sum_{j=1}^{K} p_j \left(a + bE_j^R\right) = a\sum_{i=1}^{T} p_i^t \sum_{j=1}^{L} p_j + b\sum_{i=1}^{T} p_i^t \sum_{j=1}^{L} p_j E_j^R \qquad (11)$$

поскольку $\sum_{i=1}^{L} p_i^t \sum_{j=1}^{K} p_j = 1$, то

$$BA = a + b\sum_{i=1}^{L} p_i^t \sum_{j=1}^{K} p_j E_j^R \qquad (12)$$

В рамках алгоритма BiS формируется модельный рецептор, единый для всех таутомеров и конформеров всех молекул. Рассчитывается энергия взаимодействия каждого конформера и каждого таутомера с модельным рецептором. Ориентация каждого таутомера и конформера в поле модельного рецептора оптимизируется до достижения минимума потенциальной энергии взаимодействия. Вычисляется энергия взаимодействия данной молекулы (**E**) во всем множестве ее таутомерных и конформационых форм с модельным рецептором

$$\mathbf{E} = \sum_{i=1}^{L} p_i^t \sum_{j=1}^{K} p_j E_j^R \qquad (13)$$

Такие операции проводятся со всеми прочими молекулами выборки и устанавливается зависимость вида $BA = a + b\mathbf{E}$, что при подстановке (13) приводит к уравнению (12).

Проводится варьирование параметров псевдоатомов до достижения минимальной погрешности уравнения (12) по оценке “скользящий контроль”. Изменение параметров r_m и q_m изменяет силовое поле взаимодействия, при этом в новом поле молекулы могут переориентироваться, поэтому после каждого изменения параметров проводится оптимизация расположения каждого таутомера и конформера в поле измененного модельного рецептора. Варьирование r_m и q_m проводится методом градиентного спуска.

Оценка биологической активности стереоизомеров в рамках алгоритма BiS

Зависимость биологической активности от стереоизомерной формы соединения хорошо известна. Синтез хиральных лекарств является весьма перспективным направлением. В ходе синтеза зачастую образуется рацемат, биологически активным компонентом которого является лишь один из стереоизомеров. Теоретическое выявление активного стереоизомера является важным, поскольку в этом случае есть основание для стереоспецифического синтеза. При взаимодействии с рецептором только один из стереоизомеров является активным. Данный изомер будет определять биологическое действие соединения, поскольку процессы конфигурационной изомеризации являются, как правило, достаточно медленными. Тогда биологическая активность соединения (A) может быть определена как суперпозиция активностей изомеров (a_i), то есть:

$$A = \sum_{i=1}^{I} a_i p_i^{is}$$

где p_i^{is} – вероятность нахождения i-го изомера; I – число возможных изомеров.

В случае энантиомеров их вероятности одинаковы. При рассмотрении же прочих стереоизомеров (диастереомеров, Z, E – изомеров и т.д.) вероятности следует учитывать.

Если синтез проводится в термодинамических условиях, то вероятность нахождения *i*-го изомера может быть рассчитана по формуле:

$$p_i^{is} = \frac{\exp(-E_i / kT)}{\sum_{j=1}^{T} \exp(-E_j / kT)}$$

где E_i, E_j – энергии *i*-го и *j*–го изомеров, соответственно.

В ином случае вероятность пропорциональна выходу каждого изомера в синтезе.

Если считать, что биологическая активность каждого изомера пропорциональна энергии взаимодействия “рецептор – лиганд”, то

$$a_i = b_0 + b_1 E_i^R \qquad (14)$$

где E_i^R – энергия взаимодействия рецептора с *i*-ым изомером; b_0 и b_1 – константы.

Считая, что активность вещества есть суперпозиция активностей изомеров с учетом их доли в веществе, формула (14) может быть переписана в виде

$$BA = \sum_{i=1}^{I} p_i^{is} \left(a + bE_i^R\right) = a\sum_{i=1}^{I} p_i^{is} + b\sum_{i=1}^{I} p_i^{is} E_i^R \qquad (15)$$

поскольку $\sum_{i=1}^{I} p_i^{is} = 1$, то

$$BA = a + b\sum_{i=1}^{I} p_i^{is} E_i^R \qquad (16)$$

Для использования уравнения (16) необходимо определить число изомеров, долю каждого из них в веществе, рассчитать энергию взаимодействия с рецептором. Тогда, метод анализа биологической активности может быть создан в виде комбинации метода поиска изомеров с определением их доли в веществе и 3D-QSAR алгоритма, обеспечивающего построение комплекса “модельный рецептор – лиганд” и определение энергий взаимодействия в комплексе. Коэффициенты *a* и *b* могут быть определены методами регрессионного анализа. Для учета различных стереоизомеров соединения возможнен поиск всех хиральных центров и кратных связей с обращением конфигурации. Несколько проще произвести генерацию возможных энантиомеров путем изменения знаков одной из координат всех атомов молекулы. Каждый найденный стереоизомер следует сопоставлять с ранее обнаруженными. Для этого могут быть использованы процедуры, описанные для поиска конформеров.

