БИООРГАНИЧЕСКАЯ ХИМИЯ, 1995, том 21, № 10, с. 809 - 815

УДК 547.814.03;577.1.01

ЭВОЛЮЦИОННОЕ ПРОГРАММИРОВАНИЕ ДЛЯ ВЫЯВЛЕНИЯ ЗАКОНОМЕРНОСТЕЙ "СТРУКТУРА–АКТИВНОСТЬ" В РЯДУ ПРОИЗВОДНЫХ 3-ФЕНОКСИХРОМОНА И 3-ФЕНОКСИ-4-ГИДРОКСИКУМАРИНА

© 1995 г. И. В. Тетко, В. Ю. Танчук, С. А. Васильев*, В. П. Хиля*, Г. И. Пода, А. И. Луйк

Институт биоорганической химии и нефтехимии НАН Украины, 253660, Киев, ул. Мурманская, 1, e-mail: tetko@bioorganic.kiev.ua;

** Кафедра органической химии Киевского университета им. Тараса Шевченко, 252601, Киев, ул. Владимирская, 64*

Поступила в редакцию 15.09.94 г. После доработки 17.01.95 г.

На примере ряда соединений, обладающих гиполипидемической активностью, показано, что для успешного решения проблемы выбора информативного набора параметров молекулы могут использоваться эволюционные алгоритмы. Наборы параметров, найденные для метода потенциальных функций, показали хороший прогноз активности молекул из контрольной выборки.

Ключевые слова: структура–активность; эволюционное программирование; метод "k ближайших соседей"; потенциальные функции; флавоноиды.

Проблема предсказания биологической активности (БА) химических соединений и создания новых веществ с заданной активностью является одной из наиболее важных в современной органической химии [1, 2]. Применение методов трехмерных (3D) количественных соотношений "структура–активность" (КССА) позволяет анализировать тонкие молекулярные механизмы взаимодействия физиологически активных веществ с биорецепторами [3 - 5]. Однако эти методы достаточно сложны, требуют значительного опыта работы и поэтому не всегда доступны широкому кругу исследователей. С другой стороны, имеется очень много работ, в которых сообщается о хороших результатах предсказания БА, полученных с помощью представления молекулы как вектора в пространстве параметров [6 - 8] или 1D-КССА. Такое моделирование легко автоматизируется, позволяет осуществлять скрининг большого числа молекул для разных типов активности. Поэтому, несмотря на некоторую "идейную старость" этих методов, об их применении по-прежнему сообщается в значительном числе публикаций.

Методы 1D-КССА за последние годы сильно изменились. Это связано в основном с разработкой и применением для выявления соотношений "структура–активность" новых современных методов теории распознавания образов, таких, как adaptive least squares (ALS), fuzzy adaptive least squares (FALS), нейронные сети [9 - 11]. Эти методы позволяют проводить сложные нелинейные интерполяции и, как неоднократно сообщалось, дают лучшие прогнозы активности новых веществ по сравнению с традиционными методами множественного регрессионного анализа и линейного дискриминантного анализа [9, 10]. Однако проблема выбора наиболее информативных параметров для этих методов остается актуальной. Правильный выбор небольшого числа информативных признаков позволяет повысить эффективность классификации, поскольку включение малоэффективных параметров в решающее правило резко ухудшает прогноз. Заранее, как правило, неизвестно, какой набор признаков лучше всего описывает исследуемые ССА. Прямой перебор всех возможных вариантов неприемлем, так как требует проверки огромного числа наборов параметров $2^Q - 1$ (Q – число анализируемых параметров). Традиционные подходы к определению оптимального набора признаков заключаются в использовании методов снижения размерности. Существует два общепринятых принципиально отличных друг от друга подхода. В первом отбор наилучших признаков производится на основе критерия информативности, для чего вводятся сильные математические предположения о характере исследуемого распределения (т.е. предполагаются форма и параметры исследуемого

Используются сокращения: ЭА – эволюционные алгоритмы, ЭС – эволюционная стратегия, ЭП – эволюционное программирование, ГА – генетический алгоритм, БА – биологическая активность, КССА – количественное соотношение "структура–активность", ТГ – триглицериды.

распределения). Во втором подходе специальных предположений не делается, а используются некоторые эвристические итеративные процедуры, каждый шаг которых понятен, но общий результат их применения осмыслить и изучить трудно. Пошаговые процедуры чаще используются в методах регрессионного анализа, а методы первой группы применяются в различных вариантах линейного дискриминантного анализа [12].

