ПРОГНОСТИЧЕСКАЯ МОДЕЛЬ ДЛЯ ОТБОРА СУБСТРАТОВ ИЗОФОРМЫ ЦИТОХРОМА P450 3A4

[1]Погребной А.А., [2]Гришина М.А., [2]Потемкин В.А., [2]Зракова Т.Ю.

[1]Челябинская государственная медицинская академия, г. Челябинск

[2]Челябинский государственный университет, г. Челябинск

Разработана 3D-QSAR модель, характеризующая взаимодействие субстратов изоформы цитохрома P450 3A4 с активным сайтом изоформы и позволяющая прогнозировать величины констант Михаэлиса. Проведён анализ молекулярных параметров, определяющих возможность метаболизма веществ с участием данной изоформы. Использование расчётных величины констант Михаэлиса в сочетании с величинами некоторых молекулярных параметров позволяет проводить отбор веществ, для которых возможен процесс метаболизма с участием изоформы цитохрома P450 3A4.

Цитохром P450 имеет большое значение в проявлении биологической активности различных ксенобиотиков, в том числе и лекарственных веществ. Образующиеся с участием цитохрома P450 продукты метаболизма во многих случаях обладают свойствами, существенно отличающимися от свойств исходных веществ. В результате таких процессов, в различных случаях, может происходить, например, увеличение скорости выделения ксенобиотика из организма, а также усиление исходного либо приобретение совершенно иного типа биологической активности. Поэтому учёт метаболических свойств веществ относительно изоформ цитохрома P450 представляется весьма важным при разработке перспективных лекарственных средств.

В настоящее время для многих разрабатываемых лекарств проводится выяснение роли отдельных изоформ цитохрома P450 в их метаболизме. Получаемая при этом информация весьма важна для безопасного и рационального применения лекарственных веществ с учётом метаболического фенотипа пациента, а также при использовании этого лекарственного вещества совместно с другими препаратами.

Очевидно, что исследование метаболической судьбы одного конкретного вещества, в том числе и с учётом роли изоформ цитохрома P450, проще всего и с большей степенью достоверности осуществимо экспериментальными методами. Однако при разработке лекарственных веществ возможна постановка задачи по подбору вещества, заранее обладающего определёнными метаболическими свойствами. В этом случае использование экспериментальных методов окажется малоэффективным, поскольку число потенциальных кандидатов даже из числа веществ обладающих требуемым фармакологическим эффектом может оказаться достаточно большим.

Для решения поставленной таким образом задачи рациональным подходом может быть использование какой-либо прогностической модели, позволяющей отобрать из большого числа веществ наиболее перспективные кандидаты. При этом предполагается, что окончательная проверка наличия требуемых метаболических свойств будет проводиться экспериментальными методами. В этом случае является вполне допустимым, хотя и не оптимальным, то, что прогностическая модель будет способна отбирать не все вещества, способные обладать требуемыми метаболическими свойствами, а лишь некоторую часть таких веществ.

Более важным для такой прогностической модели будет качество прогноза, которое может быть оценено по соотношению в выделенной по прогнозу группе веществ, действительно обладающих требуемыми метаболическими свойствами, и веществ, случайно отнесённых к этой группе.

Другой не менее важной задачей при разработке подобной прогностической модели является выяснение различных факторов, влияющих на процесс метаболизма вещества, и

выяснение механизма взаимодействия субстрата с участвующей в его метаболизме изоформой.

В настоящей работе предложена прогностическая модель для отбора субстратов изоформы цитохрома P450 3A4.

Построение модели было проведено с использованием выборки субстратов изоформы цитохрома P450 3A4, для которых из литературных данных были известны величины констант Михаэлиса. Данная выборка включала в себя 25 веществ. Трёхмерная структура молекул была предварительно оптимизирована в силовом поле MM+ [1,2] и окончательно с использованием полуэмпирического квантово-химического приближения PM3. В качестве величины, характеризующей субстраты, использовались логарифмы констант Михаэлиса.

