Journal of Medicinal Chemistry

Volume 17, Number 7 July 1974

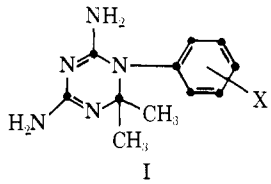
Quantitative Structure–Activity Relationship of Reversible Dihydrofolate Reductase Inhibitors.† Diaminotriazines

Corwin Hansch* and Carlo Silipo‡

Department of Chemistry, Pomona College, Claremont, California 91711. *Received February 25, 1974*

A quantitative structure–activity relationship (QSAR) has been formulated for 4,6-diamino-1,2-dihydro-2,2-dimethyl-1-(X-phenyl)-*s*-triazines inhibiting dihydrofolate reductase isolated from Walker 256 tumor. Using substituent constants and regression analysis it is shown that the substituents X on the phenyl ring are placed into two different types of space in or on the enzyme. Substituents in the 3 position show typical hydrophobic interaction while substituents in the 4 position bring about inhibition in a fashion more closely related to their molecular volume as characterized by molecular refractivity. The electronic effects of X as measured by σ do not appear to have a significant role. The QSAR for 83 inhibitors is described by log $1/C$ = $0.89(\pi\text{-}3) + 0.15(\text{MR-}4) - 0.13(\pi\text{-}3)^2 + 6.62$ where π-3 is the hydrophobic effect of substituents in the 3 position and MR-4 is the molecular refractivity of 4 substituents. The correlation coefficient for this equation is 0.905. The design of better inhibitors is discussed in the light of the above equation.

One of the most profitable enzymic systems for study by drug researchers in the past quarter century is that of the folate reductases.[1] Out of this work have come some of the first effective cancer chemotherapeutic agents as well as the highly effective antimicrobial, trimethoprim. Although a large effort has gone into the synthesis and testing of hundreds of compounds to regulate the folate reductases, we believe that there is much more "gold" to be found in this important lode. This report reviews the extensive studies of the late B. R. Baker and his students on the inhibition of dihydrofolate reductase isolated from Walker 256 tumor. Our goal is to more clearly define the structure–activity relationship in order to obtain better guidance in the development of antitumor compounds. In preliminary work it was found that dihydrofolate reductase inhibitors are amenable to quantitative correlation.[2] Since this first study a large amount of new data from Baker's laboratory has appeared.[3] These results on triazines of type I, along with the necessary substituent constants, are assembled in Table I. We have formulated the QSAR of eq 1-3 from the data in Table I.



I

Method. C in log $1/C$ in Table I represents the molar concentration of inhibitor necessary for 50% reversible inhibition of dihydrofolate reductase from Walker 256 carcinoma of rat when assayed at pH 7.4.

To estimate the role of hydrophobic interaction of the inhibitors with the enzyme, π-3 and π-4 have been employed for substituents in the meta and para positions of the *N*-phenyl ring. The hydrophobic constants, except for certain of those determined in this report, are from the benzene system. σ constants and molecular refractivity values were taken from our recent compilation.[4]

Table II lists log P and π values not previously reported. These were measured by the usual procedure.[5] The listed value is the mean of at least four determinations made over a concentration range of tenfold. The other π values of Table I were estimated from additivity principles.[6] For example, log P for 4-Me-$C_6H_4SO_2OC_6H_4$-2-C(=O)N(Me)$_2$ can be calculated as follows: log $P(C_6H_5SO_2OC_6H_5)$ + $\pi(CH_3)$ + $\pi[CON(Me)_2]$ = 3.06 + 0.50 − 1.51 = 2.05. The observed log P of 2.01 falls very close to the estimated value. The π values for substituents of the type 4-$OCH_2C_6H_4$-4′-$SO_2OC_6H_4$-X were calculated according to log $P(C_6H_5SO_2OC_6H_5)$ + $\pi(CH_3)$ + $\pi(X)$ + $\Delta\pi(0)$. [$\Delta\pi(0)$ = log $P(C_6H_5OCH_2C_6H_5)$ − log $P(C_6H_5CH_2C_6H_5)$.] For Cl and Br in aliphatic moieties aliphatic values for π have been used.[6b] Thus, $\pi(OCH_2C_6H_5NHCOCH_2Br)$ is the sum of $\pi(OCH_2C_6H_5)$ + $\pi(NHCOCH_3)$ + $\pi(Br)$ = 1.66 − 0.97 + 0.60 = 1.29. For the examples of $\pi[(CH_2)_nCOCH_2Cl]$ with n = 0, the calculation is $\pi(COCH_3)$ + $\pi(Cl)$ = −0.55 + 0.39 = −0.16. When n = 2 or 4, $\pi(COCH_3)$ is replaced by $\pi(CH_2COCH_3)$ and $(n-1)\pi(CH_2)$ is added; *e.g.*, for n = 2, −0.69 + 0.50 + 0.39 = 0.20. It has been found[6a] that when a CH_2 unit is placed between two aromatic rings or two electronegative groups or one of each, then such CH_2 groups show no hydrophobic effect. For this reason $\pi(CH_2NHCOCH_2Br)$ is assumed to be equal to log $P(H_2NCOCH_2Br)$ (*i.e.*, −0.52).§

†This investigation was supported by Public Health Service Research Grant No. CA-11110 from the National Cancer Institute.

‡NATO Fellow and Visiting Professor of Chemistry from the Institute of Pharmaceutical and Toxicological Chemistry, University of Naples.

§R. Kerley and C. Hansch, unpublished analysis.