Первый подход может оказаться неприемлемым в случае неправильности математического представления о структуре исследуемого распределения. Вторая группа методов не гарантирует нахождение глобального минимума, т.е. выбор наилучшего с точки зрения оценки классификации набора признаков.

В последнее время в качестве альтернативных методов выбора наиболее информативных параметров стали использоваться методы, моделирующие законы биологического отбора, сформулированные Дарвиным. Эти методы носят общее название "эволюционные алгоритмы" (ЭА) [13]. Они формально подобны методам пошагового отбора параметров, однако в отличие от них обладают способностью преодолевать локальные минимумы (неоптимальные наборы параметров). ЭА с успехом применялись для решения таких сложных задач, как составление оптимального расписания уроков, создание эффективной системы управления газопроводом, конструирование турбин реактивных двигателей, выбор параметров, описывающих инфракрасные спектры, и др. [14]. Поэтому нам показалось небезынтересным исследовать применимость ЭА к проблеме выбора наиболее информативных параметров для решения задач поиска соотношений "структура–активность". Остановимся более подробно на описании ЭА.

ЭВОЛЮЦИОННЫЕ АЛГОРИТМЫ

ЭА используют модели эволюционных процессов как ключевые элементы в конструировании и воплощении компьютерных вычислительных систем. Существует много различных модификаций ЭА. Они используют общую концептуальную базу о возможности компьютерного моделирования эволюции индивидуумов через процессы отбора (селекции) и воспроизводства. Эти процессы зависят от приспособленности (функции качества) индивидуальных структур, которая определяется окружающей средой. Термины "популяция", "селекция", "приспособленность" и др., используемые в ЭА, являются искусственными функциональными аналогами биологических терминов.

Иначе говоря, ЭА поддерживают такую популяцию структур, которая развивается соответственно правилам селекции и другим операциям, именующимся поисковыми операторами (генетическими операторами), такими, как рекомбинации и мутации. Каждый индивидуум в популяции оценивается в соответствии с его приспособленностью к данной среде. Селекция выявляет наиболее приспособленные индивидуумы, используя таким образом информацию об их приспособленности. Случайные рекомбинации и мутации изменяют индивидуумы, обеспечивая возникновение новых экземпляров для последующего отбора. Каждый индивидуум характеризуется набором "генов" (в нашем случае – признаков), которые кодируются цепочкой битов (0, 0, 1, 0, ..., 0, 1). Единичка определяет наличие, а нуль – отсутствие соответствующего признака в анализируемом наборе параметров. В качестве генетических операторов используются операции кроссинговера, делеции, вставки, мутации, объединения, разрыва, которые были определены по аналогии с биологическими процессами, происходящими в реальных живых объектах. Ниже приведен список основных операторов.

Делеция состоит в замене одной единицы в наборе на нуль, т.е. удалении из набора одного из признаков:

$$(1, 0, 1, 0, \ldots, 1, \ldots, 0, 1) \Rightarrow (1, 0, 1, 0, \ldots, 0, \ldots, 0, 1).$$

Вставка – антипод делеции. Она добавляет в набор новый признак:

$$(1, 0, 1, 0, \ldots, 0, \ldots, 0, 1) \Rightarrow (1, 0, 1, 0, \ldots, 1, \ldots, 0, 1).$$

Мутация состоит в случайном изменении положения одной из единиц в наборе

$$(1, 0, 1, 0, \ldots, 0, \ldots, 0, 1) \Rightarrow (1, 0, 0, 0, \ldots, 1, \ldots, 0, 1)$$

и может рассматриваться как сочетание вставки и делеции.

Операция **кроссинговера** состоит в обмене частью признаков между двумя родителями, что приводит к появлению двух потомков. Точка обмена выбирается случайно:

$$\begin{array}{c} (1, 0, 1, 0, \ldots, 1, \ldots, 0, 0) \\ \otimes \qquad\qquad \Rightarrow \\ (0, 0, 1, 1, \ldots, 0, \ldots, 0, 1) \end{array}$$

$$\Rightarrow \begin{cases} (1, 0, 1, 1, \ldots, 0, \ldots, 0, 1) \\ (0, 0, 1, 0, \ldots, 1, \ldots, 0, 0) \end{cases}$$

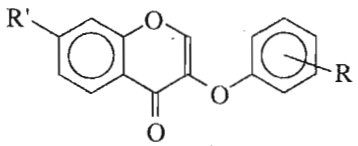
Существуют три большие группы эволюционных алгоритмов: эволюционное программирование (ЭП), эволюционные стратегии (ЭС) и генетические алгоритмы (ГА). Они отличаются интенсивностью использования генетических операторов и реализациями функции селекции. Наиболее важной в методах ЭП и ЭС является операция мутации, а в ГА – операция кроссинговера [13]. Хотя все три метода могут использоваться для определения наиболее эффективного набора параметров, мы использовали ЭС.