Каждый стереоизмер может присутствовать виде нескольких таутомеров и конформеров, поэтому более полным будет описание, объединяющее рассмотрение стереоизомеров, мультитаутомерный и мультиконформационный подходы. Тогда уравнение (15) с учетом уравнения (11) будет иметь следующий вид:

$$BA = \sum_{i=1}^{I} p_i^{is} \sum_{j=1}^{L} p_j^t \sum_{l=1}^{K} p_l \left(a + bE_l^R\right) = a\sum_{i=1}^{I} p_i^{is} \sum_{j=1}^{L} p_j^t \sum_{l=1}^{K} p_l + b\sum_{i=1}^{I} p_i^{is} \sum_{j=1}^{L} p_j^t \sum_{l=1}^{K} p_l E_l^R$$

поскольку $\sum_{i=1}^{I} p_i^{is} \sum_{j=1}^{L} p_j^t \sum_{l=1}^{K} p_l = 1$, то

$$BA = a + b\sum_{i=1}^{I} p_i^{is} \sum_{j=1}^{L} p_j^t \sum_{l=1}^{K} p_l E_l^R \qquad (17)$$

В рамках алгоритма BiS формируется модельный рецептор, единый для всех стереозомеров, таутомеров и конформеров всех молекул. Рассчитывается энергия взаимодействия каждого стереоизомера, конформера и таутомера с модельным рецептором. Ориентация каждого стереозоимера, таутомера и конформера в поле модельного рецептора оптимизируется до достижения минимума потенциальной энергии взаимодействия.

Вычисляется энергия взаимодействия данной молекулы (**E**) во всем множестве стереоизомеров, таутомеров и конформеров с модельным рецептором

$$\mathbf{E} = \sum_{i=1}^{I} p_i^{is} \sum_{j=1}^{L} p_j^{t} \sum_{l=1}^{K} p_l E_l^R \quad (18)$$

Аналогичные операции проводятся со всеми прочими молекулами выборки и устанавливается зависимость вида

$BA = a + b\mathbf{E}$,

что при подстановке выражения (18) приводит к искомому уравнению (17).

Проводится варьирование параметров пробных сфер до достижения минимальной погрешности уравнения (17) по оценке “скользящий контроль”. Изменение величин r_m и q_m изменяет силовое поле взаимодействия. В новом поле молекулы могут переориентироваться, поэтому после каждого изменения параметров проводится оптимизация расположения каждого стереозомера, таутомера и конформера в поле измененного модельного рецептора. Варьирование r_m и q_m производится методом градиентного спуска. Полученные закономерности позволяют выделить наиболее активные стереоизомеры, конформационные и таутомерные формы соединения, что показано в работах [10-12,25].

Использование алгоритма BiS для классификации

В многообразии методов анализа и прогноза биологической активности важную роль играют методы классификации, позволяющие решить вопрос активно ли соединение без уточнения количественной величины активности. Для классификации используются различные методы: статистический, дискриминантный и кластер-анализ, нейронные сети и методы распознавания образов. Особый подход на основании анализа атомных окрестностей используется в алгоритме PASS [26-28]. Несмотря на существенные успехи алгоритмов QSAR, методы классификации не теряют своего значения. Связано это с рядом причин.

Зачастую не представляется возможным количественное сопоставление результатов биологических испытаний, проведенных в разных лабораториях (различные методики, инструменты исследования, квалификация реактивов, культур, подготовка персонала). Качественное сравнение результатов (активно или неактивно) при этом возможно.

Использование алгоритмов QSAR (и особенно 3D-QSAR) имеет смысл для активных соединений, поскольку причиной активности соединений является взаимодействие с целевым рецептором. Причины же отсутствия активности могут быть самыми различными (низкая липофильность, диффузионная способность, низкая стабильность и т.д.). Поэтому перед количественным прогнозом биологической активности, имеет смысл выяснить, является ли соединение потенциально активным. При положительном решении можно произвести попытку количественного прогноза. Кроме того, для широкого круга заболеваний на настоящий момент имеются эффективные лекарственные средства и стоит задача не столько создания более эффективных препаратов, сколько лекарств, обладающих меньшими побочными или аллергическими действиями, меньшей токсичностью, себестоимостью, большим удобством применения и т.д. Таким образом, алгоритмы классификации не только не утеряли свою значимость, но и приобрели новый смысл.

