

Ю. А. Шестеренко, А. Г. Артеменко, П. Г. Полищук,
Е. Н. Муратов, О. В. Севастьянов, И. И. Романовская,
В. Е. Кузьмин, академик НАН Украины С. А. Андронати

QSPR-анализ реакционной способности субстратов тирозины методом случайного леса

За допомогою виділеного частково очищеного препарату тирозинази в розроблених умовах здійснено окиснення широкого спектра фенольних сполук з різним ступенем біоконверсії. Для вивчення зв'язку структура — властивість був використаний метод випадкового лісу; отримано адекватну консенсусну модель з 16%-ю помилкою внутрішньої тестової виборки. Показано, що найбільш значущим фактором у процесі біоконверсії фенолів є геометричні параметри молекули фенолу (~ 70%). Дипольний момент, ліпофільність і електронна поляризованість молекул відіграють менш важливу роль (15, 7 і 6% відповідно).

В настоящее время для разработки современных ферментативных технологий очистки сточных вод от высокотоксичных фенольных соединений используются окислительно-восстановительные ферменты (пероксидаза, тирозиназа) [1]. Тирозиназа (КФ 1.14.18.1) катализирует процесс окисления широкого ряда фенольных соединений, однако структура субстратов оказывает существенное влияние на степень их трансформации.

Известно, что *o*-дифенолы окисляются в присутствии фермента значительно быстрее монофенольных субстратов, поскольку лимитирующей стадией в процессе тирозиназного катализа является введение ОН-группы в *o*-положение молекулы монофенолов. Электроноакцепторные заместители в *n*-положении ароматического кольца субстрата, в отличие от электронодонорных, затрудняют биоконверсию [2]. Так, производные нитрофенола не окисляются в присутствии тирозиназы [3], за исключением *n*-нитропирокатехина, степень трансформации которого значительно ниже, чем для большинства *n*-замещенных *o*-дифенолов [2].

Известна работа, в которой приведены результаты QSAR-анализа¹ с применением CoMFA² и CoMSIA³ методов влияния структуры 48 ингибиторов (производных бензальдегида, бензойной и коричной кислот) на активность тирозиназы — ключевого фермента в биосинтезе меланина — для дизайна новых ингибиторов меланогенеза. С использованием дескрипторов, описывающих стерические и электростатические поля молекул, получены адекватные и прогнозирующие модели CoMFA и CoMSIA. Были также получены 2D QSAR-модели с использованием дескрипторов, включающих электронные (константы Гаммета), гидрофобные (π) и структурные (MR) параметры, водородные связи (H) с их вкладом

¹QSAR (quantitative structure — activity relationships) — количественные соотношения структура — активность.

²CoMFA (comparative molecular field analysis) — сравнительный анализ молекулярных полей.

³CoMSIA (comparative molecular similarity indices analysis) — сравнительный анализ индексов подобия молекул.

в изменение связывания субстрата с активным центром: π 34,9%, MR 31,6%, H 26,7%. Однако в литературе отсутствуют сведения о связи структура фенольного субстрата — степень конверсии в тирозиназном биокатализе с применением QSPR-анализа⁴ [4].

Результаты QSAR-анализа, позволяющие для набора ароматических субстратов пероксидазы хрена прогнозировать относительные изменения степени их биоконверсии, приведены в статье [5]. На основании энергий граничных орбиталей было построено двухпараметровое уравнение и подтверждено влияние электронных свойств заместителей в ароматическом кольце фенолов на их биоконверсию [5].

В настоящей работе мы попытались провести аналогичное исследование связи структура фенольного соединения — степень биоконверсии в реакциях тирозиназного катализа.

Материалы и методы исследований. Частично очищенный препарат тирозиназы из грибов *Agaricus bisporus* выделяли методом Коэна [6]; содержание белка определяли методом Лоури–Хартри [7], меди — с использованием оксалилдигидразида [8], фенолоксидазную активность — по тирозину [1]. За единицу активности принимали количество фермента, вызывающего увеличение оптической плотности на 0,001 (λ 280 нм) 0,25 ммоль/дм³ раствора l-тирозина за 1 мин при pH 6,5 и $t = 25$ °C. Степень трансформации фенолов определяли по их убыли спектрофотометрически [9]. Окисление исследуемых фенольных соединений, катализируемое тирозиназой, проводили при концентрации фермента 120 ед/см³ (pH 6,5, $t = 25$ °C, $\tau = 15$ мин).