Table I. Inhibition Constants and Physicochemical Parameters for the Inhibition of Dihydrofolate Reductase by 4,6-Diamino-1,2-dihydro-2,2-dimethyl-1-(X-phenyl)-*s*-triazine

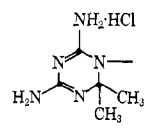
No.	X	Log 1/C			c				
		Obsd[b]	Calcd[c]	Δ log 1/C	π-3	π-4	MR-3	MR-4	σ-3,4
1[a]	4-C_6H_5	4.70	6.998	2.30	0.0	1.96	0.103	2.536	−0.01
2	3-OCH_2CO-N$(CH_2CH_2)_2O$	4.85	5.149	0.30	−1.39	0.0	3.488	0.103	0.12
3[a]	4-CN	5.14	6.713	1.57	0.0	−0.57	0.103	0.633	0.66
4	3-OCH_2CONMe_2	5.44	5.187	0.25	−1.36	0.0	2.583	0.103	0.12
5	3-OCH_3	6.17	6.615	0.45	−0.02	0.0	0.787	0.103	0.12
6[a]	4-$OCH_2CON(Me)C_6H_5$	6.17	7.300	1.13	0.0	0.12	0.103	4.554	−0.27
7	3-$COCH_2Cl$	6.21	6.487	0.28	−0.16	0.0	1.618	0.103	0.38
8	4-OCH_2CONMe_2	6.26	7.005	0.74	0.0	−1.36	0.103	2.583	−0.27
9	4-$COCH_2Cl$	6.45	6.860	0.41	0.0	−0.16	0.103	1.618	0.50
10	3-$CH_3NHCOCH_2Br$	6.58	6.136	0.44	−0.52	0.0	2.743	0.103	−0.07
11	4-CH_2CONMe_2	6.63	6.972	0.34	0.0	−1.70	0.103	2.365	−0.17
12	4-OCH_2CO-N$(CH_2)_4$	6.66	7.114	0.45	0.0	−0.72	0.103	3.312	−0.27
13	3-$OCH_2CON(Me)C_6H_5$	6.68	6.738	0.06	0.12	0.0	4.554	0.103	0.12
14	4-OCH_2CONEt_2	6.72	7.144	0.42	0.0	−0.36	0.103	3.513	−0.27
15	4-CH_2CONEt_2	6.77	7.111	0.34	0.0	−0.70	0.103	3.294	−0.17
16	3-$OCH_2CONHC_6H_5$	6.85	7.122	0.27	0.60	0.0	4.092	0.103	0.12
17[a]	3-C_6H_5	6.85	7.889	1.04	1.96	0.0	2.536	0.103	0.06
18	4-CH_2CN	6.92	6.769	0.15	0.0	−0.57	0.103	1.011	0.01
19	H	6.92	6.633	0.29	0.0	0.0	0.103	0.103	0.0
20	3-$OCH_2C_6H_4$-3′-$NHCOCH_2Br$	6.92	7.570	0.65	1.29	0.0	5.394	0.103	0.12
21	4-$CH_2CON(Me)C_6H_5$	7.00	7.268	0.27	0.0	−0.19	0.103	4.336	−0.17
22	4-$(CH_2)_2CONMe_2$	7.05	7.042	0.01	0.0	−1.20	0.103	2.830	−0.17
23	3-NO_2	7.07	6.374	0.70	−0.28	0.0	0.736	0.103	0.71
24	3-$(CH_2)_2COCH_2Cl$	7.10	6.806	0.29	0.20	0.0	2.471	0.103	−0.07
25	3-$(CH_2)_4COCH_2Cl$	7.10	7.518	0.42	1.20	0.0	3.400	0.103	−0.07
26	4-OCH_2CO-N$(CH_2)_5$	7.12	7.184	0.06	0.0	−0.32	0.103	3.777	−0.27
27	4-CH_2CO-N$(CH_2CH_2)_2O$	7.12	7.108	0.01	0.0	−1.70	0.103	3.270	−0.17
28	3-Cl, 4-OCH_2CONMe_2	7.16	7.573	0.41	0.71	−1.36	0.603	2.583	0.10
29	4-$(CH_2)_2CONEt_2$	7.28	7.181	0.10	0.0	−0.21	0.103	3.759	−0.17
30	3-Cl, 4-OCH_2CO-N$(CH_2)_4$	7.29	7.682	0.39	0.71	−0.72	0.603	3.312	0.10
31	4-OCH_2CO-N$(CH_2CH_2)_2O$	7.29	7.141	0.15	0.0	−1.39	0.103	3.488	−0.27
32	4-$CH_2CON(Me)CH_2C_6H_5$	7.30	7.337	0.04	0.0	0.43	0.103	4.801	−0.17
33	4-$(CH_2)_2CON(Me)CH_2C_6H_5$	7.31	7.407	0.10	0.0	0.93	0.103	5.266	−0.17
34	4-$(CH_2)_2CO$-N$(CH_2CH_2)_2O$	7.32	7.178	0.14	0.0	−1.20	0.103	3.735	−0.17
35	4-$(CH_2)_2CON(C_3H_7)_2$	7.35	7.320	0.03	0.0	0.80	0.103	4.688	−0.17
36	3-$(CH_2)_4C_6H_3$-2′,4′-Cl_2	7.45	7.648	0.20	5.55	0.0	5.394	0.103	−0.07
37	3-Cl, 4-OCH_2CO-N$(CH_2)_5$	7.47	7.752	0.28	0.71	−0.32	0.603	3.777	0.10
38	3-Cl, 4-$OCH_2C_6H_4$-4′-SO_2NMe_2	7.48	7.975	0.49	0.71	0.88	0.603	5.268	0.10
39	3-Cl, 4-$OCH_2C_6H_4$-3′-CN	7.51	7.748	0.24	0.71	1.09	0.603	3.749	0.10
40	3-Cl, 4-$OCH_2C_6H_5$	7.52	7.668	0.15	0.71	1.66	0.603	3.219	0.10
41	3-$O(CH_2)_3OC_6H_4$-4′-$NHCOCH_2Br$	7.55	7.810	0.26	1.77	0.0	6.555	0.103	0.12
42	4-$(CH_2)_2CON(Me)C_6H_5$	7.56	7.337	0.22	0.0	0.31	0.103	4.801	−0.17
43	3-Cl, 4-OCH_2CONEt_2	7.64	7.712	0.07	0.71	−0.36	0.603	3.513	0.10
44	3-$O(CH_2)_2OC_6H_4$-3′-$NHCOCH_2Br$	7.64	7.810	0.17	1.77	0.0	6.555	0.103	0.12
45	3-$O(CH_2)_2OC_6H_4$-2′-$NHCOCH_2Br$	7.66	7.558	0.10	1.27	0.0	6.090	0.103	0.12
46	3-$O(CH_2)_2OC_6H_4$-3′-$NHCOCH_2Br$	7.66	7.558	0.10	1.27	0.0	6.090	0.103	0.12
47	4-$O(CH_2)_2OC_6H_4$-4′-$NHCOCH_2Br$	7.70	7.530	0.17	0.0	1.27	0.103	6.090	−0.27
48	3-Cl, 4-$OCH_2C_6H_4$-3′-$CONMe_2$	7.72	7.938	0.22	0.71	0.15	0.603	5.021	0.10
49	3-Cl, 4-$OCH_2C_6H_4$-4′-$SO_3C_6H_4$-3″-Cl	7.72	8.278	0.56	0.71	3.92	0.603	7.286	0.10
50	3-Cl	7.76	7.201	0.56	0.71	0.0	0.603	0.103	0.37
51	3-CF_3	7.76	7.318	0.44	0.88	0.0	0.502	0.103	0.43
52	3-Cl, 4-$OCH_2C_6H_4$-4′-$SO_3C_6H_4$-4″-Cl	7.77	8.278	0.51	0.71	3.92	0.603	7.286	0.10
53	3-Cl, 4-OCH_2CO-N$(CH_2CH_2)_2O$	7.85	7.708	0.14	0.71	−1.39	0.603	3.488	0.10
54	3-Cl, 4-$OCH_2C_6H_4$-3′-$CON(CH_2CH_2)_2O$	7.85	8.074	0.22	0.71	0.13	0.603	5.926	0.10
55	3-Cl, 4-$OCH_2C_6H_4$-3′-CO-N$(CH_2)_4$	7.85	8.047	0.20	0.71	0.80	0.603	5.750	0.10
56	3-Cl, 4-$OCH_2CON(Me)C_6H_5$	7.89	7.868	0.02	0.71	0.12	0.603	4.554	0.10
57	4-$OCH_2CONHC_6H_5$	7.89	7.231	0.66	0.0	0.60	0.103	4.092	−0.27
58	4-$(CH_2)_2COCH_2Cl$	7.92	6.988	0.93	0.0	0.20	0.103	2.471	−0.17
59	3-OC_6H_4-4′-$NHCOCH_2Br$	7.92	7.783	0.14	1.71	0.0	4.943	0.103	0.25
60	3-Cl, 4-$(CH_2)_4C_6H_5$	7.96	7.844	0.12	0.71	4.13	0.603	4.394	−0.17
61	3-Cl, 4-$OCH_2C_6H_4$-3′-$CONHC_6H_5$	8.00	8.164	0.16	0.71	2.15	0.603	6.530	0.10
62	3-$CH_2C_6H_5$	8.00	7.908	0.09	2.01	0.0	3.001	0.103	−0.08
63	3-Cl, 4-$OCH_2C_6H_4$-3′-CO-N$(CH_2)_5$	8.02	8.117	0.10	0.71	1.20	0.603	6.214	0.10
64[a]	4-$CH_2C_6H_5$	8.05	7.068	0.98	0.0	2.01	0.103	3.001	−0.09
65	3-Cl, 4-$OCH_2C_6H_4$-4′-$SO_3C_6H_4$-3″-CF_3	8.09	8.262	0.17	0.71	4.09	0.603	7.185	0.10
66	3-Cl, 4-$OCH_2C_6H_4$-3′-$CON(Me)C_6H_5$	8.12	8.234	0.11	0.71	2.15	0.603	6.992	0.10

