

УДК 615.35.015.11

А. Л. Осипов, Р. Д. Семенов, Д. А. Грозенок

(Новосибирск)

МОДЕЛИРОВАНИЕ ЗАВИСИМОСТИ КОНЦЕНТРАЦИЯ — ЭФФЕКТ  
ДЛЯ СИСТЕМНЫХ ФУНГИЦИДОВ-ИНГИБИТОРОВ  
СИНТЕЗА ЭРГОСТЕРИНА

Рассматривается математическая модель получения унифицированных индексов активности путем компьютерной обработки патентных данных по биологическим испытаниям химических веществ.

**Введение.** Системные фунгициды-ингибиторы синтеза эргостерина в настоящее время являются предметом интенсивных исследований, приведших к появлению высокоэффективных препаратов с малыми дозами применения [1].

Спектр активности и системные свойства этих препаратов варьируют в широких пределах. Они зависят от структурных элементов, входящих в соединения, причем рассматриваемую группу фунгицидов отличает наличие одного биохимического механизма действия и разнообразие химических классов, охватывающих широкий спектр *N*-содержащих гетероциклических соединений.

Анализ связи структура — фунгицидная активность на основе имеющейся патентной и литературной информации с целью направленного конструирования новых фунгицидов осложнен разнородностью биологических данных: использованием в различных литературных источниках разных наборов фиксированных концентраций или доз действующих веществ, разных тест-объектов и методик тестирования, разных шкал (градаций) биологической активности. Указанная сложность является общей при использовании патентных и литературных данных для проблемы анализа связи структура — биологическая активность.

1. Математическая модель зависимости концентрация — эффект. В настоящей работе предложена феноменологическая модель связи концентрация (доза) — эффект, позволяющая получать унифицированные индексы активности (типа  $СК_{50}$ ,  $ЕД_{50}$ ) путем обработки патентных данных для выборок химических соединений с одинаковым местом действия. Подход применен к данным по системным фунгицидам различных химических классов [2].

Статистический анализ зависимостей активности ( $E$ ) — концентрация ( $C$ ) или доза ( $D$ ) для биологически активных соединений (БАС) показал, что для каждой отдельной методики тестирования пестицидной или фармакологической активности в интервале  $E = 20 \div 80 \%$  и в более широком интервале  $10 \div 90 \%$  выполняется статистическая гипотеза

$$\frac{dE_{\alpha}(C)}{d \lg C} = \text{const.} \quad (1)$$

Выполнение этой гипотезы позволяет выдвинуть более общую гипотезу

$$E_\alpha = \varphi_\alpha(\lg C_i - \lg C_{Pi}) + \Pi, \quad (2)$$

где  $\varphi_\alpha$  — некоторая универсальная функция для конкретного механизма действия и методики испытания на биологическую активность (различаемые индексом  $\alpha$ );  $C_i$  — концентрация  $i$ -го БАС;  $C_{Pi}$  — величина концентрации, отвечающей стандартной величине отклика (активности)  $E = \Pi$ .

В случае монотонных зависимостей активности от концентрации, которые в дальнейшем и будут анализироваться, существует обратная функция  $F_\alpha = \varphi_\alpha^{-1}$ , и модель (2) можно переписать в виде (индекс  $\alpha$  опустим)

$$\lg C_i = \lg C_{Pi} + F(E - \Pi). \quad (3)$$

Поскольку при  $E = \Pi$   $C_i = C_{Pi}$ , то должно выполняться условие

$$F(0) = 0. \quad (4)$$

Обозначим параметры функции  $F$ , подлежащие определению по экспериментальным данным, через  $a_m$ ,  $m = \overline{1, n}$ , т. е.  $F = F(E - \Pi; \{a_m\})$ , и пусть индексом  $k_i$  нумеруются величины концентраций для  $i$ -го БАС. Тогда модель (3) примет вид

$$\lg C_{ik_i} = \lg C_{Pi} + F(E_{k_i} - \Pi; \{a_m\}). \quad (5)$$

Параметры  $a_m$ ,  $m = \overline{1, n}$ , и  $\lg C_{Pi} = a_{0i}$  можно определить, например, минимизацией целевой функции

$$S = \sum_{i=1}^N \sum_{k_i=1}^{N_i} [\lg C_{ik_i} - a_{0i} - F(E_{k_i} - \Pi; \{a_m\})]^2, \quad (6)$$

где  $C_{ik_i}$  и  $E_{k_i}$  — экспериментальная концентрация и соответствующая ей биологическая активность для  $i$ -го БАС;  $N$  — число БАС в обрабатываемом массиве экспериментальных данных;  $N_i$  — число концентраций для  $i$ -го БАС в этом массиве.

