Comparison of the Inhibition of *Escherichia coli* and *Lactobacillus casei* Dihydrofolate Reductase by 2,4-Diamino-5-(Substituted-benzyl)pyrimidines: Quantitative Structure-Activity Relationships, X-ray Crystallography, and Computer Graphics in Structure-Activity Analysis†

Corwin Hansch,*‡ Ren-li Li,‡§ Jeffrey M. Blaney,*⊥ and Robert Langridge⊥

Department of Chemistry, Pomona College, Claremont, California 91711 and Department of Pharmaceutical Chemistry, School of Pharmacy, University of California at San Francisco, San Francisco, California 94143. *Received October 26, 1981*

The inhibition constants ($K_{i\,app}$) obtained from the action of 44 2,4-diamino-5-(substituted-benzyl)pyrimidines on dihydrofolate reductase (DHFR) from *Escherichia coli* and *Lactobacillus casei* bacteria are used to derive quantitative structure-activity relationships (QSAR). These equations bring out a number of differences in the DHFR which can be understood at the atomic level by studying color stereo computer graphics models constructed from the X-ray coordinates of the enzyme-inhibitor complexes. The combination of QSAR and X-ray crystallography interpreted via high-performance computer graphics offers a new level of sophistication to extend our understanding of enzyme-ligand interactions, which, when the crystallography is known, opens up a more scientific approach to drug development.

We describe our preliminary efforts to combine multiple regression analysis, X-ray crystallography, and computer graphics with the biochemistry of dihydrofolate reductase (DHFR) to develop what we believe will be an exceedingly powerful tool for drug design.

The problem of ligand-macromolecule interaction has been studied extensively in recent years using a variety of physicochemical approaches. X-ray crystallography has become increasingly powerful in the elucidation of macromolecular structures and is used not only to establish the structure of the macromolecule but also to include bound ligands of intrinsic biochemical importance. The recent technique of X-ray cryoenzymology[1] has resulted in the first direct observation of a true enzyme-substrate intermediate rather than an enzyme-inhibitor complex. Although these studies greatly enhance our knowledge of the interaction of ligand and macromolecule at the atomic level, at present they provide only a static view of the dynamic process of interaction.

Another major approach to the problem of union of ligand and macromolecule is to study the binding of well-defined molecular probes (substrates or inhibitors) to macromolecular receptors. It is only in the last 2 decades that large computers have allowed thorough multivariate analyses of these studies in terms of the physicochemical properties of the probes. Although this approach does not provide the atomic resolution which is possible with X-ray crystallography, it permits the study of the dynamic ligand-macromolecule interaction under "normal operating conditions" in solution. In fact, it is now possible to make inferences about the nature of ligand-macromolecule reactions in the living cell[2,3] or whole animal[4,5] in addition to the purified macromolecule in vitro. This approach to the study of macromolecular systems is often referred to as quantitative structure-activity relationships (QSAR) and has recently been reviewed by Martin.[6]

Interactive color three-dimensional computer graphics has recently made spectacular advances, allowing one to visualize and quickly grasp the exceedingly complex interactions between ligand and macromolecule.[7] These displays of macromolecule and ligand hold great promise for correlating results from X-ray crystallography and QSAR.

Since DHFR plays a crucial role in DNA synthesis, its inhibition can be used to control growth in any organism: animal, plant, insect, or microorganism. The fact that DHFR from different sources reacts quite differently and often very selectively with inhibitors makes it of unusual interest.[8,9] By studying enzyme from host and from pathogen, one can establish intrinsic therapeutic indexes before commencing expensive animal testing. It is for this reason that DHFR is of great interest in the search for better antibacterials,[10] as well as anticancer agents.[11,12] We have been making comparative studies of the inhibition of DHFR from various sources to gain a deeper understanding of this remarkable enzyme.[4,13-16] Recent X-ray crystallographic studies have established the structures of DHFR isolated from *E. coli*[17,18] and *L. casei*[19] at 2.5 Å

†This work was supported by Grants CA-11110 (CH) and RR-1081 (RL) from the National Institutes of Health.

‡Pomona College.

§Visiting Professor from Beijing Medical College, Beijing, China.

⊥University of California at San Francisco.

(1) Alber, T.; Petsko, G. A.; Tsernoglou, D. *Nature (London)* **1976**, *263*, 297.

(2) Coats, E. A.; Genther, C. S.; Dietrich, S. W.; Guo, Z. R.; Hansch, C. *J. Med. Chem.* **1981**, *24*, 1422.

(3) Selassie, C. D.; Guo, Z. R.; Hansch, C.; Khwaja, T. A.; Pentecost, S. *J. Med. Chem.* **1982**, *25*, 157.

(4) 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.

(5) Denny, W. A.; Cain, B. F.; Atwell, G. J.; Hansch, C.; Panthananickal, A.; Leo, A. *J. Med. Chem.* **1982**, *25*, 276.

(6) Martin, Y. C. "Quantitative Drug Design"; Marcel Dekker: New York, 1978.

(7) (a) Langridge, R.; Ferrin, T. E.; Kuntz, I. D.; Connolly, M. L. *Science* **1981**, *211*, 661. (b) Feldman, R. J.; Bing, D. H.; Furie, B. C.; Furie, D. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5409.

(8) Kisliuk, R. L.; Brown, G. M., Eds. "Chemistry and Biology of Pteridines"; Elsevier: Amsterdam, 1979.

(9) (a) Hitchings, G. H.; Smith, S. L. *Adv. Enzyme Regul.* **1980**, *18*, 349. (b) Burchall, J. J. *J. Infect. Dis.* **1973**, *128*, 437S.

(10) Albert, A. "Selective Toxicity", 6th ed.; Halsted Press: London, 1979; Chapter 9.

(11) Eys, J. V. *Cancer Bull.* **1981**, *33*, 40.

(12) Wang, Y. M.; Loo, T. L. *Cancer Bull.* **1981**, *33*, 49.

(13) Li, R. L.; Dietrich, S. W.; Hansch, C. *J. Med. Chem.* **1981**, *24*, 538.

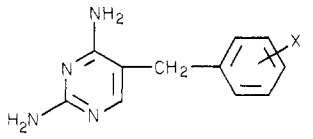
(14) Hansch, C.; Dietrich, S. W.; Fukunaga, J. Y. *J. Med. Chem.* **1981**, *24*, 544.

(15) Dietrich, S. W.; Smith, R. N.; Brendler, S.; Hansch, C. *Arch. Biochem. Biophys.* **1979**, *194*, 612.

(16) Li, R. L.; Hansch, C.; Kaufman, B. T. *J. Med. Chem.* **1982**, *25*, 435.

0022-2623/82/1825-0777$01.25/0

resolution. In the present paper we apply the techniques of QSAR, X-ray crystallography, and computer graphics to analysis of the interactions of 2,4-diamino-5-(X-benzyl)pyrimidines (I) with dihydrofolate reductase isolated from *E. coli* and *L. casei*.



