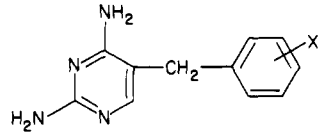
Inhibition of Chicken Liver Dihydrofolate Reductase by 5-(Substituted benzyl)-2,4-diaminopyrimidines.[1] A Quantitative Structure–Activity Relationship and Graphics Analysis

Cynthia Dias Selassie,[2a] Zhao-Xia Fang,[2b] Ren-li Li,[2b] Corwin Hansch,*[2a] Teri Klein,[2c] Robert Langridge,[2c] and Bernard T. Kaufman[2d]

Department of Chemistry, Pomona College, Claremont, California 91711, Beijing Medical College, Beijing, China, Department of Pharmaceutical Chemistry, School of Pharmacy, University of California at San Francisco, California 94143, and National Institute of Arthritis, Digestive, Diabetes, and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20114.
Received June 28, 1985

The inhibition of chicken liver dihydrofolate reductase by a series of substituted benzylpyrimidines has been investigated. From the inhibition constants a quantitative structure–activity relationship has been formulated. This mathematical model is compared with molecular graphics models constructed from the X-ray crystallographic coordinates of trimethoprim and 5-(3,4-dimethoxy-4-isopropenylbenzyl)-2,4-diaminopyrimidine bound to the enzyme. There is good correspondence between the two types of models.

The major goal of studying the interaction of ligands with bioreceptors is to gain understanding of the forces that mediate such processes. The best systems to study are those receptors whose X-ray crystallographic structures have been established. An excellent example is the enzyme dihydrofolate reductase (DHFR) whose X-ray structure has been well defined with DHFR from the following three sources: chicken liver,[3] *Lactobacillus casei*,[4] and *Escherichia coli*[4,5] bacteria. Inhibitors of DHFR have proved to be highly valuable in antibacterial and antitumor chemotherapy.[6] For these reasons we have been interested in studying the action of a variety of inhibitors on DHFR from various sources.[6] In this report we extend our study of the inhibition of DHFR from chicken liver by benzylpyrimidines I. A major problem for the medicinal chemist



I

is that of making the minimum number of structural variations in a lead compound to achieve a drug that is uniquely selective and potent. Quantitative structure–activity relationships (QSAR) has reached the stage of development where it can play an important role in this process.[7-9]

The mathematical models of QSAR constitute crude maps of receptor topography and thus have various degrees of predictability for judging the potency of congeners not yet tested. Although most investigators formulating QSAR have some feeling for the limits of predictability, the subject has been little discussed in the literature of QSAR. It has been pointed out that there are two types of predictions: those within spanned substituent space and those in unspanned substituent space.[10] By spanned substituent space (SSS) we mean that region encompassed by a given set of substituent constants. For example, if substituents with values of σ from 0.0 (H) to 0.78 (p-NO_2) have been tested, the spanned σ space is 0.0–0.78. If a reasonable number of substituents with values well spread within this range have been tested and a good QSAR has resulted, one can expect to make good predictions of new untested congeners with σ values in this range *provided other factors are constant* or do not intervene. However this is not a trivial matter, when several variables are involved. The SAR of drugs is a multidimensional problem in which even sorting out the appropriate variables constitutes a formidable process.

Predicting activity outside of SSS is in a general sense impossible. QSAR can be considered as receptor maps and thus can be compared with geographical maps. For example, if a contour map were made of the state of Kansas by taking 100 well-spaced sample points, one could derive an equation in X and Y coordinates that would predict rather well most of the elevations in other parts of this rather flat state. In taking more points and using higher order equations, the quality of the predictions could be greatly increased. Reasonable predictions about eastern Colorado could also be made. However predictions about the contours of the Rocky Mountain region of Colorado would be terrible. If the problem is put on a multidimensional basis, more like drug design, and the course of rivers, ore deposits, and nature of the soil, etc., are included, it is immediately seen to be horrendous. Mapping a bioreceptor is in some ways worse because of what is becoming increasingly more apparent—its flexibility[11-13] and the ability to bind similar ligands in more than one way.[14-16] In the field one is sure that a given hill is not

(1) This research supported by Grant GM-30362 (C.H.) and RR-1081 (R.L.) from the National Institutes of Health.
(2) (a) Pomona College. (b) Beijing Medical College. (c) University of California at San Francisco. (d) National Institutes of Health.
(3) Volz, K. W.; Matthews, D. A.; Alden, R. A.; Freer, S. T.; Hansch, C.; Kaufman, B. T.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 2528.
(4) Filman, D. J.; Bolin, J. T.; Matthews, D. A.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13663.
(5) Kuyper, L. F.; Roth, B.; Baccanari, D. P.; Ferone, R.; Beddell, C. R.; Champness, J. N.; Stammers, D. K.; Dann, J. G.; Norrington, F. E. A.; Baker, D. J.; Goodford, P. J. *J. Med. Chem.* **1982**, *25*, 1120.
(6) Blaney, J. M.; Hansch, C.; Silipo, C.; Vittoria, A. *Chem. Rev.* **1984**, *84*, 333.
(7) Kim, K. H.; Hansch, C.; Fukunaga, J. Y.; Steller, E. E.; Jow, P. Y. C.; Craig, P. N.; Page, J. *J. Med. Chem.* **1979**, *22*, 366.
(8) "Drug Design: Fact or Fantasy?"; Jolles, G., Wooldridge, K. R. H., Eds.; Academic Press: New York, 1984.
(9) "Quantitative Structure–Activity Relationships of Drugs"; Topliss, J. G., Ed.; Academic Press: New York, 1983.
(10) Hansch, C. *Pharmacochem. Libr.* **1977**, *2*, 47.
(11) McCammon, J. A. *Rep. Prog. Phys.* **1984**, *47*, 1.
(12) Roberts, G. C. K. *Pharmacochem. Libr.* **1983**, *6*, 91.
(13) Hansch, C.; Hathaway, B. A.; Guo, Z. R.; Selassie, C. D.; Dietrich, S. W.; Blaney, J. M.; Langridge, R.; Volz, K. W.; Kaufman, B. T. *J. Med. Chem.* **1984**, *27*, 129.
(14) Martin, Y. C. "Quantitative Drug Design"; Marcel Dekker: New York, 1978; p 237.

