# Structure–Activity Correlations among Rifamycin B Amides and Hydrazides

Frank R. Quinn, John S. Driscoll,*

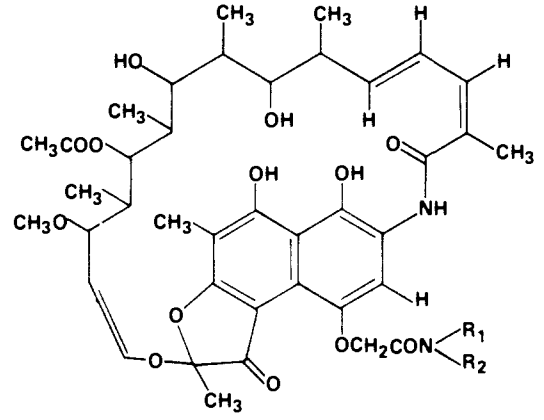
*Drug Development Branch, Drug Research and Development, Division of Cancer Treatment, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20014*

and Corwin Hansch

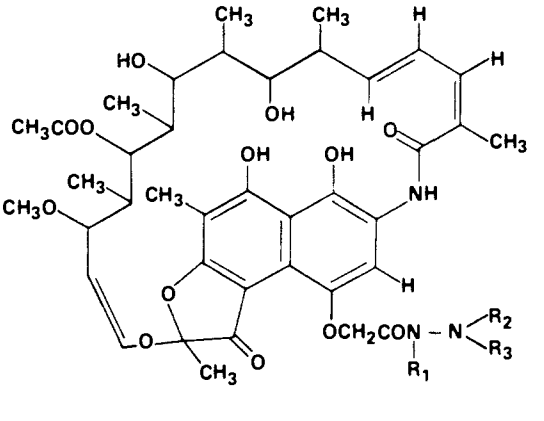
*Department of Chemistry, Pomona College, Claremont, California 91711* *Received August 26, 1974*

Structure–antibacterial activity correlation equations have been developed for a series of 44 amides and 25 hydrazides of rifamycin B in five bacterial systems. The best amide equations show that activity is a parabolic function of log $P$. A wide variation in log $P_0$ was found for the various bacterial systems. The most important correlation parameter in the hydrazide equations is $\sigma^*$. The significance of this finding is somewhat obscured by the high degree of collinearity among the parameters evaluated ($\sigma^*$, $E_s$, log $P$). Two rifamycin B amides were prepared and evaluated as a result of this study. The correlation equations quantitatively predicted their activity in five of six tests.

The rifamycins, a group of metabolites of *Streptomyces mediterranei*, inhibit gram-positive bacteria and have been used in the treatment of bacterial infections. One member of this family, rifamycin B, contains a free carboxyl group. In an attempt to increase the activity in this series, Sensi, *et al.*,[1] prepared and tested a number of amide (I) and hydrazide (II) derivatives of rifamycin B. These 75 derivatives were tested against nine bacterial systems. A wide variation in activity was observed in five of the bacterial systems and the results appeared to constitute a good opportunity for a study of quantitative structure–activity relationships (QSAR) using substituent constants[2,3] and regression analysis.[4]




I




II

A recent treatment of 16 N-disubstituted rifamycin B amides indicated a parabolic dependence of activity on lipophilicity in one of the five bacterial systems treated in the analysis described here.[5] It was felt that if quantitative correlations could be demonstrated for this large series (50 amides and 25 hydrazides) in bacterial systems, then there was good promise that quantitative correlations could be developed among rifamycin derivatives as reverse transcriptase inhibitors[6] when the appropriate test data become available.

**Methods.** The biological activity, expressed as log (1/C), is given for the rifamycin amides in Table I and for the hydrazides in Table II. These data are derived from that reported by Sensi, *et al.*,[1] and the compound numbers used here are consistent with those used in that study. The molar concentration, $C$, is defined as the lowest concentration of antibiotic that prevents visible growth after 18-hr incubation. The physiocochemical parameters used are given in Tables III and IV for 70 of the 75 compounds in the original investigation.[1] Because of the uncertainty in the estimation of steric and electronic constants for compounds **7**, **36**, **37**, **41**, and **42**, these were omitted from this study. The equations were obtained using the method of least squares and an IBM 370/165 computer. All equations designated as best for each bacterial system were determined to be statistically significant by a stepwise $F$ test.