I

We obtained eq 1 in our previous study of the inhibition of *E. coli* DHFR by I.[13] This equation is almost identical with an earlier one formulated with only 23 congeners of I.[20]

$$\log 1/K_{i\,app} = 1.33\ (\pm 0.24)\ MR'_{3,5} + 0.94\ (\pm 0.31)\ MR'_4 + 5.69\ (\pm 0.24) \quad (1)$$

$$n = 34;\ r = 0.904;\ s = 0.281$$

The inhibition constant (K_i) in eq 1 was determined by our reported method of assay,[20] n represents the number of data points used to derive eq 1, r is the correlation coefficient, and s is the standard deviation from regression. The figures in parentheses are for the construction of the 95% confidence limits. The MR effect has been factored into two terms—one for 3,5- and one for 4-substituents of I. The prime symbol with MR indicates an arbitrary use of this parameter. After considerable analysis it was concluded that only a fraction of a substituent in the 3-, 4-, or 5-position was capable of producing an MR-related effect. (MR, taken from our recent compilations, is scaled by 0.1 to make it more nearly equiscalar with the hydrophobic parameter π.[21,22]) This value was concluded to be 0.79 from a computerized series of trial and error calculations. Larger substituents having greater values did not appear to be more effective. It was surprising to us that we could not establish the importance of any hydrophobic terms for eq 1, especially since X-ray crystallographic analysis of the binding of trimethoprim [I, X = 3,4,5-$(OCH_3)_3$] by a group at the Wellcome Laboratories in England[18] reveals that the trimethoxybenzyl moiety is partially enclosed by the hydrophobic residues Phe-31, Ile-50, and Leu-28.

We have now made[16] and tested a number of new benzylpyrimidines in order to further evaluate the hypotheses behind eq 1. As before, we have used DHFR from MB1428 *E. coli* generously supplied by Martin Poe of the Merck Institute.

Results

We have formulated correlation equations 2–5 for the inhibition of *E. coli* DHFR by the benzylpyrimidines using the data in Table I. The single most important variable in the stepwise development of eq 2–5 is $MR'_{3,5}$; this is also true for eq 1. The parameters of eq 3 are very similar to those of eq 1, although the quality of fit is not quite as good. Since $MR'_{3,5}$ is independent of $\pi_{3,5}$ ($r^2 = 0.03$), the MR′ terms do not reflect a hydrophobic interaction. As we noted earlier,[4] substituents in these positions appear to produce their inhibitory effect sterically; that is, we believe that MR is a surrogate for molar volume and as such contains a small component of polarizability for the substituent. Adding a term in π to eq 3 does not result in a significant reduction in the variance; however, adding two terms—either as the bilinear model[23] of eq 5 or a parabolic model ($\pi + \pi^2$)—does result in an improved correlation. Using π^2 in eq 4 in place of log $(\beta \cdot 10^{\pi_{3,4,5}} + 1)$ gives almost the same quality of fit for the data. The addition of the σ^-_R term to eq 4 yields eq 5, which is a significant but not large improvement. The electronic parameter is defined as $\sigma^-_R = \sigma^- - \mathcal{F}$[22] and represents only the resonance of the substituent, and as we have used it is position independent; that is, we have used the same value for 3-X as for 4-X. The major effect of this term is to produce a good fit for the two NO_2-containing congeners in Table I. As we noted in our first study, these compounds are badly fit without the electronic correction. The electronic term in eq 5 might be an artifact, since there is not a good distribution of σ^- values among the present set of substituents.

$$\log 1/K_{i\,app} = 0.97\ (\pm 0.36)\ MR'_{3,5} + 6.15\ (\pm 0.28) \quad (2)$$

$$n = 43;\ r = 0.645;\ s = 0.489;\ F_{1,41} = 29.2$$

$$\log 1/K_{i\,app} = 1.21\ (\pm 0.32)\ MR'_{3,5} + 0.89\ (\pm 0.41)\ MR'_4 + 5.65\ (\pm 0.33) \quad (3)$$

$$n = 43;\ r = 0.781;\ s = 0.404;\ F_{1,40} = 19.9$$

$$\log 1/K_{i\,app} = 0.75\ (\pm 0.26)\ \pi_{3,4,5} + 1.36\ (\pm 0.24)\ MR'_{3,5} + 0.88\ (\pm 0.29)\ MR'_4 - 1.07\ (\pm 0.34) \log (\beta \cdot 10^{\pi_{3,4,5}} + 1) + 6.20\ (\pm 0.30) \quad (4)$$

$$n = 43;\ r = 0.903;\ s = 0.290;\ \log \beta = 0.12;\ \pi_0 = 0.25\ (\pm 1.3);\ F_{3,37} = 13.6$$

$$\log 1/K_{i\,app} = 0.43\ (\pm 0.15)\ \pi_{3,4,5} + 1.23\ (\pm 0.23)\ MR'_{3,5} + 0.80\ (\pm 0.27)\ MR'_4 - 0.88\ (\pm 0.26) \log (\beta \cdot 10^{\pi_{3,4,5}} + 1) - 0.45\ (\pm 0.28)\ \sigma^-_R + 5.81\ (\pm 0.22) \quad (5)$$

$$n = 43;\ r = 0.923;\ s = 0.263;\ \log \beta = -0.67;\ \pi_0 = 0.64\ (\pm 0.64);\ F_{1,36} = 8.88$$

One data point [3,5-$(OH)_2$] has been omitted in the development of eq 2–5. It is about 6000 times less active than expected and about 200 000 times less active than the 3,5-$(OCH_3)_2$. Fitting all data points to eq 4 yields eq 4a.

$$\log 1/K_{i\,app} = 1.13\ (\pm 0.47)\ \pi_{3,4,5} + 1.55\ (\pm 0.43)\ MR'_{3,5} + 1.11\ (\pm 0.53)\ MR'_4 - 1.47\ (\pm 0.62) \log (\beta \cdot 10^{\pi_{3,4,5}} + 1) + 6.20\ (\pm 0.56) \quad (4a)$$

$$n = 44;\ r = 0.806;\ s = 0.530;\ \log \beta = 0.19;\ \pi_0 = 0.332\ (\pm 0.531)$$

As we have increased the number of substituents from 23 in our first analysis to 34 in our second and finally to 43 in the present, we have consistently found the major

(17) Matthews, D. A.; Alden, R. A.; Freer, S. T.; Xuong, N. H.; Kraut, J. *J. Biol. Chem.* **1979**, *254*, 4144.
(18) Baker, D. J.; Beddell, C. R.; Champness, J. N.; Goodford, P. J.; Norrington, F. E. A.; Smith, D. R.; Stammer, D. K. *FEBS Lett.* **1981**, *126*, 49.
(19) Matthews, D. A.; Alden, R. A.; Bolin, J. T.; Filman, D. J.; Freer, S. T.; Hamlin, R.; Hol, W. G. J.; Kisliuk, R. L.; Pastore, E. J.; Plante, L. T.; Suong, N.; Kraut, J. *J. Biol. Chem.* **1978**, *253*, 6946.
(20) Dietrich, S. W.; Blaney, J. M.; Reynolds, M. A.; Jow, P. Y. C.; Hansch, C. *J. Med. Chem.* **1980**, *23*, 1205.
(21) Hansch, C.; Leo, A.; Unger, S. H.; Kim, K. H.; Nikaitani, D.; Lien, E. J. *J. Med. Chem.* **1973**, *16*, 1207.
(22) Hansch, C.; Leo, A. "Substituent Constants for Correlation Analysis in Chemistry and Biology"; Wiley-Interscience: New York, 1979.
(23) Kubinyi, H. *J. Med. Chem.* **1977**, *20*, 623.