АНАЛИЗИРУЕМЫЕ МОЛЕКУЛЫ

Для анализа был использован ряд молекул флавоноидов, 3-феноксильных производных хромона и кумарина. Как известно, производные флавоноидов обладают широким спектром биологической активности, являются аналогами природных соединений и считаются перспективными веществами для поиска новых высокоэффективных лекарственных средств. Для всех соединений изучалась гипогликемическая активность, оцениваемая по уровню снижения триглицеридов (ТГ) в крови.

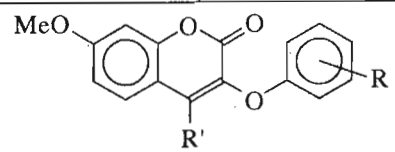
Обучающая выборка состояла из 25, а контрольная – из 27 соединений, синтезированных нами преимущественно в последнее время [15 - 17] (см. табл. 1). Все молекулы были отнесены к двум классам: активным (уровень снижения ТГ > 25%) и неактивным (уровень снижения ТГ ≤ 25%).

Таблица 1. Биологическая активность 3-феноксихромонов и кумаринов



(1) - (20) R' = OH
(21) - (40) R' = OC(O)Me

Соединение	R	Снижение ТГ, %	Активность*	Соединение	R	Снижение ТГ, %	Активность*
1	H	−23.0	−	30**	*м*-OMe	−3.4	−
2	*о*-F	−27.0	+	31	*п*-OMe	−14.3	−
3	*п*-F	−44.5	+	32	*о*-NO_2	−24.8	−
4	*п*-Cl	−12.0	−	33	*м*-NO_2	−33.0	+
5**	*о*-Br	−9.4	−	34	*п*-NO_2	−31.7	+
6**	*м*-Br	−11.7	−	35**	*о*-OC(O)Me	−38.4	+
7**	*п*-Br	−7.1	−	36**	*м*-OC(O)Me	−44.2	+
8**	*п*-I	−19.9	−	37**	*п*-OC(O)Me	−36.0	+
9	*о*-OMe	−4.9	−	38**	*п*-OEt	−40.4	+
10	*м*-OMe	−22.7	−	39	*п*-CO_2Me	−37.8	+
11	*п*-OMe	−22.1	−	40**	*п*-OCH(Me)CO_2Et	−37.2	+
12	*о*-NO_2	−20.1	−				
13	*м*-NO_2	−24.0	−				
14	*п*-NO_2	−28.3	+				
15**	*о*-OH	−22.5	−				
16**	*м*-OH	−14.5	−				
17**	*п*-OH	−24.7	−				
18**	*п*-OEt	−46.9	+	41	H	−23.0	−
19	*п*-CO_2Me	−37.7	+	42	*о*-F	−24.5	−
20**	*п*-OCH(Me)CO_2Et	−27.0	+	43**	*п*-Cl	−11.4	−
21	H	−31.7	+	44**	*п*-Br	−14.6	−
22	*о*-F	−45.8	+	45**	*о*-I	−24.4	−
23	*п*-F	−32.0	+	46**	*п*-OMe	−39.2	+
24	*п*-Cl	−31.3	+	47**	H	−33.4	+
25**	*о*-Br	−29.5	+	48	*о*-F	−49.9	+
26**	*м*-Br	−31.9	+	49	*п*-Cl	−22.0	−
27**	*п*-Br	−34.3	+	50**	*п*-Br	−27.4	+
28**	*п*-I	−6.7	−	51**	*о*-I	−29.1	+
29	*о*-OMe	−23.3	−	52**	*п*-OMe	−42.5	+



(41) - (46) R' = OH
(47) - (52) R' = OC(O)Me

* Соединения, обладающие уровнем снижения ТГ ≤ 25%, рассматриваются как активные, а соединения с уровнем снижения ТГ > 25% – как неактивные.