Table I (*continued*)

No.	X	Log 1/C			c				
		Obsd[b]	Calcd[c]	Δ log 1/C	π-3	π-4	MR-3	MR-4	σ-3,4
67	3-O$(CH_2)_2$OC$_6$H$_4$-4′-NHCOCH$_2$Br	8.13	7.558	0.57	1.27	0.0	6.090	0.103	0.12
68	3-Cl, 4-OCH$_2$C$_6$H$_4$-3′-CONEt$_2$	8.14	8.077	0.06	0.71	1.15	0.603	5.950	0.10
69	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_5$	8.20	8.203	0.00	0.71	3.21	0.603	6.786	0.10
70	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-3″-CN	8.24	8.282	0.04	0.71	2.64	0.603	7.316	0.10
71	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_3$-3″,4″-Cl$_2$	8.25	8.353	0.10	0.71	4.63	0.603	7.786	0.10
72	3-$(CH_2)_2$C$_6$H$_4$-4′-NHCOCH$_2$Br	8.26	8.004	0.26	2.29	0.0	5.640	0.103	−0.07
73	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-2″-CF$_3$	8.33	8.262	0.07	0.71	4.09	0.603	7.185	0.10
74	3-$(CH_2)_4$C$_6$H$_4$-4′-NHCOCH$_2$Br	8.38	8.184	0.20	3.67	0.0	6.632	0.103	0.07
75	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-4″-CN	8.39	8.282	0.11	0.71	2.64	0.603	7.316	0.10
76	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-4″-OCH$_3$	8.40	8.305	0.10	0.71	3.19	0.603	7.470	0.10
77	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-4″-F	8.40	8.201	0.20	0.71	3.35	0.603	6.775	0.10
78	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-2″-OCH$_3$	8.40	8.305	0.10	0.71	3.19	0.603	7.470	0.10
79	3-$(CH_2)_4$C$_6$H$_4$-3′-NHCOCH$_2$Br	8.41	8.184	0.23	3.67	0.0	6.632	0.103	−0.07
80	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-3″-CH$_3$	8.44	8.272	0.17	0.71	3.77	0.603	7.248	0.10
81	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-3″-F	8.46	8.201	0.26	0.71	3.35	0.603	6.775	0.10
82	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-3″-OCH$_3$	8.52	8.305	0.21	0.71	3.19	0.603	7.470	0.10
83[a]	3,4-Cl$_2$	8.54	7.276	1.26	0.71	0.71	0.603	0.603	0.60
84	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-2″-Cl	8.62	8.278	0.34	0.71	3.92	0.603	7.286	0.10
85	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-4″-CON$(CH_3)_2$	8.62	8.473	0.15	0.71	1.70	0.603	8.588	0.10
86	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-2″-CON$(CH_3)_2$	8.63	8.473	0.16	0.71	1.70	0.603	8.588	0.10
87	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-2″-CN	8.70	8.282	0.42	0.71	2.64	0.603	7.316	0.10
88	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-2″-F	8.74	8.201	0.54	0.71	3.35	0.603	6.775	0.10
89	3-Cl, 4-OCH$_2$C$_6$H$_4$-4′-SO$_3$C$_6$H$_4$-3″-CON$(CH_3)_2$	8.76	8.473	0.29	0.71	1.70	0.603	8.588	0.10
90[a]	4-$(CH_2)_4$C$_6$H$_3$-2′,4′-Cl$_2$	9.21	7.426	1.78	0.0	5.55	0.103	5.394	−0.17