Table I. Parameters Used in the Derivation of Equations 2–8 for the Inhibition of DHFR by Congeners

		log $1/K_{i\,app}$											
		E. coli			*L. casei*								
no.	X	obsd	calcd[a]	|Δ|	obsd	calcd[b]	|Δ|	$MR'_{3,5}$	$MR'_{3,4}$	MR'_{4}	$\pi_{3,4,5}$	$\pi_{3,4}$	σ^{-}_{R}
1	3,5-$(OH)_2$[c]	3.04 ± 0.03	6.00	2.96	3.38 ± 0.05	5.48	2.10	0.54	0.38	0.10	−1.34	−0.67	−0.90
2	4-$O(CH_2)_6CH_3$	5.60 ± 0.04	6.05	0.45	5.36 ± 0.04	5.51	0.15	0.21	0.89	0.79	3.17	3.17	−0.42
3	4-$O(CH_2)_5CH_3$	6.07 ± 0.04	6.22	0.15	5.73 ± 0.04	5.81	0.08	0.21	0.89	0.79	2.63	2.63	−0.42
4	H	6.18 ± 0.05	6.19	0.01	5.20 ± 0.03	5.50	0.30	0.21	0.21	0.10	0.0	0.0	0.0
5	4-NO_2	6.20 ± 0.06	6.69	0.49	6.00 ± 0.03	6.03	0.03	0.21	0.84	0.74	−0.28	−0.28	0.57
6	3-F	6.23 ± 0.03	6.19	0.04	5.38 ± 0.02	5.52	0.14	0.20	0.20	0.10	0.14	0.14	−0.38
7	3-$O(CH_2)_7CH_3$	6.25 ± 0.04	6.19	0.05	5.30 ± 0.04	5.20	0.10	0.89	0.89	0.10	3.71	3.71	−0.42
8	3-CH_2OH	6.28 ± 0.03	6.59	0.31	5.67 ± 0.04	5.79	0.12	0.82	0.82	0.10	−1.03	−1.03	0.08
9	4-NH_2	6.30 ± 0.01	6.01	0.29	5.47 ± 0.02	5.55	0.08	0.21	0.64	0.54	−1.23	−1.23	−0.17
10	3,5-$(CH_2OH)_2$	6.31 ± 0.03	6.71	0.40	5.73 ± 0.04	5.79	0.06	1.44	0.82	0.10	−2.06	−1.03	0.16
11	4-F	6.35 ± 0.03	6.19	0.16	5.67 ± 0.04	5.52	0.15	0.21	0.20	0.09	0.14	0.14	−0.38
12	3-$O(CH_2)_6CH_3$	6.39 ± 0.05	6.36	0.03	5.60 ± 0.03	5.51	0.09	0.89	0.89	0.10	3.17	3.17	−0.42
13	4-$OCH_2CH_2OCH_3$	6.40 ± 0.05	6.69	0.29	6.05 ± 0.03	6.04	0.01	0.21	0.79	0.60	−0.40	−0.40	−0.42
14	4-Cl	6.45 ± 0.01	6.60	0.15	6.19 ± 0.04	6.12	0.07	0.21	0.70	0.60	0.71	0.71	−0.14
15	3,4-$(OH)_2$	6.46 ± 0.07	5.95	0.51	5.84 ± 0.03	5.46	0.38	0.39	0.58	0.28	−1.34	−1.34	−0.90
16	3-OH	6.47 ± 0.03	6.21	0.26	5.82 ± 0.03	5.48	0.34	0.39	0.39	0.10	−0.67	−0.67	−0.45
17	4-CH_3	6.48 ± 0.02	6.60	0.12	5.83 ± 0.03	6.07	0.24	0.21	0.67	0.57	0.56	0.56	−0.11
18	3-$OCH_2CH_2OCH_3$	6.53 ± 0.05	7.01	0.48	6.12 ± 0.03	6.04	0.08	0.89	0.89	0.10	−0.40	−0.40	−0.42
19	3-$CH_2O(CH_2)_3CH_3$[c]	6.55 ± 0.03	7.06	0.51	5.49 ± 0.02	6.32	0.83	0.89	0.89	0.10	0.84	0.84	0.01
20	3-OCH_2CONH_2	6.57 ± 0.04	6.46	0.11	5.96 ± 0.03	5.75	0.21	0.89	0.89	0.10	−1.37	−1.37	−0.42
21	4-OCF_3	6.57 ± 0.01	6.70	0.13	6.30 ± 0.02	6.33	0.03	0.21	0.89	0.79	1.04	1.04	−0.11
22	3-CH_2OCH_3	6.59 ± 0.03	6.83	0.24	5.64 ± 0.04	5.93	0.29	0.89	0.89	0.10	−0.78	−0.78	0.01
23	3-Cl	6.65 ± 0.00	6.83	0.18	5.90 ± 0.02	6.12	0.22	0.70	0.70	0.10	0.71	0.71	−0.14
24	3-CH_3	6.70 ± 0.02	6.81	0.11	5.78 ± 0.03	6.07	0.29	0.67	0.67	0.10	0.56	0.56	−0.11
25	4-$N(CH_3)_2$	6.78 ± 0.03	6.81	0.03	6.17 ± 0.02	6.20	0.03	0.21	0.89	0.79	0.18	0.18	−0.22
26	4-Br	6.82 ± 0.01	6.74	0.08	6.21 ± 0.03	6.32	0.11	0.21	0.89	0.79	0.86	0.86	−0.16
27	4-OCH_3	6.82 ± 0.02	6.79	0.08	6.25 ± 0.04	6.15	0.10	0.21	0.89	0.79	−0.02	−0.02	−0.42
28	3-$O(CH_2)_3CH_3$	6.82 ± 0.03	6.87	0.05	6.13 ± 0.03	6.27	0.14	0.89	0.89	0.10	1.55	1.55	−0.42
29	3-$O(CH_2)_5CH_3$	6.86 ± 0.03	6.54	0.38	5.77 ± 0.04	5.81	0.04	0.89	0.89	0.10	2.63	2.63	−0.42
30	4-$O(CH_2)_3CH_3$	6.89 ± 0.03	6.55	0.34	6.37 ± 0.03	6.27	0.10	0.21	0.89	0.79	1.55	1.55	−0.42
31	4-$NHCOCH_3$	6.89 ± 0.00	6.40	0.49	6.05 ± 0.02	5.87	0.18	0.21	0.89	0.79	−0.97	−0.97	−0.26[d]
32	3-OSO_2CH_3	6.92 ± 0.03	6.77	0.15	5.92 ± 0.03	5.90	0.02	0.89	0.89	0.10	−0.88	−0.88	−0.26[d]
33	3-OCH_3	6.93 ± 0.02	7.11	0.18	5.93 ± 0.02	6.15	0.22	0.89	0.89	0.10	−0.02	−0.02	−0.42
34	3-Br	6.96 ± 0.03	7.06	0.10	6.23 ± 0.03	6.32	0.09	0.89	0.89	0.10	0.86	0.86	−0.16
35	3-NO_2, 4-$NHCOCH_3$	6.97 ± 0.02	7.08	0.11	6.00 ± 0.03	6.39	0.39	0.84	1.53	0.79	−1.25	−1.25	0.31
36	3-$OCH_2C_6H_5$	6.99 ± 0.05	6.84	0.15	6.15 ± 0.03	6.24	0.09	0.89	0.89	0.10	1.66	1.66	−0.42
37	3-CF_3	7.02 ± 0.01	6.67	0.35	6.16 ± 0.02	6.05	0.11	0.61	0.60	0.10	0.88	0.88	0.27
38	3,4-$(OCH_2CH_2OCH_3)_2$	7.22 ± 0.04	7.43	0.21	6.51 ± 0.02	6.57	0.06	0.89	1.58	0.79	−0.80	−0.80	−0.84
39	3-I	7.23 ± 0.04	6.99	0.23	6.67 ± 0.02	6.33	0.24	0.89	0.89	0.10	1.12	1.12	−0.10
40	3-CF_3, 4-OCH_3	7.69 ± 0.08	7.28	0.40	7.30 ± 0.05	6.70	0.60	0.61	1.29	0.79	0.86	0.86	−0.15
41	3,4-$(OCH_3)_2$	7.72 ± 0.07	7.73	0.01	6.92 ± 0.02	6.83	0.09	0.89	1.58	0.79	0.08	0.08	−0.84
42	3,5-$(OCH_3)_2$, 4-$O(CH_2)_2OCH_3$	8.35 ± 0.08	8.38	0.03	6.26 ± 0.02	6.59	0.33	1.58	1.58	0.79	−0.78	0.76	−0.84
43	3,5-$(OCH_3)_2$	8.38 ± 0.08	8.07	0.31	6.42 ± 0.03	6.16	0.26	1.58	0.89	0.10	0.08	0.04	−0.84
44	3,4,5-$(OCH_3)_3$	8.87 ± 0.05	8.47	0.40	6.88 ± 0.04	6.64	0.24	1.58	1.58	0.79	−0.60	−0.58	−0.84

