was stirred at room temperature for 2 h and evaporated to dryness. The residue was extracted with $CHCl_3$. The organic layer was washed with H_2O, dried ($MgSO_4$), and evaporated. The residue was crystallized from EtOH to give the title styrene (1.3 g, 80%), mp 125–127 °C. Anal. ($C_{20}H_{14}Cl_2$) C, H.

3,5-Bis(4-chlorophenyl)styrene Oxide (3d). A solution of the above styrene (1.7 g) and *m*-chloroperbenzoic acid (1 g) in $CHCl_3$ (50 mL) was kept at 40 °C for 4 h and held at room temperature overnight. The solution was washed with $NaHCO_3$, dried (K_2CO_3), and evaporated to yield the crude epoxide (1.7 g) containing about 10% of starting material as shown by TLC (silica; dichloromethane–petroleum ether, 1:1). The crude product was used directly in the next step.

α-[(Di-*n*-butylamino)methyl]-3,5-bis(4-chlorophenyl)-phenylcarbinol Hydrochloride (3e). The above crude epoxide (1.1 g) and di-*n*-butylamine (3 mL) in EtOH (10 mL) were refluxed 3 h. The solution was evaporated to dryness and the mixture was dissolved in 10% HCl–MeOH. The solution was evaporated to dryness again and the residue was suspended in Et_2O. The colorless precipitate was filtered, washed with H_2O, and recrystallized from acetone–petroleum ether to give the target compound (800 mg, 51%), mp 231–233 °C. Anal. ($C_{28}H_{34}Cl_3NO$) C, H, Cl, N.

α-[(*n*-Butylamino)methyl]-3,5-bis(4-chlorophenyl)-phenylcarbinol Hydrochloride (3f). A solution of **3d** (700 mg) and *n*-butylamine (3 mL) in EtOH (20 mL) was refluxed for 4 h. The mixture was processed as for **3e** to give the target compound (450 mg, 48%), mp 248–250 °C. Anal. ($C_{24}H_{26}Cl_3NO$) C, H, Cl, N.

Acknowledgment. This work was supported by the U.S. Army Medical Research and Development Command under Contract DADA17-69-C-9065. This is contribution no. 1562 from the Army Research Program on Antiparasitic Drugs. The advice and timely suggestions of Dr. R. E. Strube of the Walter Reed Army Institute of Research are gratefully acknowledged.

Confidence Interval Estimators for Parameters Associated with Quantitative Structure–Activity Relationships

Stephen W. Dietrich, Nicholas D. Dreyer, Corwin Hansch,*

Department of Chemistry

and Donald L. Bentley

Department of Mathematics, Pomona College, Claremont, California 91711. *Received January 14, 1980*

Use of the "jackknife" as a statistical tool for construction of both point and confidence interval estimators for parameters which are functions of regression coefficients and/or iteratively estimated parameters is described. Examples are presented demonstrating its utility in constructing estimators for log P_0 and π_0 using the parabolic and bilinear quantitative structure–activity relationship (QSAR) models relating nonlinear dependence of activity on hydrophobicity and for log $[1/K_{i(app)}]$ assuming competitive enzyme inhibition.

Medicinal chemists are increasingly attempting to come into line with Lord Kelvin's famous dictum: "If you cannot measure, your knowledge is meager and unsatisfactory". Quantifying results accurately in medicinal chemistry is considerably more difficult than in the field of physics. The biological data obtained from animal experiments, or even in vitro experiments with purified enzymes, contain considerable noise; therefore, simply expressing results in numerical terms is not enough. One must have some idea of how reliable the numbers are that one obtains. This is particularly important in the area of QSAR. In the formulation of these mathematical models, one wants to know where shortcomings in the model are most likely to reside. It is necessary, therefore, to be aware not only of error (experimental variance) in the observed biological data but also of the error asssociated with predicted values derived from the mathematical model using these data.