[a] These points not used in deriving eq 1-3. [b] See Baker, *et al.*[3] [c] Calculated using eq 2. [d] See section on Method for sources of these constants.

Table II. Partition Coefficients and New π Constants

No.	Compound	Log *P*	Mp, °C	Substituent	π[a]
I	Phenoxyacetic amide	0.76 ± 0.00	101.5[b]	OCH$_2$CONH$_2$	−1.37
II	Phenoxyacetic *N,N*-dimethylamide	0.77 ± 0.01	48 ÷ 48.5[c]	OCH$_2$CON$(CH_3)_2$	−1.36
III	Phenoxyacetic anilide	2.73 ± 0.01	101.5[b]	OCH$_2$CONHC$_6$H$_5$	0.60
IV	Phenoxyacetic *N*-methylanilide	2.25 ± 0.02	93.5 ÷ 94[d]	OCH$_2$CON(Me)C$_6$H$_5$	0.12
V	*N*-Phenoxyacetylmorpholine	0.74 ± 0.01	89.0[e]	OCH$_2$CO-N$(CH_2CH_2)_2$O	−1.39
VI	*N*-Phenoxyacetylpiperidine	1.81 ± 0.02	50 ÷ 50.5[e]	OCH$_2$CO-N$(CH_2)_5$	−0.32
VII	Benzenesulfonic *N,N*-dimethylamide	1.35 ± 0.00	47.5[e]	SO$_2$N$(CH_3)_2$	−0.78
VIII	Benzyl phenyl ether	3.79 ± 0.02	39.0[f]	OCH$_2$C$_6$H$_5$, CH$_2$OC$_6$H$_5$	1.66
IX	*N,N*-Dimethyl-*p*-toluenesulfonylsalicyl amide	2.01 ± 0.02	(Liquid)[c]	SO$_3$C$_6$H$_4$-2-CON$(CH_3)_2$	−0.68[g]
X	4,6-Diamino-1,2-dihydro-2,2-dimethyl-1(*p*-biphenyl)-*s*-triazine hydrochloride	−1.08 ± 0.01[h]	205 ÷ 207[i]		−5.17[j]
XI	*N,N*-Dimethylbenzamide	0.62		CON$(CH_3)_2$	−1.51[k]
XII	Benzylamine hydrochloride	−1.96[l]		CH$_2$NH$_2$·HCl	−4.09[k]

[a] π = log *P* − log *P*(benzene) = log *P* − 2.13. [b] I. Heilbron, "Dictionary of Organic Compounds," Vol. 4, 4th ed, Oxford University Press, New York, N. Y., 1965, p 2663. [c] This report. [d] C. A. Bischoff, *Ber.*, **34,** 2125 (1901). [e] A. Ginzberg, *ibid.*, **36,** 2706 (1903). [f] R. L. Merker and M. J. Scott, *J. Org. Chem.*, **26,** 5180 (1961). [g] π = log *P* − log *P*(CH$_3$C$_6$H$_5$) = 2.01 − 2.69 = −0.68. [h] In this determination 0.1 *M* pH 7.4 phosphate buffer was used as the aqueous phase. [i] B. R. Baker and B. T. Ho, *J. Heterocycl. Chem.*, **2,** 335 (1965). [j] π = log *P* − log *P*(biphenyl) = −1.08 − 4.09 = −5.17. [k] C. Hansch and K. Kim, unpublished analysis. [l] In this determination 0.1 *N* HCl was used as the aqueous phase.

Table III. Squared Correlation Matrix for Parameters Used in the Correlation Study[a]

	π-3	π-4	MR-3	MR-4	σ-3	σ-4	π-3,4	MR-3,4	σ-3,4
π-3	1.00	0.01	0.37	0.02	0.00	0.02	0.30	0.12	0.00
π-4		1.00	0.04	0.50	0.31	0.17	0.77	0.51	0.05
MR-3			1.00	0.38	0.09	0.32	0.02	0.01	0.02
MR-4				1.00	0.28	0.63	0.27	0.52	0.00
σ-3					1.00	0.21	0.20	0.16	0.47
σ-4						1.00	0.08	0.25	0.09
π-3,4							1.00	0.59	0.05
MR-3,4								1.00	0.00
σ-3,4									1.00

[a] Numbers in Table III show the per cent correlation (r^2) between each of the variables.

