Formulation of *De Novo* Substituent Constants in Correlation Analysis: Inhibition of Dihydrofolate Reductase by 2,4-Diamino-5-(3,4-dichlorophenyl)-6-substituted Pyrimidines

CORWIN HANSCH [x], CARLO SILIPO [*], and EDWARD E. STELLER

Abstract □ A correlation equation based solely on *de novo* constants was formulated for 105 2,4-diamino-5-(3,4-dichlorophenyl)-6-substituted pyrimidines acting as inhibitors of dihydrofolate reductase. An equation with seven indicator variables gives a correlation with a correlation coefficient of 0.903 and a standard deviation of 0.229. The technique used is a modification of the Free–Wilson approach. The results indicate that correlation equations with fewer parameters than the theoretical required to account for all molecular changes may often be encountered. It is also shown that cross-product terms can be used to establish the significance of cooperative substituent effects.

Keyphrases □ Correlation analysis—formulation of *de novo* substituent constants, inhibition of dihydrofolate reductase by 2,4-diamino-5-(3,4-dichlorophenyl)-6-substituted pyrimidines, equation with seven indicator variables □ Dihydrofolate reductase—inhibition by 2,4-diamino-5-(3,4-dichlorophenyl)-6-substituted pyrimidines, correlation analysis, formulation of *de novo* substituent constants, cross-product terms □ Pyrimidines, 2,4-diamino-5-(3,4-dichlorophenyl)-6-substituted—inhibitors of dihydrofolate reductase, correlation analysis, formulation of *de novo* substituent constants, cross-product terms

During the past decade, the use of computers and statistical techniques in the correlation of chemical structure with biological activity has been developing at a continually increasing rate. It is difficult now even for specialists in structure–activity work to keep up with new techniques. Medicinal chemists are beginning to work with tools developed in the fields of psychology and economics—areas long plagued by multivariable problems. Discriminate analysis (1), pattern recognition (2), factor analysis (3), cluster analysis (4), multidimensional scaling (5), and regression analysis (6, 7) are being applied to various structure–activity problems. These techniques will be coupled with the computerized manipulation of organic structural data to develop increasingly more powerful analytical tools (8) for drug design.

BACKGROUND

Regression analysis has been extensively employed in two ways. The most experience has been obtained in relating biological activity of sets of congeners to physicochemical properties using constants from model systems (6, 7). An alternative method, first used by Bruice *et al.* (9) and then more fully developed by Free and Wilson (10), derives *de novo* substituent constants directly from the biological activities of a set of congeners. The Free–Wilson approach was succinctly stated as (11):

$$A_n = \sum_p \sum_s a_{n.ps} + \mu \qquad \text{(Eq. 1)}$$

where A is a biological response, μ is the activity contributed by the constant portion (parent structure) of the congeneric set, and a is the contribution of activity by each substituent s located on position p of the parent structure. Equation 1 can be used to account for any structural feature in a set of congeners contributing to the total activity of each congener in a strictly additive fashion. The use of Free–Wilson-type constants, which can be regarded in a general sense as discrete variables [indicator variables (12)], can be combined with continuous variables such as π or σ to extend greatly one's ability to formulate correlation equations (13–17)[1].

Following the Free–Wilson approach strictly, one would need a discrete variable for every specific structural change present in a set of congeners. This requirement places one in a dilemma; one must either make a rather large number of derivatives to be certain of formulating meaningful constants or use relatively few derivatives to obtain rather unreliable constants. The authors have found that, in some cases, changes that might seem rather different from a chemist's point of view can be covered by a single indicator variable (14)[1]; this opens up new possibilities in the development of correlation equations.

It was found in the present study of the congeners listed in Table I that the continuous variables π, MR (molar refractivity), and σ were of no value in the formulation of the structure–activity relationship. However, a quite good correlation equation for 105 of the congeners of Table I could be formulated (Eq. 7)[2] using only seven indicator variables. The distinct structural features of the compounds of Table I are summarized in Table II.

By taking certain liberties with Eq. 1, matters can be simplified greatly by assuming that a given substituent Z or SO_2F always makes the same contribution to A, regardless of the length and geometry of the two bridges connecting the two phenyl rings to the parent structures. Even with this simplification, to follow Eq. 1 strictly, one would need to formulate 36 *de novo* constants, with many based on one or two data points. Previous experience has shown that enzyme inhibitors do not always show such high specificity or, to put it another way, test systems are not always sensitive enough to detect small biological responses.

After first discovering that the functions π and MR were of no value in correlating the data of Table I, the indicator variables of Table III were studied. The features of Table III were parameterized by assigning a value of 1 for their presence or of 0 for their absence. Some attempts were made to combine certain variables; *I*-18 was used to account for *I*-12 + *I*-13, *I*-21 for *I*-8 + *I*-9, and *I*-22 for *I*-10 + *I*-11. These combinations proved to be less significant than the single variables.

Three variables (*I*-19, *I*-20, and *I*-27) were studied to test the possibility that enzymes from different sources showed different activity. Only *I*-20 was necessary in the final analysis; that is, for practical purposes, enzyme from liver, L-1210/FR8, or L-1210/DF8 behaved in the same way.

The irreversible activating effect of the various inhibitors is also listed in Table I. Log $1/C$ values are for reversible inhibition. No obvious relationship is seen between the two types of inhibition.