Two parameters of importance in QSAR work are log P_0 (or π_0) for nonlinear dependence of activity on hydrophobicity and log $1/C$ = log $[1/K_{i(app)}]$ [or log $(1/K_i)$] for enzyme inhibition. This report discusses a technique for placing confidence limits on these parameters.

The analysis of experimental data in the biological sciences frequently entails the use of linear regression analysis:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_m X_m + \text{error}$$

In many instances the researcher then finds it necessary to calculate an estimate for a parameter θ, which is a function of the coefficients of the regression equation:

$$\theta = f(\alpha_0, \alpha_1, \ldots, \alpha_m)$$

The complexity of such a relationship often makes it difficult, if not impossible, to directly provide a confidence interval associated with the calculated parameter. In addition, analysis of other experimental data involves nonlinear regression techniques which may require iterative solutions for estimates $\{B_j\}$ of parameter(s) $\{\beta_j\}$, as well as estimates $\{A_i\}$ of regression coefficients $\{\alpha_i\}$. Such regression analyses provide approximate confidence intervals for the regression coefficients $\{\alpha_i\}$ directly, assuming that the iteratively derived $\{B_j\}$ values are the true $\{\beta_j\}$. (This assumption could create problems in the confidence interval statements on the $\{\alpha_i\}$ due to inaccuracies in the estimation of the $\{\beta_j\}$. The standard statistical procedure in the literature[1-3] is to assume that the iteratively derived $\{B_j\}$ values are the true $\{\beta_j\}$ and then to calculate the confidence intervals for the $\{\alpha_i\}$ based on the least-squares linear regression using a Student's t value with $(n - K)$ degrees of freedom, where K = the total number of $\{\alpha_i\}$ and $\{\beta_j\}$.) However, such regression analyses do not provide simple confidence intervals for the iteratively derived $\{B_j\}$. Derivation of confidence intervals for a parameter θ, which is a function of both the α_i and β_j values [i.e., $\theta = g(\{\alpha_i\},\{\beta_j\})$], is therefore also difficult.

(1) Kubinyi, H. *J. Med. Chem.* **1977**, *20*, 625.
(2) Kubinyi, H. *Arzneim.-Forsch.* **1976**, *26*, 1991.
(3) Kubinyi, H.; Kehrhahn, O.-H. *Arzneim.-Forsch.* **1978**, *28*, 598.

0022-2623/80/1823-1201$01.00/0

The use of the jackknife[4-7] as a statistical tool for directly establishing the confidence intervals of parameters which are functions of the coefficients of linear or nonlinear regression analyses is presented. Actual applications of the jackknife technique to examples of published QSAR studies and to some of our own experimental enzyme inhibition data are utilized to illustrate this procedure.

Optimal Independent Variable Value Estimation for the Parabolic and Bilinear QSAR Models. QSAR regression studies in medicinal chemistry attempt to correlate the relative activities of a series of molecules with the physicochemical properties of the molecules. It is frequently found, however, that activity is nonlinearly rather than linearly related to such properties: as the value of a physicochemical property of the molecules is increased, activity increases at first, peaks, and then decreases. Such effects are most often seen for hydrophobicity, as parameterized by log P or π. The ability to set reasonable confidence limits on the value of an independent variable at which activity maximizes for a data set is valuable for the design of new compounds and for the comparison of similar optima for different data sets.

The parabolic model[8-10] relates this nonlinear dependence of activity to a parabolically shaped relationship:

$$\log 1/C = a + bX + cX^2 \qquad (1)$$

$$b > 0, c < 0$$

To determine that value X_0 where activity is maximized, one sets the first derivative with respect to X of the right-hand side of eq 1 equal to 0 and solves:

$$X_0 = -b/(2c) \qquad (2)$$

Hansch et al.[11] introduced a method for the calculation of confidence intervals for X_0 which is currently used in the literature. An important assumption in the derivation of the confidence interval for X_0 is the additivity of the error in the model; i.e., $\log 1/C_i = a + bX_i + cX_i^2 + \epsilon_i$, where ϵ_i = error.