[a] Calculated using eq 4. [b] Calculated using eq 7. [c] These points not used in the formulation of eq 4 or eq 7, as the case may be. [d] Estimated value.

factor relating to inhibition to be MR′.

We have developed eq 6–8 from the data in Table I for

$$\log 1/K_{i\,app} = 0.84\ (\pm 0.32)\ MR'_{3,4} + 5.29\ (\pm 0.30) \quad (6)$$

$$n = 42;\ r = 0.643;\ s = 0.340;\ F_{1,40} = 26.1$$

$$\log 1/K_{i\,app} = 0.31\ (\pm 0.11)\ \pi_{3,4} + 0.95\ (\pm 0.21)\ MR'_{3,4} - 0.88\ (\pm 0.24)\ \log(\beta \cdot 10^{\pi_{3,4}} + 1) + 5.32\ (\pm 0.20) \quad (7)$$

$$n = 42;\ r = 0.876;\ s = 0.222;\ \log \beta = -1.33;\ \pi_0 = 1.05\ (0.77\text{–}1.33);\ F_{3,37} = 7.10$$

$$\log 1/K_{i\,app} = 0.31\ (\pm 0.10)\ \pi_{3,4} + 0.89\ (\pm 0.21)\ MR'_{3,4} - 0.91\ (\pm 0.23) \log(\beta \cdot 10^{\pi_{3,4}} + 1) - 0.22\ (\pm 0.22)\ \sigma^{-}_{R} + 5.31\ (\pm 0.19) \quad (8)$$

$$n = 42;\ r = 0.889;\ s = 0.214;\ \log \beta = -1.34;\ \pi_0 = 1.05\ (0.80\text{–}1.30);\ F_{1,36} = 3.95$$

the inhibition of *L. casei* DHFR for comparison with eq 2–5 for the inhibition of *E. coli* DHFR. In eq 6–8, 5-substituents have not been included in the MR and π terms as in eq 2–5. This results in an improved correlation even though there are only five such examples in Table I. Two data points have been dropped in forming eq 6–8: 3,5-$(OH)_2$ and 3-$CH_2O(CH_2)_3CH_3$. The former is about 100 times less active than expected, whereas the latter is 5 times less active than predicted. Including all data points and fitting the data to eq 7 yields eq 7a.

$$\log 1/K_{i\,app} = 1.16\ (\pm 0.38)\ MR'_{3,4} + 0.34\ (\pm 0.19)\ \pi_{3,4} - 0.91\ (\pm 0.43)\ \log(\beta \cdot 10^{\pi_{3,4}} + 1) + 5.07\ (\pm 0.35) \quad (7a)$$

$$n = 44;\ r = 0.753;\ s = 0.407;\ \log \beta = -1.33;\ \pi_0 = 1.11\ (\pm 0.326)$$

A salient feature of difference between these two bacterial enzymes is the greater sensitivity of *E. coli* to inhibitors I. The range in log $1/K_i$ for the 44 congeners of Table I is 5.83 for *E. coli* and 3.92 for *L. casei*. Both enzymes have about the same lower limit in $1/K_i$; it is the difference in the upper limit set by compounds with 3,5-$(OCH_3)_2$ groups that accounts for the greater range.

Trimethoprim [3,4,5-$(OCH_3)_3$] is the most inhibitory compound with the *E. coli* reductase, while the 3-CF_3-4-OCH_3 analogue is the most potent congener with the *L. casei*. It is fascinating to find that while the 3,5-$(OH)_2$ is invariably poorly correlated by all equations, the 3,4-$(OH)_2$ behaves reasonably in both sets of equations.

A more direct comparison of the response of the two enzymes to the set of inhibitors is given by eq 9. The eight

$$\log 1/K_{i(E.\,coli)} = 0.75\ (\pm 0.19)\ \log 1/K_{i(L.\,casei)} + 2.18\ (\pm 1.2) \quad (9)$$

$$n = 36;\ r = 0.805;\ s = 0.237;\ F_{1,33} = 6.27$$

poorest-fit data points were omitted in deriving eq 9 (see Table II). This procedure shows up the most widely different features of the structure–activity relationship of the two enzymes. Except for 3,5-$(CH_2OH)_2$, all 3,5-$(X)_2$ analogues are poorly correlated by eq 9. Three compounds containing the $OCH_2CH_2OCH_3$ moiety are poorly matched; one of these, the clinically useful drug tetroxoprim, also contains the 3,5-$(OCH_3)_2$ as well. Two large substituents—4-$O(CH_2)_6CH_3$ and 3-$O(CH_2)_6CH_3$—are also mavericks. The aberrance of the 3-alkoxy congener seems odd in view of the fact that the 3-$O(CH_2)_7CH_3$ analogue is well fit. Actually, eq 9 is not a very sharp correlation, which is not surprising considering that only 43 of the amino acid residues in the two enzymes are identically conserved.