0022-2623/86/1829-0621$01.50/0

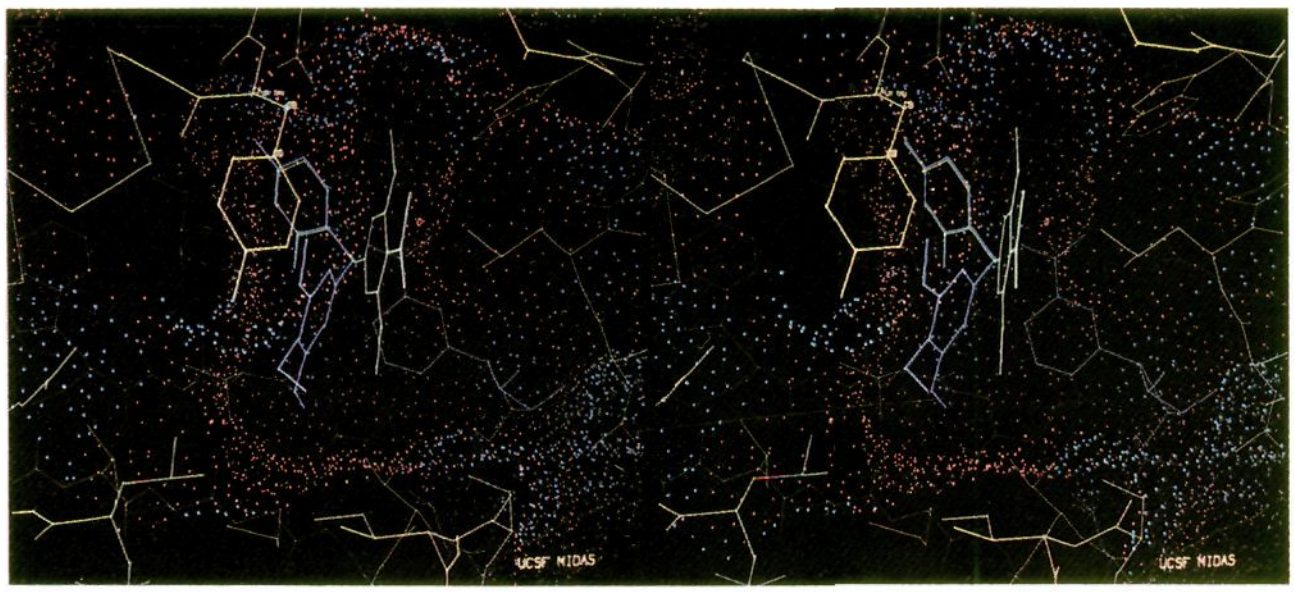


Figure 1. Trimethoprim (purple) and compound II (green) are shown bound to chicken liver DHFR. The positions of both compounds were determined from X-ray crystallographic coordinates of the substances bound separately to the enzyme (Matthews et al.). To the right of these molecules tyrosine-31 (yellow) is displayed. Tyrosine is in the "down" position it occupies with methotrexate, triazines, and compound II. In this position the 5-methoxy group of trimethoprim collides with tyrosine, causing it to move to the position shown in Figure 2. Red dots code for hydrophobic surface (carbon) and blue dots represent hydrophilic surface (oxygen and nitrogen).

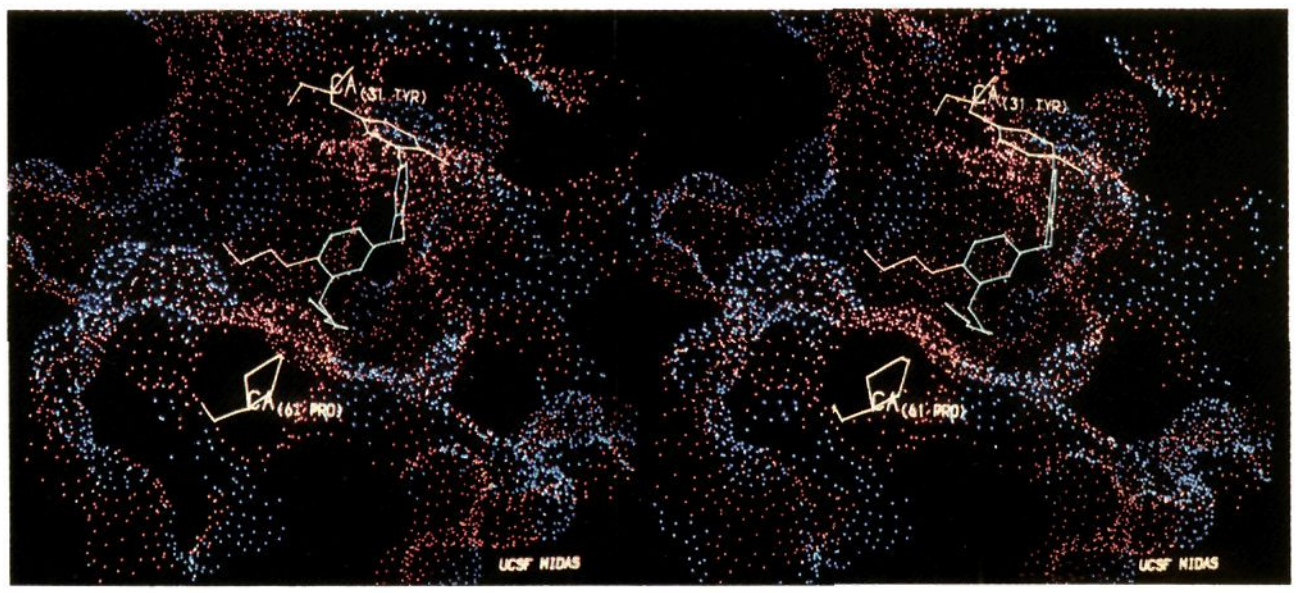


Figure 2. The "wire" model (green) is that of [3-(hexyloxy)-4-propoxybenzyl]pyrimidine bound to chicken DHFR. Tyrosine is shown in the "up" position for illustrative purposes. This does not imply that 3,4-disubstituted benzylpyrimidines would cause tyrosine to shift. Proline-61 in the foreground produces a steric effect that is seen with the long 3-alkoxy groups.

