[Chem. Pharm. Bull. 31(7)2432—2437(1983)]

Pyrimidine Derivatives. VII.[1] Structure–Activity Relationship of Hypoglycemic 4-Amino-2-(1-piperazinyl)pyrimidines investigated by the Adaptive Least-Squares Method

TETSUO SEKIYA,*,[a] SHUNSUKE HATA,[a] and SHUN-ICHI YAMADA[b]

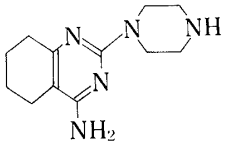
Research Laboratory, Mitsubishi Yuka Pharmaceutical Co., Ltd.,[a] 500 Furuki, Wakaguri, Inashiki, Ibaraki 300–03, Japan, and Faculty of Pharmaceutical Sciences, Josai University,[b] 1–1 Keyakidai, Sakado, Saitama 350–02, Japan

(Received December 8, 1982)

Structure–activity studies of 36 hypoglycemic 4-amino-2-(1-piperazinyl)-5,6-polymethylenepyrimidine derivatives were performed by the adaptive least-squares method. In the analysis, the activity was classified into four groups. The best recognition was obtained with the following equation: $L = -0.00286\mathrm{MW} + 0.241\mathrm{p}K_a + 0.741(\mathrm{I\text{-}4}) + 2.099(\mathrm{I\text{-}5}) + 0.358(\mathrm{I\text{-}6}) - 0.602$. The number of misclassified compounds was 5, and the Spearman rank correlation coefficient was 0.921. In the equation, MW is molecular weight, and $\mathrm{p}K_a$ stands for the basicity of the piperazinyl group. The indicator variables I-4, I-5, and I-6 are assigned a value of 1 when the 4-acylated 1-piperazinyl group (not including 4-acryloyl-1-piperazinyl group), 4-acryloyl-1-piperazinyl group and any group having an $>NCH_2CH_2O-$ structure are present, respectively. From the equation, it was concluded that the 2-(1-piperazinyl)pyrimidine moiety is an essential structure for the activity, and the basicity of the 1-piperazinyl group is also important. The structure–activity relationships are analyzed and discussed in detail.

Keywords——pattern recognition; QSAR; ALS method; hypoglycemic; 2-(1-piperazinyl)pyrimidine; 2-(1-piperazinyl)-5,6,7,8-tetrahydroquinazoline

In a previous paper,[2] we reported the synthesis of 4-amino-2-(1-piperazinyl)-5,6-tetramethylenepyrimidine I and various structurally related compounds, and described their hypoglycemic activity. Consideration of the structure–activity relationships revealed that the



I

presence of a piperazinyl group is important for the activity and *N*-substitution by any group possessing CO or CS on the piperazinyl moiety reduces the activity. These results prompted us to attempt a quantitative structure–activity analysis.

Quantitative structure–activity studies usually utilize the Hansch–Fujita equations,[3] which has been applied widely with great success;[4] however, in the screening stage of research into a new drug, it is often the case that the Hansch–Fujita method is not suitable in practice, for example when the activity data are given by classes or as percent values at a single dose. In latter case, though an expedient method for the analysis has been proposed using log $(A/100-A)$ (where A is % activity),[5] this convenient method cannot be applied if the data contain non-active or 100% active compounds. In such cases, the adaptive least-squares (ALS) method, which was proposed by Moriguchi *et al.*,[6] is the proper procedure, where the

relation can be expressed as a single equation called the discriminant function.

$$L=w_0+w_1x_1+w_2x_2+\cdots+w_kx_k \quad (1)$$

In this paper, we describe quantitative structure-activity studies of hypoglycemic 4-amino-2-(1-piperazinyl)pyrimidine derivatives by using the ALS method.

Method

Activity Classes——Thirty-six 4-amino-2-(1-piperazinyl)-5,6-polymethylenepyrimidine derivatives (**1**—**36**) were classified into 4 groups based on the hypoglycemic activity at a dose of 30 mg/kg *p.o.* as shown in Table I.

Descriptors——The structural descriptors finally used are shown in Table II, and the correlation matrix is given in Table III; however, the sixteen descriptors shown in Fig. 1 were employed for ALS analysis in the first stage.