Both enzymes have a major common feature and that is that MR′ accounts for most of the variance in K_i. This confirms our earlier finding of eq 1 for *E. coli*. Adding eight new quite varied analogues to produce eq 3 gives an equation with essentially the same parameters as eq 1. However, with more strongly hydrophobic substituents we can now establish in eq 4 and 5 the presence of a hydrophobic effect. This seems independent of MR′ as seen by the fact that the coefficients with MR′ and the intercepts of eq 4 and 5 do not differ significantly from those of eq 1 and 2.

MR comes from the Lorentz–Lorenz equation (MR = $n^2 - 1/n^2 + 2 \cdot M_w/d$) where n is the refractive index of a substance, d its density, and M_w the molecular weight. It is an additive-constitutive property, and since n varies little among organic compounds, it is mainly a measure of volume (M_w/d). It is, of course, also a measure of the polarizability: MR = $4\pi N\alpha/3$, where N is Avogadro's number and α is the polarizability and thus related to London dispersion forces. The determination of MR has been reviewed by Bauer et al.[24] At this point it is difficult to see exactly how volume and/or dispersion forces are involved in the inhibition process.

Molecular Modeling

An important advance in the interpretation of structures obtained via X-ray crystallography is the development of computer graphics. Three-dimensional structural models, complete with their attendant Van der Waals molecular surfaces, can be color coded and manipulated in real time and stereo in order to fully visualize and optimize intermolecular contacts.

Models of the interaction of 2,4-diamino-5-(X-benzyl)-pyrimidines with *E. coli* DHFR were model-built based on the structure determined by the group at Wellcome Laboratories in England for the trimethoprim (TMP)–*E. coli* DHFR complex. Binding of these inhibitors to *L. casei* DHFR was based on the structure for the methotrexate (MTX)–NADPH–*L. casei* DHFR complex determined by Matthews et al.[19] by (1) superimposing the 2,4-diaminopyrimidine ring of the inhibitor onto the corresponding atoms of MTX or (2) aligning the *L. casei* DHFR structure with the *E. coli* DHFR–TMP structure (as described by Matthews et al.[19]) to estimate the location of bound TMP in *L. casei*; either method resulted in essentially the same location for the 2,4-diaminopyrimidine moiety.

For analogues other than TMP, the benzyl substituents were constructed using standard bond lengths and angles and then fit into the molecular surface of the DHFR active site by adjustment of their torsional angles using the molecular modeling programs CHEM (written by Andrew Dearing) and MS (written by Michael Connolly[25]). CHEM allows the user to change the location of different molecules relative to each other, adjust torsional angles, and monitor interatomic distances while displaying both the skeletal and molecular surface (as calculated by MS) models in color and stereo on an Evans and Sutherland Picture System 2.

Figure 1 shows MTX (green) and NADPH (yellow) bound to *L. casei* DHFR (blue). The red dots outline the molecular surface of the active site with part of the surface cut away to show the bound MTX inhibitor. The surface is formed by "rolling" a 1.4-Å radius probe sphere (an idealized water molecule) over the active-site residues and

(24) Bauer, N.; Fajans, K.; Lewin, S. Z. In "Techniques of Organic Chemistry", 3rd ed.; Weissberger, A., Ed., Interscience: New York, 1960; Vol. I, Part II, p 1139.

(25) Connolly, M., submitted to *Acta Crystallogr.*

Table II. Data Used for the Derivation of Equation 9

		log $1/K_{i\,app}$			
		L. casei	*E. coli*		
no.	X	obsd	obsd	calcd	$\lvert\Delta\rvert$
1	3,5-$(OH)_2$ [a]	3.38	3.04	4.72	1.68
2	4-$O(CH_2)_6CH_3$ [a]	5.36	5.60	6.20	0.60
3	3-$O(CH_2)_7CH_3$	5.30	6.25	6.16	0.09
4	3,5-$(CH_2OH)_2$	5.73	6.31	6.48	0.17
5	3,4-$(OH)_2$	5.84	6.46	6.56	0.10
6	3,5-$(OCH_3)_2$, 4-$OCH_2CH_2OCH_3$ [a]	6.26	8.35	6.88	1.47
7	3-OH	5.82	6.47	6.55	0.08
8	4-NH_2	5.47	6.30	6.28	0.02
9	3,4-$(OCH_2CH_2OCH_3)_2$ [a]	6.51	6.44	7.06	0.62
10	3-CH_2OH	5.67	6.28	6.43	0.15
11	4-$N(CH_3)_2$	6.17	6.78	6.81	0.03
12	4-CH_3	5.83	6.48	6.55	0.07
13	3-OCH_2CONH_2	5.96	6.19	6.65	0.46
14	4-$OCH_2CH_2OCH_3$ [a]	6.05	6.02	6.72	0.70
15	4-OCH_3	6.25	6.82	6.87	0.05
16	4-Br	6.21	6.82	6.84	0.02
17	4-OCF_3	6.30	6.57	6.91	0.34
18	4-NO_2	6.00	6.20	6.68	0.48
19	3-CH_2OCH_3	5.64	6.59	6.41	0.18
20	4-F	5.67	6.35	6.43	0.08
21	3,5-$(OCH_3)_2$ [a]	6.42	8.38	7.00	1.38
22	3,4,5-$(OCH_3)_3$ [a]	6.88	8.87	7.34	1.53
23	4-$O(CH_2)_5CH_3$	5.73	6.07	6.48	0.41
24	3-NO_2, 4-$NHCOCH_3$	6.00	6.97	6.68	0.29
25	3-$O(CH_2)_6CH_3$ [a]	5.60	5.14	6.38	1.24
26	3-OSO_2CH_3	5.92	6.92	6.62	0.30
27	3-OCH_2CH_2OCH	6.12	6.37	6.77	0.40
28	4-Cl	6.19	6.45	6.82	0.37
29	3-OCH_3	5.93	6.93	6.63	0.30
30	3,4-$(OCH_3)_2$	6.92	7.72	7.37	0.35
31	H	5.20	6.18	6.08	0.10
32	3-F	5.38	6.23	6.22	0.01
33	3-CH_3	5.78	6.70	6.52	0.18
34	4-$O(CH_2)_3CH_3$	6.37	6.89	6.96	0.07
35	3-CF_3	6.16	7.02	6.80	0.22
36	3-Cl	5.90	6.65	6.61	0.04
37	3-Br	6.23	6.96	6.85	0.11
38	4-$NHCOCH_3$	6.05	6.89	6.72	0.17
39	3-$CH_2O(CH_3)_3CH_3$	5.49	6.55	6.30	0.25
40	3-I	6.67	7.23	7.18	0.05
41	3-CF_3, 4-OCH_3	7.30	7.69	7.66	0.03
42	3-$O(CH_2)_5CH_3$	5.77	6.87	6.51	0.36
43	3-$O(CH_2)_3CH_3$	6.13	6.82	6.78	0.04
44	3-$OCH_2C_6H_5$	6.15	6.99	6.79	0.20

[a] These points were not used in deriving eq 9.

placing a dot at each point of tangency of the probe sphere with the Van der Waals surface of the protein. This provides a feeling for what the entire DHFR molecule looks like. In particular, it is impressive to see how large a fraction of the enzyme is taken up by the active site.