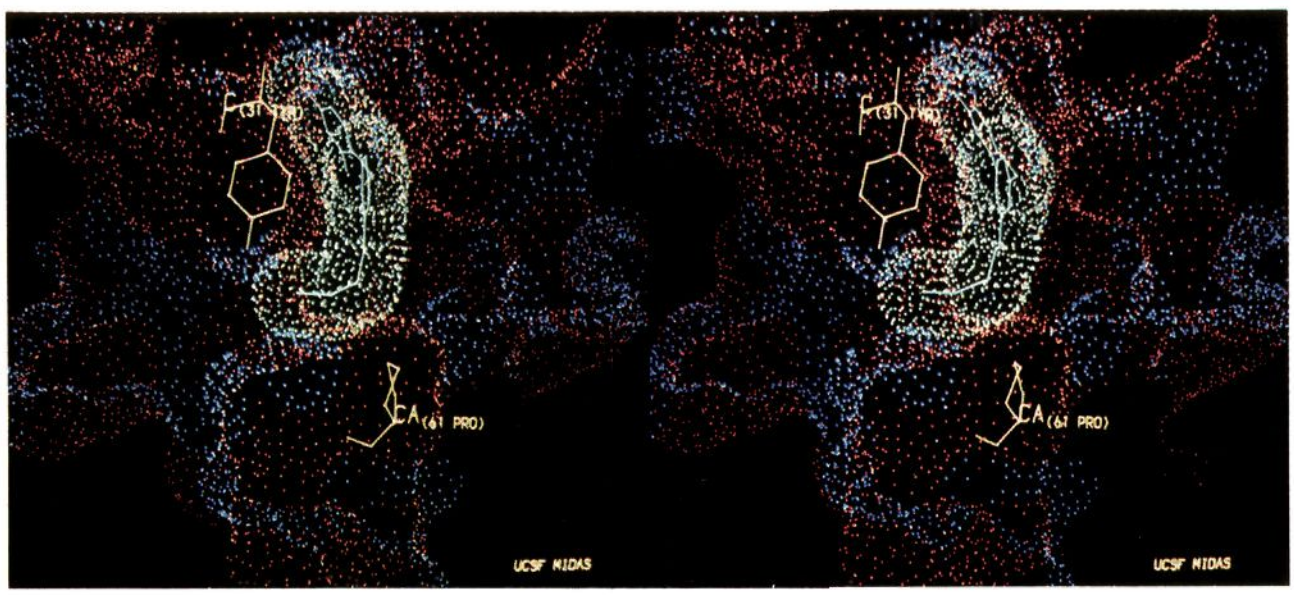


Figure 3. This figure shows tyrosine in the down position with a surface over the 3-OC_7 substituent.

going to change its shape at least during the lifetime of a normal map. With enzymes we have as yet, no way of judging exactly how much "give" a certain active site will have for a ligand. Roberts[12] has written of "flexible keys and deformable locks" in talking about DHFR. Thus even having the X-ray crystallographic structure in hand and high-powered stero computer graphics for visualizing the structure of receptors does not assure one of success in making quantitative predictions about ligand fit to receptor.[13] Nevertheless, X-ray crystallography, QSAR, and computer graphics in combination can do much to make lead modification much more efficient. In the present report our previously derived QSAR for inhibitors I acting on chicken liver DHFR is extended to 14 new benzyl-

(15) Li, R. L.; Hansch, C.; Matthews, D. A.; Blaney, J. M.; Langridge, R.; Delcamp, T. J.; Susten, S. S.; Freisheim, J. H. *Quant. Struct–Act. Relat.* **1982**, *1*, 1.
(16) Langridge, R.; Ferrin, T. E.; Kuntz, I. D.; Connolly, M. L. *Science (Washington, D.C.)* **1981**, *211*, 661.

pyrimidines and the quality of the results is discussed.

Results

From the data in Table I, eq 1–7 for the inhibition of chicken liver DHFR have been derived. In deriving eq 1–7, one data point, 4-N$(CH_3)_2$, has been omitted. This derivative is 3.7 times less active than eq 7 predicts. Including this point yields essentially the same equation but with a lower correlation (r = 0.909, s = 0.222).

$$\log 1/K_i = 0.27(\pm0.10)\pi_3 + 4.44(\pm0.11) \quad (1)$$

n = 53, r = 0.593, s = 0.393, $F_{1,51}$ = 27.7, $F_{1,40\alpha\cdot001}$ = 12.6

$$\log 1/K_i = 0.53(\pm0.15)\pi_3 - 1.30(\pm0.63)\log(\beta\cdot10^{\pi_3} + 1) + 4.45(\pm0.10) \quad (2)$$

n = 53, r = 0.721, s = 0.346, π_3^0 = 2.40(±1.05)log β = −2.57, $F_{2,49}$ = 8.51, $F_{2,40\alpha\cdot001}$ = 8.25

$$\log 1/K_i = 0.56(\pm0.14)\pi_3 - 1.26(\pm0.50)\log(\beta\cdot10^{\pi_3} + 1) - 0.49(\pm0.26)MR_5 + 4.59(\pm0.12) \quad (3)$$

n = 53, r = 0.793, s = 0.307, π_3^0 = 2.25(±0.89) log β = −2.36, $F_{1,48}$ = 14.3

$$\log 1/K_i = 0.54(\pm0.13)\pi_3 - 1.26(\pm0.47)\log(\beta\cdot10^{\pi_3} + 1) - 0.46(\pm0.23)MR_5 + 0.18(\pm0.10)\pi_4 + 4.55(\pm0.10) \quad (4)$$

n = 53, r = 0.843, s = 0.274, π_3^0 = 2.32(±1.37) log β = −2.45, $\pi_{1,47}$ = 13.1

$$\log 1/K_i = 0.51(\pm0.11)\pi_3 - 1.17(\pm0.40)\log(\beta_3\cdot10^{\pi_3} + 1) - 0.53(\pm0.21)MR_5 + 0.67(\pm0.27)\pi_4 - 0.75(\pm0.39)\log(\beta_4\cdot10^{\pi_4} + 1) + 4.77(\pm0.15) \quad (5)$$