Система  $N + n$  уравнений метода наименьших квадратов (МНК) получается дифференцированием  $S$ :

$$\begin{cases} \frac{\partial S}{\partial a_{0i}} = 0, & i = \overline{1, N}, \\ \frac{\partial S}{\partial a_m} = 0, & m = \overline{1, n}. \end{cases} \quad (7)$$

Рассмотрение приведенных в литературе разнообразных биологических данных по пестицидной активности химических соединений показывает, что зависимости  $E = f(C)$ , а с ними и зависимости  $C = f_1(E)$ ,  $\lg C = f_2(E)$  имеют, как правило, пороговый характер, т. е. значениям  $E = 0$  и  $E = 100\%$  отвечают конечные значения концентраций БАС:  $C_{0i} > 0$ ,  $C_{ii} < \infty$  ( $i$  — номер БАС). Зависимости являются гладкими, что позволяет аппроксимировать их в рам-

ках эмпирического подхода полиномами либо другими подходящими функциями. Простейшей функцией является полиномиальная:

$$\lg C_i = \sum_{m=1}^n a_m x^m + a_{0i}, \quad (8)$$

где

$$x = E - \Pi, \quad a_{0i} = \lg C_{\Pi i}. \quad (9)$$

Для этой модели система уравнений (7) является линейной и решается с использованием программного комплекса, описанного в [3].

В качестве стандартного уровня активности  $\Pi$  может быть взята любая величина в интервале  $0 \div 100\%$  (нами использовано значение  $\Pi = 50\%$ ). После определения коэффициентов модели (8) по ней могут быть рассчитаны величины  $\lg C_i(\tilde{\Pi})$  для любого заданного уровня активности  $\tilde{\Pi}$ , которые образуют набор унифицированных индексов для этого уровня активности.

Порядок полинома в модели (8) определяется объемом исходных данных (числом соединений, изученных концентраций или доз). Минимальный порядок, приводящий к наблюдаемому на практике сигмоидному характеру зависимостей (2), использованный в данной работе, равен 3.

В патентных и литературных источниках, наряду с данными для различных концентраций со значениями активности в интервале  $0 \div 100\%$ , встречаются данные для соединений со значениями активности 0 или 100% для нескольких различных концентраций. При обработке данных для таких соединений следует использовать максимальную концентрацию из подмножества с  $E = 0\%$  и минимальную из подмножества с  $E = 100\%$ . Для соединений, имеющих исходные данные по биологической активности, заданные в баллах, требуется производить преобработку их по формуле

$$E = 100 \frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}}, \quad (10)$$

где  $E$  — нормированная активность (%);  $Y$  — значение активности в баллах;  $Y_{\max}$  — балл максимальной активности;  $Y_{\min}$  — балл минимальной активности.

Т а б л и ц а 1

Номер соединения	Число концентраций	Концентрация	Исходная активность	Нормированная активность (%)
1	4	249,999	100	100
		99,999	98	98
		40,0	69	69
		2,5	0	0
2	1	99,999	83	83
3	1	99,999	90	90
4	1	99,999	94	94
5	1	99,999	100	100
6	3	249,999	89	89
		99,999	80	80
		40,0	68	68
7	1	99,999	0	0
8	1	99,999	83	83
9	1	99,999	54	54

Таблица 2

Номер соединения	Логарифм концентрации	Концентрация
1	0,804	6,367
2	0,820	6,608
3	0,666	4,630
4	0,607	4,045
5	0,567	3,688
6	0,965	9,231
7	2,406	254,705
8	0,820	5,608
9	1,849	70,569