MW is the molecular weight, and NC (or NC^2) is the number of carbon atoms. The pK_a is the value for the piperazinyl group; the pK_a values of substituted pyrrolidines are used instead of those of piperazines. The π (or π^2) value was calculated by the method of Leo;[8] and the value for 2-(1-piperazinyl)-5,6,7,8-tetrahydroquinazoline was defined as O. NM is the number of methylenes in the 5,6-polymethylene group. The indicator variables I-1, I-2, I-3, I-4, and I-5 were assigned a value of 1 corresponding to the presence of –NH–, $>NCH_2-$, >NPh, >N-acyl (not including >N-acryloyl), and >N-acryloyl moieties in the piperazinyl group, respectively. I-6 was assigned a value of 1 corresponding to the presence of any group having the $>NCH_2CH_2O-$ structure. The indicator variables I-7, I-8, and I-9 were assigned a value of 1 corresponding to the presence of primary, secondary, and tertiary amines, respectively, as the 4-substituent of the pyrimidines.

ALS Calculation——The calculation was performed according to the method of Moriguchi.[6] The starting score a_j ($j=1, 2, \cdots, m$), and the cutting point, b_j ($j=1, 2, \cdots, m-1$), were assigned in the same manner as described previously (eqs. 2 and 3):[6a]

$$a_j=2\left(2\sum_{i=1}^{j-1} n_i+n_j\right)\Big/n-2 \quad (2)$$

$$b_j=(a_j+a_{j+1})/2 \quad (3)$$

where n is the number of whole substances, and n_i and n_j refer to the sizes of groups i and j, respectively. With the whole data set of 36 compounds in this study, the starting score a_j and the cutting point b_j are as follows; $a_1=-1.333$,

Table I. Classification of Activity

Class	Activity (%)[a]	No. of compounds
1	Act. < 20	12
2	20 ≦ Act. < 50	10
3	50 ≦ Act. < 80	11
4	80 ≦ Act.	3