n = 53, r = 0.884, s = 0.243, $F_{2,45}$ = 7.32, π_3^0= 2.27, π_4^0 = 1.23, $F_{2,40\alpha\cdot005}$ = 6.07, log β_3 = −2.38, log β_4 = −0.28

$$\log 1/K_i = 0.45(\pm0.11)\pi_3 - 1.13(\pm0.43)\log(\beta_3\cdot10^{\pi_3} + 1) - 0.63(\pm0.20)MR_5 + 0.75(\pm0.29)\pi_4 - 0.80(\pm0.39)\log(\beta_4\cdot10^{\pi_4} + 1) + 0.49(\pm0.32)\pi_5 + 4.87(\pm0.17) \quad (6)$$

n = 53, r = 0.905, s = 0.224, $F_{1,44}$ = 9.16, π_3^0= 2.36, π_4^0 = 1.20, $F_{1,40\alpha\cdot01}$ = 7.31, log β_3 = −2.54, log β_4 = −0.07

$$\log 1/K_i = 0.43(\pm0.11)\pi_3 - 1.13(\pm0.35)\log(\beta_3\cdot10^{\pi_3} + 1) - 0.66(\pm0.19)MR_5 + 0.59(\pm0.22)\pi_4 - 0.63(\pm0.33)\log(\beta_4\cdot10^{\pi_4} + 1) + 0.48(\pm0.30)\pi_5 + 0.14(\pm0.09)MR_3 + 4.64(\pm0.14) \quad (7)$$

n = 53, r = 0.921, s = 0.208, $F_{1,43}$ = 8.07, π_3^0= 2.00, π_4^0 = 1.53, $F_{1,40\alpha\cdot01}$ = 7.31, log β_3 = −2.21, log β_4 = −0.40

The squared correlation matrix for the variables is

	π_3	π_4	π_5	MR_3	MR_5
π_3	1	0.00	0.00	0.52	0.01
π_4		1	0.01	0.04	0.00
π_5			1	0.00	0.12
MR_3				1	0.00
MR_5					1

The eigenvalues and the fraction of variance accounted for by each are 1.78 (35.51%), 1.35 (26.92%), 1.02 (20.31%), 0.62 (12.35%), and 0.25 (4.90%).

For meta substituents a single substituent is called 3 and when two are present one is labeled 5. Since all of the examples in which two meta substituents are present are symmetrical, there is no ambiguity about which is 5. MR refers to the molar refractivity of the substituent and the subscripts refer to the position of substituent attachment to the phenyl ring.

In the stepwise development of eq 7 all of the terms are significant at >0.99 level of significance. There are two bilinear parts to eq 7 that establish the optimum hydrophobicity (π_0) for substituents in the 3- and 4-position. It is of special interest that the initial slopes of π_3, π_4, and π_5 are all about 0.5, which suggests about 50% desolvation of X in partitioning onto the enzyme. However, in the case of π_3 the right-hand side of the bilinear curve has a significant negative slope (0.43 − 1.13 = −0.70), which suggests a possible steric effect of large groups (beyond π_0, e.g., OC_5H_{11}). For π_4 the right-hand slope is 0.59 − 0.63 = −0.04, indicating that parts of hydrophobic substituents longer than OC_4H_9 (π = 1.55) do not contact the enzyme but project into the aqueous phase.

The effects of substituents in the 5-position are accounted for by MR_5 and π_5. The negative coefficient with MR_5 suggests steric hindrance to binding, which more than offsets the positive hydrophobic effect modeled by π_5. Since MR is scaled by 0.1 for alkyl groups, it is equiscalar with π. Thus binding appears to occur in sterically hindered hydrophobic space.

The fraction of variance accounted for by the eigenvalues shows that one vector is of marginal importance, accounting for just under 5% of the variance. Martin[14] has suggested that a reasonable cut-off point for significance might be 5%. From the correlation matrix it seems likely that MR_3, which is highly collinear with π_3 and which is the last variable to enter in the stepwise development of eq 7, is the weak variable, π_3, which is the most significant variable, is thus the first to turn up in eq 1. MR_3 is the last to turn up and has a low positive coefficient, suggesting a favorable inhibitory effect from gross bulk of 3-substituents.

Equation 7 is a complex expression having 10 disposable parameters and we would be reluctant to consider it were it not for the fact that conclusions from eq 7 correspond well with what is to be seen in the molecular graphics. While many might be inclined to regard eq 7 as being too complex, we feel, as for most QSAR, that it is in fact too simple. The general weakness of QSAR is that it attempts to accomplish too much with too few variables. The result is that only the high points of the terrain of the receptor can be outlined—the subtleties escape our attention.

Molecular Graphics and Discussion

The models of the DHFR receptor and the pyrimidines have been built as previously described.[15] The display and manipulation of the coordinates was done on an Evans and Sutherland picture system 2[16] by means of the program MIDAS.[17] Solvent accessible surfaces for the enzyme were generated with MS[18] and interactively as van der Waals surfaces.[19]

The structure of the chicken liver dihydrofolate reductase was determined by X-ray crystallography by Voltz et al.[20] Coordinates for the chicken liver DHFR–NADPH–

(17) Huang, C.; Jarvis, L.; Ferrin, T. E.; Langridge, R. 1983, UCSF MIDAS: Molecular Interactive and Display.
(18) Connolly, M. L. *Science (Washington, D.C.)* **1983**, *221*, 709.
(19) Bash, P. A.; Pattabiraman, N.; Huang, C.; Ferrin, T. E.; Langridge, R. *Science (Washington, D.C.)* **1983**, *222*, 1325.