The bilinear model of Kubinyi[1-3] relates nonlinear dependence of activity to linear ascending and descending curves (slopes not necessarily equal) and with a parabolic transition within the range of X_0:

$$\log 1/C = a + bX + c \log (\beta \times 10^X + 1) \qquad (3)$$

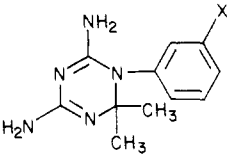
$$\beta > 0, -c > b > 0$$

The parameters of this nonlinear model are estimated by least-squares regression, iterating on β. Setting the first derivative with respect to X of the right-hand side of eq 3 equal to 0 and solving yields that value X_0 where activitity is maximized:

$$X_0 = \log \left[\frac{-b}{\beta(b + c)} \right] \qquad (4)$$

No formula for constructing a confidence interval for X_0 has been described in the literature.

Log $[1/K_{i(app)}]$ Estimation for Enzyme Inhibition. We have recently examined[12,13] the inhibition of the enzyme dihydrofolate reductase (DHFR, EC 1.5.1.3) from various species by the 2,4-diamino-1,2-dihydro-2,2-dimethyl-1-(3-substituted phenyl)-s-triazines (I). Many of



I

these molecules act as competitive nonstoichiometric inhibitors of the substrate dihydrofolate (FAH_2), assuming a rapid equilibrium bireactant system and saturating concentration for NADPH, the cofactor. The relationship between inhibitor concentration and inhibitory activity for competitive inhibition can be expressed[14] as:

$$\frac{[I_t]}{1 - (V_i/V_0)} = K_{i(app)}\left(\frac{V_0}{V_i}\right) \qquad (5)$$

where $[I_t]$ = total inhibitor concentration; V_0 = reaction velocity in absence of inhibitor; V_i = reaction velocity in presence of inhibitor; $K_{i(app)}$ = the apparent enzyme–inhibitor dissociation constant. For the case described above

$$K_{i(app)} = \beta K_i \left[\frac{[S] + \alpha K_{FAH_2}}{\alpha K_{FAH_2}} \right] \qquad (6)$$

where [S] = substrate concentration; K_{FAH_2} = FAH_2–enzyme complex dissociation constant; K_i = the actual enzyme–inhibitor complex dissociation constant; α = factor by which binding of FAH_2 changes K_{NADPH} (NADPH–enzyme dissociation constant) and by which binding of NADPH changes K_{FAH_2}; β = factor by which binding of I changes K_{NADPH} and by which binding of NADPH changes K_i (For the case of simple competitive inhibition of a rapid equilibrium unireactant system, eq 5 still applies and

$$K_{i(app)} = K_i \left[\frac{[S] + K_m}{K_m} \right]$$

where K_m = Michaelis–Menten constant for substrate–enzyme complex. In either case, $K_{i(app)} = [I_{50}]$ = that value of $[I_t]$ which causes 50% inhibition of the enzyme [obtained by substituting $[I_{50}]$ for $[I_t]$ and $^1/_2V_0$ for V_i in eq 5].) Therefore, a least-squares fit of $[I_t]/[1 - (V_i/V_0)]$ vs. V_0/V_i (eq 5) should directly provide an estimator for $K_{i(app)}$ and its associated confidence interval. This procedure, however, assumes that the error in $[I_t]/[1 - (V_i/V_0)]$ is additive and normally distributed and that V_0/V_i is known within negligible error. In reality, however, these assumptions do not hold; plots of $[I_t]/[1 - (V_i/V_0)]$ vs. V_0/V_i for apparently good experimental data consistently yield large deviations in $[I_t]/[1 - (V_i/V_0)]$ for $V_i/V_0 > \sim 0.8$ and in V_0/V_i for $V_i/V_0 < \sim 0.3$. It is rather the error in V_i/V_0, the form in which the original data were collected, that seems to be nearly normally distributed. Rearranging eq 5 provides