Для описания структуры исследуемых соединений применяли интегральные структурные параметры, описывающие геометрические параметры молекул фенольных соединений (моменты инерции вдоль осей координат I_x , I_y , I_z и их отношения), их электронную структуру (E_{HOMO} , E_{LUMO} , дипольный момент), молекулярную рефракцию и липофильность [10]. Энергию граничных орбиталей фенолов рассчитывали полуэмпирическим методом PM3 [11].

Для построения статистических зависимостей структура — свойство использовались методы множественной линейной регрессии [12] и метод случайного леса (СЛ) — сравнительно новый статистический метод анализа, который находит все большее применение для построения QSAR/QSPR моделей [13]. Модель СЛ представляет собой ансамбль отдельных деревьев решений и способа представления правил в иерархической, последовательной структуре, где каждому объекту соответствует единственный узел с решением [14].

Таким образом, с помощью этого метода возможно решение как классификационных, так и регрессионных задач. Каждое из деревьев в модели СЛ строится в соответствии со следующими правилами:

1. Из всего набора соединений обучающей выборки формируется новая, включающая такое же число соединений, которая является обучающей для конкретного дерева. Соединения в новую выборку выбираются случайным образом, при этом любое из них может быть выбрано неоднократно. Не вошедшие в обучающую выборку соединения помещаются во внутреннюю тестовую выборку и используются при оценке качества и прогнозирующей способности модели СЛ.

2. При разделении данных в каждом узле дерева применяется алгоритм CART [14], однако рассматриваются не все переменные, а только небольшая их часть, которая в каждом узле выбирается случайно.

⁴QSPR (quantitative structure — property relationships) — количественные соотношения структура — свойство.

3. Каждое дерево строится до максимально возможной глубины, процедура отсечения ветвей отсутствует.

Для регрессионных задач прогноз осуществляется усреднением всех прогнозов отдельных деревьев в лесе. Для классификационных задач прогноз проводят по наибольшему числу голосов, поданных за какой-либо класс. Аналогичным образом осуществляется прогноз и для внутренней тестовой выборки, каждое “дерево” выдает прогноз только для тех соединений, которые не вошли в обучающую выборку этого конкретного дерева. Величина ошибки классификации внутренней тестовой выборки — определяющий параметр при выборе конечной модели.

Результаты и их обсуждение. Из грибов *Agaricus bisporus* выделен частично очищенный препарат тирозиназы с выходом по белку 0,67 мг/г грибов, содержанием меди 0,19%, удельной активностью 500 ед/мг белка в минуту. Показано, что 90% общего белка обладают фенолоксидазной активностью (нативный электрофорез в ПААГ).

Частично очищенный препарат тирозиназы в разработанных условиях (рН 6,5, $t = 25$ °С, $\tau = 15$ мин) катализировал окисление широкого спектра фенольных соединений с различной степенью биоконверсии, зависящей от их структуры (табл. 1).

Попытка обнаружить связь структура — свойство с помощью метода пошаговой множественной линейной регрессии не увенчалась успехом: статистически адекватных уравнений построить не удалось. Поэтому для решения данной задачи использовали нелинейный метод СЛ, обладающий следующими преимуществами: отсутствием проблемы переобучения моделей и необходимости предварительного отбора переменных; наличием адекватной внутренней процедуры оценки качества и прогнозирующей способности моделей; устойчивостью моделей к наличию “шума” в исходной выборке; эффективностью работы с большими базами данных; интерпретируемостью получаемых моделей; возможностью корректно анализировать выборки, включающие соединения с различным механизмом действия.

Исследуемые соединения (табл. 1, структуры 1–25) в зависимости от степени биоконверсии были разделены на два класса: с высокой (больше 80%) и низкой (меньше 80%).

В результате была получена адекватная консенсусная модель (200 “деревьев”) с 16%-й ошибкой внутренней тестовой выборки.

Важной особенностью метода СЛ является возможность оценить степень влияния различных структурных параметров на исследуемое свойство. Для нашей задачи такой анализ показал (рис. 1), что наиболее значимым фактором в процессе биоконверсии фенолов тирозиназой являются геометрические характеристики молекулы фенола, оцененные с помощью дескрипторов эллипсоида инерции ($\sim 70\%$), что, по всей вероятности, обусловлено необходимостью соответствия молекул фенолов полости фермента. Дипольный момент, липофильность и электронная поляризуемость молекул играют менее важную роль (15, 7 и 6% соответственно). Хотя влияние параметров, отражающих энергии граничных орбиталей, весьма мало (меньше 1%), при их исключении из леса увеличивается ошибка классификации для внутренней тестовой выборки.