** Соединения, входящие в контрольную выборку.

Ранее нами [16] был проведен анализ пространственного строения остовной части молекул хромона и кумарина. Этот анализ показал большую лабильность и возможность адаптации исследуемых молекул по отношению к вероятным местам взаимодействия. Учитывая большую гибкость молекул, для кодирования их образов использовали параметры, не зависящие или слабо зависящие от пространственного строения молекул. Бо́льшую часть этих параметров (46) составляют топологические индексы. Наряду с классическими индексами Виннера, Рандича, Кира, Балабана и др. были использованы оригинальные модификации этих индексов [18]. Значительное число параметров было обусловлено распределением электронной плотности молекул. Заряды были рассчитаны по программе MNDO (MOPAC 5.0) для конформации молекул, соответствующей минимуму энергии. В число использованных индексов были включены значения зарядов на атомах остова флавоноидов, сумма квадратов зарядов на атомах молекул, коэффициент распределения смеси октанол–вода, ван-дер-ваальсов объем молекул и др. Общее число индексов составило 63. Поскольку некоторые из топологических индексов оказались сильно скоррелированными, мы отобрали для последующего анализа только 43, для которых парные коэффициенты корреляции R были <0.95.

Набор дескрипторов молекулы кодировался цепочкой битов, длина которой соответствовала количеству анализируемых индексов (43). Чтобы получить качественный прогноз методами распознавания образов, пользуются эмпирическим правилом: отношение количества молекул (N) к количеству используемых параметров (Q) должно быть >4 - 5 [11], т.е. число молекул должно быть приблизительно в 4 - 5 раз больше числа используемых параметров. Поскольку обучающая выборка состояла из 25 молекул, размер "генома" (число "генов" или индексов в наборе) был ограничен шестью индексами. Если в процессе применяемых генетических операций размер генома превышал эту цифру, мы игнорировали этот набор и создавали новый. Для такого "разреженного" генома кроссинговер, а также ряд других генетических операторов теряют смысл. Мы ограничились в наших исследованиях только использованием операторов мутации (вероятность мутации принималась равной 50%), делеции и вставки (по 25% соответственно).

Алгоритм для работы ЭС (одна генерация) заключается в следующем:

1) стартовать со временем: $t = 0$;

2) выбрать случайным образом начальную популяцию индивидуумов (т.е. создать случайный набор из m индивидуумов, где m – размер популяции, который задается исследователем): **population** $P(t)$;

3) оценить приспособленность первоначальной популяции (т.е. оценить качество каждого из наборов параметров): **evaluate** $P(t)$;

4) проверить на критерий завершения (время, приспособленность и т.д.): **test the end** $P(t)$;

5) увеличить время на единицу: $t = t + 1$;

6) провести случайным образом генетические операции (популяция увеличивается): $P'(t)$ = **mutate** $P(t)$;

7) оценить приспособленность новых индивидуумов $P'(t)$: **evaluate** $P'(t)$;

8) выбрать новую популяцию "выживших" индивидуумов на основе приспособленности и затем перейти на шаг 4 (размер популяции возвращается к предыдущему значению): P = **survive** $(P, P'(t))$.

Наша первоначальная популяция состояла из $m = 50$ индивидуумов (наборов параметров), которые были созданы случайным образом. На шаге 6 за счет генетических операций каждый раз порождалось $\lambda = 350$ потомков. Приспособленность потомков оценивалась на шагах 3 и 7, а на шаге 8 из суммарной популяции (родители плюс потомки, всего 400 индивидуумов) отбиралось $m = 50$ наилучших экземпляров (так называемый "элитный" отбор [13]). Отношение $m/\lambda = 1/7$ является оптимальным для задач ЭС [13]. В качестве критерия завершения использовалось достижение определенного числа поколений – 50 (длительность одного поколения составляют шаги с 4-го по 8-й) или отсутствие увеличения приспособленности наилучшего индивидуума на протяжении последних 10 поколений. При этом каждый раз за одну генерацию программы оценивалось не более 17000 индивидуумов. Практически всегда за время 20 - 30 поколений удается найти наилучшие наборы. Для сравнения: полный перебор всех возможных комбинаций из 43 по шесть и менее параметров составляет:

$$\sum_{i}^{6} C_{43}^{i} \approx 10^{7}.$$

ПРОЦЕДУРА ОЦЕНКИ ПРИСПОСОБЛЕННОСТИ ИНДИВИДУУМОВ

Для работы ЭС требуются достаточно большая популяция индивидуумов и выполнение колоссального объема вычислений. Даже получить 17000 оценок наборов параметров, необходимых только для одной генерации программы, такими "медленными" алгоритмами, как ALS, FALS, нейронные сети и др., требующими значительное время для обучения, не представляется возможным. В этой связи в качестве классификатора