На основании исходных данных 3D-QSAR алгоритмом BiS [3] был создан модельный рецептор, характеризующий взаимодействие субстрата с активным сайтом изоформы цитохрома P450 3A4. Коэффициент корреляции экспериментальных и расчётных величин логарифмов констант Михаэлиса составил 0.96. Коэффициент корреляции при «скользящем контроле» составил 0.88. Данная модель, построенная с использованием выборки из 25 субстратов, получила название «модель 1».

Полученный таким образом модельный рецептор представляет собой набор изолированных псевдоатомов, по своим характеристикам близких к атомам углерода, водорода, кислорода, азота. В процессе создания рецептора оптимизируется количество и расположение, а также характеристики таких псевдоатомов. Целью оптимизации является достижение максимальной корреляции экспериментальных и расчётных величин биологической активности, в случае субстратов изоформы цитохрома P450 3A4 — логарифмов констант Михаэлиса.

Очевидно, что построенный описанным способом рецептор, моделирующий активный сайт изоформы, не воспроизводит в точности действительную структуру сайта. Например, один из псевдоатомов может моделировать сразу несколько реальных атомов со схожими характеристиками. Либо, наоборот, один реальный атом, обладающий достаточной подвижностью, может моделироваться сразу несколькими псевдоатомами, расположенными вблизи друг друга. Таким образом, псевдорецептор моделирует не конкретную структуру активного сайта, а некоторое эффективное поле, создаваемое сайтом и обеспечивающее проявление метаболических свойств.

Ещё одной причиной, приводящей к различию модели и реальной структуры сайта, может быть то, что субстраты изоформ цитохрома P450, в том числе и субстраты 3A4, характеризуются значительным структурным разнообразием. Вполне возможно, что среди субстратов изоформы 3A4 существует несколько подгрупп субстратов, для каждой из которых характерен свой тип взаимодействия с активным сайтом. Поэтому представляется рациональным создавать отдельный модельный рецептор для каждой из таких подгрупп. В том случае, если модельный рецептор будет построен с использованием полной выборки, он, очевидно, будет представлять собой суперпозицию нескольких рецепторов, соответствующих отдельным подгруппам субстратов. Кроме того, образующие суперпозицию отдельные рецепторы могут быть искажены в результате оптимизации их структуры с учётом взаимодействия этих рецепторов с субстратами других групп. Либо, как вариант, при оптимизации структуры рецептора могут преимущественно учитываться структурные особенности одной из подгрупп субстратов и игнорироваться особенности субстратов других подгрупп.

Для модельных рецепторов, построенных с использованием алгоритма BiS, возможно выделение подгрупп веществ, для которых имеются различия в характере их взаимодействия с рецептором. Такое выделение может быть осуществлено, например, на основании анализа величин энергий взаимодействия лигандов с рецептором, а также сил, действующих вдоль трех ортогональных осей.

Подобный анализ, проведённый для модели активного сайта изоформы 3A4, показал, что наиболее существенные различия имеются для величин сил F_x, действующих вдоль оси X, направленной от места входа в модельный рецептор к его донной части. На основании данного анализа из исходной выборки субстратов были выделены две группы веществ, для первой из которых величины сил составили от –1.86 до –0.30, а для второй — от 0.13 до 1.23. С точки зрения взаимодействия веществ с модельным рецептором можно трактовать вещества первой группы как выталкивающиеся из рецептора, а вещества второй группы как втягивающиеся в него.

Субстраты изоформы 3A4, использованные для построения модели активного сайта, соответствующие им экспериментальные величины логарифмов констант Михаэлиса LogKmExp и величины действующих вдоль оси X сил F_x представлены в таблице 1.