Таблица 3

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Среднее значение суммы квадратов
Регрессия	2,912	11	0,265
Остаток	0,104	2	0,052
Итого	3,016	13	

$$F = 5,076$$

Квадрат смешанной корреляции 0,965  
 Коэффициент множественной корреляции 0,983  
 Стандартное отклонение оценки 0,228

2. Пример обработки патентных данных по фунгицидной активности по предложенному методу. Исходные данные по биологической активности были взяты из патента США 4512995 и представлены в табл. 1. Табл. 2 содержит унифицированные концентрации для уровня активности  $I = 50\%$ . Качество регрессионной модели характеризуется табл. 3. Табл. 4 демонстрирует вычисленные по предложенной модели логарифмы концентраций для любого заданного уровня активности. В ней приведены унифицированные индексы для уровней активности от 0 до 100 с шагом 10.

Расчет унифицированных индексов активности по приведенному алгоритму возможен при выполнении следующих условий: наличие в выборке биологических данных подмножества соединений, испытанных при двух (и более) концентрациях с уровнями активности, равномерно распределенными в интервале  $0 \div 100\%$ , объемом  $\sum_i N_i \geq 12-15$ .

Таблица 4

Номер соединения	Логарифм концентрации										
	0	10	20	30	40	50	60	70	80	90	100
1	0,397	0,126	0,076	0,203	0,460	0,804	1,188	1,568	1,900	2,138	2,237
2	0,414	0,143	0,093	0,219	0,476	0,820	1,204	1,585	1,916	2,154	2,253
3	0,259	-0,011	-0,061	0,064	0,322	0,666	1,050	1,430	1,762	2,000	2,098
4	0,200	-0,070	-0,120	0,006	0,263	0,607	0,991	1,371	1,703	1,941	2,040
5	0,160	-0,110	-0,160	-0,033	0,223	0,567	0,951	1,331	1,663	1,901	2,000
6	0,559	0,288	0,238	0,364	0,621	0,965	1,349	1,730	2,061	2,299	2,398
7	1,999	1,728	1,679	1,805	2,062	2,406	2,790	3,171	3,502	3,740	3,839
8	0,414	0,143	0,093	0,219	0,476	0,820	1,204	1,585	1,916	2,154	2,253
9	1,442	1,171	1,121	1,247	1,505	1,849	2,233	2,613	2,945	3,183	3,281

**Заключение.** Применение предложенного подхода особенно эффективно при наличии базы данных по биологическим испытаниям химических соединений [2, 4]: появляется возможность получения унифицированных индексов активности для представительных выборок БАС с одинаковым местом действия и последующим использованием этих данных для машинного анализа связи молекулярная структура — активность [3, 5].

#### СПИСОК ЛИТЕРАТУРЫ

1. Brown I. F., Taylor H. M. Synthesis-directing structure-activity relationships of some fungicides inhibiting ergosterol biosynthesis // Pesticide Synthesis Through Rational Approaches /Eds. P. S. Magee, G. K. Kohn. Washington: Amer. Chem. Soc, 1984.
2. Нигматуллин Р. С., Осипов А. Л., Карпова Н. А. Система поиска и анализа патентной информации о химических соединениях // ИТИ. Сер. 2. 1989. № 1.
3. Зацепин В. М., Осипов А. Л., Семенов Р. Д. Система компьютерного предсказания физико-химических и биологических свойств веществ // Автометрия. 1995. № 5.
4. Ватолин Ю. Н., Зацепин В. М., Осипов А. Л. База данных по химическим средствам защиты растений из патентных источников // Системы баз данных и знаний. Калинин, 1989.
5. Валуева Л. Н., Зацепин В. М., Осипов А. Л. и др. Система анализа и прогнозирования пестицидно-активных соединений // Использование вычислительных машин в спектроскопии молекул и химических исследованиях. Новосибирск, 1989.

*Поступила в редакцию 2 сентября 1996 г.*

---

**Реклама продукции в нашем журнале — залог Вашего успеха!**