The degree of independence of the important variables is given in Table IV; collinearity is not a problem. The biological data are from Refs. 18–26.

RESULTS

A reasonable alternative in starting an analysis with such a large number of variables, where it is impossible to derive equations with all possible linear combinations of variables, is to first formu-

[1] C. Silipo and C. Hansch, to be published.
[2] Equations 2–12 are shown in Table V.

Table I—Inhibition Constants and Indicator Variables for the Inhibition of Dihydrofolate Reductase by 2,4-Diamino-5-(3,4-dichlorophenyl)-6-substituted Pyrimidines

Number	Y	Substituents[a]	Log 1/C Obs.[b]	Log 1/C Calc.[c]	$\lvert\Delta \text{Log } \frac{1}{C}\rvert$	I-1	I-2	I-4	I-7	I-8	I-9	I-10	I-13	I-15	I-17	I-20	Inactivation[d], %	Enzyme Source[e]	Reference for Inhibition Data
1	2-Cl	4-NHCONHC₆H₄—4'-SO₂F	5.48	5.941	0.46	0	1	0	1	1	0	0	0	1	0	0	58	2	20
2f	2,6-(CH₃)₂	4-NHCONHC₆H₄—3'-SO₂F	5.48	6.161	0.68	0	0	0	1	1	0	0	0	0	0	0	17	2	19
3		4-NHCONHC₆H₄—4'-SO₂F	5.68	5.941	0.26	0	0	0	0	1	0	0	0	1	0	0	82	4	20
4		4-NHCONHC₆H₃—3'-CH₃,4'-SO₂F	5.70	5.941	0.24	0	0	0	0	1	0	0	0	1	1	0	45	4	20
5	3-Cl	4-NHCONHC₆H₃—3'-Cl,4'-SO₂F	5.74	5.941	0.20	0	1	0	1	1	0	0	0	1	1	0	82	2	20
6	2-Cl	4-NHCONHC₆H₄—3'-SO₂F	5.85	6.161	0.31	0	0	0	0	1	0	0	0	1	0	0	98	2	19
7		4-NHCONHC₆H₄—3'-SO₂F	6.00	6.161	0.16	0	0	0	1	1	0	0	0	0	0	0	50	4	19
8	3-Cl	4-NHCONHC₆H₄—4'-SO₂F	6.01	5.941	0.07	1	0	0	1	1	0	0	0	1	1	0	90	2	20
9	3-Cl	4-NHCONHC₆H₃—3'CH₃,4'-SO₂F	6.01	5.941	0.07	1	0	0	0	1	0	0	0	1	1	0	87	2	20
10		4-NHCONHC₆H₃—3'-Cl,4'-SO₂F	6.04	6.306	0.27	0	0	0	0	1	1	0	0	1	1	0	76	2	21
11		4-NHCONHC₆H₄—4'-SO₂F	6.05	6.306	0.26	0	1	0	1	1	0	0	0	0	0	0	76	2	21
12		4-NHCOC₆H₄—3'-SO₂F	6.09	6.390	0.30	0	0	0	1	1	1	0	0	0	0	0	94	2	18
13	3-Cl	4-NHCONHC₆H₃—2'-Cl,5'-SO₂F	6.17	6.161	0.01	0	0	0	0	0	0	0	0	1	1	0	100	2	19
14		4-NHCONHC₆H₃—2'-Cl,5'-SO₂F	6.25	6.161	0.09	0	0	0	1	1	1	0	0	0	0	0	50	4	19
15		4-NHCOC₆H₄—3'-SO₂F	6.28	6.390	0.11	0	1	0	1	0	0	0	0	0	0	0	82	2	21
16	2-Cl	4-NHCONHC₆H₃—3'-OCH₃,4'-SO₂F	6.28	5.941	0.34	0	0	0	1	1	0	0	0	1	1	0	63	2	20
17	3-Cl	4-NHCOC₆H₃—4'-CH₃,3'-SO₂F	6.30	6.390	0.09	0	1	0	1	0	0	0	0	0	1	0	51	2	21
18	2-Cl	4-NHCONHC₆H₃—2'-OCH₃,5'-SO₂F	6.32	6.161	0.16	0	0	0	0	1	0	0	0	0	1	0	50	4	19
19		4-NHCONHC₆H₃—4'-OC₂H₅,3'-SO₂F	6.35	6.161	0.19	0	0	0	0	0	0	1	0	0	1	0	55	4	19
20f		4-CH₂NHCONHC₆H₃—2'-Cl,5'-SO₂F	6.37	7.174	0.80	0	0	0	0	0	0	1	0	0	1	0	100	2	22
21		4-CH₂NHCONHC₆H₃—2'-CH₃,4'-SO₂F	6.37	6.954	0.58	0	0	0	0	0	0	0	0	0	1	0	71	2	22
22f		4-CH₂NHCOC₆H₃—3'-SO₂F	6.43	7.174	0.74	0	0	0	0	0	0	1	0	0	1	0	100	2	22
23f		4-CH₂NHCOC₆H₄—4'-CH₃,3'-SO₂F	6.46	7.174	0.71	0	0	0	0	0	0	0	0	0	1	0	70	4	22
24	3-Cl	4-NHCONHC₆H₄—3'-SO₂F	6.47	6.161	0.31	0	0	1	1	1	1	0	0	1	1	0	100	2	19
25f		3-CH₂NHCONHC₆H₃—2'-Cl,5'-SO₂F	6.48	7.174	0.69	0	0	0	0	0	0	0	0	0	1	0	93	1	25
26		4-CH₂NHCONHC₆H₃—2'-CH₃,5'-SO₂F	6.49	7.174	0.68	0	0	1	0	1	0	0	0	0	1	0	84	2	22
27	3-CH₃	4-NHCOC₆H₄—4'-SO₂F	6.51	6.615	0.11	0	0	0	1	0	0	0	0	1	1	0	50	4	20
28	2-Cl	4-NHCOC₆H₄—3'-SO₂F	6.51	6.684	0.17	0	1	0	1	1	1	0	0	1	0	1	39	3	21
29	3-CH₃	4-NHCOC₆H₄—3'-SO₂F	6.55	6.390	0.16	0	0	1	1	1	1	0	0	0	1	0	94	2	21
30	3-Cl	4-NHCONHC₆H₃—2'-OCH₃,5'-SO₂F	6.55	6.161	0.39	0	0	0	1	0	0	0	0	1	1	0	55	2	19
31	3-Cl	4-NHCONHC₆H₃—4'-OCH₃,3'-SO₂F	6.68	6.161	0.52	0	0	1	1	1	1	0	0	0	1	0	95	2	19
32		4-NHCOC₆H₄—4'-SO₂F	6.74	6.684	0.06	0	0	0	0	0	0	0	0	1	0	1	33	3	21
33		4-NHCONHC₆H₄—4'-OCH₃,3'-SO₂F	6.77	6.674	0.10	0	0	0	0	1	0	0	0	1	1	0	80	3	19
34		4-(CH₂)₂NHCOC₆H₄—4'-SO₂F	6.77	6.954	0.18	0	0	0	0	0	0	0	0	0	1	0	50	4	22
35		4-(CH₂)₂NHCOC₆H₄—4'-SO₂F	6.80	6.954	0.15	0	0	0	1	0	0	0	0	0	0	0	76	1	25
36	3-CH₃	4-CH₂NHCOC₆H₄—4'-SO₂F	6.80	6.954	0.15	0	0	1	1	0	1	0	0	0	0	0	98	1	24
37	2-Cl	4-NHCOC₆H₃—2'-Cl,5'-SO₂F	6.80	6.904	0.10	0	1	0	1	1	0	0	0	1	1	0	53	3	21
38	3-CH₃	4-NHCONHC₆H₃—3'-CH₃,4'-SO₂F	6.80	6.615	0.19	0	0	1	0	0	0	1	0	0	1	0	84	4	20
39		4-(CH₂)₂NHCONHC₆H₄—4'-SO₂F	6.82	6.954	0.13	0	0	0	1	0	0	1	0	1	0	0	73	1	25
40	2-Cl	4-CH₂NHCONHC₆H₃—4'-CH₃,3'-SO₂F	6.82	7.174	0.35	0	1	1	0	0	0	0	0	0	1	0	62	1	24
41		4-CH₂NHSO₂C₆H₄—4'-SO₂F	6.82	6.954	0.13	0	0	0	0	0	0	0	1	0	0	0	85	2	22