**Derivation of Log *P* Values.** The log $P$ of the parent compound, rifamycin SV, was determined experimentally in the octanol–water system.[7] The log $P$ values of the congeners in this series were calculated from this base value. Partition coefficients for the rifamycin B amide derivatives (**1**–**49**) were calculated using the measured values for the partition coefficient of rifamycin SV and acetamide.[8]

$$\underset{+0.78}{\log P\ (\text{rifamycin SV})} + \underset{-1.46}{\log P\ (\text{acetamide})} = \underset{-0.68}{\log P\ (\text{compound 1})} \quad (1)$$

From the above relationship, the calculated log $P$ value for the unsubstituted amide (compound 1) in the amide series (**1**–**49**) is −0.68. The log $P$ values for the N-substituted derivatives were then calculated using eq 2.

$$\log P = -0.68 + \pi R_1 + \pi R_2 \quad (2)$$

As a test of the accuracy of this method for calculating rifamycin log $P$ values, a small sample of rifamycin B monoethylamide (**3**) was obtained and its log $P$ value was determined. The found value of 0.13 was in satisfactory agreement with the 0.32 value calculated.

The log $P$ values for the hydrazides were obtained in a similar manner. The calculated log $P$ value for unsubstituted rifamycin B hydrazide was obtained by adding the previously calculated base value for rifamycin B amide