Таблица 1. Субстраты изоформы 3A4, использованные для построения моделей

Субстраты	LogKmExp	F_x
Celecoxib	2.90	-1.86
Acetaminophen	5.63	-1.14
Oxybutynin	2.80	-0.70
9-cis-Retinal	5.91	-0.64
Pranidipine	2.16	-0.55
Clozapine	4.79	-0.53
Midazolam	0.44	-0.51
all-trans-Retinal	5.54	-0.30
Ziprasidone	5.46	-0.06
Theophylline	10.13	-0.03
Quinidine	4.31	-0.02
Estrone	5.15	0.03
Quinine	4.66	0.04
Indinavir	-3.22	0.06
Brotizolam	3.89	0.07
Terfenadine	3.40	0.13
Fentanyl	4.76	0.18
Tazofelone	2.01	0.24
Delavirdine	1.69	0.27
Salicylate	6.24	0.41
Amitriptyline	4.42	0.52
Tropisetron	5.15	0.54
Zileuton	6.71	0.60
Irinotecan	2.91	0.73
Azelastine	3.93	1.23

Сопоставление экспериментальных величин логарифмов констант Михаэлиса и величин действующих вдоль оси X сил F_x показывает, что данные группы нельзя рассматривать как группы, различающиеся по аффинности субстратов. Для субстратов каждой из выделенных групп имеется примерно одинаковый разброс по величинам LogKmExp, и в то же время не наблюдается существенных различий между группами в целом. Скорее всего, субстраты, относящиеся к разным группам характеризуются различным типом взаимодействия с активным сайтом изоформы.

С использованием выделенных групп веществ алгоритмом BiS также были построены модельные рецепторы. Модель, построенная с использованием выборки субстратов, для которых величины констант составили от –1.86 до –0.30, получила название «модель 2». Модель, построенная с использованием выборки субстратов, для которых величины констант составили от 0.13 до 1.23, получила название «модель 3».

Для построения модели 2 было использовано 8 веществ, для построения модели 3 — 10 веществ. Коэффициенты корреляции экспериментальных и расчётных величин логарифмов констант Михаэлиса для моделей 2 и 3 составили 1.0. Коэффициенты корреляции при «скользящем контроле» также составили 1.0.

Далее была исследована возможность использования расчётных величин логарифмов констант Михаэлиса каждой из моделей для отбора субстратов изоформы цитохрома P450 3A4. При этом использовалась контрольная выборка из 211 веществ, 112 из которых были известны как субстраты изоформы 3A4. Вещества, использованные для создания модельных рецепторов, были исключены из контроля.

Данное исследование проводилось методом распознавания образов с использованием алгоритма «Идеальный эталон» [4]. В качестве дескрипторов были использованы расчётные величины логарифмов констант Михаэлиса (LogKmBiS), молекулярная масса (M), молекулярный объём (V), радиусы вдоль главных осей эллипсоида, описанного вокруг молекулы (A1, A2, A3), энергии ВЗМО и НСМО (HOMO и LUMO), а также разность этих энергий (DeltaE).

Согласно алгоритму отбора, использованному в данной работе, к числу субстратов относились те вещества, для которых расстояние до эталона не превышало заданную пороговую величину ε. Пороговая величина ε выражалась в долях единицы, поскольку дескрипторы использованных для распознавания предварительно нормировались так, чтобы их значения по модулю находились в интервале от 0 до 1.

Для оценки качества прогноза при проведении отбора использовались два количественных показателя, которые вычислялись исходя из фактической и прогнозируемой принадлежности веществ к числу субстратов.

Одним из количественных показателей прогноза являлось качество распознавания, которое определялось как отношение числа объектов, верно отнесённых по прогнозу к классу субстратов, к общему числу объектов, отнесённых по прогнозу к этому классу.

Другим количественным показателем являлся охват выборки, который определялся как отношение числа объектов, верно отнесённых по прогнозу к классу субстратов, к общему числу объектов, фактически принадлежащих классу субстратов в контрольной выборке.