Table I. Parameters Used in the Derivation of Equations 1–7 for the Inhibition of Chicken Liver Dihydrofolate Reductase

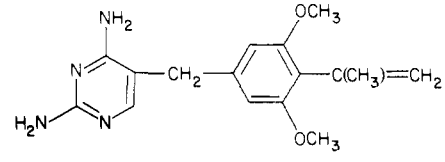
no.	X	log 1/C obsd	log 1/C calcd[a]	Δ log 1/C	π_3	π_4	π_5	MR_5	MR_3
1	3,4,5-$(CH_2CH_3)_3$	5.25	5.00	0.25	0.86	0.86	0.86	1.03	1.03
2	3,5-$(OCH_3)_2$, 4-$(OCH_2CH_2OCH_3)$	3.64	3.75	−0.11	0.01	−0.80	0.01	0.79	0.79
3	3,4,5-$(OCH_3)_3$	3.98	3.89	0.09	0.01	−0.54	0.01	0.79	0.79
4	3,5-$(OCH_3)_2$, 4-$N(CH_3)_2$	4.15	4.11	0.04	0.01	−0.09	0.01	0.79	0.79
5	3,5-$(OCH_3)_2$, 4-Br	4.54	4.29	0.25	0.01	0.44	0.01	0.79	0.79
6	3,5-$(OCH_3)_2$, 4-SCH_3	4.29	4.14	0.15	0.01	0.00	0.01	0.79	0.79
7	3,5-$(OCH_3)_2$, 4-$C(CH_3)=CH_2$	4.17	4.30	−0.13	0.01	0.47	0.01	0.79	0.79
8	3,5-$(OCH_2CH_3)_2$, 4-C_4H_4N	4.33	4.45	−0.12	0.47	0.38	0.47	1.25	1.25
9	3,5-$(CH_2OH)_2$	3.23	3.24	−0.01	−1.03	0.00	−1.03	0.72	0.72
10	3,5-$(OCH_3)_2$	4.12	4.14	−0.02	0.01	0.00	0.01	0.79	0.79
11	3,5-$(OCH_2CH_3)_2$	4.14	4.32	−0.18	0.47	0.00	0.47	1.25	1.25
12	3,5-$(CH_3)_2$	4.61	4.75	−0.14	0.56	0.00	0.56	0.57	0.57
13	3,4-$(OH)_2$	3.59	3.91	−0.32	−0.67	−0.67	0.00	0.10	0.29
14	3-NO_2, 4-$NHCOCH_3$	4.34	4.00	0.34	−0.28	−0.91	0.00	0.10	0.74
15	3,4-$(OCH_2CH_2OCH_3)_2$	3.91	4.37	−0.46	−0.43	−0.43	0.00	0.10	1.93
16	3,4-OCH_2O-	4.68	4.52	0.16	−0.03	−0.03	0.00	0.10	0.45
17	3,4-$(OCH_3)_2$	4.46	4.34	0.12	−0.29	−0.29	0.00	0.10	0.79
18	3-CF_3, 4-OCH_3	4.99	4.95	0.04	1.25	−0.20	0.00	0.10	0.50
19	3-$O(CH_2)_7CH_3$, 4-OCH_3	4.71	4.78	−0.07	3.78	−0.20	0.00	0.10	3.97
20	3-OCH_2CONH_2	4.27	4.12	0.15	−1.37	0.00	0.00	0.10	1.60
21	3-CH_2OH	4.31	4.14	0.17	−1.03	0.00	0.00	0.10	0.72
22	3-OSO_2CH_3	4.33	4.34	−0.01	−0.88	0.00	0.00	0.10	1.70
23	3-CH_2OCH_3	4.37	4.32	0.05	−0.78	0.00	0.00	0.10	1.21
24	3-OH	3.87	4.23	−0.36	−0.67	0.00	0.00	0.10	0.29
25	3-$OCH_2CH_2OCH_3$	4.83	4.62	0.21	−0.30	0.00	0.00	0.10	1.93
26	3-OCH_3	4.45	4.63	−0.18	0.11	0.00	0.00	0.10	0.79
27	3-F	4.70	4.59	0.11	0.23	0.00	0.00	0.10	0.09
28	3-CH_3	4.72	4.77	−0.05	0.52	0.00	0.00	0.10	0.57
29	3-Cl	5.01	4.84	0.17	0.67	0.00	0.00	0.10	0.60
30	3-Br	5.03	4.95	0.08	0.86	0.00	0.00	0.10	0.89
31	3-CF_3	4.92	4.91	0.01	0.88	0.00	0.00	0.10	0.50
32	3-$CH_2O(CH_2)_3CH_3$	5.17	5.18	−0.01	0.84	0.00	0.00	0.10	2.60
33	3-I	4.79	5.12	−0.33	1.12	0.00	0.00	0.10	1.39
34	3-$O(CH_2)_3CH_3$	5.22	5.35	−0.13	1.55	0.00	0.00	0.10	2.17
35	3-$OCH_2C_6H_5$	5.63	5.49	0.14	1.56	0.00	0.00	0.10	3.17
36	3-$O(CH_2)_5CH_3$	5.67	5.40	0.27	2.63	0.00	0.00	0.10	3.07
37	3-$O(CH_2)_6CH_3$	4.79	5.16	−0.37	3.23	0.00	0.00	0.10	3.52
38	3-$O(CH_2)_7CH_3$	5.08	4.86	0.22	3.79	0.00	0.00	0.10	3.97
39	4-NH_2	3.73	3.80	−0.07	0.00	−1.32	0.00	0.10	0.10
40	4-$NHCOCH_3$	4.26	4.03	0.23	0.00	−0.91	0.00	0.10	0.10
41	4-$OCH_2CH_2OCH_3$	4.26	4.36	−0.10	0.00	−0.30	0.00	0.10	0.10
42	4-NO_2	4.37	4.49	−0.12	0.00	0.00	0.00	0.10	0.10
43	4-OCH_3	4.29	4.40	−0.11	0.00	−0.20	0.00	0.10	0.10
44	4-F	4.79	4.55	0.24	0.00	0.14	0.00	0.10	0.10
45	4-$N(CH_3)_2$[b]	4.01	4.58	−0.57	0.00	0.24	0.00	0.10	0.10
46	4-CH_3	4.56	4.67	−0.11	0.00	0.56	0.00	0.10	0.10
47	4-Cl	4.83	4.70	0.13	0.00	0.71	0.00	0.10	0.10
48	4-Br	4.79	4.72	0.07	0.00	0.86	0.00	0.10	0.10
49	4-OCH_3	4.32	4.74	−0.42	0.00	1.04	0.00	0.10	0.10
50	4-$O(CH_2)_3CH_3$	4.67	4.75	−0.08	0.00	1.55	0.00	0.10	0.10
51	4-$OCH_2C_6H_5$	4.83	4.75	0.08	0.00	1.66	0.00	0.10	0.10
52	4-$O(CH_2)_5CH_3$	4.71	4.72	−0.01	0.00	2.63	0.00	0.10	0.10
53	4-$O(CH_2)_6CH_3$	4.73	4.70	0.03	0.00	3.23	0.00	0.10	0.10
54	H	4.71	4.49	0.22	0.00	0.00	0.00	0.10	0.10

[a] Calculated by using eq 7. [b] Not used in the derivation of the equations.