(continued)

Table I—Continued

Structure heading for Substituents column: Z—[bridge]—(C₆H₄)—SO₂F

Number	Y	Substituents[a]	Log 1/C Obs.[b]	Log 1/C Calc.[c]	|Δ Log 1/C|	I-1	I-2	I-4	I-7	I-8	I-9	I-10	I-13	I-15	I-17	I-20	Inactivation[d], %	Enzyme Source[e]	Reference for Inhibition Data
42		4-CH₂NHCONHC₆H₄—3'-SO₂F	6.82	7.174	0.35	0	0	0	0	0	0	1	0	0	0	0	100	2	22
43		4-NHCOC₆H₃—4'-CH₃,3'-SO₂F	6.82	6.904	0.08	0	0	0	0	0	1	0	0	0	1	1	35	3	21
44		4-CH₂NHCOC₆H₃—4'-CH(CH₃)₂,3'-SO₂F	6.85	7.174	0.32	0	0	0	0	0	0	0	0	0	1	0	85	2	22
45	3-CH₃	4-NHCONHC₆H₃—2'-OCH₃,5'-SO₂F	6.85	6.835	0.02	0	0	1	1	1	0	0	0	0	1	0	100	4	19
46	2-Cl	4-CH₂NHSO₂C₆H₄—4'-SO₂F	6.89	6.954	0.06	0	1	0	1	0	0	0	0	1	0	0	12	1	24
47	3-CH₃	4-NHCONHC₆H₃—2'-Cl,5'-SO₂F	6.89	6.835	0.06	0	0	1	1	1	0	0	0	0	1	0	100	4	19
48	2-Cl	4-CH₂NHCONHC₆H₃—3'-CH₃,4'-SO₂F	6.92	6.954	0.03	0	1	0	1	1	0	0	0	0	1	0	78	1	24
49	2-Cl	4-CH₂NHCOC₆H₃—4'-CH₃,3'-SO₂F	6.92	7.174	0.25	0	0	1	1	0	1	1	0	0	0	0	77	3	24
50	3-CH₃	4-NHCOC₆H₄—4'-SO₂F	6.92	6.684	0.24	0	0	0	1	0	0	0	0	0	1	1	45	1	21
51		3-CH₂NHSO₂C₆H₄—3'-SO₂F	6.96	7.174	0.21	0	0	0	0	0	0	1	0	1	0	0	25	1	25
52		4-(CH₂)₂NHCONHC₆H₄—3'-SO₂F	6.96	7.174	0.21	0	0	0	0	0	0	0	0	0	0	0	48	1	25
53		4-(CH₂)₂NHSO₂C₆H₄—4'-SO₂F	6.96	6.954	0.01	0	0	0	0	0	0	0	0	1	0	0	60	1	25
54	3-CH₃	4-CH₂NHSO₂C₆H₄—4'-SO₂F	6.96	6.954	0.01	1	0	1	1	0	0	0	0	0	1	0	94	3	24
55		4-NHCONHC₆H₃—2'-Cl,5'-SO₂F	6.96	6.954	0.01	0	0	0	0	0	0	0	0	1	0	0	82	2	21
56		4-CH₂NHCONHC₆H₃—4'-CH₃,3'-SO₂F	6.96	7.039	0.08	0	0	0	0	1	0	0	0	0	1	1	81	3	22
57	3-CH₃	4-NHCOC₆H₃—2'-Cl,5'-SO₂F	7.00	7.174	0.17	0	0	1	0	1	0	0	0	0	1	0	91	1	21
58	2-OCH₃	4-CH₂NHCOC₆H₃—2'-Cl,5'-SO₂F	7.00	6.904	0.10	0	0	1	1	0	1	1	0	0	0	0	77	2	24
59		4-CH₂NHCONHC₆H₃—2'-Cl,5'-SO₂F	7.02	6.954	0.07	0	0	0	0	0	0	0	0	0	1	0	90	1	22
60	3-CH₃	4-CH₂NHSO₂C₆H₄—4'-SO₂F	7.04	7.174	0.13	0	0	0	1	0	0	0	0	0	1	0	100	1	25
61	2-Cl	4-CH₂NHCOC₆H₃—2'-Cl,5'-SO₂F	7.05	7.174	0.12	0	1	0	0	0	0	0	0	0	1	0	65	1	24
62	2-OCH₃	4-CH₂NHCOC₆H₃—2'-Cl,5'-SO₂F	7.07	6.954	0.12	0	0	0	0	0	0	0	0	0	1	1	37	1	24