Due to the different QSAR for benzylpyrimidine inhibition of the two enzymes, we expected to see significant differences in the active-site structures of *L. casei* and *E. coli* DHFR. The active-site region of the *E. coli* DHFR–TMP complex (red) is shown superimposed onto the model-built *L. casei* DHFR–TMP complex (blue) in Figure 2. NADPH (yellow) is also visible with the nicotinamide moiety near the benzyl group of TMP. Phe-30 and -31 provide the hydrophobic floor and Leu-27 and -28 the hydrophobic walls for the active sites of each enzyme. The major difference in the two active sites is the replacement of Leu-19 and Phe-49 in *L. casei* DHFR by the corresponding Met-20 and Ile-50 in *E. coli* DHFR, resulting in a smaller and more constrained active site in the *L. casei* enzyme than in the *E. coli* enzyme. The side chain of Met-20 in the *E. coli* DHFR–MTX binary complex points away from the active site, but in the ternary *L. casei* DHFR–MTX–NADPH complex the corresponding Leu-19 side chain fills up a substantial portion of the active site. It is not clear whether this is an intrinsic structural difference between the two bacterial enzymes or is due to a conformational change in the *L. casei* DHFR induced by cofactor (NADPH) binding (the structure of the ternary complex for *E. coli* DHFR–MTX–NADPH has not been determined). Preliminary modeling suggests that such a conformational change is not necessary for *E. coli* DHFR to accommodate NADPH; so we make the assumption that these structural differences are indeed due to an intrinsic difference between the two enzymes. A comparison of the molecular surfaces of the active sites is shown in Figure 3. The *E. coli* DHFR–TMP complex is on the right and the corresponding *L. casei* structure on the left. The larger active site volume of the *E. coli* enzyme is immediately apparent, but note that trimethoprim appears to fit well *sterically* in both enzymes.

The real difference between eq 4 and 7 is that neither a π_5 or MR_5 term appears in eq 7, indicating that while a 5-substituent interacts favorably with *E. coli* DHFR, the same substituent has essentially no effect on *L. casei* DHFR. The 3- and 4-substituents behave similarly with each enzyme. Why do 5-substituents interact so differently

with the two enzymes? The Leu-19 of *L. casei* DHFR intrudes into the active site near the 5-OCH_3 (methoxy groups are shown in positions 3–5 in Figure 3; the upper OCH_3 is taken as 5-OCH_3) of trimethoprim, resulting in a tighter steric fit of trimethoprim to the *L. casei* enzyme than the *E. coli* by burying the 5-OCH_3 group (Figure 3). However, burying the 5-OCH_3 group in this hydrophobic pocket of the enzyme must be accompanied by desolvation of the oxygen—there is no room in the complex for water molecules to remain hydrogen bonded to the oxygen lone pair (Figure 4). In Figures 4 and 5, the oxygen surface of the 5-OCH_3 is colored red. It is clear from fragment constants for partitioning between octanol and water and from the data of Wolfenden et al.[26] for partitioning between the gas phase and water that such desolvation is energetically unfavorable and will increase the free energy of binding to the enzyme. Thus, the tight steric fit of the 5-OCH_3 group into *L. casei* DHFR is offset by the desolvation of the oxygen, so that the net effect of the 5-substituent is negligible. The situation with *E. coli* DHFR is quite different, since the oxygen of the 5-OCH_3 group can remain solvated when bound to the "looser" active site of *E. coli* DHFR (Figure 5) while placing the CH_3 into a hydrophobic pocket, resulting in a favorable interaction of this substituent with the enzyme due to Van der Waals and "hydrophobic" forces. This model predicts that a 5-CH_2CH_3 substituent should result in increased activity with *L. casei* DHFR, since the full effect of the tight steric fit and Van der Waals interactions can be realized without having to pay a penalty for desolvation. A 5-CH_2CH_3 group should have essentially the same effect as the 5-OCH_3 group on *E. coli* DHFR.

The 3-CH_2OH analogue behaves differently with each enzyme (Table II); the 3-CH_2OH group has little effect on *E. coli* DHFR and contributes only an extra 0.5 log unit to log $1/K_i$ for *L. casei*. In both enzymes, binding of the hydroxyl group of the 3-substituent in the same orientation as shown for trimethoprim in Figures 3–5 is unlikely, since this would force the very hydrophilic hydroxyl group into a hydrophobic portion of the active site. Instead, rotation about the phenyl–CH_2 bond probably occurs to rotate this group *away* from the active site so that the hydroxyl group remains solvated, resulting in little contact of the substituent with the enzyme and an activity very close to that of the unsubstituted (3-H) analogue. The slight positive effect of the 3-CH_2OH with *L. casei* DHFR is due to the tighter active site in the region of Leu-19, which can form a Van der Waals contact with the CH_2 group while the OH is in contact with the solvent. Addition of a second CH_2OH group to the 5-position has essentially no effect on either enzyme due to a similar rotation of this group to place its hydroxyl group in solvent.

The activity of the 3,5-$(OH)_2$ congener is greatly overpredicted by both eq 4 and 7. Although a single hydroxyl group (3-OH) is well fit by eq 4 and 7 and apparently tolerated by both enzymes, the addition of a second hydroxyl group reduces the log $1/K_i$ by 2.4 log units for *L. casei* and 3.4 log units for *E. coli*. A single phenolic hydroxyl group can remain solvated upon binding by a slight rotation of the phenyl ring, but there is no way to prevent partially burying the second hydroxyl group of the 3,5-$(OH)_2$ analogue in a hydrophobic region, since as the ring rotates to place the 3-OH into the solvent, the 5-OH is pushed deeper into the enzyme. The 3,5-$(OH)_2$ groups lack the rotatable bond which is present in the 3,5-$(CH_2OH)_2$ groups and are therefore unable to "escape" the hydrophobic surface of the enzyme as easily. Forced desolvation of the very hydrophilic phenolic hydroxyl group results in an increase in the free energy of binding by ~3–4 kcal/mol (assuming K_i can be treated as an equilibrium constant), approximately the energy of a hydrogen bond. Addition of the second OH group at the 4-position is tolerated with a negligible effect on log $1/K_i$, since the 4-position is more readily accessible to solvent; rotation of the ring does not move the 4-substituent as the 3-OH moves toward solvent.

Why do we observe correlation primarily with MR rather than π, since the active sites of both enzymes contain large amounts of hydrophobic surface? Recent QSAR results on binding of phenyl hippurates to papain[27] show strong correlation with π for a variety of substituents. Molecular modeling reveals that these substituents are almost totally buried (desolvated) within a hydrophobic region of this enzyme, whereas with the bacterial DHFR we see a smaller portion of each substituent buried; much more of the substituent remains solvated. These preliminary results suggest that correlation with π occurs when desolvation (as modeled by partitioning from water into octanol) is nearly complete upon binding to a hydrophobic surface.