использовались "быстрые" методы потенциальных функций и "k-ближайших соседей" [12]. В качестве функции приспособленности применялась оценка решающего правила $\hat{P}_e(U)$, согласно методу скользящего контроля:

$$f(U) = \hat{P}_e(U) = COR/N, \qquad (1)$$

где COR – количество правильно предсказанных молекул по методу скользящего контроля, U – анализируемый набор параметров. Метод скользящего контроля состоит в следующем.

Допустим, мы рассматриваем набор параметров $U = (u_1, \ldots, u_q)$, т.е., в терминологии ЭС, один индивидуум. Каждая молекула i характеризуется своими конкретными значениями этих параметров и представляется вектором $\mathbf{x}_i$ в пространстве параметров U. Размерность вектора параметров $\mathbf{x}_i$ соответствует q – числу параметров в U. Для каждого из анализируемых классов активности молекул θ_j с $j = 1, \ldots, K$ (в данной работе, как указано выше, молекулы были отнесены к двум классам активности – активные и неактивные) известен свой набор молекул $\mathbf{x}_i^j$, или обучающая выборка класса θ_j. Объединение всех наборов молекул $\mathbf{x}_i^j$ представляет собой полную обучающую выборку. Для вычисления оценки (1) каждая из молекул последовательно удаляется из обучающей выборки и прогнозируется на основе оставшихся молекул.

В качестве метрики пространства $d(\mathbf{x}, \mathbf{y})$ мы использовали эвклидово расстояние.

МЕТОД "k БЛИЖАЙШИХ СОСЕДЕЙ"

Вокруг точки $\mathbf{x}$, которую необходимо классифицировать, для каждого исследуемого класса активности θ_j строится сфера минимального радиуса $R = \min\limits_{i \in \theta_j} d(\mathbf{x}_i, \mathbf{x})$, внутрь которой попадает ровно k точек из обучающей выборки данного класса. Анализируемая молекула относится к классу, для которого радиус построенной сферы будет минимальным. Оптимальное число соседей оценивалось в диапазоне $k = 1, \ldots, N/4$ и определялось в процессе работы программы. При равенстве значения функции качества для разного количества соседей $k = k_1, \ldots, k_l$ мы использовали для дальнейшего анализа наименьшее количество соседей $k = \min(k_1, \ldots, k_l)$.

МЕТОД ПОТЕНЦИАЛЬНЫХ ФУНКЦИЙ

Все точки $\mathbf{x}_i$ из обучающей выборки рассматриваются как центры потенциалов:

$$F(R = d(\mathbf{x}_i, \mathbf{x})) = fun(R/\sigma) = \varphi_i.$$

Здесь σ – параметр сглаживания, аналогичный по смыслу параметру k в методе "k ближайших соседей"; φ_i – потенциал в анализируемой точке $\mathbf{x}$ относительно точки сравнения i; $fun(R/\sigma)$ – любая гладкая функция (использовалось нормальное распределение $f(R/\sigma) = \text{const}\, e^{-(R/\sigma)^2}$). Отнесение молекулы производится к тому классу θ_j, который создает максимальный потенциал для анализируемой точки:

$$\Phi_j = \frac{1}{n_j} \sum_{i \in \theta_j} \varphi_i,$$

$$I = \max_j \Phi_j.$$

Нормировка производится на полное число молекул n_j в наборе обучения, относящихся к классу θ_j. Предварительный анализ с использованием разных σ показал, что наилучшие результаты оценки метода потенциальных функций методом скользящего контроля достигались при $\sigma = 1 - 7$ и слабо зависели от параметра сглаживания σ. Конкретное значение σ с точностью до 0.5 определялось в процессе вычислений. Аналогично методу "k ближайших соседей" при равенстве функции качества для различных значений σ выбиралось наименьшее из них. Более подробное описание использованных методов распознавания образов можно найти в [12].