Изменение значения пороговой величины ε позволяет получать различные варианты отбора субстратов, различающиеся по своим количественным показателям. Например, может быть увеличен охват выборки за счёт некоторого снижения качества распознавания. Представляется, что такая возможность может быть полезна при проведении предварительного отбора субстратов из большого числа веществ.

В случае такого отбора окончательная проверка метаболических свойств будет осуществляться экспериментально, и роль отбора будет сводиться к выделению из исходной выборки веществ подгруппы меньшего объёма, но содержащей относительно большее количество перспективных веществ. Очевидно, что возможность получения нескольких вариантов прогноза позволит в случае выделения подгруппы недостаточного объёма увеличить его при снижении качества распознавания в приемлемых пределах.

Проведённое в настоящей работе исследование с использованием модели 1 показало, что отбор субстратов изоформы 3A4 с удовлетворительными количественными показателями может быть получен при использовании в качестве дескрипторов расчётных величин логарифмов констант Михаэлиса в сочетании с некоторыми молекулярными параметрами. В частности при использовании в качестве дескрипторов величин LogKmBiS, энергий ВЗМО (HOMO) и молекулярных объёмов (V) оказалось возможным отобрать 21.8% действительных субстратов этой изоформы с качеством распознавания 100%. Различные варианты отбора веществ при использовании этих дескрипторов представлены в таблице 2.

Использование данных дескрипторов по отдельности позволяет получить такое же качество распознавания только для молекулярного объёма. Охват выборки при этом

составляет 12.5%. Взятые по отдельности величины LogKmBiS и энергии ВЗМО не позволяли осуществлять отбор субстратов с удовлетворительным качеством распознавания.

Таблица 2. Количественные характеристики отбора субстратов при использовании дескрипторов LogKmBiS (модель 1), HOMO и V

ε	Качество распознавания, %	Охват выборки, %
0.1	100.0	21.8
0.15	82.1	26.4
0.2	79.4	31.0
0.25	72.9	40.2
0.3	70.7	47.1

При снижении качества распознавания возможно увеличение охвата выборки, однако в этом случае роль дескриптора LogKmBiS снижается, и при использовании только лишь энергии ВЗМО и молекулярного объёма могут быть получены лучшие показатели прогноза. Количественные характеристики вариантов отбора веществ при использовании этих двух дескрипторов представлены в таблице 3.

Таблица 3. Количественные характеристики отбора субстратов при использовании дескрипторов HOMO и V

ε	Качество распознавания, %	Охват выборки, %
0.1	100.0	4.6
0.15	92.3	13.8
0.2	90.5	21.8
0.25	89.3	28.7
0.3	88.2	34.5
0.4	79.6	44.8
0.5	71.0	56.3

Сопоставление результатов, приведённых в таблицах 2 и 3, позволяет сделать вывод, что дескриптор LogKmBiS улучшает характеристики отбора для ограниченного подмножества веществ, но оказывается малопригодным для выборки большего объёма.

Необходимость использования дополнительных дескрипторов для получения удовлетворительных результатов, очевидно, может быть истолкована следующим образом. Модельный рецептор характеризует только лишь взаимодействие субстрата с активным сайтом изоформы. В то же время для осуществления метаболического процесса молекула также должна удовлетворять некоторым дополнительным условиям.

Например, некоторые молекулярные параметры могут играть роль на этапе транспорта молекулы до момента её попадания в активный сайт изоформы. Такими параметрами могут быть молекулярный объём и радиусы вдоль главных осей эллипсоида, описанного вокруг молекулы. Несмотря на то, что данные параметры в неявном виде учитываются при создании модели активного сайта, их влияние на процесс транспорта может быть иным.

Было проведено сравнение средних величин ряда молекулярных параметров с использованием t-критерия Стьюдента, для веществ, относящихся и не относящихся к числу субстратов изоформы цитохрома P450 3A4. Данное сравнение показало, что для многих молекулярных параметров имеются статистически достоверные различия. Некоторые из таких молекулярных параметров представлены в таблице 4.