63	3-CH₃	4-CH₂NHSO₂C₆H₄—4'-SO₂F	7.07	6.954	0.12	0	0	0	0	0	0	0	1	0	1	0	68	2	24
64		4-CH₂NHCONHC₆H₄—4'-SO₂F	7.07	6.954	0.12	0	0	0	0	0	0	0	0	0	1	0	95	2	25
65f		3-CH₂NHCONHC₆H₄—4'-SO₂F	7.07	6.954	0.12	0	0	0	0	0	1	1	0	0	0	0	58	1	25
66	2-OCH₃	3-NHCOC₆H₃—2'-Cl,5'-SO₂F	7.08	6.390	0.68	1	0	1	1	1	0	1	0	0	0	0	93	3	21
67		4-CH₂NHCONHC₆H₄—4'-SO₂F	7.12	7.319	0.20	0	0	0	0	0	0	0	0	0	1	0	83	1	25
68		3-CH₂NHCONHC₆H₄—3'-SO₂F	7.17	7.174	0.00	0	0	0	0	0	0	0	0	0	1	0	46	1	24
69		3-CH₂NHCOC₆H₄—4'-SO₂F	7.18	6.954	0.23	0	0	0	1	0	0	1	0	0	0	0	78	1	24
70		4-NHCONHC₆H₄—3'-SO₂F	7.20	7.319	0.12	0	0	0	1	1	0	0	0	0	1	1	69	3	19
71	3-CH₃	4-CH₂NHCONHC₆H₄—3'-SO₂F	7.20	7.174	0.03	0	0	0	0	0	0	0	0	0	1	0	51	1	24
72	2-Cl	4-CH₂NHCONHC₆H₄—4'-SO₂F	7.21	6.954	0.26	0	1	0	1	0	0	0	0	0	1	0	46	1	24
73	3-CH₃	4-CH₂NHCONHC₆H₄—4'-SO₂F	7.22	7.174	0.05	0	0	0	1	0	0	0	0	0	1	0	56	1	25
74		4-CH₂NHCOC₆H₄—4'-SO₂F	7.22	7.174	0.05	1	0	0	1	0	0	0	0	0	0	0	55	1	25
75		3-CH₂NHSO₂C₆H₄—3'-SO₂F	7.22	7.174	0.05	0	0	0	0	0	0	0	0	1	0	0	17	2	22
76		4-CH₂NHSO₂C₆H₄—4'-SO₂F	7.24	7.174	0.07	0	0	0	0	0	0	0	0	1	0	0	73	1	25
77		3-CH₂NHCOC₆H₄—3'-SO₂F	7.24	7.269	0.03	1	0	0	0	0	1	1	0	0	0	0	82	3	21
78		4-NHCOC₆H₄—3'-SO₂F	7.25	7.174	0.08	0	0	0	1	0	0	0	0	0	1	1	93	1	25
79		4-(CH₂)₂NHCONHC₆H₄—3'-SO₂F	7.25	6.954	0.30	0	0	0	0	0	0	0	0	0	0	0	46	1	25
80		3-CH₂NHSO₂C₆H₄—3'-SO₂F	7.25	7.539	0.29	1	0	0	0	0	0	0	0	1	0	0	64	1	25
81		3-CH₂NHSO₂C₆H₄—3'-SO₂F	7.26	7.468	0.21	0	0	0	0	0	0	0	1	1	0	0	23	3	22
82		4-CH₂NHCOC₆H₄—4'-SO₂F	7.27	7.174	0.10	1	0	0	1	0	0	0	0	0	0	0	82	2	25
83	3-CH₃	3-CH₂NHCONHC₆H₃—4'-CH₃,3'-SO₂F	7.28	7.373	0.09	0	0	0	0	0	0	0	1	0	1	1	46	1	25
84	2-OCH₃	4-NHSO₂C₆H₄—4'-SO₂F	7.28	7.174	0.11	0	1	0	1	0	0	0	1	0	0	0	64	1	24
85		4-CH₂NHCOC₆H₄—3'-SO₂F	7.31	7.174	0.14	0	0	0	0	0	0	0	0	0	1	0	23	1	25
86		4-(CH₂)₂NHSO₂C₆H₄—3'-SO₂F	7.31	7.373	0.06	0	0	0	0	0	0	0	1	1	0	0	93	3	21
87	2-Cl	4-NHSO₂C₆H₄—4'-SO₂F	7.32	7.174	0.15	0	1	0	1	0	0	0	0	0	0	0	33	2	24
88	2-Cl	4-CH₂NHCOC₆H₄—3'-SO₂F	7.33	7.373	0.04	0	1	0	1	0	0	0	1	0	0	0	57	1	25