TMP complex at 2.2-Å resolution were kindly supplied by D. A. Matthews of the University of California, San Diego. Our models of the benzylpyrimidine inhibitors were based on the coordinates of this complex. Models of the substituted benzylpyrimidines were built by adding the substituents to the phenyl ring in an extended conformation using standard geometries. The active-site walls are formed by Asn-48 to Trp-57, Ser-59 to Asn-67, Trp-113 to Tyr-121, Thr-136, Ile-138, Asp-145, and Thr-146. In the stereo views (Figures 1–3) the red dots represent hydrophobic surfaces (carbon) while the blue dots code for hydrophilic surfaces (oxygen and nitrogen).

One of the fascinating features of chicken DHFR established by Matthews and his colleagues[21] via X-ray crystallography is that when trimethoprim binds to DHFR, Tyr-31 moves about 180° from the position it occupies when methotrexate or triazine inhibitors are bound. This has been illustrated with a color stereo graphics model.[6]

In Figure 1 the benzyl portion of TMP is shown in purple and Tyr-31 (yellow) is shown not as it appears with TMP but in the "down" position it takes when the closely related compound II is bound. The coordinates of II



II

bound to chicken DHFR were generously supplied by D. A. Matthews. It would appear that the 5-OCH_3 of TMP causes the movement of Tyr-31, but in the presence of II the hydrophobic contact with the surface of the isopropenyl group holds Tyr-31 down. The position of both compounds in the active site are those determined by the X-ray crystallographic studies (Matthews et al.[21]), and it is most interesting to see how different the binding position of II (green) is compared to TMP (purple). The red surface between Tyr-31 and II is that defined by the Try-31–II complex. The flat isopropenyl group is well placed to make good hydrophobic contact.

It is surprising that TMP and II, only one of which

(20) Stenbuck, D.; Baltzly, R.; Hood, H. M. *J. Org. Chem.* **1963**, *28*, 1983.
(21) Freisheim, J. H.; Matthews, D. A. In "Folate Antagonists as Therapeutic Agents"; Sirotnak, F. M., Burchall, J. J., Emsinger, W. B., Montgomery, J. A., Eds.; Academic Press: New York, 1985; Vol. 1.

causes Tyr-31 to move, have almost the same log $1/K_i$ values of 3.98 and 4.17, respectively. The steric effect of the 5-OCH_3 must be almost completely offset by the hydrophobic effect of the 4-$C(CH_3)=CH_2$ group.

Matthews and his colleagues have shown by X-ray crystallography that the 3,4,5-triethyl analogue of I causes Tyr-31 to move just as TMP does despite the possibility of hydrophobic interaction between the 4-CH_2CH_3 and Tyr-31. The triethyl analogue turns out to be one of the more active compounds in Table I and is reasonably well predicted by eq 7 as are TMP and II.

It is clear from the graphics that Tyr-31 must move if TMP is to take the position it is known to have in Figure 1. This is accounted for in the QSAR by the negative MR_5 term, which shows a steric effect when two meta substituents are present, and the positive π_5 term, which shows a positive hydrophobic interaction. This is evident from Figure 2, which shows Tyr-31 in the moved position when the phenyl ring containing 4-$O(CH_2)_2CH_3$ and 3-$O(CH_2)_5CH_3$ side chains.

It is of interest to note that behind the position where the tyrosine residue resided there is hydrophobic space sufficient for binding by the 5-CH_2CH_3 group as well as for most of a 4-$O(CH_2)_3CH_3$ substituent. This latter substituent represents the cut off in effective hydrophobic binding for 4-substituents as delineated by the bilinear terms of π_4 in eq 7. Substituents longer than 4-$O(CH_2)_3CH_3$ are seen to contact polar space and then to protrude beyond the enzyme. Since the last carbon of 4-$O(CH_2)_3CH_3$ is held against polar space and not hydrophobic surface, there may be enough effect in terms of dispersion forces alone to account for the good activity of this congener.

Figure 2 shows the large hydrophobic pocket open to solvent on one side so that X would be only about 50% desolvated in the complex as one might surmize from the 0.5 coefficient that is found with π_3, π_4, and π_5. Essentially the same situation prevails with Tyr-31 in the "down" position as can be seen in Figure 3 where a van der Waals surface has been placed on 3-X.

In Figures 2 and 3 one can see that the first part of a long hydrophobic 3-substituent can bind to a hydrophobic surface without hindrance up to $\pi_0 = 2.00$ (3-$O(CH_2)_4CH_3$). At this point unfavorable contact occurs due to a rise in the hydrophobic surface caused by Pro-61. It would seem that it is this steric effect by Pro-61 that causes the negative slope of the right-hand side of the π_3 bilinear relationship. This steric effect is most pronounced with 3-$O(CH_2)_6CH_3$, which is less active than predicted, but begins to fall off at 3-$O(CH_2)_7CH_3$, which is slightly more active than predicted by eq 7. It is likely that eq 7 would mispredict longer chains by even larger amounts since the end carbons of these chains would not contact the enzyme, causing neither steric effects nor positive hydrophobic binding effects.