Groups of the type $(CH_2)_nCONR_2$ (n = 1 or 2) were estimated from the appropriate $\pi[(CH_2)_nCONH_2]$ and $\pi[OCH_2CON(R)_2]$ values. For example, it is assumed that the value $\Delta\pi$ = log $P(C_6H_5CH_2CH_2CONH_2)$[7] − log $P(C_6H_5OCH_2CONH_2)$ = 0.91 − 0.76 = 0.15 is the same as log $P[C_6H_5CH_2CH_2CON(CH_3)_2]$ − log $P[C_6H_5OCH_2CON(CH_3)_2]$. Hence, $\pi[CH_2CON(C_2H_5)_2]$ is the sum of $\pi[OCH_2CON(CH_3)_2]$ + $\Delta\pi$ + $2\pi(CH_3)$ = − 1.36 + 0.15 + 1.00 = −0.21. To estimate $\pi[OCH_2CON(CH_2)_4]$, the value of 0.4 (π for cyclic CH_2) was subtracted from $\pi[OCH_2CON(CH_2)_5]$ = −0.32 − 0.40 = −0.72. $\pi(C_6H_5OCH_2CH_2O)$ is estimated to be the sum of $\pi(C_6H_5OCH_2)$ + $\pi(CH_3O)$ = 1.66 − 0.22 = 1.64. $\pi(OCH_2CH_2CH_2OC_6H_5)$ = 1.64 + 0.50 = 2.14.

It has also been assumed that the same value of π can be employed for meta and para substituents[5] and no attempt was made to correct for the folding effect or small differences[6b] for groups adjacent to each other.

Although there is a certain amount of collinearity[8] between MR and π (see Table III), they do appear to be independent enough to be of value in the same equation to assess different aspects of the enzymic binding areas. MR has been used as a measure of the bulk effects of substituents. The MR values in Table I have been scaled by 0.1 for convenience in calculation. E_s values are not available for most of the complex substituents and, in any case, would seem to be inappropriate. The MR values have been calculated as previously reported.[4]

For many of the functions, σ constants are not available and we have estimated the values for those groups shown in Table I. In addition to the constants of Table I, two "dummy"[9] parameters were studied. A significant role for electronic effects of the substituents could not be detected using σ. One dummy (indicator) variable was used to explore the possibility that the $CON(CH_2CH_2)_2O$ function might have a special effect. In preliminary studies this seemed to be true; however, it was discovered that the use of MR eliminated the need for this variable. In the course of his work Baker made a change in the mode of his enzymic assay. A second indicator variable was used to see if a significant difference in the two sets of congeners tested by the two different methods could be uncovered. A significant difference could not be found.

Compounds I–VII in Table II were prepared by reaction of phenoxyacetyl or benzenesulfonyl chlorides with the appropriate amine. Compound VIII was obtained by reaction of benzyl chloride and phenol in triethylamine.[10] The 4,6-diamino-1,2-dihydro-2,2-dimethyl-1-(*p*-biphenyl)-*s*-triazine hydrochloride was synthesized by the "three-component" method of Modest,[11] *i.e.*, the direct condensation of 4-phenylaniline hydrochloride with dicyandiamide and acetone in ethanol. The *N,N*-dimethyl-*p*-toluenesulfonylsalicyl amide was prepared in good yield by refluxing *p*-toluenesulfonyl chloride with *N,N*-dimethylsalicyl amide[12] in benzene containing triethylamine. The product was purified by column chromatography followed by vacuum distillation.

The purity of all compounds was confirmed by tlc. All new compounds gave carbon and hydrogen analyses which agreed within 0.3% of the theoretical.

Results

Constructing the "best" QSAR for a highly complex set of congeners such as those in Table I is a very difficult problem and there are many different ways to go about it. In the present instance we first examined separately the four sets of congeners in the four papers by Baker[3] using the obvious parameters $\Sigma\sigma$, $\Sigma\pi$, $\Sigma\pi^2$, π-3, and π-4. From these studies it appeared that most of the congeners were behaving in the same fashion and that the sets could probably be merged. During this preliminary work it was also apparent that, when a simple rigid phenyl group was attached in either the 3 or 4 position of the phenyl ring of I, activity was much lower than one would expect. These two data points were withheld in subsequent work; derivatives containing the COOH and CH_2NH_2 functions were also withheld. Log P or π values for functions ionized at physiological pH are largely the result of ion-pair partitioning. Since it is not known what the counterion in the buffer system of Baker would be, it is not possible to place log P values for ionized and neutral functions on the same scale with any degree of confidence. As of the present, QSAR must be formulated separately for these two classes of compounds. Eventually, with more experience, it may be possible to treat such mixed sets of molecules *via* one QSAR.

In the first runs with about 90 data points, those substituents containing the grouping $C(=O)N(CH_2CH_2)_2O$ stood out as being poorly fit. Adding a dummy variable corrected this problem; however, later on it was found that the use of MR-4 (MR = molecular refractivity[4]) eliminated, for practical purposes, this special role for the morpholine group.

It became evident early on that there was a large difference in the hydrophobic roles of substituents in the 3 and 4 positions. Coefficients with π-3 were rather large (~0.8) while coefficients with π-4 were small (~0.2). This suggested that substituents in the 4 position were producing inhibition because of their bulkiness rather than through simple hydrophobic interaction. To explore this possible bulk effect the variables MR-3 and MR-4 were studied. The use of MR-4 eliminated the need for π-4. From the beginning a term in $\Sigma\sigma$ was found to be of no value in reducing the variance. After considerable experimenting with the dropping of poorly fit points, compounds 1, 3, 6, 17, 64, 83, and 90 were omitted from the data set and all possible equations for the linear combination of the ten variables π-3, π-4, π-3,4, MR-3, MR-4, MR-3,4, σ-3,4, $(\pi\text{-}3)^2$, $(\pi\text{-}4)^2$, and $(\pi\text{-}3,4)^2$ were derived. Theoretically, this should yield 2^{10} − 1 = 1023 equations; actually, because of singular matrices resulting from near collinearity among some of the variables, only 783 equations resulted. Of these, the equation with two variables having the lowest standard deviation in its class and the three-variable

equation with the lowest standard deviation in its class are

$$\log 1/C = 0.464\ (\pm 0.10)\ (\pi\text{-}3) + 0.181\ (\pm 0.03)\ (\text{MR-}4) + 6.613\ (\pm 0.17) \quad (1)$$

