

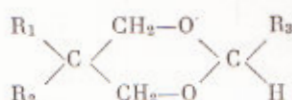
Академик АН ЛатвССР С. А. ГИЛЛЕР, А. Б. ГЛАЗ,
Л. А. РАСТРИГИН, А. Б. РОЗЕНБЛИТ

РАСПОЗНАВАНИЕ ФИЗИОЛОГИЧЕСКОЙ АКТИВНОСТИ ХИМИЧЕСКИХ СОЕДИНЕНИЙ НА ПЕРЦЕПТРОНЕ СО СЛУЧАЙНОЙ АДАПТАЦИЕЙ СТРУКТУРЫ

Из-за отсутствия достаточных знаний о геометрическом и электронном соответствии действующего соединения со строением активного центра фермента, ответственного за проявление физиологической активности, взаимосвязь между структурой химического соединения и его биологическими свойствами устанавливается в настоящее время, в основном эмпирическим путем. Известны попытки, например (^{1, 2}), установить корреляции между физико-химическими параметрами и физиологической активностью. Из-за сложного характера этой взаимосвязи полученные уравнения регрессии обладают низкой статистической достоверностью и малой прогнозирующей способностью.

В работе И. И. Иоффе с сотрудниками (³) предложено применить для предсказания некоторых свойств химических соединений, например активности катализаторов, методы теории распознавания образов. Однако в качестве признаков в этой работе использовались лишь физико-химические параметры, большинство из которых может быть найдено только в результате тщательного экспериментального исследования уже синтезированного вещества.

В настоящей работе нами впервые предпринимается попытка предсказания физиологической активности неисследованного в этом отношении вещества непосредственно по структуре молекулы, исходя из известных данных об этой активности в ряду близких по строению соединений. Это сделано на примере ряда алкил- и алкоксиалкилзамещенных 1,3-диоксанов (⁴). Рассматриваемые соединения представляются общей структурной формулой



и могут существовать в форме цис- и транс-изомеров. По виду заместителей рассматриваемый ряд распадается на 4 группы: 2,5-диалкил-1,3-диоксаны; 2,5-диалкил-5-алкоксиметил-1,3-диоксаны; 2,5-диалкил-5-α-алкоксиэтил-1,3-диоксаны и 2,5-диалкил-5-β-алкоксиэтил-1,3-диоксаны.

Задача распознавания в данном случае может быть сформулирована как классификация соединений по физиологической активности в зависимости от вида радикалов и изомерии диоксанового кольца. В качестве критерия классификации принят антагонизм с коразолом. Соединения, обладающие соответствующей активностью, относятся к первому классу, а соединения, не обладающие активностью, — ко второму.

Кодирование. Структурная формула каждого радикала может быть представлена сочетаниями из 5 простейших структурных компонентов: O, H, CH, CH₂, CH₃. Для кодирования каждого из этих компонентов на сетчатке перцептрона отводилось по 3 рецептора, а для кодирования цис- и транс-изомерии — 2 рецептора (см. табл. 1). Связи между компонентами не кодировались, так как они однозначно определялись видом со-

седних компонентов. На рис. 1 приведены примеры кодирования некоторых радикалов на сетчатке. Как видно, для кодирования радикалов R_1 отведено 3 строки, а для радикалов R_2 и R_3 по 2 строки. В действительности, в целях экономии элементов сетчатки, строки располагались последовательно. Размер сетчатки был выбран равным $8 \times 8 = 64$.

Эксперимент. Для решения данной задачи на ЭЦВМ «Минск-22» моделировался трехслойный перцептрон с сетчаткой $S = 64$ рецептора

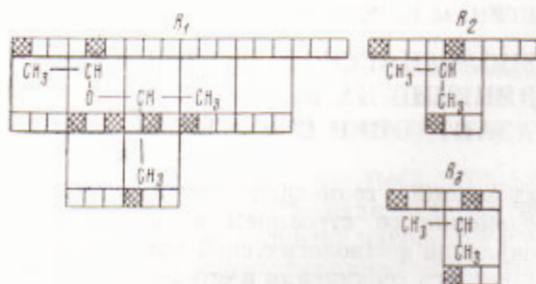


Рис. 1. Кодирование радикалов R_1 , R_2 и R_3 на сетчатке

и числом ассоциативных элементов, равным 100. Каждый элемент имел 5 входов, из них 3 входа для возбуждающих и 2 для тормозящих связей с сетчаткой. Исходная конфигурация $S - A$ -связей выбиралась случайным образом. Порог θ каждого A -элемента принимался равным 1. Из 46 приведенных в (4) соединений одна часть выбиралась для составления обучающей последовательности,

причем в нее входили представители всех 4 групп названных выше соединений, а оставшаяся часть использовалась только для экзамена.