In summary then, except for the rather vague MR_3 term, comparison of QSAR and graphics models finds excellent concordance between the two. Although eq 7 is very complicated, this intricacy is dictated by the complexity of the geometry of the active site and the large amount of variation in the substituents attached to the benzyl moiety. We are particularly pleased to find another example where, when a hydrophobic substituent binds to a hydrophobic region so that it is desolvated on only one side, one finds a coefficient of about 0.5 with the π term. Although this was not entirely unforseen, there is no theoretical way to estimate just what the coefficient would be between octanol–water partitioning and protein–water partitioning. It seems reasonable to expect that such coefficients may become firmly enough established to be of diagnostic value in the mapping of unknown receptors.[22,23]

The larger number of data points on which eq 7 is based has led to a different QSAR from that previously reported (eq 8) for the benzylpyrimidine inhibitors acting on chicken DHFR.[24]

$$\log 1/K_i = 0.55\pi_{3,4,5} - 0.42 \log (\beta \cdot 10^{\pi_{3,4,5}} + 1) + 0.20MR_3 + 0.32\sigma + 4.46 \quad (8)$$

$$n = 39,\ r = 0.900,\ s = 0.241$$

At the time of development of eq 8, insufficient congeners with long hydrophobic chains were present in the data set to show the separate roles for π_3 and π_4. Also there were only two compounds where 5-substituents were present, and these were both methoxyl groups. The parameter of major importance in eq 8 is $\pi_{3,4,5}$, and its coefficient of essentially 0.5 is what we now find with the larger, better constructed data set. Moreover, eq 7 now defines cut-off points for hydrophobic interactions of 3- and 4-substituents independently. The weak MR_3 term is present in both equations, but the marginal σ term has disappeared in eq 7.

A satisfying aspect of eq 7 is the modeling of the steric effect by MR_5 that 5-substituents clearly encounter with Tyr-31 when the benzyl group binds as with TMP. Also it is striking that π_5 accounts for the possible hydrophobic interaction with the hydrophobic space that is exposed when Tyr-31 moves. We have found other instances where substituents appear to encounter steric effects with hydrophobic surfaces,[25,26] which can be accounted for with a combination of hydrophobic and steric terms.

In attempting to delineate the forces involved in the interaction of ligand and receptor, one is faced with a dilemma. Should one attempt to develop parameters that would ultimately provide an atom by atom description of the forces between ligand and receptor or should one compromise and accept a crude description, which is an approximation of reality, based on relatively few variables? The problem with being rigorous at this stage of development in QSAR is that because of the noise inherent in the test data and the uncertainty of our models as well as the unknown flexibility of the receptor we must use statistical means for the validation of parameters as well as for the overall mathematical model. Since one needs a minimum of about five data points to check out a given variable, the amount of experimental work in terms of organic synthesis and bioassay, to say nothing of devising higher quality parameters or using more sophisticated calculations such as molecular dynamics or quantum mechanics, becomes formidable. Nevertheless, the field is bound to move in this direction, and only the largest industrial laboratories will be able to undertake such comprehensive and expensive studies. Now that the cost of developing a new drug is in the range of 70 to 100 million dollars, a few million set aside for careful analysis to insure that one has indeed come somewhere near ideality in effectiveness for the class of drug is not unreasonable.

In spite of its complexity, eq 7 is on the side of attempting to do too much with too few parameters. The use of hydrophobic constants for rather large substituents

(22) Carotti, A.; Smith, R. N.; Wong, S.; Hansch, C.; Blaney, J. M.; Langridge, R. *Arch. Biochem. Biophys.* **1984**, *229*, 112.
(23) Hansch, C.; McClarin, J.; Klein, T.; Langridge, R. *Mol. Pharmacol.* **1985**, *27*, 493.
(24) Li, R. L.; Hansch, C.; Kaufman, B. T. *J. Med. Chem.* **1982**, *25*, 435.
(25) Ames, M. M.; Selassie, C. D.; Woodson, L. C.; Van Loon, J. A.; Hansch, C.; Weinshilboum, R. M. *J. Med. Chem.* **1986**, *29*, 354.
(26) Hansch, C.; Fukunaga, J. Y.; Jow, P. Y. C.; Hynes, J. B. *J. Med. Chem.* **1977**, *20*, 96.

Table II. 2,4-Diamino-5-(X-benzyl)pyrimidines

no.	X	mp, °C	yield,[a] %	formula[b]	ref
11	3,5-$(OC_2H_5)_2$	177–180	16	$C_{15}H_{20}N_4O_2$	30
12	3,5-$(CH_3)_2$	207–208	9	$C_{13}H_{16}N_4$	20
16	3,4-OCH_2O-	254–256	13	$C_{12}H_{12}N_4O$	30
51	4-$OCH_2C_6H_5$	204–205	21	$C_{18}H_{18}N_4O$	30

[a] Yield based on the amount of starting benzaldehyde. [b] Analyzed for C, H, and N.

Table III. Acrylonitrile Intermediates

	mp or bp (mmHg), °C	yield,[a] %
3,5-$(CH_3)_2$-2-$C_6H_3CH_2C(CN)$=$CHOCH_3$	100–145 (2)	68
3,4-$OCH_2OC_6H_3CH_2C(CN)$=$CHNHC_6H_5$	127–129	33
4-$OCH_2C_6H_5C_6H_4CH_2C(CN)$=$CHNHC_6H_5$	143–144	30
3,5-$(OEt)_2C_6H_3CH_2C(CN)$=$CHNHC_6H_5$	131–132	23

[a] Yield based on the amount of starting benzaldehyde.

Table IV. Partition Coefficients of Some Benzylpyrimidines

no.	X	log P	ref	$\pi_{3,4,5}$
1	3,4,5-$(CH_2CH_3)_3$	1.55[a]	34	2.58
2	3,5-$(OCH_3)_2$, 4-$OCH_2CH_2OCH_3$	-1.81[a]	34	-0.78
3	3,4,5-$(OCH_3)_3$	-1.55[a]	34	-0.52
4	3,5-$(OCH_3)_2$, 4-$N(CH_3)_2$	1.51[b]	34	-0.07
5	3,5-$(OCH_3)_2$, 4-Br	2.04[c]	35	0.46
6	3,5-$(OCH_3)_2$, 4-SCH_3	1.60[c]	35	0.02
7	3,5-$(OCH_3)_2$, 4-$C(CH_3)$=CH_2	3.07[b]	34	0.49
8	3,5-$(OCH_2CH_3)_2$, 4-NCH=CHCH=CH	2.90[b]	34	1.32
10	3,5-$(OCH_3)_2$	1.60[c]	35	0.02
11	3,5-$(OCH_2CH_3)_2$	2.52[b]	34	0.94
15	3,4-$(OCH_2CH_2OCH_3)_2$	-1.89[a]	34	-0.86
16	3,4-OCH_2O-	1.52[c]	35	-0.06
17	3,4-$(OCH_3)_2$	1.00[c]	35	-0.58
19	3-$O(CH_2)_7CH_3$, 4-OCH_3	2.55[a]	34	3.58
25	3-$OCH_2CH_2OCH_3$	-1.33[a]	34	-0.30
26	3-OCH_3	-0.92[a]	34	0.11
27	3-F	-0.80[a]	34	0.23
28	3-CH_3	-0.51[a]	34	0.52
29	3-Cl	-0.36[a]	34	0.67
35	3-$OCH_2C_6H_5$	0.53[a]	34	1.56
39	4-NH_2	0.26[c]	35	-1.32
40	4-$NHCOCH_3$	-1.94[a]	34	-0.91
42	4-NO_2	-1.03[a]	34	0
43	4-OCH_3	1.34[d]	36	-0.20
45	4-$N(CH_3)_2$	1.82[b]	34	0.24
54	H	-1.03[a]	34	
	H	1.58[b]	34	
	H	1.54[d]	36	