n	r	s
83	0.834	0.422

$$\log 1/C = 0.890\ (\pm 0.14)\ (\pi\text{-}3) + 0.150\ (\pm 0.03)\ (\text{MR-}4) - 0.127\ (\pm 0.03)(\pi\text{-}3)^2 + 6.618\ (\pm 0.13) \quad (2)$$

n	r	s
83	0.905	0.328

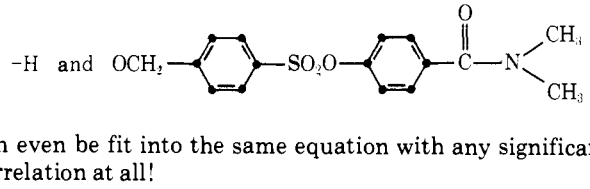
In the above equations, n is the number of data points used to derive the equations, r is the correlation coefficient, s is the standard deviation, and the figures in parentheses are the 95% confidence intervals.

The exponential term in eq 2 makes it a significantly better result than eq 1 ($F_{1,79} = 53.3$). No four-variable equations were found to have a standard deviation lower than eq 2.

Given the present set of data and substituent constants, it would appear unreasonable and possibly misleading to attempt a higher resolution of the QSAR. Several five-variable equations had standard deviations of 0.32 with correlation coefficients of 0.912. While these were justified by the F statistic, it was difficult to find any consistent meaning in these results. Since they result in only a 1.3% reduction of the variance, they do not merit consideration. None of the 783 equations had a standard deviation lower than 0.3195. This is very little improvement over eq 2.

It is of course out of the question to consider even a small fraction of the cross product terms such as (π-3·π-4), (π-3·π-4·MR-3·MR-4), etc., for the above ten variables. However, the term (π-3·MR-4) for the two most important variables was studied in a variety of combinations without success.

Equation 2 accounts for 82% of the variance in log $1/C$ of the 83 complex modifications of I; hence, 18% of the information in log $1/C$ cannot be accounted for. This is not at all surprising when one considers that we are dealing with truly enormous changes in an already complex parent structure. It is astonishing that substituents as grossly different as



can even be fit into the same equation with any significant correlation at all!

Discussion

As medicinal chemists venture deeper and deeper into the land of computerized statistical studies of the relationships between chemical structure and biological activity, they are finding rather strange monsters in the form of chance correlations. Viewing such encounters, medicinal chemists who are considering entering the new land may consider the risks too great and the results too uncertain. As we stumble toward the parameterization of biological activity and chemical structure so that large masses of data can be examined rapidly from literally thousands of points of view, many pitfalls and false turns will result in disagreeable surprises. However, there is no turning back to the "good old days" when elaborate pictures of the fit of molecules to an unknown and probably unknowable receptor were the primary result of a structure-activity study. The chance correlations and grossly wrong conclusions drawn from structure-activity studies in the "good old days" were horrendous; however, living with them constantly dulled one's senses to their outrageous simplicity. The gross overemphasis on geometry by medicinal chemists was in part a legacy from the lock and key theory of biochemistry and in part stemmed from the self-defeating feeling that biological systems are too complicated to deal with in numerical terms. However, there is little choice for the medicinal chemist doomed to work with large numbers of variables—either he attempts regression analysis by the "seat of his pants" or he turns to statistics and computers.

The chances of drawing false conclusions from computerized analysis are of course still present but far less likely than from unaided inspection of the data. In the present instance we have looked at several thousand equations. Each equation constitutes a different view of the data with statistical checks to help one decide in an objective manner which is the "best view." This "walking around the data" and studying it from thousands of viewpoints yields entirely new perspectives and suggests new hypotheses for testing.

The present example, as for most known QSAR, has been formulated after work on the project has ceased and with data obtained from structures which were not planned from the start. Structural modifications were made with little thought as to the availability of suitable substituent parameters. Now that some progress has been made in the organization of suitable parameters and techniques for their selection, we can look forward to better sets of data upon which to sharpen our tools.[8-13]

In the formulation of a QSAR it becomes more and more apparent that there is no substitute for a meaningful model. No amount of fancy computerized screening of large numbers of parameters of a *miscellaneous* nature can alone produce correlations in which one can place much confidence.[14] The model which we and others[15] have been testing for the past decade assumes that in varying substituents on a parent structure, one changes its hydrophobic, steric, and electronic characteristics and that these perturbations which are reflected in the biological response of a standard system can be more or less accounted for in the physicochemical properties of the substituents. At this point in time log P and π constants from octanol-water partition coefficients[6b] appear to be the most expeditious for operationally defining the hydrophobic character of a substituent. Hammett-Taft σ constants still appear to be superior to calculated MO indices for characterizing the electronic effect of substituents.

To factor electronic effects into inductive and resonance components $\mathcal{F}$ and $\mathcal{R}$ of Swain and Lupton, recently redefined,[4] are most generally useful simply because a very much larger set is available than the more tightly defined σ_I and σ_R of Taft[16] or **S** and **P** of Unger[17] and Swain. However, from our own studies and those of Taft,[16] it is clear that one cannot expect $\mathcal{R}$ to model well the types of interactions correlated by σ^- and σ^+.

Steric effects are the most complex to deal with and there are several types of situations one can expect to encounter. Intramolecular effects of substituents on reactions such as $XCH_2C(=O)OCH_3$ are often well correlated by Taft's steric parameter E_s. E_s has also been shown to account for intermolecular effects such as the interaction of a hapten with an antibody.[18] There is another type of steric effect for which E_s is not suitable. This is deter-

mined by the bulk tolerance of the receptor site for bulky substituents such as large alkyl groups or, in the present study, substituted aromatic rings. E_s as defined by Taft[19] using the above-mentioned X-substituted acetates shows a leveling-off effect as atoms in X become more remote from the reaction center. For example, in the series H, CH_3, C_2H_5, C_3H_7, C_4H_9, C_5H_{11}, and C_8H_{17}, E_s is 1.24, 0.00, −0.07, −0.36, −0.39, −0.40, and −0.33, respectively. The values of propyl to octyl are the same within experimental error. While total bulk increases greatly in a linear fashion from propyl to octyl, the effect of X on the reaction center in, say, hydrolysis is constant. To force such large substituents into a portion of an enzyme would be much more a function of molar volume than E_s. To seek out such situations in substituent space, molecular refractivity (MR) has proven to be of value.[20] It is in this sense that we have employed MR in this study. We are of course aware that this is not the ideal way to define effective molar volume, but it seems to be the best available at present.