Для обучения перцептрона использовались следующие алгоритмы:

1. Алгоритм Розенблатта (α -система подкрепления с коррекцией ошибок) (5).

Таблица 1

Кодирование компонентов структурной формулы

№№	1-й рецептор	2-й рецептор	3-й рецептор	Компонент или изомер	Вид на сетчатке
0	0	0	0	Ноль	
1	0	0	1	H	
2	0	1	0	CH	
3	0	1	1	CH2	
4	1	0	0	CH3	
5	1	0	1	0	
6	1	1	0	—	
7	1	1	1	—	
1	0	0	—	цис	
2	1	1	—	транс	

2. Алгоритм Розенблатта в комбинации с алгоритмом случайного поиска (6) конфигурацией $S - A$ -связей тех A -элементов, которые не удовлетворяют заданным требованиям. Суть этого алгоритма заключается в следующем.

Перцептрону с исходной случайной конфигурацией $S - A$ -связей предъявляется обучающая последовательность и проводится обучение по

алгоритму Розенблатта. После этого обучающая последовательность предъявляется для экзамена и определяется показатель качества для данной конфигурации S — A-связей — число неправильных ответов перцептрона, а для каждого A-элемента определяется также число правильных ответов. Если величина показателя качества оказывается больше заданной, делается первый шаг поиска: для A-элементов, у которых число правильных ответов меньше некоторого порога β , случайным образом перестраиваются их связи с сетчаткой. После этого снова проводится обучение по Розенблатту и, в случае необходимости, делается новый случайный шаг поиска. Обучение продолжается до тех пор, пока величина показателя качества не станет равной или меньше заданной, или все A-элементы не будут давать число правильных ответов больше порога β . После этого проводится экзамен, но уже на экзаменационной последовательности.

Результаты эксперимента * Таблица 2

Длина обучающей последовательности	Обучающая последовательность		Экзаменационная последовательность	
	алгоритм Розенблатта	комбинирован. алгоритм	алгоритм Розенблатта	комбинирован. алгоритм
$n_1 = 22$	65	86	49	69
	$\frac{17}{}$	$\frac{7}{}$	$\frac{27}{}$	$\frac{10}{}$
$n_2 = 24$	66	89	51	71
	$\frac{16}{}$	$\frac{6}{}$	$\frac{21}{}$	$\frac{13}{}$
$n_3 = 26$	68	85	57	76
	$\frac{14}{}$	$\frac{5}{}$	$\frac{22}{}$	$\frac{9}{}$

* Числа над чертой — процент правильных ответов, под чертой — длина доверительного интервала.

Результаты эксперимента. Было выбрано три различных обучающих последовательности с числом объектов $n_1 = 22$, $n_2 = 24$, $n_3 = 26$. Для каждой из них выбиралось по одному значению порога β и проводилось 10 независимых экспериментов. Осредненные результаты и доверительные интервалы, соответствующие доверительной вероятности 0,95, приведены в табл. 2. Как видно из табл. 2, значительно более эффективным является комбинированный алгоритм со случайной адаптацией структуры S — A-связей.

Полученные результаты дают основания считать перспективным кибернетический подход к проблеме прогнозирования физиологической активности свойств органических соединений.

Институт органического синтеза
Академии наук ЛатвССР

Поступило
15 IV 1971

Институт электроники и вычислительной техники
Академии наук ЛатвССР
Рига

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

- ¹ C. Hansch, *Farmaco*, 23, 4, 293 (1968). ² W. P. Purcell, John M. Clayton, *J. Med. Chem.*, 11, 2, 199 (1968). ³ И. И. Иоффе, В. С. Федоров и др., *ДАН*, 189, № 6, 1290 (1969). ⁴ А. В. Богатский, Ю. И. Вихляев и др., *Хим.-фармацевтич. журн.*, 2, 9, 3 (1968). ⁵ Ф. Розенблатт, *Принципы нейродинамики*, М., 1965. ⁶ Л. А. Растринин, *Статистические методы поиска*, М., 1968.