**Table I.** Antibacterial Activity of Rifamycin Amides [Log (1/*C*)]

| | *M. aureus* | | *S. faecalis* | | *S. hemolyticus* | | *B. subtilis* | |
|---|---|---|---|---|---|---|---|---|
| Compd | Obsd | Pred[a] | Obsd | Pred[b] | Obsd | Pred[c] | Obsd | Pred[d] |
| 1 | 5.70 | 5.81 | 4.91 | 4.88 | 6.62 | 6.02 | 4.78 | 4.74 |
| 2 | 6.68 | 6.10 | 5.89 | 5.29 | 5.09 | 6.31[e,f] | 7.11 | 4.96[e,f] |
| 3 | 6.99 | 6.35 | 6.20 | 5.66 | 7.50 | 6.54[f] | 5.40 | 5.17 |
| 4 | 6.50 | 6.57 | 5.82 | 5.97 | 6.42 | 6.72[e] | 5.41 | 5.36 |
| 5 | 6.70 | 6.48 | 5.90 | 5.85 | 7.12 | 6.65 | 5.41 | 5.28 |
| 6 | 6.73 | 6.58 | 5.83 | 5.99[e] | 6.43 | 6.73[e] | 5.42 | 5.37 |
| 8 | 6.92 | 6.80 | 6.04 | 6.29 | 6.84 | 6.88 | 5.43 | 5.57 |
| 9 | 6.54 | 6.99 | 6.39 | 6.56 | 6.86 | 6.96 | 5.64 | 5.79 |
| 10 | 6.48 | 7.03 | 6.56 | 6.60 | 6.66 | 6.96 | 5.78 | 5.83 |
| 11 | 6.90 | 7.08 | 6.78 | 6.67 | 6.98 | 6.96 | 5.80 | 5.90 |
| 12 | 7.03 | 5.72[e,f] | 6.12 | 4.74[e,f] | 7.25 | 5.92[f] | 5.11 | 4.67 |
| 13 | 7.12 | 7.65 | 6.02 | 6.45 | 7.89 | 7.72 | 5.72 | 5.92 |
| 14 | 7.91 | 8.06 | 7.01 | 7.02 | 8.01 | 8.03 | 6.03 | 6.29 |
| 15 | 8.15 | 8.33 | 7.15 | 7.40 | 8.15 | 8.14 | 6.53 | 6.59 |
| 16 | 8.46 | 8.48 | 7.46 | 7.57 | 7.86 | 8.05 | 7.16 | 6.84 |
| 17 | 8.86 | 8.44 | 7.86 | 7.53 | 8.16 | 8.11 | 7.28 | 6.75 |
| 18 | 8.65 | 8.49 | 7.65 | 7.54 | 7.55 | 7.76 | 7.05 | 7.03 |
| 19 | 8.22 | 8.18 | 7.22 | 7.20 | 8.44 | 8.10[e] | 6.67 | 6.42 |
| 20 | 8.49 | 8.47 | 7.49 | 7.48 | 7.89 | 7.60 | 6.89 | 7.08 |
| 21 | 7.95 | 7.87 | 7.06 | 6.76 | 7.95 | 7.90 | 6.03 | 6.11 |
| 22 | 7.91 | 8.06 | 7.31 | 7.02 | 8.13 | 8.03 | 6.65 | 6.29 |
| 23 | 8.61 | 7.99 | 7.61 | 6.92 | 8.13 | 7.99 | 6.96 | 6.22[e,f] |
| 24 | 8.44 | 8.21 | 7.22 | 7.24 | 8.62 | 8.11[e] | 6.35 | 6.45 |
| 25 | 8.62 | 8.07 | 7.62 | 7.04 | 8.31 | 8.04 | 6.66 | 6.30 |
| 26 | 8.01 | 8.21 | 7.22 | 7.24 | 8.22 | 8.11 | 6.35 | 6.45 |
| 27 | 8.62 | 8.33 | 7.62 | 7.40 | 8.15 | 8.14 | 7.27 | 6.59[e] |
| 28 | 8.44 | 8.21[e] | 7.14 | 7.23 | 8.62 | 8.11[e] | 6.97 | 6.44[e] |
| 29 | 8.14 | 8.25 | 7.22 | 7.29 | 8.22 | 8.13 | 6.36 | 6.49 |
| 30 | 8.15 | 8.34 | 7.53 | 7.40 | 7.98 | 8.14 | 6.68 | 6.60 |
| 31 | 8.46 | 8.36[e] | 7.33 | 7.43 | 8.16 | 8.14[e] | 6.37 | 6.63 |
| 32 | 8.24 | 8.37 | 7.09 | 7.45 | 8.16 | 8.14 | 6.37 | 6.65 |
| 33 | 6.93 | 7.32 | 5.85 | 5.96 | 7.15 | 7.40 | 5.23 | 5.65 |
| 34 | 7.52 | 7.58 | 6.18 | 6.34 | 6.96 | 7.65 | 5.73 | 5.86 |
| 35 | 6.97 | 7.49 | 6.05 | 6.21 | 6.97 | 7.57 | 5.74 | 5.79 |
| 38 | 7.24 | 8.28[f] | 5.77 | 7.33[e,f] | 7.47 | 8.14 | 6.38 | 6.53 |
| 39 | 7.26 | 7.91 | 6.34 | 6.82 | 8.21 | 7.93 | 5.73 | 6.15 |
| 40 | 8.22 | 8.12 | 7.45 | 7.11 | 8.22 | 8.07 | 6.67 | 6.35 |
| 43 | 8.01 | 8.06 | 6.92 | 7.02 | 8.22 | 8.03 | 6.04 | 6.29 |
| 44 | 8.22 | 8.15 | 7.27 | 7.16 | 7.92 | 8.09 | 6.36 | 6.39 |
| 45 | 8.22 | 8.15 | 7.45 | 7.16 | 7.92 | 8.09 | 6.67 | 6.39 |
| 46 | 8.32 | 8.18 | 7.45 | 7.20 | 7.92 | 8.10 | 6.45 | 6.42 |
| 47 | 6.44 | 7.54[e,f] | 5.84 | 6.29 | 7.26 | 7.62 | 5.74 | 5.83 |
| 48 | 7.92 | 7.69 | 6.22 | 6.49 | 8.36 | 7.75[e] | 6.67 | 5.95[e,f] |
| 49 | 7.93 | 7.82 | 6.98 | 6.68 | 7.76 | 7.86 | 6.15 | 6.06 |

[a] Predicted by eq 18. [b] Predicted by eq 20. [c] Predicted by eq 22. [d] Predicted by eq 24. [e] Compounds omitted in the development of eq 5–16. [f] Compounds omitted in the development of eq 17–24.