With this brief view of our model in mind, what generally useful information can be gained from eq 2? In the first place, it is important to remember that the terms which do not appear in eq 2 are almost as important as those that do. The lack of significance for σ means that one does not have to worry greatly about the electronic effects of substituents on the phenyl ring in the design of new inhibitors. However, since the variation in σ for the functions of Table I is not great, a study of a better selection of groups should be made before *completely* discounting σ. Another point which makes the role of σ somewhat ambivalent is the rather high collinearity between σ_4 and MR-4 (r^2 = 0.63). However, since there is no significant correlation between σ-3,4 and MR, one can take confidence in the view that σ does not play a significant role in the QSAR. Adding a term in $(MR\text{-}4)^2$ did not improve the correlation; hence, still bulkier functions should be studied in this position. A term in (π-3·MR-4) is not significant; therefore, cooperative action between 3 and 4 substituents appears to be absent at least in any important way. Addition of the terms $\Sigma\pi + \Sigma\pi^2$ to eq 2 does not reduce the variance so that gross overall lipophilicity has not reached a limiting value in the design of inhibitors.

Equations 1 and 2 suggest that there are two kinds of substituent space (meta and para) in or on the enzyme. Functions in the 3 position appear to be placed into a typical hydrophobic milieu. The coefficient of about 1 for this term is observed quite commonly.[20,21] Substituents in the 4 position appear to be thrust into a more apolar region which π-4 does not model well. It seems likely that groups in the 4 position cause inhibition by producing conformational changes in dihydrofolate reductase by more firmly attaching the inhibitor through dispersion forces or by a combination of both.

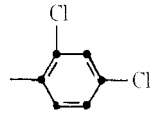
Unfortunately, π-3 and MR-4 are not as cleanly separated from MR-3 and π-4 as one would like. This can be seen from the correlation matrix of Table III. Even though there is considerable collinearity between these terms which, to a certain extent, confounds our attempts to separate them, eq 2 indicates two types of substituent space. This can be seen by comparing eq 3 with eq 2. Not only is eq 3 a much poorer correlation than eq 2, but also, the much smaller coefficient with π-4 than with π-3 indicates the different characters of meta and para space. By a more judicious choice of substituents it is possible to circumvent the above problem, but such steps must be taken before the synthetic program has started.[8]

$$\log 1/C = 0.852\ (\pm 0.18)\ (\pi\text{-}3) + 0.212\ (\pm 0.06)\ (\pi\text{-}4) - 0.144\ (\pm 0.04)\ (\pi\text{-}3)^2 + 7.042\ (\pm 0.12) \quad (3)$$

n = 83, r = 0.851, s = 0.405

The limit to the size of groups which can be positioned in para space does not yet appear to have been reached by the functions of Table I. This can be surmised from the fact that addition of a term in $(MR\text{-}4)^2$ to eq 3 does not improve the correlation. In the case of 3 substituents, the significant role for $(\pi\text{-}3)^2$ indicates that it is not profitable to place functions in this position with π values larger than $\pi(0)$ [$\pi(0)$ = 3.5(3.0–4.3)]. However, this point requires further testing since only three functions in the meta position have π values greater than 3.5.

The inhibitors of Table I span a remarkable range of 4.5 log units or a 30,000-fold range of activity. Considering this and the enormous variation in the substituents, one feels a sense of elation in viewing the order brought to this mass of data by eq 2. Nevertheless, there is still a good deal of information which escapes our net. Seven points in Table I marked by the footnote *a* have not been used in the derivation of eq 1–3. The reason for the lack of fit of three of these seems easy to understand in qualitative terms. Compounds 1 and 17 contain a stiff phenyl group attached directly to the phenyl nucleus. These molecules are 100- and 10-fold less active than our QSAR predicts. Providing some flexibility for the ring by the insertion of a CH_2 unit gives, in compounds 62 and 64, molecules 2000 and 200 times more active, respectively. This cannot be a hydrophobic or electronic effect of CH_2; hence, its origin must be steric in nature. Unless the large groups are placed on a kind of atomic swivel they are prohibited from making proper contact with their respective substituent spaces. Compound **90**, the most active in the set, is 60 times more active than eq 2 estimates. Here we have the extremely lipophilic moiety

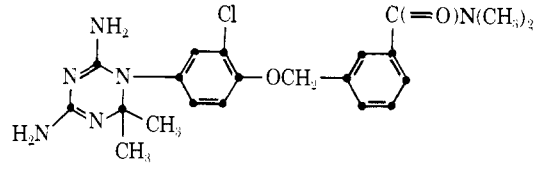


on a four-carbon chain so that it could conceivably locate itself in meta space. This hypothesis can be tested by means of eq 2. If the π value of 3.4 is substituted into eq 2 (as π-3), and if one uses the scaled MR of 1.86 for the four CH_2 units with the MR-4 term in eq 2, one calculates a log $1/C$ of 8.5 which is in better agreement with the observed value of 9.2 although still more than twice the standard deviation of eq 2.