$$\frac{V_i}{V_0} = \frac{K_{i(app)}}{K_{i(app)} + [I_t]} \qquad (7)$$

(4) Miller, R. G. *Biometrika* **1974**, *61*, 1.
(5) Mosteller, F.; Tukey, J. W. "Data Analysis and Regression"; Addison-Wesley: Reading, Mass., 1977; pp 133–163.
(6) Duncan, G. T. *Technometrics* **1978**, *20*, 123.
(7) Miller, R. G. *Ann. Stat.* **1974**, *2*, 880.
(8) Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616.
(9) Hansch, C. *Acc. Chem. Res.* **1969**, *2*, 232.
(10) Hansch, C. in "Drug Design"; E. J. Ariëns, Ed.; Academic Press: New York, 1971; Vol. I, p 271.
(11) Hansch, C.; Steward, A. R.; Anderson, S. M.; Bentley, D. *J. Med. Chem.* **1968**, *11*, 1.
(12) Dietrich, S. W.; Smith, R. N.; Fukunaga, J. Y.; Hansch, C. *Arch. Biochem. Biophys.* **1979**, *194*, 600.
(13) Dietrich, S. W.; Smith, R. N.; Brendler, S.; Hansch, C. *Arch. Biochem. Biophys.* **1979**, *194*, 612.
(14) Segel, I. H. "Enzyme Kinetics", Wiley: New York, 1975; pp 22–24, 100–111, 150–159, 273–291.

The least-squares estimate of $K_{i(app)}$ in eq 7 can be solved for iteratively by minimizing the sum of squares deviations for V_i/V_0 as a function of $K_{i(app)}$:

$$\sum[(V_i/V_0)_{obsd} - (V_i/V_0)_{calcd}]^2$$

This procedure provides (for subsequent use in QSAR regression studies) an estimate of log $[1/K_{i(app)}]$ which more correctly reflects the distribution of the error in the V_i/V_0 experimental data but provides no simple confidence interval for the true parameter.

The Jackknife.[4-7] Let $Y_1, ..., Y_n$ be a sample of variable values and let $\hat{\theta}$ be an estimator for the parameter θ based on all n observations. For this study in particular, let $\hat{\theta}$ be an estimate of a function of the regresssion coefficients and/or iteratively estimated parameters associated with a least-squares linear or nonlinear regression. Let $\hat{\theta}_{-i}$ be the estimator for θ using all but the ith observation. Then,

$$\tilde{\theta} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\theta}_i$$

is the jackknife estimator for θ where $\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{-i}$ = the ith pseudovalue. The statistic

$$\frac{\sqrt{n}(\tilde{\theta} - \theta)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\tilde{\theta}_i - \tilde{\theta})^2}}$$

should have an approximate Student's t distribution with $n-1$ degrees of freedom.[4] Hence, the $100(1-\alpha)$ percent confidence interval for the jackknife estimator $\tilde{\theta}$ can be calculated as:

$$\tilde{\theta} \pm t^{\alpha}_{n-1}\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(\tilde{\theta}_i - \tilde{\theta})^2}$$

The term "jackknife" stems from the usefulness of this technique when more sophisticated tools may be unavailable, in much the same manner as the everyday pocketknife can serve as a useful substitute for more sophisticated mechanical tools.[5] The jackknife evaluates the contribution of each datum point to an estimator for θ by considering the resulting estimator when that datum is omitted from the computation.

Iteration Procedures. The stepwise procedure of Kubinyi and Kehrhahn[3] was utilized for the iterative solutions of the bilinear equations (eq 3) except that when the middle log β value (log β_2) yielded the lowest sum of squares deviations (SS) in log $1/C$, a new log β_2 was calculated using a three-point parabolic fit for:

$$(SS1, \log \beta_1), (SS2, \log \beta_2), (SS3, \log \beta_3)$$

where log β_1 = log β_2 − increment and log β_3 = log β_2 + increment; the new increment value was then set to one-tenth of the old value. This procedure allows extremely rapid convergence for the iterations considered. The initial log β_2 value was set equal to −log P_0 (or $-\pi_0$) calculated using the parabolic model, and the initial increment value was set equal to the standard deviation for log P (or π). Iteration was terminated when SS2 was less than min(SS1, SS3) and the increment was <0.001. (Identical results were obtained in every case when iteration was instead terminated when increment was <0.0001.)