Полученная модель была использована для качественного предсказания степени биоконверсии соединений, описанных ранее в литературе (см. табл. 1, структуры 26–35) [3, 15]. Для этих соединений модель также показала достаточно хорошую прогнозирующую способность: ошибка классификации для данного набора (внешний тест) составила 20%.

Таблица 1. Наблюдаемые и предсказанные значения степени трансформации фенолов тирозиназой

Номер структуры	Структура	Степень конверсии, %	Класс	
			наблюдаемый	предсказанный
1	1,2,3-Триоксибензол	100,0	1	1
2	1,3-Диоксибензол	0,0	2	2
3	2,4,5-Трихлорофенол	28,6	2	2
4	2,4,6-Трихлорофенол	33,1	2	2
5	2,4-Дихлорофенол	12,3	2	2
6	2-Аминофенол	36,8	2	2
7	2-Крезол	23,1	2	2
8	2-Хлорофенол	16,7	2	2
9	3,4-Дигидроксифенил-пропионовая кислота	100,0	1	1
10	3,4-Дигидроксифенил-уксусная кислота	100,0	1	1
11	3-Крезол	54,5	2	2
12	3-Хлорофенол	47,2	2	2
13	4-Аминофенол	0,0	2	2
14	4-Бромфенол	41,3	2	2
15	4-Гидроксифенил-пропионовая кислота	100,0	1	1
16	4-Гидроксифенилуксусная кислота	100,0	1	1
17	4-Метил-1,2-диоксибензол	100,0	1	1
18	4-Метоксифенол	99,0	1	1
19	4-Трет-бутил-1,2-диоксибензол	100,0	1	1
20	4-Хлорофенол	59,3	2	2
21	4-Этоксифенол	100,0	1	1
22	1-Нафтол	10,7	2	2
23	Пентахлорофенол	34,2	2	2
24	Тирозин	66,7	2	2
25	Фенол	63,1	2	2
26	L- α -Метилтирозин [15]	—	2	1
27	L-Изопреналин [15]	—	2	1
28	L-Диоксифенилаланин [15]	—	1	1
29	L- α -Метилдиоксифенилаланин [15]	—	2	2
30	Дофамин [15]	—	1	1
31	Метилловый эфир [15] диоксифенилаланина	—	1	1
32	Тирамин [15]	—	1	1
33	o-Нитрофенол [3]	—	2	2
34	n-Нитрофенол [3]	—	2	2
35	m — Нитрофенол [3]	—	2	2

Таким образом, полученные результаты позволяют предположить, что в отличие от процесса окисления фенолов пероксидазой, где решающую роль играют электронные

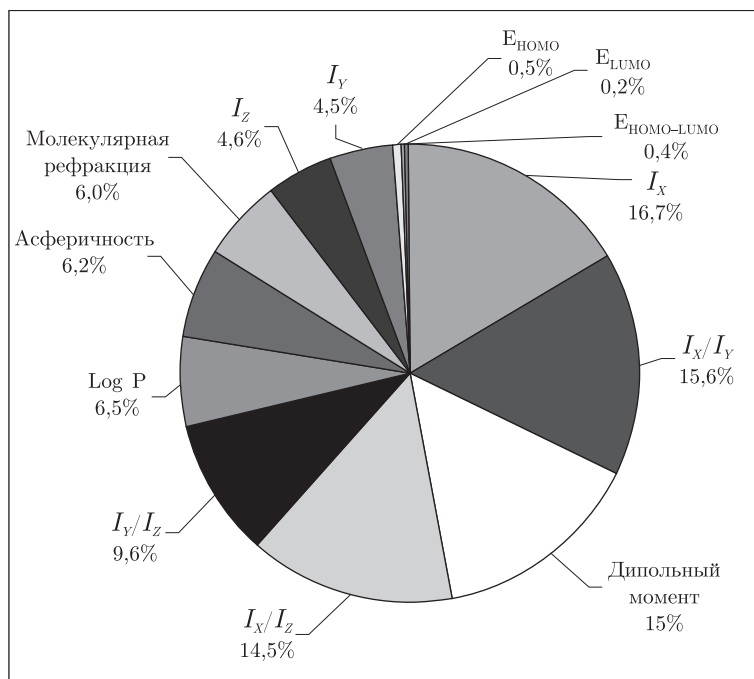


Рис. 1. Оценка влияния различных структурных параметров на степень конверсии фенолов, полученная на основе интерпретации модели СЛ.