There are four compounds, **3**, **6**, **64**, and **83**, which are quite poorly predicted. The 4-CN (**3**) is 30 times less active than expected. It behaves somewhat like phenyl in that when removed from the ring by a CH_2 group, its activity rises dramatically (18). For such a small group it seems unlikely that this is a simple steric problem. The CN function is usually well behaved hydrophobically so that one tends to look for electronic difficulties. It may be that its high dipole moment has a negative effect. The 4-$OCH_2CON(CH_3)C_6H_5$ group is ten times less active than expected. This function is well correlated when in the 3 position (13). The most serious failure of eq 2 is with the quite active 3,4-dichloro derivative **83**. This congener is more than ten times as active as expected. The 3-chloro and 3-trifluoromethyl derivatives **50** and **51** are also somewhat underpredicted. This suggests that substituent space very near the ring may be more sensitive to perturbations than the average sensitivity characterized by the coeffi-

cients in eq 2. Such discontinuities in substituent space could be studied by the study of a better selection of small substituents in the 3 and 4 positions. One should not be shocked by such findings. Large areas of enzymes cannot really be homogeneous. With the exception of the 3,4-dichloro congener, the serious failures in Table I occur among the least active derivatives.

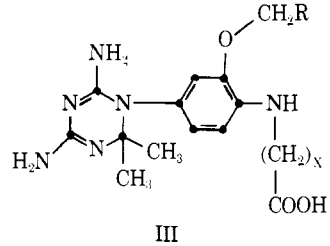
Baker's reason for intensively studying inhibitors of dihydrofolate reductase from Walker 256 tumor was to develop drugs for cancer chemotherapy. In this he achieved notable success in that compound II is now undergoing



II

clinical trials after having demonstrated great activity against L1210 leukemia in mice. Many of the most active dihydrofolate reductase inhibitors turned out to have little antitumor activity *in vivo*. It now seems clear that one of the reasons for this is that many of the compounds are much too lipophilic and that congeners with superoptimal log *P* values are simply lost in the *in vivo* random walk to the sites of action.[21] It is not by accident that II contains the $CON(CH_3)_2$ function. The π value for this group is −1.51.

Using eq 2, one should be able to beat the problem of superoptimum lipophilicity. It would seem best to take advantage of meta space by the introduction of the most lipophilic function (π-3 = 3.5-4.0). To counterbalance this large amount of lipophilicity, one could place a highly hydrophilic group in the 4 position. At present we are synthesizing compounds of type III. By varying X it should be possible for the ionized carboxyl group not only to remain outside of hydrophobic space, but also for it to find



III

an electron-deficient center with which to associate for better binding. It is hoped that studies of the variation of III will lead to an extension of eq 2 which will be of better predictive value.

Finally, it should be reemphasized that even with what might appear from superficial inspection to be an extremely diverse set of substituents, there is as mentioned above considerable collinearity between some of the variables. This problem is so complex and so important that it must receive thorough and proper study before a set of congeners is prepared; otherwise one is left with an indeterminable which cannot be resolved.

References

(1) V. R. Bertino, *Ann. N. Y. Acad. Sci.*, **186**, 5 (1971).
(2) C. Hansch, *Ann. N. Y. Acad. Sci.*, **186**, 235 (1971).
(3) (a) B. R. Baker and G. J. Lourens, *J. Med. Chem.*, **11**, 26 (1968); (b) B. R. Baker, *ibid.*, **11**, 483 (1968); (c) B. R. Baker and W. T. Ashton, *ibid.*, **15**, 945 (1972); (d) B. R. Baker and W. T. Ashton, *ibid.*, **16**, 209 (1973).
(4) C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien, *J. Med. Chem.*, **16**, 1207 (1973).
(5) T. Fujita, J. Iwasa, and C. Hansch, *J. Amer. Chem. Soc.*, **86**, 5175 (1964).
(6) (a) C. Hansch, A. Leo, and D. Nikaitani, *J. Org. Chem.*, **37**, 3090 (1972); (b) A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, **71**, 525 (1971).
(7) J. Iwasa, T. Fujita, and C. Hansch, *J. Med. Chem.*, **8**, 150 (1965).
(8) C. Hansch, S. H. Unger, and A. B. Forsythe, *J. Med. Chem.*, **16**, 1217 (1973).
(9) C. Daniel and F. S. Wood, "Fitting Equations to Data," Wiley-Interscience, New York, N. Y., 1971, pp 55, 169, 203.
(10) R. L. Merker and M. J. Scott, *J. Org. Chem.*, **26**, 5180 (1961).
(11) (a) E. J. Modest, *J. Org. Chem.*, **21**, 1 (1956); (b) B. R. Baker and B. T. Ho, *J. Heterocycl. Chem.*, **2**, 335 (1965).
(12) H. Schindlbauer, *Monatsh. Chem.*, **99**, 1799 (1968).
(13) (a) J. G. Topliss, *J. Med. Chem.*, **15**, 1006 (1972); (b) Y. C. Martin and W. J. Dunn III, *ibid.*, **16**, 578 (1973).
(14) (a) J. G. Topliss and R. J. Costello, *J. Med. Chem.*, **15**, 1066 (1972);(b) C. L. Perrin, *Science*, **183**, 551 (1974).
(15) (a) C. Hansch and T. Fujita, *J. Amer. Chem. Soc.*, **86**, 1616 (1964); (b) A. Verloop in "Drug Design," Vol. III, E. J. Ariëns, Ed., Academic Press, New York, N. Y., 1972, p 133; (c) R. Franke and P. Oehme, *Pharmazie*, **28**, 489 (1973).
(16) S. K. Dayal, S. Ehrenson, and R. W. Taft, *J. Amer. Chem. Soc.*, **94**, 9113 (1972).
(17) S. H. Unger and C. Hansch, *J. Med. Chem.*, **16**, 745 (1973).
(18) E. Kutter and C. Hansch, *Arch. Biochem. Biophys.*, **135**, 126 (1969).
(19) R. W. Taft in "Steric Effects in Organic Chemistry," M. S. Newman, Ed., Wiley, New York, N. Y., 1956, p 556.
(20) C. Hansch and E. Coats, *J. Pharm. Sci.*, **59**, 731 (1970).
(21) C. Hansch and J. M. Clayton, *J. Pharm. Sci.*, **62**, 1 (1973).