late an equation with the largest number of terms possible. Often all variables cannot be considered because high collinearity of certain vectors yields singular matrixes which cannot be solved for the regression coefficients. In the present instance, 22 variables of Table III (I-3, I-6, I-16, I-18, I-21, I-22, and cross-product variables not included) gave a correlation equation with $s = 0.240$ and $r = 0.910$. Adding four cross-product terms (I-4·I-8, I-4·I-17, I-8·I-17, and I-4·I-8·I-17) gave a 26-variable equation with $s = 0.212$ and $r = 0.934$. This equation constitutes a kind of upper limit, which one would hope to approach with fewer terms. Equations 8 and 11 come close to this type of approach.

Such brute force methods cannot be used with cross-product terms. The enormous number of such terms is given by the expression $2^n - (n + 1)$, where n is the number of primary terms. Hence, for only 10 variables, there would be 1013 possible cross-product terms! The possibilities are truly enormous for interactions of a complex inhibitor with a macromolecule.

To search for patterns of cooperative activity among the different structural units, the authors explored the 15 cross-product terms of Table III. These possibilities were selected simply by study of the residuals of the best equations. It seems unlikely that a mechanical machine approach can be developed to uncover the significant interactions among such a large number of variables. The human mind is the most efficient pattern recognizer at this level of complexity.

After considering thousands of regression equations, the following variables were found to be the most significant: I-1, I-2, I-7, I-8, I-9, I-10, I-13, I-15, I-20, I-4·I-8, and I-8·I-17. Equations with all possible linear combinations of these variables were derived (2047). The "best" equation (that with the lowest standard deviation) in each class is listed in Table V. The terms appear quite independent and fall into place in regular order. All of the terms in the equations of Table V are significant as judged by the F test, except Eq. 11 ($F_{1,60\,\alpha 0.001} = 12.0$; $F_{1,60\,\alpha.05} = 4.1$; $F_{1,60\,\alpha.1} = 2.8$). There is little reduction in the variance in the equations above Eq. 8:

$$\log 1/C = 0.365(\pm 0.12)I\text{-}1 - 1.013(\pm 0.12)I\text{-}8 - 0.784(\pm 0.19)I\text{-}9 +$$
$$0.419(\pm 0.20)I\text{-}13 - 0.220(\pm 0.09)I\text{-}15 + 0.513(\pm 0.18)I\text{-}20 +$$
$$0.674(\pm 0.23)I\text{-}4\cdot I\text{-}8 + 7.174(\pm 0.07)$$

n	r	s	
105	0.903	0.229	(Eq. 8)

The F statistic also drops sharply at this point. Equation 8 would seem to be the best point from which to make projections.

It is interesting that the cross-product term I-4·I-8 plays a significant role in Eq. 8. The positive coefficient with this term means that the cooperative activity of a 3-CH₃ and a 4-NHCONH— bridge adds significantly to the activity. This is not a "pure" cross-product term, since I-4 does not occur in Eq. 8 except in the product form. The 3-CH₃ appears to make a significant contribution to activity only when the 4-NHCONH— bridge is present. When the 3-CH₃ is present with other bridges, the data points are well fit (except Compound 100), although there are no variables in Eq. 8 to account for this function. If I-4 is used alone, the correlation is poorer than when I-4·I-8 is used; use of this cross-product with I-4 gives an equation with a low coefficient with I-4. The 3-Cl substituent does not behave in this fashion.

In Table V, it is seen that the two most significant variables reflect structural changes that produce lower activity (I-8 and I-9). A most interesting point is Compound 96 (an alternative parent compound), which is the only congener showing only reversible activity; this highlights the unusual specificity of the SO₂F function.