[a] Aqueous phase = 0.1 N HCl. [b] Aqueous phase = 0.1 N NaOH. [c] Aqueous phase = phosphate buffer, pH 7.40. These values were calculated by extrapolation such that log P(0.1 N NaOH) = log $(10^{\log P(\text{pH } 7.4)}/0.6)$. [d] Aqueous phase = methanol–water, pH 8.40.

assumes a homogeneity of protein surface that is simplistic. The surprising fact is that it works as well as it does. Equation 7 accounts for about 85% of the variance in log $1/K_i$ for 53 multisubstituted complex compounds. One would like to do better and above all, one would like to feel more confident about the basic assumptions in the mathematical model.

The bilinear model that we have employed in eq 7 has shortcomings in the way we have used it. This model was developed by Kubinyi[27] to account for the passive movement of organic compounds through biological membranes. Under such conditions a specific hydrophobic interaction of one part of a congener, not shown by other members in the series, with a specific part of a lipophilic protein or membrane would be smoothed out by countless other random hydrophobic interactions with the large variety of such sites in a cell or living organism, unless it was of unusual intensity. With an isolated protein such as DHFR, the situation is different. For example, imagine a rather flat limited hydrophobic space just wide and long enough to accommodate a phenyl group of π 1.96. On such a surface, an *n*-butyl group in extended form would project beyond the surface so in spite of its slightly higher π value (2.13) it would produce a smaller hydrophobic effect. An adamantyl group (π = 3.30) because of its shape would not produce the hydrophobic effect its higher π value would suggest.

In traversing the series from methyl to hexyl, the bilinear model would delineate the point at which the hydrophobic chain extends beyond the surface (π_0); however, a substituent such as $(CH_2)_4NHC(O)CH_3$ would produce the same hydrophobic effect as butyl since its polar moiety would project beyond hydrophobic space and hence have no effect on binding. However, a bilinear QSAR developed on the basis of alkyl chains would predict much lower activity because its overall π value is only 0.24.

Recently, atomic hydrophobic constants have been defined[28,29] with which one could deal more effectively with the hydrophobic heterogeneity of a protein surface. The difficulty of this approach, beyond that of obtaining a reliable set of such parameters, is that it introduces the need for many more data points to support the statistical validation of the more numerous vectors.

We have not touched on the problems of defining the most effective conformation of flexible compounds or the great difficulties of dealing with highly specific steric effects between receptor and ligands. In the face of all these adversities in fitting data with imperfect parameters and crude mathematical models, eq 7 constitutes a better result than one might expect before undertaking such an investigation.

Experimental Section

The assay procedures for the inhibition of chicken liver dihydrofolate reductase have been reported.[38] The syntheses of most of the benzylpyrimidines have been reported with the following exceptions:[24,37] Compounds **4–8**, Table I, were generously provided by Dr. H. Gutmann and P. Weber of the Hoffman-La Roche Co. in Basel, Switzerland. Compound **1** was a gift from Barbara Roth of Borroughs Wellcome, Research Triangle Park, NC. Compounds **11**, **12**, **16**, **19**, and **51** were synthesized as described below. Compound **12** was prepared by the method of Stenbuck et al., which we have previously used.[20] Compounds **11**, **16**, and **51** were prepared by the general method of Kompis et al.[30] These results are summarized in Table II. The necessary benzaldehydes were prepared according to known procedures: 3,5-$(CH_3)_2$,[31] 4-$OCH_2C_6H_5$,[32] and 3,5-$(OCH_2CH_3)_2$.[33] The melting points are uncorrected (Büchi capillary apparatus). Microanalyses are within 0.4% of the theoretical values, and thin-layer chromatography was used to check the purity of the final products. Table III gives the data on the acrylonitrile intermediates. Table IV lists all the measured partition coefficients that were used to determine the π values of the appropriate substituents.

(27) Kubinyi, H. *Drug Res.* **1979**, *23*, 97.
(28) Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1985**, *28*, 333.
(29) Broto, P.; Moreau, G.; Vandycke, C. *Eur. J. Med. Chem.* **1984**, *19*, 71.
(30) Kompis, I.; Then, R.; Boehni, E.; Ray-Belljt, G.; Zanetti, G.; Montavon, M. *Eur. J. Med. Chem.–Chim. Ther.* **1980**, *15*, 17.
(31) Marvel, C. S.; Saunder, J. H.; Overberger, C. G. *J. Am. Chem. Soc.* **1946**, *68*, 1085.
(32) Stoermer, R.; Wodarg, F. *Ber.* **1929**, *61B*, 2323.
(33) Szilagyi, G.; Kasztreiner, E.; Vargha, L. *Chem. Abstr.* **1971**, *75*, P140518Y.
(34) Pomona College Medchem Database Issue 26.
(35) Seiler, P.; Bischoff, O.; Wagner, R. *Arzneim-Forsch.* **1982**, *32*, 711.
(36) Calas, M.; Barbier, A.; Giral, L.; Balmayer, B.; Despaux, E. *Eur. J. Med. Chem.* **1982**, *17*, 497.
(37) Coats, E. A.; Genther, C. S.; Selassie, C. D.; Strong, C. D.; Hansch, C. *J. Med. Chem.*, in press.
(38) Dietrich, S. W.; Blaney, J. M.; Reynolds, M. A.; Jow, P. Y. C.; Hansch, C. *J. Med. Chem.* **1980**, *23*, 1201.