Why does there appear to be a fixed limit on MR (MR′) for inhibition of bacterial DHFR? The QSAR indicates that there is essentially no additional interaction for substituents larger than OCH_3 at the 3-, 4-, or 5-position; only the first two atoms of the substituent appear to contact the enzyme, with additional atoms having a negligible effect on activity. Although the active sites are reasonably "tight" in the immediate vicinity of the trimethoxyphenyl portion of trimethoprim, they become wider and looser near the entrance. That portion of substituents beyond the first two atoms will be located in this "loose" region where there is no opportunity for tight contact with the active site; the larger substituents tend to remain in solvent and have little interaction with the enzyme.

Conclusions

We have used QSAR together with the X-ray crystallographically determined structure of a drug receptor to provide a detailed molecular model for the drug–receptor interaction. The use of three-dimensional real-time color computer graphics is extremely powerful in interpreting and modeling the many complex features involved in drug–receptor recognition. The ability of the computer to simplify and make tractable an intermolecular interaction which can involve several hundreds or even thousands of atoms is essential.

The model presented here for the differential inhibition of *E. coli* and *L. casei* dihydrofolate reductase by 2,4-diamino-5-(X-benzyl)pyrimidines is admittedly speculative. The real power of the method is its ability to generate experimentally testable predictions of the activity of new analogues, the ultimate goal of "rational" drug design. It is most gratifying to see that deductions from QSAR can be reinforced by the more established discipline of X-ray crystallography, without which one has such a limited view (i.e., imagination) of the topography of the binding region. However, without the aid of QSAR the subtle dynamic features of the drug–receptor interaction may be overlooked. A synergistic approach which combines the techniques of QSAR, X-ray crystallography, and high-performance computer graphics holds much promise for becoming an extremely powerful tool for drug design.

(26) Wolfenden, R.; Anderson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849.

(27) Smith, R. N.; Hansch, C.; Kim, K. H.; Omiya, B.; Fukumura, G.; Selassie, C. D.; Jow, P. Y. C.; Blaney, J. M.; Langridge, R. *Arch. Biochem. Biophys.* **1982**, *215*, 319.

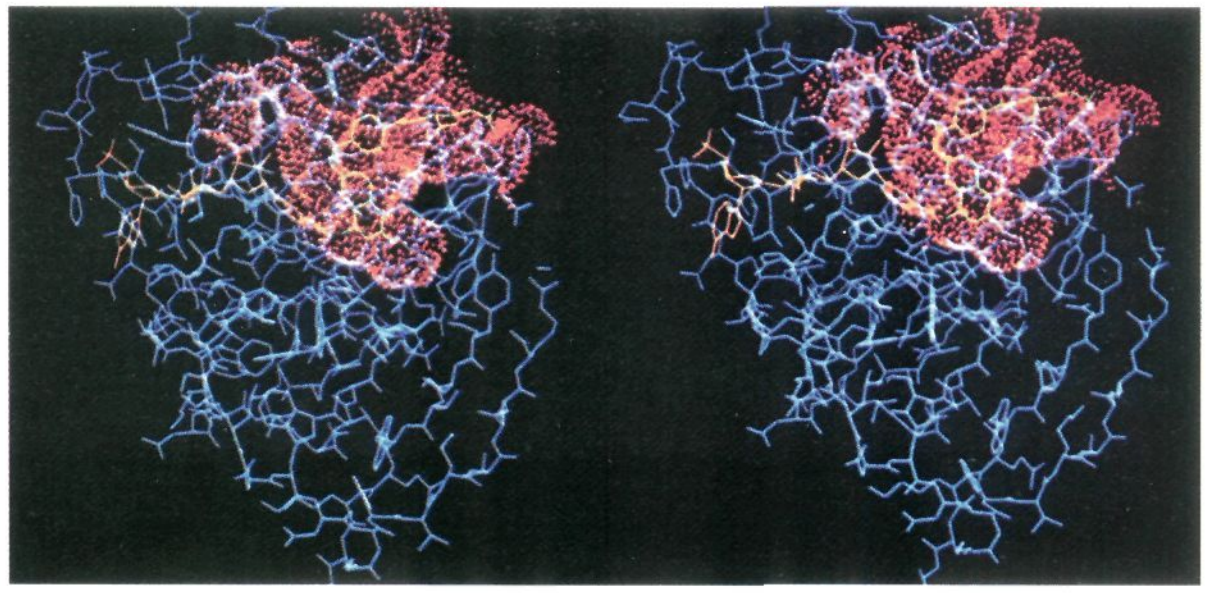


Figure 1. Van der Waals molecular surface of *L. casei* DHFR active site (red) with bound MTX inhibitor (green), NADPH (yellow), and the rest of the DHFR molecule (blue).

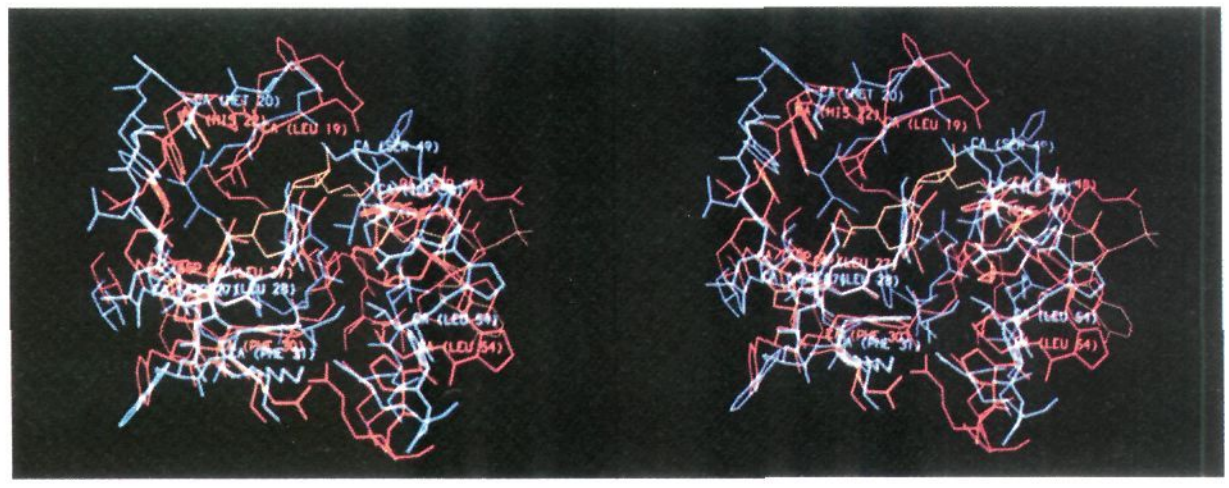


Figure 2. Active site of the binary *E. coli* DHFR–TMP complex (red) superimposed on the active site of the model-built ternary *L. casei* DHFR–TMP complex (blue) with NADPH (yellow).