(−0.68) to an average of the experimental values[8] of log *P* for dimethylamine (−0.25) and then subtracting 1.0 for the removal of two methyl groups (eq 3). Log *P* values for substituted hydrazides were calculated using eq 4.

$$\log P \text{ (rifamycin B hydrazide)} = -0.68 + \log P\ (Me_2NH) - 1.0 = -1.93 \quad (3)$$

$$\log P = -1.93 + \pi R_1 + \pi R_2 + \pi R_3 \quad (4)$$

The value for the *tert*-butyl group (**6**) was obtained from the experimental value[9] for methyl *tert*-butyl ether by subtracting −0.97 for an ether oxygen and 0.5 for one methyl group. The value for the diallyl derivative **19** was obtained by adding −0.3 (for a double bond) to the value of a propyl radical. The value for the 1-propynyl group (**28**) was obtained by subtracting 1.0 (for ethyl) from the experimentally obtained value of log *P* for 1-pentyne.[8] The value for the $-CH_2CH_2CN$ derivative **35** was obtained using a value of −0.84 for the CN group. Compounds **39–46** represent cyclized amides in which pyrrolidine, piperidine, or azepine rings have been formed. In the case of the unsubstituted pyrrolidine **39**, the log *P* was calculated by adding 4 × 0.4 (ring $CH_2$) to the base value of −0.68. The log *P* values for compounds **40–46** were calculated in a similar manner, adding 0.5 for each methyl substituent and subtracting 0.2 for each chain branch.

The log *P* values of the morpholine derivatives of rifamycin B amide (**47–49**) were obtained by adding the log *P* value[8] for diethyl ether (0.77) to the base amide value. The values for the cyclized hydrazides (**68–75**) were calculated