Six points in Table I were not used in the regression analysis. Four of these (Compounds 20, 22, 23, and 25) contain the bridge —CH₂NHCO— or —CH₂NHCONH—; however, the use of I-10, I-11, and I-22 and the study of many cross-products failed to improve the correlation of these compounds since other congeners containing these functions are well fit. It seems strange that these four are so poorly fit and one wonders if it could be due to experimental error. Data point 2, which is also poorly fit, is unique in that it contains di-*ortho*-substitution.

If Compound 108 is taken as the parent structure, it can be seen from Table I that only three congeners with slightly greater activi-

---

Table (Compounds 89–111)

No.	w-subst	R	log 1/C (obs)	log 1/C (calc)[c]	Δlog 1/C	% [d]	e
89	2-OCH₃	4-CH₂NHCOC₆H₄—4'-SO₂F	7.33	6.954	0.38	68	1
90		4-CH₂NHCOC₆H₃—4'-Cl,3'-SO₂F	7.33	7.174	0.16	100	2
91	2-Cl	4-CH₂NHCOC₆H₄—4'-SO₂F	7.34	6.954	0.39	56	1
92		4-(CH₂)₂C₆H₄—4'-SO₂F	7.34	7.319	0.02	97	2
93	2-Cl	3-CH₂NHCOC₆H₄—4'-SO₂F	7.38	7.319	0.06	82	1
94		4-CH₂NHSO₂C₆H₄—4'-SO₂F	7.39	7.174	0.22	20	1
95		4-NHCOC₆H₄—4'-SO₂F	7.39	7.049	0.34	49	3
96		4-SO₂F	7.40	7.539	0.14	0	1
97	2-OCH₃	4-CH₂NHSO₂C₆H₄—3'-SO₂F	7.41	7.174	0.24	37	1
98	2-Cl	4-CH₂NHCONHC₆H₃—2'-Cl,5'-SO₂F	7.44	7.174	0.27	60	1
99		4-(CH₂)₂C₆H₄—3'-SO₂F	7.44	7.539	0.10	92	2
100	3-CH₃	4-CH₂NHCONHC₆H₄—3'-SO₂F	7.52	7.174	0.35	100	2
101		4-CH₂NHCONHC₆H₄—4'-SO₂F	7.52	7.319	0.20	80	2
102		4-CH₂NHCONHC₆H₄—3'-SO₂F	7.54	7.539	0.02	75	1
103		4-NHSO₂C₆H₄—3'-SO₂F	7.57	7.593	0.05	23	2
104	2-Cl	4-CH₂NHCOC₆H₃—4'-CH₃,3'-SO₂F	7.60	7.539	0.03	94	2
105		4-CH₂NHCOC₆H₄—3'-SO₂F	7.62	7.539	0.06	100	2
106	3-(CH₂)₂	C₆H₄—4'-SO₂F	7.66	7.319	0.30	82	2
107		4-CH₂NHCOC₆H₄—4'-SO₂F	7.68	7.319	0.34	68	1
108		4-SO₂F	7.70	7.174	0.51	43	2
109	3-CH₃	4-NHSO₂C₆H₄—3'-SO₂F	7.70	7.593	0.11	97	3
110		4-CH₂NHCOC₆H₃—2'-Cl,5'-SO₂F	7.74	7.539	0.16	80	
111		4-NHSO₂C₆H₄—3'-SO₂F	7.74	7.593	0.15	42	

[a] The function w is either CH₂ or O; this is parameterized by $w - 1 = 1$ for $w = $ CH₂. [b] From Refs. 18–26. [c] Calculated from Eq. 8. [d] Percent irreversible inactivation from Refs. 18–26. [e] 1 = enzyme from liver, 2 = enzyme L-1210/DF8 leukemia, 3 = L-1210/FR8 leukemia, and 4 = L-1210/0 leukemia. [f] These points not used in deriving Eqs. 2–12.
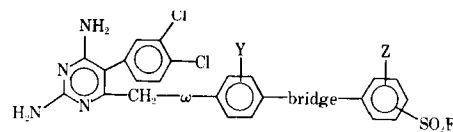
Table II—Structural Features of Congeners of Table I

w	Number of Occurrences	Y	Number of Occurrences	Bridge	Number of Occurrences	Z	Number of Occurrences	SO$_2$F	Number of Occurrences
O	92	2-Cl	20	4-NHCONH—	27	2'-Cl	13	5'-	17
CH$_2$	19	3-Cl	8	4-NHCO—	13	3'-Cl	2	4'-	45
		2-CH$_3$	1	4-CH$_2$NHCONH—	15	2'-CH$_3$	1	3'-	47
		3-CH$_3$	16	4-CH$_2$NHCO—	20	3'-CH$_3$	5		
		6-CH$_3$	1	3-CH$_2$NHCONH—	5	4'-CH$_3$	7		
		2-OCH$_3$	6	3-CH$_2$NHCO—	3	2'-OCH$_3$	3		
		4-SO$_2$F	2	4-CH$_2$CH$_2$NHCO—	2	3'-OCH$_3$	1		
				4-CH$_2$CH$_2$NHCONH—	2	4'-OCH$_3$	2		
				4-NHSO$_2$—	6	4'-OCH$_2$CH$_3$	1		
				4-CH$_2$NHSO$_2$—	7	4'-CH(CH$_3$)$_2$	1		
				4-CH$_2$CH$_2$NHSO$_2$—	2				
				3-CH$_2$NHSO$_2$—	4				
				3-CH$_2$CH$_2$—	1				
				4-CH$_2$CH$_2$—	2				

ty were discovered and these differences are probably not significant. The conclusion from the analysis is that a significantly more potent inhibitor is not likely to be found by working with the set of structural features of Table II. New departures are needed for further advances with this type of inhibitor.