a) Hypoglycemic activity at 30 mg/kg *p.o.*

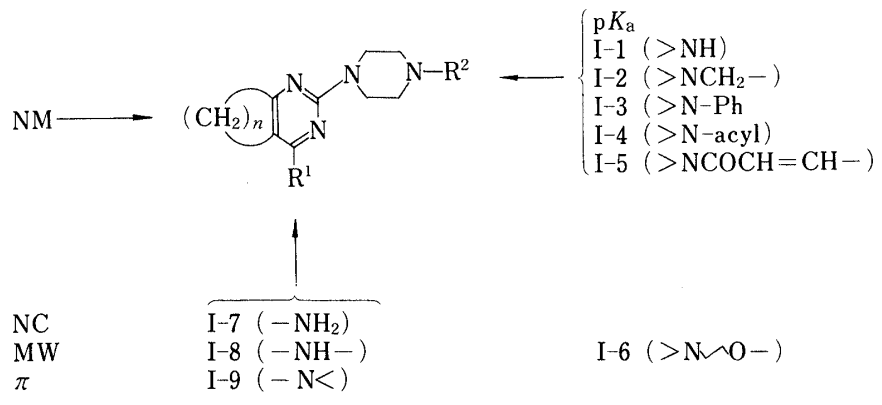


Fig. 1

TABLE II. Descriptor Variables

No.	n	R^1	R^2	Act.	MW	pK_a	I-1	I-2	I-3	I-4	I-5	I-6
1	4	NH_2	COPh	1	337	0	0	0	0	1	0	0
2	4	NH_2	COOiso-Bu	1	333	0	0	0	0	1	0	0
3	4	NH_2	CONHEt	1	304	0	0	0	0	1	0	0
4	4	NH_2	$CONHC_{10}H_7$[b]	1	411	0	0	0	0	1	0	0
5	4	NH_2	CSSEt	1	337	0	0	0	0	1	0	0
6	4	NH_2	$CSSCH_2Ph$	1	399	0	0	0	0	1	0	0
7	4	NH_2	SO_2(4-MePh)	1	387	0	0	0	0	1	0	0
8	4	NH_2	Ph	1	309	4.30	0	0	1	0	0	0
9	4	NH_2	Pyrrolidinyl	1	218	0	0	0	0	0	0	0
10	4	NH_2	Piperidino	1	232	0	0	0	0	0	0	0
11	4	NH_2	4-Bz-piperidino	1	367	0	0	0	0	0	0	0
12	4	NH_2	Morpholino	1	234	0	0	0	0	0	0	1
13	4	NH_2	CHO	2	261	0	0	0	0	1	0	0
14	4	NH_2	$COCH_3$	2	275	0	0	0	0	1	0	0
15	4	NH_2	COCH=CHPh	2	363	0	0	0	0	0	1	0
16	4	NH_2	CONHPh	2	419	0	0	0	0	1	0	0
17	4	NH_2	4-ClPh	2	343	4.30[c]	0	0	1	0	0	0
18	4	NH_2	2-MePh	2	323	4.30[c]	0	0	1	0	0	0
19	4	NHBu	H	2	781	11.11	1	0	0	0	0	0
20	4	NEt_2	H	2	289	11.11	1	0	0	0	0	0
21	4	$NHCH_2CH_2OH$	H	2	735	11.11	1	0	0	0	0	1
22	5	NH_2	H	2	247	11.11	1	0	0	0	0	0
23	4	NH_2	CH_3	3	338	10.17	0	1	0	0	0	0
24	4	NH_2	$CH_2CH=CH_2$	3	364	9.50[d]	0	1	0	0	0	0
25	4	NHEt	H	3	396	11.11	1	0	0	0	0	0
26	4	NMe_2	H	3	441	11.11	1	0	0	0	0	0
27	4	Pyrrolidino	H	3	287	11.11	1	0	0	0	0	0
28	4	NH_2	CH_2Ph	3	414	9.50	0	1	0	0	0	0
29	4	NH_2	COCH=CH (2-Fu)[e]	3	353	0	0	0	0	0	1	0
30	4	NH_2	4-MePh	3	323	4.30	0	0	1	0	0	0
31	3	NHMe	H	3	232	11.11	1	0	0	0	0	0
32	4	NH_2	H	3	360	11.11	1	0	0	0	0	0
33	4	NHMe	H	3	247	11.11	1	0	0	0	0	0
34	3	NH_2	H	4	219	11.11	1	0	0	0	0	0
35	3	Morpholino	H	4	289	11.11	1	0	0	0	0	1
36	4	Morpholino	H	4	303	11.11	1	0	0	0	0	1

a) The values for *N*-substituted pyrrolidine were used. *b*) $C_{10}H_7$ is a naphthyl group. *c*) The value used was that of *N*-phenylpyrrolidine. *d*) The value used was that of *N*-benzylpyrrolidine. *e*) Fu is a furyl group.

TABLE III. Correlation Matrix

	MW	pK_a	I-1	I-2	I-3	I-4	I-5	I-6
MW	1.000	0.182	0.159	0.066	−0.066	0.000	0.024	0.132
pK_a		1.000	0.849	0.259	−0.068	−0.639	−0.250	0.208
I-1			1.000	−0.226	−0.265	−0.446	−0.182	0.286
I-2				1.000	−0.106	−0.187	−0.073	−0.106
I-3					1.000	−0.219	−0.085	−0.125
I-4						1.000	−0.150	−0.085
I-5							1.000	0.219
I-6								1.000

$a_2=-0.111$, $a_3=1.055$, $a_4=1.833$, $b_1=-0.722$, $b_2=0.472$, and $b_3=1.444$. In the iterative adaptation the correction term, $C_i(t)$, for substance i misclassified at the tth iteration is given by eq. 4, [6b]

$$C_i(t)=0.1/(\delta_i(t)+0.45)^2+0.1 \tag{4}$$

where $\delta_i(t)$ is the deviation defined by eq. 5.

$$\delta_i(t)=|L_i(t)-b_k| \tag{5}$$

In eq. 5, $L_i(t)$ is the discriminant score for a substance i misclassified at the tth iteration, and b_k is the cutting point (nearer to $L_i(t)$) for the observed class for substance i. The adaptive iteration was performed 20 times,[6b] and the best discriminant function was selected. As the criteria of the best discrimination, the Spearman rank correlation coefficient with many ties, R_S, and the ε value[6] were used.

Computation——All computations were done with a UNIVAC 1100/81 computer, and all of the ALS programs used were written in ASCII FORTRAN. The original ALS programs (1980 version, written in JIS FORTRAN) were supplied by Prof. I. Moriguchi.