В данном пункте показано использование количественного алгоритма BiS для классификации. Достаточно простым подходом к классификации соединений является предположение пороговой величины активности, ниже которой соединение считается неактивным, выше – признается активным. Если предположить значение активности неактивных соединений равной 0, а активных – равной 1, то можно использовать алгоритм BiS, принимая, что значение BA уравнениях (6, 12, 17) принимает значения 0 или 1. В общем случае расчетные значения BA будут не целыми и не будут соответствовать номерам классов (0 или 1). В то же время для активных соединений обучающей выборки можно вычислить среднее значение расчетной величины BA (BA_1) и величину стандартного отклонения (σ_1). Для неактивных соединений рассчитываются BA_0 и σ_0. Исходя из гауссова распределения,

вычисляется вероятность отнесения соединения к классу активных и неактивных. Данная процедура производится с учетом стереоизомерии, конформационного и таутомерного состояния. Данный алгоритм получил название Cl_BiS. Проведен виртуальный скрининг базы данных соединений компании TOSLAB. Результаты скрининга представлены на сайте компании www.toslab.com . На настоящий момент алгоритм Cl_BiS обеспечивает возможность прогноза спектра из 43 видов биологической активности.

Заключение

Создана комбинированная вычислительная технология BiS с возможностями количественного и качественного анализа и прогноза биологической активности с учетом стереоизомерии, конформационного и таутомерного состояния. Ведется создание параллельной версии комбинированного алгоритма на кластере СКИФ Союзного Государства России и Белоруссии.

Литература

1. *Shapiro G., Floersheim P., Boelsterli J. et al.* Muscarinic Activity of the Thiolactone, Lactam, Lactol, and Thiolactol Analogues of Pilocarpine and a Hypothetical Model for the Binding of Agonists to the m1 Receptor// J. Med. Chem. 1992. V. 35. P. 15-27.
2. *Masek B.B., Merchant A., Matthew J.B.* Molecular Shape Comparison of Angiotensin II Receptor Antagonists// J. Med. Chem. 1993. V. 36, P. 1230-1238.
3. *Miller M.D., Sheridan R.P., Kearsley S.K.* SQ: A Program for Rapidly Producing Pharmacophorically Relevent Molecular Superpositions// J. Med. Chem. 1999. V. 42. P. 1505-1514
4. *Grant J.A., Gallardo M.A., Pickup B.T.* A Fast Method of Molecular Shape Comparison: A Simple Application of a Gaussian Description of Molecular Shape// J. Comput. Chem. 1996. V. 17. 1653-1666.
5. *Potemkin V. A., Bartashevich E. V., Grishina M. A., Guccione S.* An Alternative Method for 3D-QSAR and the Alignment of Molecular Structures: BiS (Biological Substrate Search)// in: Rational Approaches to Drug Design. (H.-D. Holtje and W.Sippl eds.). Barcelona: Prous Science Publishers, 2001. P. 349–353.
6. *Потемкин В.А., Гришина М.А., Белик А.В., Чупахин О.Н.* Исследование количественной взаимосвязи структура – антибактериальная активность производных хинолона// Хим.-фарм. журн. 2002. Т. 36. № 1. С. 22–25.
7. *Mikuchina K., Potemkin V., Grishina M., Laufer S.* 3D QSAR analysis and pharmacophore modelling of p38 MAP kinase inhibitors using BiS algorithm// Arch. Pharm. Pharm. Med. Chem. 2002. V. 335. № 1. P. C74.
8. *Mikouchina K., Potemkin V., Laufer S.* Modellstudie zurOrientierung von Hemmstoffen im aktiven Zentrum der p38 MAP Kinase – Anwendung von BIS-Algorithmen// Doktorandentagung, Book of Abstracts.- Dusseldorf, 2003. P. 54.
9. *Grishina M. A., Potemkin V. A., Rusinov G.L.,Bartashevich E. V., Guccione S., Perspicace S, Chupakhin O.N.* Comparing a BiS (Biological Substrate Search) alignment of anticancer DNA and DNA/RNA antimetabolites into their active sites with .pdb co-crystals// “From Genes to Drugs via Crystallography”, 33rd crystallographic course at the E. Majorana Centre, Erice, Italy, 23 May – 2 June 2002. Book of Abstracts. Poster Absracts. P. 48.
10. *Потемкин В.А., Арсламбеков Р.М., Барташевич Е.В. и др.* Мультиконформационный метод анализа биологической активности молекулярных структур// Журн. структ. химии. 2002. Т. 43. № 6. С. 1134–1138.
11. *Гришина М.А., Потемкин В.А., Арсламбеков Р.М., Белик А.В.* 3D-QSAR метод мультиконформационного анализа биологической активности соединений// “Информационно-вычислительные технологии в решении фундаментальных