Роль молекулярного объёма для субстратов подсемейства 3A хорошо известна. Считается, что субстраты этого подсемейства характеризуются большими величинами молекулярного

объёма, и это обстоятельство может быть использовано для различения веществ, способных и неспособных являться субстратами подсемейства 3А [5].

Молекулярная масса также может иметь значение вследствие её корреляции с молекулярным объёмом. Однако с другой стороны она влияет на динамические характеристики молекулы, что может быть важным для сохранения правильной ориентации молекулы в активном сайте в том случае, если ориентация за счёт межмолекулярного взаимодействия осуществляется недостаточно жёстко.

Энергетические параметры, такие как энергии ВЗМО и НСМО, а также разность этих энергий, определяют возможность переноса электронов, что важно для процессов окисления и восстановления.

Таблица 4. Молекулярные параметры, достоверно различающиеся по средним величинам для веществ, специфичных по отношению к изоформе цитохрома P450 3A4

Молекулярные параметры	t
Молекулярная масса	6.80
Молекулярный объем	6.40
Энергия 1-4 взаимодействий	6.06
Площадь поверхности молекулы	5.34
Максимальное собственное значение матрицы ассоциации	3.77
Энергия перекрестных взаимодействий «связь – торсионный угол»	3.60
Энергия внутримолекулярных ван-дер-ваальсовых взаимодействий	3.14
Энергия ВЗМО	3.01

Максимальное качество распознавания для 100%-ного охвата выборки составило 81.7 % и оказалось достижимым только в факторном пространстве 8 дескрипторов (значения расчетных констант Михаэлиса, полученных в рамках алгоритмов BiS и ConGO, молекулярная масса, два максимальных собственных значения матрицы ассоциации, энергия напряжения связей, энергия внутримолекулярных ван-дер-ваальсовых взаимодействий, энергия перекрестных взаимодействий «связь – торсионный угол»). Использование нейронной сети для классификации со 100%-ным охватом выборки показало, что наилучший вариант наблюдается при использовании пары факторов: молекулярная масса, минимальный радиус описанного вокруг молекулы эллипсоида. Однако качество распознавания в этом случае составляет лишь 66.7%.

В случае модели 2 наиболее оптимальные показатели прогноза были получены при использовании в сочетании с величинами LogKmBiS энергии НСМО (LUMO) и минимального радиуса эллипсоида, описанного вокруг молекулы (A1). Количественные характеристики прогноза при использовании этих дескрипторов представлены в таблице 5. Качество распознавания, равное 100%, оказалось в данном случае недостижимо. Максимальный охват выборки, соответствующий качеству распознавания 100%, полученный при использовании величин LogKmBiS в сочетании с другими молекулярными параметрами, составил для модели 2 не более 14%.

Таблица 5. Количественные характеристики отбора субстратов при использовании дескрипторов LogKmBiS (модель 2), LUMO и A1

ε	Качество распознавания, %	Охват выборки, %
0.15	92.9	37.5
0.2	84.1	51.0
0.25	76.5	59.6
0.3	72.5	68.3
0.35	70.5	71.2

В данном случае, как и в случае модели 1, применение величин LogKmBiS улучшает показатели прогноза по сравнению с использованием только лишь двух оставшихся дескрипторов. Количественные характеристики прогноза при использовании только дескрипторов LUMO и A1 представлены в таблице 6.

Таблица 6. Количественные характеристики отбора субстратов при использовании дескрипторов LUMO и A1

ε	Качество распознавания, %	Охват выборки, %
0.15	100.0	4.6
0.2	90.0	10.3
0.25	85.7	20.7
0.35	73.9	39.1
0.4	73.2	47.1
0.45	71.6	55.2
0.5	70.9	64.4

Сравнение пар значений качества распознавания и охвата выборки из таблиц 5 и 6 показывает, что дескриптор LogKmBiS, рассчитанный с использованием модели 2, может быть полезным при отборе относительно большего числа субстратов, чем это было в случае модели 1.