For the iterative solutions of the enzyme inhibition equations (eq 7), initial values of $K_{i(app)}$ were estimated graphically, initial increment values were taken as one-tenth of the initial $K_{i(app)}$ estimate, the three-point parabolic fit was again used, and the iteration was terminated when SS2 was less then min(SS1, SS3) and the increment was $<10^{-6}\cdot K_{i(app)}$.

Results and Discussion

Table I contains the results of jackknifing log P_0 or π_0 for a number of literature QSAR regression studies relating the nonlinear dependence of various in vitro and in vivo activities on hydrophobicity using the parabolic and bilinear models. For those cases in which equations for both models are presented, the bilinear equation provides a statistically more significant correlation (i.e., by partial F test[3]). The jackknife does indeed appear to provide reasonable confidence intervals for log P_0 and π_0 as calculated from the bilinear model (eq 8, 10, 12–15, 17, and 19). The tightness of the intervals is, as expected, dependent in part on the degree of fit of the regression equation and the number of data points. For those cases with corresponding parabolic correlation equations (eq 9, 11, 16, and 18), the jackknife generally predicts larger confidence intervals, as expected from the better fit of the bilinear model for these data sets. The jackknife confidence intervals for the parabolic model log P_0 and π_0 values are similar to their usual estimates[11] (although slightly larger). As might be anticipated for both models, the jackknifed log P_0 and π_0 values themselves generally deviate the least from those directly calculated with all data points when n is large, the regression fit is good, and, hence, the jackknife confidence intervals are small.

A small but exemplary data set is presented in Table II with the results of jackknifing log $[1/K_{i(app)}]$ using eq 7 for some of our enzyme inhibition data. Iteration on $K_{i(app)}$ to minimize SS in eq 7 and using all 11 data points yields log $[1/K_{i(app)}]$ = 6.17 (note no confidence interval available) with r^2 = 0.9447. The corresponding 95% jackknife confidence interval estimator for log $[1/K_{i(app)}]$ is 6.18 ± 0.07; that is [6.11, 6.24]. In contrast, linear regression on eq 5 to minimize SS of $[I_i]/[1 - (V_i/V_0)]$ yields a point estimator for log $[1/K_{i(app)}]$ = 6.10 with 95% confidence interval [5.98, 6.27] and r^2 = 0.1945. Considering the quality of the data and the small sample size, the jackknife provides a reasonable confidence interval estimate for log $[1/K_{i(app)}]$. Normally, we collect three times as many data points as in this example. Our experience has been that 95% confidence intervals for log $[1/K_{i(app)}]$ are on the order of ±0.03. This example also demonstrates the preferability of using eq 7 and the jackknife technique for estimating log $[1/K_{i(app)}]$ and its confidence interval, as opposed to direct calculation from eq 5. The data in Table II (especially data points 5, 10, and 11) also illustrate that, with the jackknife, small deviations of the $\hat{\theta}_{-i}$ values from $\hat{\theta}$ lead to large deviations in the $\tilde{\theta}_i$ values and potentially in the magnitudes of the jackknife estimator $\tilde{\theta}$ and its associated confidence interval. Therefore, when utilizing the jackknife technique, it is extremely important to carry sufficient significant digits in the calculations so as not to obtain misleading results due to rounding errors.