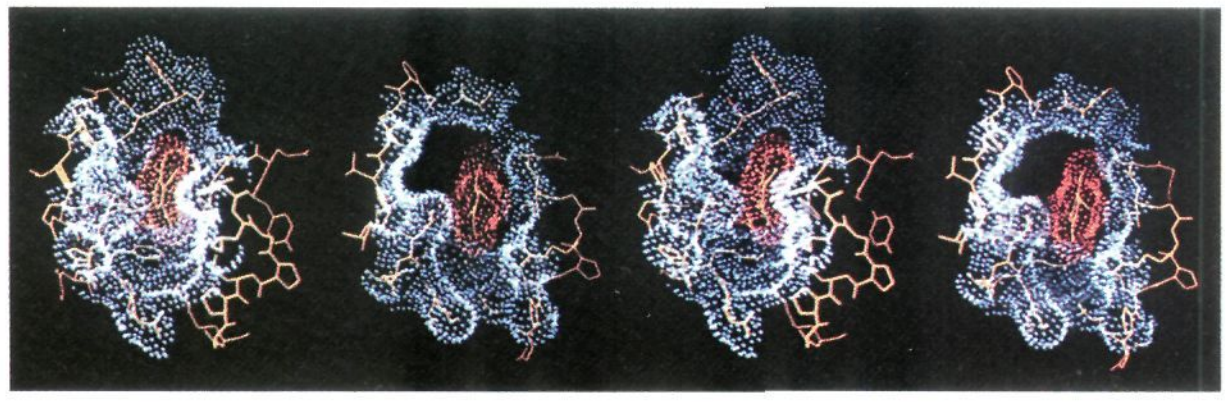


Figure 3. Molecular surfaces of TMP (red) in the active sites of *L. casei* DHFR (left) and *E. coli* DHFR (right).

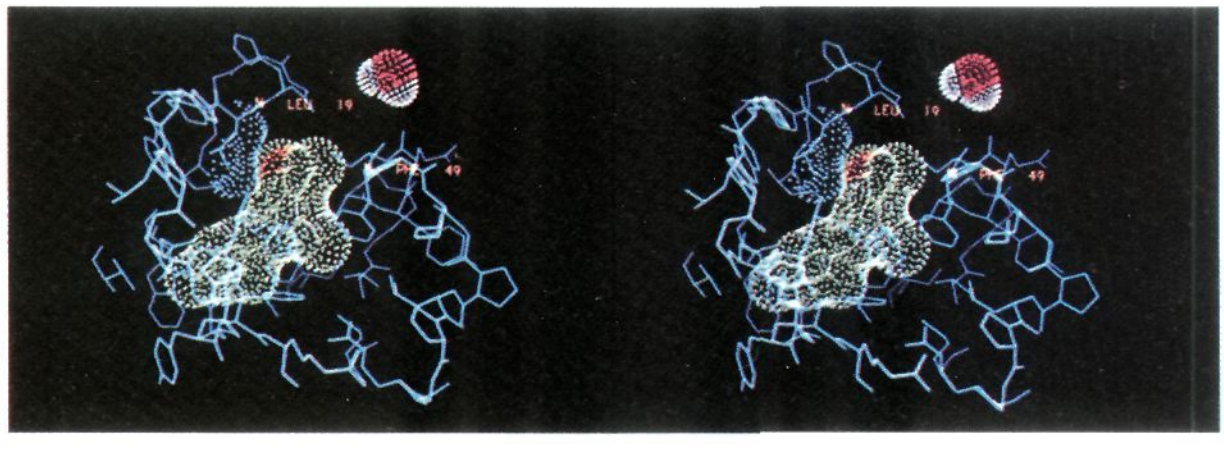


Figure 4. *L. casei* DHFR (blue)–TMP (green) model, showing how the oxygen (red) of the 5-OCH_3 group of TMP is buried by Leu-19.

Experimental Section

Enzymic Assay. Our previously described procedure for assaying inhibitors with DHFR has been used in this work.[20] The confidence limits on log $1/K_i$ have been estimated using the jackknife procedure.

Inhibitors. The syntheses of the benzylpyrimidines have been reported.[13,16,20]

QSAR Parameters. The MR values have been taken from ref 21 or calculated by procedures described therein. Other parameters come from the compilation in ref 22.

Correlation Analysis. In developing the correlation equations, we have not used stepwise regression techniques; instead, we have considered all possible equations. Our general approach to the DHFR problem is discussed in ref 28 and 29.

(28) Dietrich, S. W.; Dreyer, N. D.; Hansch, C.; Bentley, D. L. *J. Med. Chem.* **1980**, *23*, 1201.
(29) Silipo, C.; Hansch, C. *J. Am. Chem. Soc.* **1975**, *97*, 6849.

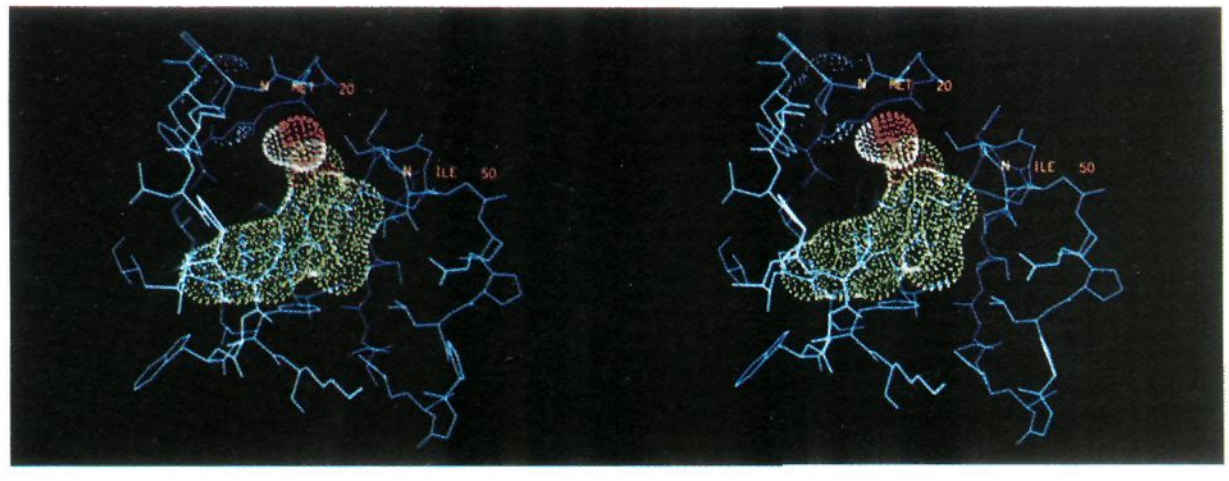


Figure 5. *E. coli* DHFR (blue)–TMP (green) model; the oxygen (red) of the 5-OCH_3 group of TMP is exposed to solvent and is shown interacting with a single water molecule (O = red, H = white).

Acknowledgment. We are much indebted to the group at the Wellcome Laboratories (ref 18) for the coordinates of *E. coli* DHFR complexed with trimethoprim and to D. A. Matthews and J. Kraut who have refined the coordinates for our use. We thank Martin Poe of the Merck Institute for the *E. coli* DHFR. The photographs were made by Paul Weiner. J.M.B. is supported in part by the American Foundation for Pharmaceutical Education.