**Table II.** Antibacterial Activity of Rifamycin Hydrazides [Log (1/C)]

| | *M. aureus* | | *S. faecalis* | | *S. hemolyticus* | | *B. subtilis* | | *M. tuberculosis* | |
|---|---|---|---|---|---|---|---|---|---|---|
| Compd | Obsd | Pred[a] | Obsd | Pred[b] | Obsd | Pred[c] | Obsd | Pred[d] | Obsd | Pred[e] |
| 50 | 6.98 | 7.00 | 6.04 | 6.06 | 7.11 | 7.15 | 5.40 | 5.27 | 6.71 | 6.84 |
| 51 | 6.93 | 6.90 | 5.93 | 5.95 | 7.23 | 7.06 | 5.44 | 5.48 | 7.36 | 7.25 |
| 52 | 8.13 | 8.41 | 7.21 | 7.47 | 8.61 | 8.37 | 6.65 | 6.68 | 6.65 | 6.51 |
| 53 | 8.62 | 8.61 | 7.62 | 7.66 | 8.62 | 8.54 | 6.96 | 6.98 | 6.35 | 6.60 |
| 54 | 8.94 | 8.64 | 7.76 | 7.69 | 8.64 | 8.57 | 6.98 | 7.14 | 6.37 | 6.75[f] |
| 55 | 8.87 | 8.66 | 7.95 | 7.72 | 8.65 | 8.59 | 7.30 | 7.31 | 7.00 | 6.91 |
| 56 | 8.62 | 8.51 | 7.44 | 7.56 | 8.62 | 8.46 | 6.62 | 6.83 | 6.62 | 6.55 |
| 57 | 8.93 | 8.70 | 7.63 | 7.76 | 8.63 | 8.62 | 7.63 | 7.13 | 6.76 | 6.71 |
| 58 | 8.87 | 8.73 | 7.77 | 7.78 | 8.87 | 8.65 | 7.29 | 7.29 | 6.69 | 6.80 |
| 59 | 8.80 | 8.76 | 7.96 | 7.81 | 8.96 | 8.67 | 7.31 | 7.46 | 7.00 | 6.96 |
| 60 | 8.43 | 8.53 | 7.60 | 7.58 | 8.60 | 8.47 | 6.65 | 6.91 | 6.73 | 6.63 |
| 61 | 8.94 | 8.72 | 7.63 | 7.77 | 8.64 | 8.64 | 7.64 | 7.22 | 6.76 | 6.72 |
| 62 | 8.48 | 8.74 | 7.65 | 7.79 | 8.48 | 8.66 | 7.17 | 7.37 | 7.00 | 6.88 |
| 63 | 8.49 | 8.78 | 7.66 | 7.83 | 8.67 | 8.69 | 7.49 | 7.55 | 7.01 | 7.03 |
| 64 | 8.93 | 8.54 | 7.68 | 7.59 | 8.33 | 8.48 | 7.28 | 6.99 | 6.76 | 6.71 |
| 65 | 8.64 | 8.73 | 7.94 | 7.78 | 7.94 | 8.65[f] | 7.64 | 7.29 | 6.69 | 6.80 |
| 66 | 8.66 | 8.76 | 7.66 | 7.81 | 8.66 | 8.67 | 7.31 | 7.46 | 7.00 | 6.96 |
| 67 | 8.64 | 8.54 | 7.94 | 7.59 | 8.09 | 8.48 | 7.28 | 7.06 | 6.68 | 6.79 |
| 68 | 8.63 | 8.63 | 7.63 | 7.68 | 7.93 | 7.90 | 6.98 | 7.01 | 6.98 | 7.04 |
| 69 | 8.64 | 8.72 | 7.64 | 7.77 | 7.94 | 7.99 | 6.98 | 7.16 | 7.28 | 7.09 |
| 70 | 8.64 | 8.74 | 7.94 | 7.79 | 7.94 | 8.00 | 7.29 | 7.25 | 7.29 | 7.16 |
| 71 | 8.65 | 8.75 | 7.95 | 7.80 | 8.11 | 8.01 | 7.30 | 7.33 | 7.00 | 7.24 |
| 72 | 8.23 | 8.24 | 7.68 | 7.29 | 7.76 | 8.22 | 6.36 | 6.41 | 6.68 | 6.44 |
| 73 | 7.64 | 8.34[f] | 6.68 | 7.39[f] | 8.86 | 8.31[f] | 6.06 | 6.56[f] | 6.37 | 6.48 |
| 74 | 8.10 | 8.35 | 7.17 | 7.41 | 8.25 | 8.32 | 6.38 | 6.65 | 6.38 | 6.56 |
| 75 | 7.65 | 8.37[f] | 6.70 | 7.42[f] | 7.65 | 8.33[f] | 5.78 | 6.73[f] | 5.78 | 6.64[f] |

[a] Predicted by eq 26. [b] Predicted by eq 25. [c] Predicted by eq 27. [d] Predicted by eq 28. [e] Predicted by eq 29. [f] Compounds omitted in the development of eq 25–29.

using the experimental values[10] of log $P$ for morpholine (−1.08) and piperidine (0.85). These were added to the base value for the amides (−0.68) to yield the hydrazide values.

**Calculation of $E_s$.** Steric constant values, except where noted, were those of Taft.[11] The $E_s$ values for phenyl and $p$-halophenyl substituents (**8–11**) were assumed to be the same ($E_s = -2.58$), since the major contribution can be attributed to the ring. The value for 1-hydroxyethyl (**12**) is taken from a series of recently measured values by Talvik and Palm.[12]

The steric contribution of the allyl group (**19**) was considered to be equivalent to the average of the values for Et and Pr. The same value (−0.36) was used for the 2-propynyl group (**28**). The $E_s$ value for the $-CH_2CH_2OH$ derivatives (**33, 34**) was estimated to be that of $n$-propyl. The $E_s$ values for the cyclic amides and hydrazides (**39–49, 68–75**) were estimated by using the approximation that the total steric contribution for these rings could be deduced by summing the contribution of those moieties obtained by a scission giving the largest acyclic segments. For example, $E_s$ for pyrrolidine (**39**) was assumed to be roughly equivalent to the sum of the contribution of two ethyl groups. Similarly, the steric contribution of 2,5-dimethylpyrrolidine was assumed to be the equivalent of 2-isopropyl groups ($2 \times -0.47$). Table V summarizes these values and the methods of calculation.