РАСЧЕТНЫЕ РЕЗУЛЬТАТЫ

Использовалось 10 генераций ЭС для обоих методов. Для обучающей выборки наибольшее число правильно классифицируемых молекул (метод скользящего контроля) составило 23 из 25. Это значение обоими методами достигалось на каждой генерации в среднем за 20 - 30 поколений. После завершения каждой генерации для анализа были взяты наилучшие наборы, полученные в этой генерации. Всего для метода потенциальных функций было обнаружено для набора, а для метода "k ближайших соседей" – 31 набор, которые по методу скользящего контроля правильно классифицировали 23 молекулы из 25 (92%). Столь большое количество наборов, полученных для метода "k ближайших соседей", указывает на недостаточность использования критерия скользящего контроля при оценке прогностической способности метода. Наилучшие наборы индексов для метода потенциальных функций приведены в табл. 2.

Метод потенциальных функций показал лучшие результаты прогноза для новых молекул по сравнению с методом "k ближайших соседей". В среднем количество правильно прогнозируемых молекул из контрольной выборки было:

23.5 ± 0.5 ($87 \pm 2\%$) – для метода потенциальных функций (по двум наилучшим результатам),

Таблица 2. Наилучшие наборы параметров, которые были получены с использованием метода потенциальных функций

Набор	Номера параметров*	Обучение, % (25 молекул)	Число генераций	Прогноз, % (27 молекул)
1	38, 25, 18, 13	92	8	85.19
2	38, 18, 17, 13	92	9	88.89

* 13, 17, 18 и 25 – топологические индексы: 13 – десятичный логарифм суммы коэффициентов характеристического полинома матрицы смежности; 17 – $M_2(G) = \sum_{ef} (v_i, v_j)$, где v_i, v_j – степени смежных вершин химического графа; сумма берется по всем ребрам графа; 18 – $(G) = (v_1 v_2 \ldots v_n)$, где v_i – число вершин степени i; n – максимальная степень вершин в данном графе; 25 – аналог индекса формы Кира 3-го порядка: $f = (N-1)(N-3)^2/n^2$, если число атомов (N) нечетное, и $f = (N-3)(N-2)^2/n^2$, если четное (n – число троек в матрице расстояний молекулярного графа). Индексы 13, 17 и 18 нормированы на число тяжелых атомов в молекуле. 38 – квадратный корень из суммы квадратов атомных зарядов $\bar{q} = \sqrt{\sum_i q_i^2}$.

21.2 ± 2 ($77 \pm 7.5\%$) – для метода "k ближайших соседей" (результаты усреднены по 31 наилучшему набору).

Метод потенциальных функций оказался более адекватным для решения данной задачи, чем метод "k ближайших соседей". Последний показал значительное смещение оценки прогностической силы наборов параметров на контрольной выборке (77% правильных прогнозов по сравнению с 92% на обучающей выборке). Возможно, для корректной работы метода "k ближайших соседей" следует использовать более строгие критерии оценки качества классификации, дающие несмещенные оценки, например метод "складного ножа" [19].

ЗАКЛЮЧЕНИЕ

ЭА представляют собой мощные оптимизационные средства, но их эффективность зависит от возможности проведения большого числа оценок за приемлемое время. Использование таких простых методов классификации, как метод потенциальных функций и метод "k ближайших соседей", делает возможным выполнение ЭА даже на персональных компьютерах типа IBM PC/486. Параметры (группы параметров), полученные с применением ЭА и простых методов классификации, могут использоваться для анализа более сложными и "медленными" методами распознавания образов, такими, как нейронные сети. Именно о таком применении ЭА, например, сообщалось в работе [20]. Наборы параметров, полученные с помощью метода "k ближайших соседей", были успешно использованы для дальнейшего обучения нейронных сетей со встречным распространением. В нашем случае, однако, даже прогнозы, полученные на стадии отбора признаков с применением простых методов, могут быть достаточно хорошими и использоваться для построения эффективного классификатора.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Исследование гиполипидемической активности [20] соединений было проведено в Пятигорском фармацевтическом институте проф. Ю.К. Василенко. Гиперлипидемию вызывали внутрибрюшинным введением тритона WR-1339 в дозе 100 мг/100 г массы тела животного. Одновременно с тритоном вводили перорально в дозе 200 мг/кг в виде суспензии исследуемые вещества. Контрольной группе животных вводили только тритон. Через 12 ч крыс подвергали декапитации. Для всех животных исследовали сыворотку крови на содержание ТГ [21]. Результаты исследований приведены в табл. 1.

Работа финансировалась Государственным комитетом Украины по вопросам науки и технологий.