Unusually large confidence intervals can result when attempting to jackknife utilizing a data set for which a minority of the data points (even just one) are crucial to the shape of the fitted regression equation, e.g., with nonlinear QSAR models, for data sets having few or no data points in the maximum activity region or having almost all of the data points with independent variable values either greater than or less than the value associated with maximum activity. These problems seem to be reduced by having data points which (1) are evenly spaced over the independent parameter range(s) of importance; (2) for nonlinear regressions, are reasonably distributed on both sides and in the region of points of maximum

Table I. Comparison of Jackknife and Currently Used Methods for Estimation of Log P_0 and π_0 and Their Associated Confidence Intervals[a]

compd	system and act.	eq no.	equation	n	r	s	jackknife estimate	ref (eq)
I	bovine liver DHFR, I_{50}	8	$\log 1/C = 6.64(\pm0.11) + 1.05(\pm0.14)\pi_3 - 1.21(\pm0.20)\log(\beta \times 10^{\pi_3} + 1)$ $\log\beta = -0.736; (\pi_3)_0 = 1.56$	28	0.955	0.210	$(\pi_3)_0 = 1.54$ [1.24, 1.84]	12 (7)
		9	$\log 1/C = 6.47(\pm0.13) + 0.63(\pm0.11)\pi_3 - 0.12(\pm0.03)\pi_3^2$ $(\pi_3)_0 = 2.72$ [2.42, 3.11][b]	28	0.921	0.268	$(\pi_3)_0 = 3.01$ [1.78, 4.24]	12 (6)
I	rat liver DHFR, I_{50}	10	$\log 1/C = 6.29(\pm0.12) + 1.11(\pm0.15)\pi_3 - 1.34(\pm0.26)\log(\beta \times 10^{\pi_3} + 1)$ $\log\beta = -0.984; (\pi_3)_0 = 1.68$	18	0.977	0.171	$(\pi_3)_0 = 1.81$ [1.41, 2.21]	12 (9)
		11	$\log 1/C = 6.29(\pm0.14) + 0.80(\pm0.13)\pi_3 - 0.19(\pm0.05)\pi_3^2$ $(\pi_3)_0 = 2.12$ [1.82, 2.56][b]	18	0.963	0.210	$(\pi_3)_0 = 2.20$ [1.85, 2.55]	12 (8)
I	*L. casei* DHFR, I_{50}	12	$\log 1/C = 3.13(\pm0.15) + 0.53(\pm0.10)\pi_3 - 0.67(\pm0.35)\log(\beta \times 10^{\pi_3} + 1) + 0.79(\pm0.25)\mathrm{MR}'$ $\log\beta = -3.461; (\pi_3)_0 = 4.03$	28	0.949	0.302	$(\pi_3)_0 = 4.04$ [3.30, 4.78]	13 (5)
I	*S. aureus*, in vitro MIC	13	$\log 1/C = 2.83(\pm0.16) + 0.59(\pm0.05)\pi_3 - 1.52(\pm0.17)\log(\beta \times 10^{\pi_3} + 1)$ $\log\beta = -5.994; (\pi_3)_0 = 5.79$	23	0.986	0.218	$(\pi_3)_0 = 5.90$ [5.56, 6.24]	13 (6)
I	*E. coli*, in vitro MIC	14	$\log 1/C = 2.57(\pm0.23) + 0.51(\pm0.07)\pi_3 - 1.09(\pm0.20)\log(\beta \times 10^{\pi_3} + 1)$ $\log\beta = -5.116; (\pi_3)_0 = 5.07$	22	0.960	0.307	$(\pi_3)_0 = 4.92$ [4.39, 5.45]	13 (7)
N'-alkylniketham ide chlorides	*S. aureus*, antibacterial act.	15	$\log 1/C = 3.27(\pm0.34) + 0.56(\pm0.09)\pi - 0.74(\pm0.19)\log(\beta \times 10^{\pi} + 1)$ $\log\beta = -5.970; \pi_0 = 6.57$	20	0.970	0.265	$\pi_0 = 6.59$ [5.98, 7.20]	1 (58)
		16	$\log 1/C = 2.91(\pm0.46) + 0.91(\pm0.20)\pi - 0.06(\pm0.02)\pi^2$ $\pi_0 = 7.62$ [6.96, 8.80][b]	20	0.961	0.292	$\pi_0 = 7.66$ [6.75, 8.57]	18 (43)
N-alkylpiperidines	Red cell dove, hemolytic act.	17	$\log 1/C = 1.30(\pm0.09) + 0.96(\pm0.04)\log P - 1.41(\pm0.12)\log(\beta \times P + 1)$ $\log\beta = -3.555; \log P_0 = 3.89$	11	0.999	0.050	$\log P_0 = 3.90$ [3.79, 4.01]	3 (3)
		18	$\log 1/C = 1.01(\pm0.30) + 1.52(\pm0.25)\log P - 0.18(\pm0.04)(\log P)^2$ $\log P_0 = 4.25$ [3.91, 4.78][b]	11	0.992	0.142	$\log P_0 = 4.20$ [3.68, 4.72]	19 (23)
alkanes	mice, LD_{100} (iv)	19	$\log 1/C = -0.65(\pm0.35) + 0.96(\pm0.11)\log P - 1.31(\pm0.13)\log(\beta \times P + 1)$ $\log\beta = -3.523; \log P_0 = 3.96$	11	0.996	0.039	$\log P_0 = 4.06$ [3.79, 4.33]	2 (43)
		20	$\log 1/C = 0.20(\pm1.11) + 0.94(\pm0.47)\log P - 0.11(\pm0.05)(\log P)^2$ $\log P_0 = 4.37$ [3.75, 4.70][b]	11	0.930	0.148	$\log P_0 = 4.63$ [3.69, 5.57]	2 (41)