Results and Discussion

The ALS analysis was carried out without physicochemical parameters such as π and pK_a at first. In the descriptor selection, care was taken to avoid possible chance correlation. The best discriminant function using seven descriptors is expressed in eq. 8,

Table IV. Development of Equation 8

	Descriptors						Recognition		Prediction		Eq. No.
							n_{mis}	R_S	n_{mis}	R_S	
MW	I-1	I-2	I-3	I-5			10 (0)	0.866	13 (0)	0.819	(6)
MW	I-1	I-2	I-3	I-5	I-4		7 (0)	0.902	13 (0)	0.811	(7)
MW	I-1	I-2	I-3	I-5	I-4	I-6	6 (0)	0.911	10 (0)	0.847	(8)

Table V. Recognition and Prediction by Eqs. 8, 10, and 11

No.	Obs.	Recognition			Prediction			No.	Obs.	Recognition			Prediction		
		(8)	(10)	(11)	(8)	(10)	(11)			(8)	(10)	(11)	(8)	(10)	(11)
1	1	1	1	1	1	1	1	**19**	2	2	2	2	2	2	2
2	1	1	1	1	1	1	1	**20**	2	3	3	3	3	3	3
3	1	1	1	1	1	1	2	**21**	2	2	2	2	3	2	3
4	1	1	1	1	1	1	1	**22**	2	3	3	3	3	3	3
5	1	1	1	1	1	1	1	**23**	3	3	3	3	3	3	3
6	1	1	1	1	1	1	1	**24**	3	3	3	3	3	3	3
7	1	1	1	1	1	1	1	**25**	3	3	3	3	3	3	3
8	1	2	2	2	2	2	2	**26**	3	3	3	3	3	3	3
9	1	1	1	1	1	1	1	**27**	3	3	3	3	3	3	3
10	1	1	1	1	1	1	1	**28**	3	3	3	3	3	3	2
11	1	1	1	1	1	1	1	**29**	3	3	3	3	2	2	2
12	1	1	1	1	2	1	2	**30**	3	2	2	2	2	2	2
13	2	2	2	2	2	2	2	**31**	3	3	3	3	3	4	3
14	2	2	2	2	2	2	2	**32**	3	3	3	3	3	3	3
15	2	2	2	2	3	3	3	**33**	3	3	3	3	3	3	3
16	2	1	1	1	1	1	1	**34**	4	3	4	4	3	3	3
17	2	2	2	2	2	2	2	**35**	4	4	3	4	4	3	4
18	2	2	2	2	2	2	2	**36**	4	4	3	4	4	3	3
								% correct.		83	81	86	72	69	64

$$L = -0.00291\mathrm{MW} + 2.685(\mathrm{I\text{-}1}) + 2.965(\mathrm{I\text{-}2}) + 1.655(\mathrm{I\text{-}3}) + 0.957(\mathrm{I\text{-}4}) + 2.341(\mathrm{I\text{-}5}) + 0.484(\mathrm{I\text{-}6}) - 0.826 \quad (8)$$

$n=36$ (4 gr.), $n_{\mathrm{mis}}=6(0)$, $p=0.833$, $R_S=0.911$, $\varepsilon=0.394$

where n stands for the number of compounds, n_{mis} is the number misclassified, the figure in parentheses after the value of n_{mis} is the number misclassified into the next class but one, and p describes the percentage of correct classifications. Table IV shows the development of eq. 8.

To confirm the validity of the ALS results, the leave-one-out technique was applied, and the descriptor set in eq. 8 gave the best predictive success as shown in Table IV. The results of recognition and prediction by eq. 8 are given in Table V.

The descriptors of eq. 8 appear to have the following implications. The MW seems to be a kind of correction term from the activity at the same weight to the activity at equal molar amount, because increase of MW caused the activity to decrease. I-6 indicates that the presence of the $\rangle NCH_2CH_2O-$ structure contributes to the effect. The indicator variables I-1, I-2, I-3, I-4, and I-5 suggest that the presence of the piperazinyl group is important for the activity. In particular, I-1 and I-2 indicate that the $-NH-$ and $\rangle NCH_2-$ moieties are favorable for hypoglycemic activity. To clarify the physicochemical meanings of the indicator variables related to the piperazinyl group, we examined the relation using the pK_a values as the basicity and the π values as the lipophilicity instead of I-1, I-2, and I-3, and obtained eqs. 10 and 11, of which the latter was the best recognition equation. The development of eq. 11 is given in Table VI.