научных проблем и прикладных задач химии, биологии, фармацевтики, медицины". Тез. Докл. М.: Научный и учебный методический центр, 2002. С. 22 – 23.
12. *Потемкин В.А., Гришина М.А., Микушина К.М., Пожиленкова Г.В.* Комбинированный подход к анализу биологической активности соединений// Информационно-вычислительные технологии в решении фундаментальных и прикладных задач (ИВТН-2003). Сборник материалов. М., 2003. С. 11.
13. *Потемкин В.А., Барташевич Е.В., Белик А.В.* Модель расчета атомных объемных характеристик в молекулярных системах// Журн. физ. химии. 1998. Т. 72. С. 650-656.
14. *Потемкин В.А., Барташевич Е.В., Белик А.В.* Новые подходы к прогнозу термодинамических параметров веществ по молекулярным данным// Журн. физ. химии. 1996. Т. 70. № 3. С. 448-452.
15. *Grishina M.A., Potemkin V.A.* Theoretical investigation of interactions in the molecular systems "receptor-ligand" for DNA and DNA/RNA-antimetabolites// IXth International Seminar on Inclusion Compounds (ISIC-9), Book of Abstracts. Novosibirsk, 2003.- P. 73.
16. *Потемкин В.А., Гришина М.А., Белик А.В.* Моделирование комплексов "рецептор-лиганд" и сопоставление с данными рентгеноструктурного анализа// III Национальная кристаллохимическая конференция, Тез. Докл. Черноголовка, 2003.- С. 71-72.
17. *Doweyko A.M.* The Hypothetical Active Site Lattice Approach to Modeling Active Sites from Data on Inhibitor Molecules// J. Med. Chem. 1988. V. 31. P. 1396-1406.
18. *Das J., Lin J., Moquin R.V. et al.* Molecular Design, Synthesis, and Structure-Activity Relationships Leading to the Potent and Selective $P56^{lck}$ Inhibitor BMS-243117// Bioorg. Med. Chem. Lett. 2003. V. 13. P. 2145-2150
19. *Doweyko A.M.* The Hypothetical Active Site Lattice - In Vitro and in Vivo Explorations Using a Three-Dimensional QSAR Technique// J. Math. Chem. 1991. V. 7. P. 273.
20. *Doweyko A.M., Mattes W.B.* An Application of 3D-QSAR to the Analysis of the Sequence Specificity of DNA Alkylation by Uracil Mustard// Biochemistry. 1992. V. 31. P. 9388-9392.
21. *Doweyko A.M.* Three Dimensional Pharmacofores from Binding Data// J. Med. Chem. 1994. V. 37. P. 1769-1778.
22. *Kaminski J.J., Doweyko A.M.* Antiulcer Agents 6. Analysis of the in vitro Biochemical and in vivo Gastric Antisecretory Activity of Substituted Imidazo[1,2-alpyridines and Related Analogues Using Corporative Molecular Field Analysis and Hypothetical Active Site Lattice // J. Med. Chem. 1997. V. 40. P. 427-436.
23. *Santagati M., Doweyko A., Santagati A. et al*. 5-HT_{1A} Receptors Mapping by Conformational Analysis (2D NOESY/MM) and "THREE WAY MODELLING" (HASL, CoMFA**,** PARM)// Proceedings of the 12th European Symposium on Quantitative Structure-Acticity Relationships. Copenhagen. 1998. P. 23-28.
24. *Guccione S.* PARM, a new genetic algoritm for 3D QSAR Studies // The QSAR and Modeling Society Newsletter. 1998. No. 9. P. 17-18.
25. *Гришина М.А., Потемкин В.А., Микушина К.М., Белик А.В.* Мультитаутомерный анализ биологической активности в рамках алгоритма BiS// I Российская школа-конференция "Молекулярное моделирование в химии, биологии и медицине", Саратов, 18 – 20 сентября 2002. Тез. докл. Саратов, 2002.- С. 42–43.
26. *Глориозова Т.А., Филимонов Д.А., Лагунин А.А., Поройков В.В.* Тестирование компьютерной системы предсказания спектра биологической активности PASS на выборке новых соединений// Хим.-фарм. журн. 1998. Т. 32. № 12. С. 33-39.
27. *Поройков В.В., Филимонов Д.А., Степанчикова А.В. и др*. Оптимизация синтеза и фармакологического исследования веществ на основе компьютерного

прогнозирования их спектров биологической активности// Хим.-фарм. журн. 1996. Т. 30. № 9. С. 20–23.
28. *Филимонов Д.А., Поройков В.В., Караичева Е.И.* Компьютерное прогнозирование спектра биологической активности химических соединений по их структурной формуле: система PASS// Эксп. клин. фармакол. 1985. Т. 58. № 2. С. 56 - 62.