Table III—Indicator Variables (=1) Studied in the Formulation of Eqs. 2–12

Indicator Variable	Substituent Studied
I-1	w = CH$_2$
I-2	Y = 2-Cl
I-3	Y = 3-Cl
I-4	Y = 3-CH$_3$
I-5	Only substituent 3-position
I-6	Only substituent 4-position
I-7	Substituent in 4 plus other substituent in 2, 3, or 6
I-8	Bridge 4-NHCONH—
I-9	Bridge 4-NHCO—
I-10	Bridge 4-CH$_2$NHCONH—, 4-CH$_2$CH$_2$NHCONH—, or 3-CH$_2$NHCONH—
I-11	Bridge 4-CH$_2$NHCO—, 4-CH$_2$CH$_2$NHCO—, or 3-CH$_2$NHCO—
I-12	Bridge 4-CH$_2$NHSO$_2$—, 4-CH$_2$CH$_2$NHSO$_2$—, or 3-CH$_2$NHSO$_2$—
I-13	Bridge 4-NHSO$_2$—
I-14	Bridge 4-CH$_2$CH$_2$— or 3-CH$_2$CH$_2$—
I-15	4'-SO$_2$F
I-16	3'-SO$_2$F
I-17	SO$_2$F plus other substituent on the last ring
I-18	I-12 + I-13
I-19	L-1210/FR8 enzyme
I-20	L-1210/0 enzyme
I-21	I-8 + I-9
I-22	I-10 + I-11
I-23	Z = 2'-Cl
I-24	Z = 3'-CH$_3$
I-25	Z = 4'-CH$_3$
I-26	3'- or 5'-SO$_2$F
I-27	Liver enzyme
I-28	Z = 2'-OCH$_3$

Cross-Product Terms Studied

I-4·I-21	I-15·I-22
I-4·I-8	I-16·I-10
I-8·I-17	I-16·I-11
I-15·I-5	I-16·I-22
I-15·I-6	I-6·I-22·I-16
I-15·I-10	I-6·I-22·I-17
I-15·I-11	I-4·I-8·I-17
I-15·I-21	

DISCUSSION

In designing the inhibitors considered in this study, Baker and coworkers (18–26) drew on their experience gained from the study of a large number of heterocyclic inhibitors of dihydrofolate reductase. The 3,4-dichlorophenyl moiety in the 5-position probably provides close to optimal lipophilic interaction with the lipophilic pocket in this area of the enzymes (13)[1]; however, this point is worthy of more study. Substituents in the 6-position of the pyrimidine ring do not appear to interact with lipophilic space (no correlation with π), nor does MR account for substituent effects in this part of enzymic space. Substituents Y and Z have little or no effect on activity. It seems that most of the substituents of the 6-position must project into a very loosely structured part of the enzyme or into the surrounding aqueous phase.

The more flexible bridges, those containing a CH$_2$ unit in addition to an amide moiety, have little or no effect on activity either way. These more flexible bridges allow maximum freedom (from negative steric effects) to the second phenyl ring and its substituents. Two of the rigid bridges, —NHCONH— and —NHCO—, have very deleterious effects on activity; this result seems to be connected to the positioning of the second phenyl ring. Strangely, the 4-NHSO$_2$— bridge (I-13) does make a small contribution to activity. Congeners with this function are, on the average, 10 times as active as those having the 4-NHCO— bridge. What property of NHSO$_2$ is responsible for this activity is not obvious. Similar hydrophobicity is shown by $\pi_{NHSO_2C_6H_5} = 0.45$ and $\pi_{NHCOC_6H_5} = 0.49$. The electronic effect of SO$_2$— on the phenyl ring can be ruled out since CH$_2$NHSO$_2$ does not show special activity. This leaves one with the conclusion that the geometry of the two groups may be the critical factor.

One must bear in mind that structural features not parameterized by specific variables do not make significant contributions to the activity μ (Eq. 1) of the parent structure. One value of the present analysis is that it clearly indicates, in an objective way, structural features that are not worth further study. While these features are often carried in the mind of the active investigator in a qualitative sense, they are quite difficult for a person new to the work to discover. Correlation equations quickly bring the salient features of a very complex set of data into focus.

The value of the present analysis does not reside in its ability to suggest new congeners for synthesis; rather, it demonstrates that a highly complex set of data, whose structural properties produce a set of discontinuous perturbations in a macromolecule, can be brought into coherent order with a relatively small number of variables. This constitutes a valuable modification of the Free–Wilson approach.

Moreover, the results of this analysis suggest that the Free–Wilson approach can be further extended by the use of cross-product terms. This use of cross-product terms may prove to be of considerable value in establishing the significance of special patterns or

Table IV—Squared Correlation Matrix for Indicator Parameters Used in the Correlation Study[a]

	I-1	I-2	I-4	I-7	I-8	I-9	I-10	I-13	I-15	I-17	I-20
I-1	1.00	0.05	0.04	0.19	0.01	0.00	0.00	0.01	0.01	0.01	0.00
I-2		1.00	0.04	0.25	0.02	0.02	0.01	0.01	0.00	0.00	0.00
I-4			1.00	0.21	0.01	0.01	0.01	0.02	0.00	0.00	0.01
I-7				1.00	0.02	0.01	0.00	0.01	0.00	0.00	0.00
I-8					1.00	0.04	0.08	0.02	0.00	0.14	0.00
I-9						1.00	0.03	0.01	0.00	0.00	0.39
I-10							1.00	0.02	0.00	0.07	0.00
I-13								1.00	0.00	0.03	0.01
I-15									1.00	0.07	0.00
I-17										1.00	0.01
I-20											1.00

[a] Numbers show the percent correlation (r^2) between each variable.