Величины LogKmBiS, рассчитанные с использованием модели 3, позволили получить наиболее оптимальные показатели прогноза при их использовании совместно с величинами молекулярного объёма (V) и минимального радиуса эллипсоида, описанного вокруг молекулы (A1). Количественные характеристики прогноза при использовании этих дескрипторов представлены в таблице 7.

Таблица 7. Количественные характеристики отбора субстратов при использовании дескрипторов LogKmBiS (модель 3), V и A1

ε	Качество распознавания, %	Охват выборки, %
0.15	100.0	5.9
0.25	94.1	15.7
0.4	83.3	24.5
0.7	81.0	50.0
0.8	76.2	59.8
0.9	71.3	70.6

Количественные характеристики прогноза при использовании только дескрипторов V и A1 представлены в таблице 8.

Таблица 8. Количественные характеристики отбора субстратов при использовании дескрипторов V и A1

ε	Качество распознавания, %	Охват выборки, %
0.1	80.8	37.2
0.15	72.5	49.0
0.2	71.0	64.7
0.25	67.3	66.7
0.3	63.7	70.6
0.35	61.4	76.5

Из результатов, представленных в таблицах 7 и 8 следует, что добавление дескриптора LogKmBiS, рассчитанного по модели 3, значительно улучшает показатели прогноза. Максимальное качество распознавания при использовании только дескрипторов V и A1 составило 80.8% с соответствующим ему охватом выборки 37.2%. Сочетание этих же дескрипторов с LogKmBiS позволяет увеличить качество распознавания до 100%, а с качеством распознавания 81% становится возможным отбор 50% действительных субстратов изоформы 3A4 из контрольной выборки.

Таким образом, представляется целесообразным выделение из исходной выборки двух подмножеств субстратов и построение с использованием каждого из таких подмножеств отдельных прогностических моделей. Структура моделей 2 и 3 значительно упростилась по сравнению с моделью 1, что следует из того, что модель 1 содержит 500 псевдоатомов, модель 2 — 198, а модель 3 — 287. В то же время, очевидно, что построенные модели активных сайтов 2 и 3, точно так же как и модель 1, воспроизводят лишь некоторое эффективное поле, создаваемое сайтом и обеспечивающее проявление метаболических свойств.

Дальнейшая модернизация моделей использованным в работе методом представляется затруднительной из-за малого объёма обучающих выборок. Поэтому совершенствование прогностических моделей, очевидно, потребует как расширения обучающей выборки, так и модернизации методов моделирования.

Работа выполнена при финансовой поддержке Суперкомпьютерной программы СКИФ Союзного Государства России и Белоруссии, а также при финансовой поддержке Правительства Челябинской области (грант № урчел_04-04-96094).

Литература

1. *Буркерт У., Эллинджер Н.Л.* Молекулярная механика. М.: Мир, 1986.
2. *Кларк Т.* Компьютерная химия. М.: Мир, 1990.
3. *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, W. Sippl (Eds.). Proceedings of the 13th European Symp. on Quantitative Structure-Activity Relationships, QSAR 2000 Heinrich-Heine-Universitat, Dusseldorf, Germany, 27 August — 1 September, 2000. Dusseldorf: Prous Science Publishers, 2001. P.349-353.
4. *Потемкин В.А., Белик А.В.* Алгоритм "идеальный эталон" для поиска соединений с заданными свойствами // I Всесоюзная конференция по теоретической органической химии (ВАТОХ): Тез. докл. Волгоград, 1991. С.89.
5. *Lewis D.F.V.* On the recognition of mammalian microsomal cytochrome P450 substrates and their characteristics: towards the prediction of human P450 substrate specificity and metabolism // Biochem. Pharmacol. 2000. Vol.60. No.3. P.293-306.