СПИСОК ЛИТЕРАТУРЫ

1. *Fujita T.* // Drug Design: Fact of Fantasy / Eds G. Jolles, K. Wooldrige. L.: Acad. Press, 1984. P. 260 - 291.
2. *Dunn W.J. III* // Chemometrics and Intell. Lab. Syst. 1989. V. 6. № 1. P. 181 - 194.
3. *Димогло А.С.* // Хим.-фармацевт. журн. 1985. Т. 19. № 8. С. 438 - 451.
4. *Doweyko A.M.* // J. Mathem. Chem. 1991. V. 7. № 2. P. 273 - 285.
5. *Gramer R.D. III, Patterson D.E., Bunce J.D.* // J. Amer. Chem. Soc. 1988. V. 110. № 18. P. 5959 - 5967.
6. *Free S.M., Wilson J.W.* // J. Med. Chem. 1964. V. 7. № 4. P. 395 - 405.
7. *Hansch C., Dunn W.T.* // J. Med. Chem. 1972. V. 62. № 1. P. 1 - 19.
8. *Moriguchi I., Komatsu K.* // Chem. Pharm. Bull. 1977. V. 25. № 7. P. 2800 - 2802.
9. *Moriguchi I., Hirono S., Liu Q., Matsushita Y., Nakagawa T.* // Chem. Pharm. Bull. 1990. V. 38. № 9. P. 3373 - 3379.
10. *Tetko I.V., Luik A.I., Poda G.I.* // J. Med. Chem. 1993. V. 36. № 7. P. 811 - 814.
11. *Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.* Классификация и снижение размерности. М.: Финансы и статистика, 1989. С. 608.
12. *Back T., Schwefel H.P.* // Evolutionary Computation. 1993. V. 1. № 1. P. 1 - 25.
13. *Holland J.H.* // Sci. Amer. 1992. V. 220. № 9. P. 44 - 51.
14. *Васильев С.А., Лукьянчиков М.С., Молчанов Г.И.* // Хим.-фармацевт. журн. 1991. Т. 25. № 7. С. 34 - 38.

15. *Васильев С.А., Боярчук В.Л., Лукьянчиков М.С., Хиля В.П.* // Хим.-фармацевт. журн. 1991. Т. 25. № 11. С. 50 - 55.

16. *Васильев С.А.* Синтез и биологическое действие 3-арилоксихромонов и 3-фенокси-4-гидроксикумаринов. Дис. ... канд. хим. наук. Киев: Киевск. ун-т, 1992. 164 с.

17. *Пода Г.И., Танчук В.Ю., Тетко И.В., Кошель М.И., Луйк А.И.* // Теорет. и экспер. химия. 1993. Т. 29. № 2. С. 122 - 125.

18. *Toussaint G.T.* // IEEE Trans. Inform. Theory. 1974. V. 20. № 4. P. 472 - 479.

19. *Brill F.Z., Brown D.E., Martin W.N.* // IEEE Trans. Neural Networks. 1992. V. 3. № 2. P. 324 - 328.

20. *Larratini S., Paoletti R., Bizzi L., Rizzi O.* // Drugs Affecting Lipid Metabolism / Ed. Paoletti R. Amsterdam: Elsevier Press, 1961. P. 144 - 147.

21. *Neri B.P., Frings C.S.* // Clin. Chem. 1973. V. 19. № 10. P. 1201 - 1203.

Evolutionary Computation for Reveal Structure–Activity Relationships in 3-Phenoxychromone and 3-Phenoxy-4-hydroxycoumarin Derivatives

I. V. Tetko*, V. Yu. Tanchuk*, S. A. Vasil'ev, V. P. Khilya**, G. I. Poda*, and A. I. Luik***

** Institute of Bioorganic Chemistry and Petrochemistry, National Academy of Sciences of Ukraine, ul. Murmanskaya 1, Kiev, 253660 Ukraine; e-mail: tetko@bioorganic.kiev.ua*

*** Organic Chemistry Department, Shevchenko University, ul. Vladimirskaya 64, Kiev, 252601 Ukraine*

Abstract – Based on a set of compounds possessing hypolipidemic activity, it was demonstrated that evolutionary algorithms can be successfully used to compile an informative set of molecular parameters. The parameter sets selected using the method of potential functions allowed correct prediction of the activity of test molecules.

Key words: structure–activity relationship, evolutionary programming, method of k nearest neighbors, potential functions, flavonoids.