Table V—Equations Correlating Log $1/C$ with 11 Most Significant Variables[a]

Equation Number	Intercept	I-8	I-9	I-20	I-4·I-8	I-1	I-15	I-13	I-8·I-17	I-7	I-2	I-10	s^b	r^c	$F_{1,x}^d$
2	7.13	−0.82											0.377	0.683	90
3	7.20	−0.89	−0.48										0.348	0.742	19
4	7.19	−0.96	−0.86	0.59									0.317	0.795	22
5	7.19	−1.08	−0.84	0.55	0.63								0.293	0.830	18
6	7.13	−1.05	−0.79	0.50	0.68	0.31							0.269	0.860	20
7	7.21	−1.05	−0.82	0.51	0.67	0.33	−0.21						0.248	0.884	18
8	7.17	−1.01	−0.78	0.51	0.67	0.37	−0.22	0.42					0.229	0.903	18
9	7.17	−1.16	−0.78	0.51	0.69	0.36	−0.21	0.42	0.23				0.223	0.909	6.2
10	7.11	−1.18	−0.82	0.54	0.63	0.42	−0.22	0.40	0.24	0.18			0.218	0.915	5.7
11	7.12	−1.18	−0.81	0.53	0.58	0.42	−0.22	0.40	0.21	0.18	−0.15		0.213	0.919	5.2
12	7.14	−1.20	−0.83	0.53	0.58	0.42	−0.22	0.38	0.21	0.18	−0.15	−0.06	0.213	0.920	1.0

[a] Figures in each column are the regression coefficients for the indicated variable. [b] Standard deviation. [c] Correlation coefficient. [d] F statistic for addition of each successive term.

constellations of atoms not normally parameterized in structure–activity relationships.

The importance of correlation equations can only be fully appreciated when X-ray crystallographic studies of inhibitor binding with the enzyme are undertaken. The terms of the correlation equations should be of great help in interpreting the interactions of micro- and macromolecules.

REFERENCES

(1) Y. C. Martin, J. B. Holland, C. H. Jarboe, and N. Plotnikoff, *J. Med. Chem.*, 17, 409(1974).

(2) B. R. Kowalski and C. F. Bender, *J. Amer. Chem. Soc.*, 96, 916(1974).

(3) M. L. Weiner and P. H. Weiner, *J. Med. Chem.*, 16, 665(1973).

(4) C. Hansch, S. H. Unger, and A. B. Forsythe, *ibid.*, 16, 1217(1973).

(5) S. S. Schiffman, *Science*, 185, 112(1974).

(6) C. Hansch, in "Drug Design," vol. I, E. J. Ariëns, Ed., Academic, New York, N.Y., 1971, p. 271.

(7) A. Verloop, in "Drug Design," vol. III, E. J. Ariëns, Ed., Academic, New York, N.Y., 1972, p. 133.

(8) C. Hansch, A. Leo, and D. Elkins, *J. Chem. Doc.*, 14, 57(1974).

(9) T. C. Bruice, N. Kharasch, and R. J. Winzler, *Arch. Biochem. Biophys.*, 62, 305(1956).

(10) S. M. Free, Jr., and J. W. Wilson, *J. Med. Chem.*, 7, 395(1964).

(11) A. Cammarata and K. S. Rogers, in "Advances in Linear Free-Energy Relationships," N. B. Chapman and J. Shorter, Eds.,

Plenum, New York, N.Y., 1972, p. 403.

(12) C. Daniel and F. S. Wood, "Fitting Equations to Data," Wiley–Interscience, New York, N.Y., 1971, pp. 55, 169, 203.

(13) C. Hansch and C. Silipo, *J. Med. Chem.*, 17, 661(1974).

(14) C. Hansch and M. Yoshimoto, *ibid.*, 17, 1160(1974).

(15) C. Silipo and C. Hansch, *Mol. Pharmacol.*, 10, 954(1974).

(16) C. Silipo and C. Hansch, *Farmaco, Ed. Sci.*, 30, 35(1975).

(17) M. Yoshimoto, C. Hansch, and P. Y. C. Jow, *Chem. Pharm. Bull.*, 23, 437(1975).

(18) B. R. Baker, G. J. Lourens, R. B. Meyer, Jr., and N. M. J. Vermeulen, *J. Med. Chem.*, 12, 67(1969).

(19) B. R. Baker and N. M. J. Vermeulen, *ibid.*, 12, 74(1969).

(20) *Ibid.*, 12, 79(1969).

(21) *Ibid.*, 12, 82(1969).

(22) *Ibid.*, 12, 86(1969).

(23) *Ibid.*, 12, 89(1969).

(24) *Ibid.*, 12, 684(1969).

(25) *Ibid.*, 13, 82(1970).

(26) *Ibid.*, 13, 1154(1970).

ACKNOWLEDGMENTS AND ADDRESSES