E_{LUMO} (Lowest Unoccupied Molecular Orbital) — энергия нижней вакантной молекулярной орбитали; E_{HOMO} (Highest Occupied Molecular Orbital) — энергия высшей занятой молекулярной орбитали; I_x , I_y , I_z — моменты инерции в главных осях координат; Log P — липофильность

факторы реакционной способности фенолов, в процессе окисления фенолов тирозиназой большее влияние оказывают стерические факторы (геометрические характеристики молекулы).

1. *Ikehata K., Nicell J.* Color and toxicity removal following tyrosinase-catalyzed oxidation of phenols // *Biotechnol. Prog.* – 2000. – **16**, No 4. – P. 533–540.
2. *Halaoui S., Asther M., Sigoillot J.-C. et al.* Fungal tyrosinases: new prospects in molecular characteristics, bioengineering and applications // *J. Appl. Microbiol.* – 2006. – **100**, No 2. – P. 219–232.
3. *Conrad J., Dawso S., Hubbard E. et al.* Inhibitor binding to the binuclear active site of tyrosinase: temperature, pH, and solvent deuterium isotope effects // *Biochemistry.* – 1994. – **33**, No 19. – P. 5739–5744.
4. *Xue C.B., Luo W.C., Ding Q. et al.* Quantitative structure-activity relationship studies of mushroom tyrosinase inhibitors // *Journal of Computer-Aided Molecular Design.* – 2008. – **22**, No 5. – P. 299–309.
5. *Романовская И. И., Муратов Е. Н., Кузьмин В. Е. и др.* Анализ влияния структуры фенольных соединений на степень их ферментативной конверсии // *Доп. НАН України.* – 2006. – № 9. – С. 159–164.
6. *Романовська І. І., Шестеренко Ю. А., Севастьянов О. В., Топтіков В. А.* Дослідження складу частково очищеного препарату тирозинази з грибів *Agaricus bisporus* // *Медична хімія.* – 2008. – **10**, № 2. – С. 79–82.
7. *Hartree E. F.* Determination of protein: a modification of the Lowry method, that gives a linear photometric response // *Anal. Biochem.* – 1972. – **48**, No 1. – P. 422–427.
8. *Stark G. R., Dawson C. R.* Spectrophotometric microdetermination of copper in copper oxidases using oxalyldihydrazide // *Analytical Chemistry.* – 1958. – **30**, No 2. – P. 191–194.
9. *Кореньман И. М.* Фотометрический анализ. Методы определения органических соединений. – Москва: Химия, 1975. – 368 с.
10. *Wang R., Fu Y., Lai L.* A new atom-additive method for calculating partition coefficients // *J. Chem. Inf. Comput. Sci.* – 1997. – **37**. – P. 615–621.

11. *Hyper Chem 6.0* – trial version. Hyperchem Software. – Hypercube Inc., FL USA. <http://www.hyper.com>.
12. Ферстер Э., Ренц Б. Методы корреляционного и регрессионного анализа. – Москва: Финансы и статистика, 1983. – 324 с.
13. Breiman L. Random Forests // Machine Learning. – 2001. – 45. – P. 5–32.
14. Breiman L., Friedman J. H., Olshen R. A., Stone C. J. Classification and Regression Trees. – Belmont, CA: Wadsworth, 1984. – 368 p.
15. Espin J. C., Varon R., Fenoll L. G. et al. Kinetic characterization of the substrate specificity and mechanism of mushroom tyrosinase // Eur. J. Biochem. – 2000. – 267. – P. 1270–1279.

Фізико-хімічний інститут ім. А. В. Богатського
НАН України, Одеса

Поступило в редакцію 23.11.2009

**Yu. A. Shesterenko, A. G. Artemenko, P. G. Polischuk, E. N. Muratov,
O. V. Sevastyanov, I. I. Romanovskaya, V. E. Kuz'min,**
Academician of the NAS of Ukraine **S. A. Andronati**

QSPR-analysis of tyrosinase substrates reactivity by the random forest method

With a help of isolated and partially purified tyrosinase preparation under the conditions developed, the oxidation of a wide range of phenolic compounds with different degrees of bioconversion is conducted. For the investigation of the structure – property relationship, the method of random forest is used; an adequate consensus model is obtained. It predicts the out-of-bag set with only 16% of misclassification errors. It is shown that the most significant factor in the bioconversion of phenols is the geometric characteristics of phenol's molecule (~ 70%). The dipole moment, lipophilicity, and electronic polarizability of a molecule plays a less significant role (15, 7, and 6%, respectively).