[a] n = number of data points; r = correlation coefficient; s = standard deviation of the regression; numbers in parentheses are for construction of 95% confidence intervals.
[b] Log P_0 and π_0 confidence intervals were calculated by the procedure of ref 11.

Table II. Jackknifing Log $[1/K_{i(app)}]$: Competitive Enzyme Inhibition[a]

no.	$[I_t]$ ($\times 10^{-7}$ M)	V_i/V_0	log $[1/K_{i(app)}]$ $-i$th ($\hat{\theta}_{-i}$)	log $[1/K_{i(app)}]$ ith pseudo ($\tilde{\theta}_i$)
1	0.365	0.9868	6.1765	6.1556
2	0.953	0.8265	6.1692	6.2287
3	1.827	0.8070	6.1784	6.1358
4	1.907	0.7287	6.1655	6.2650
5	3.654	0.5852	6.1588	6.3327
6	3.813	0.6566	6.1797	6.1230
7	5.481	0.5155	6.1648	6.2727
8	5.720	0.5569	6.1797	6.1233
9	7.308	0.4597	6.1693	6.2276
10	11.440	0.4170	6.1871	6.0494
11	19.067	0.3392	6.1907	6.0128

[a] Log $[1/K_{i(app)}]$ for all 11 data points ($\hat{\theta}$) = 6.1746.

activity; and (3) have accurately determined dependent and independent variable values.

One additional and very valuable aspect of the jackknife technique is that it permits one to examine the influence of each of the individual members of a data set on the estimates of the parameters of the equation (QSAR) as the data points are dropped one at a time. While certain cases of instability of the estimates may be quite obvious (e.g., only a single data point with log $P >$ log P_0), such is not always the situation. The $-i$th values (i.e., $\hat{\theta}_{-i}$) and jackknife estimates (i.e., θ) can be calculated for jackknifing not only log P_0 (or π_0) but also the regression coefficients for a QSAR equation. Examination of these values makes it possible to determine which, if any, data points are critical in determining the form of the derived QSAR equation (i.e., spotting of potential outliers). Similarly, the $-i$th values obtained by jackknifing log $[1/K_{i(app)}]$ for competitive enzyme inhibition data can be used to determine which, if any, data points are critical for the estimation of log $[1/K_{i(app)}]$.

In conclusion, the jackknife technique does appear to be a useful statistical tool for constructing confidence intervals for parameters which are estimated by linear or nonlinear regression techniques. In particular, the method appears well suited to confidence interval estimation of the independent variable value associated with maximum activity in the bilinear QSAR model, for log $[1/K_{i(app)}]$ from competitive enzyme inhibition data, and (perhaps) also for the parabolic QSAR model maximum activity independent variable value. Application of this technique is by no means limited to these three cases; extension to confidence intervals for other parameters which are complicated functions of other variables is possible and of general utility.