TABLE VI. Development of Equation 11

Descriptors					Recognition		Prediction		Eq. No.
					n_{mis}	R_S	n_{mis}	R_S	
MW	pK_a	I-5			10 (0)	0.866	13 (0)	0.811	(9)
MW	pK_a	I-5	I-4		7 (0)	0.902	11 (0)	0.853	(10)
MW	pK_a	I-5	I-4	I-6	5 (0)	0.921	13 (0)	0.794	(11)

$$L = -0.00220\mathrm{MW} + 0.232pK_a + 0.592(\mathrm{I\text{-}4}) + 1.911(\mathrm{I\text{-}5}) - 0.651 \quad (10)$$

$n=36$(4 gr.), $n_{\mathrm{mis}}=7(0)$, $p=0.806$, $R_S=0.902$, $\varepsilon=0.413$

$$L = -0.00286\mathrm{MW} + 0.241pK_a + 0.741(\mathrm{I\text{-}4}) + 2.099(\mathrm{I\text{-}5}) + 0.358(\mathrm{I\text{-}6}) - 0.602 \quad (11)$$

$n=36$(4 gr.), $n_{\mathrm{mis}}=5(0)$, $p=0.861$, $R_S=0.921$, $\varepsilon=0.423$

The predictive success of eq. 11 by the leave-one-out technique was slightly worse than that of eq. 10. The contribution indexes of the descriptors in eq. 11 are as follows: MW, 0.334; pK_a, 1.240; I-4, 0.332; I-5, 0.480; and I-6, 0.112. The descriptors in eq. 11 are the same as those of eq. 8 except pK_a, which was replaced by I-1, I-2, and I-3 in eq. 8.

These findings suggest that the presence of the piperazinyl group contributes to the activity, and the basicity of the piperazinyl group is the most important factor. The factors first examined, such as the variables I-7, I-8, and I-9, which were for the kinds of 4-amino substituents of pyrimidines, did not show any relation to the activity. NM, the number of methylenes in the 5,6-polymethylene group, also had no relation to the hypoglycemic effect. The findings suggest that the substitutent of pyrimidine may not be an essential factor except

in the case of the 2-(1-piperazinyl) group.

From these results, it may be concluded that 2-(1-piperazinyl)pyrimidine is the key structure for hypoglycemic activity and that the basicity of the piperazinyl group is the most important factor. Recently, we synthesized some structurally related compounds, 4-alkoxy and 4-alkyl-2-(1-piperazinyl)pyrimidines, and examined their hypoglycemic activities.[1,9] Their structure–activity relationships support the above conclusions.

Acknowledgement The authors wish to thank Prof. I. Moriguchi, Kitasato University, for supplying the ALS program and for many helpful suggestions, and they are indebted to Dr. Y. Masuda and Dr. T. Kobayashi of this laboratory for their support and encouragement throughout this work.

References

1) Part VI: T. Sekiya, H. Hiranuma, T. Kanayama, S. Hata, and S. Yamada, *Chem. Pharm. Bull.*, **31**, 2254 (1983).
2) T. Sekiya, H. Hiranuma, T. Kanayama, and S. Hata, *Eur. J. Med. Chem.*, **15**, 317 (1980).
3) C. Hansch and T. Fujita, *J. Am. Chem. Soc.*, **86**, 1616 (1964).
4) C. Hansch, "Drug Design," Vol. 1, ed. by E. J. Ariëns, Academic Press, New York, 1971, pp. 271—342.
5) J. Thomas, C. E. Berkoff, W. B. Flagg, J. J. Gallo, R. F. Haff, C. A. Pinto, C. Pellerano, and L. Savini, *J. Med. Chem.*, **18**, 245 (1975).
6) *a*) I. Moriguchi, K. Komatsu, and Y. Matsushita, *J. Med. Chem.*, **23**, 20 (1980); *b*) I. Moriguchi and K. Komatsu, Abstracts of Papers, 8th Symposium on Structure–Activity Relationships, Tokyo, Oct. 1981, p. 55.
7) L. C. Craig and R. M. Hixon, *J. Am. Chem. Soc.*, **53**, 4367 (1931).
8) C. Hansch and A. Leo, "Substituent Constants for Correlation Analysis in Chemistry and Biology," John Wiley and Sons, New York, 1979, pp. 18—43.
9) T. Sekiya, H. Hiranuma, T. Kanayama, S. Hata, and S. Yamada, *Eur. J. Med. Chem.*, **17**, 75 (1982).