It should also be noted that a number of other authors (e.g., ref 15–17) have also examined the use of the jackknife technique for parameter and confidence interval estimation in the examination of enzyme kinetic data.

Acknowledgment. This research was supported by Grant CA-11110 from the National Cancer Institute. This material is based upon work supported by the National Science Foundation under Grant SPI-7914805; S.W.D. is a NSF National Needs Postdoctoral Fellow.

(15) Dammkoehler, R. A. *J. Biol. Chem.* **1966**, *241*, 1955.
(16) Cornish-Bowden, A.; Wong, J. T. *Biochem. J.* **1978**, *175*, 969.
(17) Duggleby, R. G. *Biochem. J.* **1979**, *181*, 255.
(18) Lien, E. J.; Hansch, C.; Anderson, S. M. *J. Med. Chem.* **1968**, *11*, 430.
(19) Hansch, C.; Glave, W. R. *Mol. Pharmacol.* **1971**, *7*, 337.

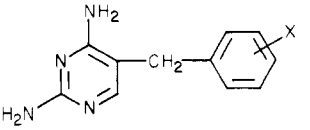
Quantitative Structure–Selectivity Relationships. Comparison of the Inhibition of *Escherichia coli* and Bovine Liver Dihydrofolate Reductase by 5-(Substituted-benzyl)-2,4-diaminopyrimidines

Stephen W. Dietrich, Jeffrey M. Blaney, Mark A. Reynolds, Priscilla Y. C. Jow, and Corwin Hansch*

Department of Chemistry, Pomona College, Claremont, California 91711. *Received January 14, 1980*

A quantitative structure–activity relationship (QSAR) has been formulated for the inhibition of purified *E. coli* dihydrofolate reductase by 23 5-(substituted benzyl)-2,4-diaminopyrimidines: log $1/C = 1.14MR'_{3,4,5} + 5.73$; $r = 0.887$; $s = 0.285$. In this expression, $MR'_{3,4,5}$ refers to the sum of MR values for X in the 3, 4, and 5 positions of the phenyl moiety. MR′ signifies that the effective value of MR is limited to 0.79. Comparison of the QSAR for *E. coli* enzyme inhibition with that previously obtained for bovine enzyme offers the first general explanation for the great selectivity of the important antibacterial agent trimethoprim. Such QSSR promise to be of value in devising more selective drugs.

This report continues our analysis of the interaction of dihydrofolate reductase (DHFR; EC 1.5.1.3) from various species with substituted pyrimidines and triazines.[1] In particular, we discuss the inhibition of DHFR from *E. coli* by benzylpyrimidines of type I.



I

Since DHFR shows such wide variation from organism to organism in its sensitivity to inhibitors, it offers an exceptional opportunity for selective inhibition of a pathogen with respect to the host. Inhibitors of DHFR have proved to be of great value as antimicrobial agents, as well as in cancer chemotherapy. It therefore is important to gain a clearer understanding of the molecular forces which determine the relative inhibitory activities of these in-

(1) (a) Blaney, J. M.; Dietrich, S. W.; Reynolds, M. A.; Hansch, C. *J. Med. Chem.* **1979**, *22*, 614. (b) Dietrich, S. W.; Smith, R. N.; Fukunaga, J. Y.; Olney, M.; Hansch, C. *Arch. Biochem. Biophys.* **1979**, *194*, 600. (c) Dietrich, S. W.; Smith, R. N.; Brendler, S.; Hansch, C. *Arch. Biochem. Biophys.* **1979**, *194*, 612. (d) Hansch, C.; Dietrich, S. W.; Smith, R. N. in "Chemistry and Biology of Pteridines"; Kisliuk, R. L.; Brown, G. M., Eds.; Elsevier: New York, 1979; Vol. 4, p 425. (e) Silipo, C.; Hansch, C. *J. Am. Chem. Soc.* **1975**, *97*, 6849.

0022-2623/80/1823-1205$01.00/0