**Calculation of $\sigma^*$.** The parameter chosen in this study to account for polar effects was the Taft aliphatic substituent constant, $\sigma^*$. These summed values are given in Table III. While the number of experimental $\sigma^*$ values is limited,[11] an analysis of the $\sigma^*$ values of similar groups permits some reasonable approximations to be made. For example, the contribution of an $\alpha$-methyl group appears to be approximately −0.1. This can be deduced from observing that the $\sigma^*$ value for a benzyl substituent is 0.215 while that for $\alpha$-methyl benzyl is 0.11. Again, the value for Et is −0.1, while that for $i$-Pr is −0.19. Similarly, the values for $i$-Pr and *sec*-butyl differ by −0.085. This approximation was used to calculate $\sigma^*$ for the 1-hydroxyethyl group (**12**). In this instance, −0.1 was added to the value for $-CH_2OH$ (0.55) to yield a value of 0.45.

The inductive effect of terminal alkyl groups becomes constant with increasing chain length. Hence the $\sigma^*$ for the pentyl group (**18**) was assumed to be approximately the same as that of the butyl group. The value for the allyl derivative (**19**) was obtained from the measured value for the vinyl group (0.65) by assuming a 50% decrease in $\sigma^*$ for the effect of an intervening methylene. Comparison of other groups in the Taft series seems to justify this approximation. The value for the 1-propynyl derivative **28** was taken from work of Hall.[13] The value for the 2-hydroxyethyl substituent (**33, 34**) was obtained from the value for $CH_2OH$ (0.555) using the factor of 0.5 as in the case of the allyl group. Similarly, the $\sigma^*$ for the 2-cyanoethyl compound **35** was obtained by taking ½ of the value for $-CH_2CN$.

The $\sigma^*$ values for the N-containing rings (**39–46, 68–70**) were calculated in a manner analogous to the calculation of the steric constants (Table V). The $\sigma^*$ values for the morpholine derivatives (**47–49, 72–75**) were derived by considering the $\sigma^*$ for morpholine to be the sum of $CH_2CH_2O$ and Et values. The $\sigma^*$ value for $-CH_2CH_2O$ (0.28) was obtained from the value for $-CH_2OH$ by using the 50% factor for the intervening methylene. The uncertainty in this approximation is reflected in the correlation difficulties with some of the ring analogs.

## Results and Discussion

**Rifamycin B Amides.** Table VI shows some of the equations generated for the rifamycin B amides in the *Micrococcus aureus* system. Equations were derived for all linear combinations of log $P$, $(\log P)^2$, $\sigma^*$, $(\sigma^*)^2$, and $E_s$. Only the most significant equations are shown. In general, the equations in Table VI are typical of those determined for *Streptococcus faecalis*, *Streptococcus hemolyticus*, and *Bacillus subtilis*. Forty-one of the 44 compounds in Table III were used to generate the data for the *M. aureus* system. The four compounds omitted from the derivation of the best equation (Table VI, eq 13) are designated in Table I.

In the *M. aureus* system, the electronic contribution ($\sigma^*$) appears to be the most important single factor (eq 7). Equation 13, which was statistically significant ($F$ test), best correlated the data in the *M. aureus* bacterial system. All three parameters are necessary with $\sigma^*$ predominant.

The best equations for the amides in the four bacterial test systems evaluated are shown in Table VII. *M. aureus*, *B. subtilis*, and *S. faecalis* (eq 13–15) all show a similar dependence on $E_s$, log $P$, and $\sigma^*$ with correlation coefficients >0.9 and standard deviations of 0.24–0.33. Only 36 of 45 amides (see Table I) were used to derive the best equation for *S. hemolyticus*. The best amide equation in this system showed a dependence only on $\sigma^*$. No adequate quantitative correlation could be obtained for the rifamycin amides in the *Mycobacterium tuberculosis* system.

The correlations obtained (eq 13–16) appear quite reasonable in that they explained about 80% of the variance in the data ($r^2$) and are statistically significant relative to the next best equations containing fewer variables. The following points, however, caused some concern about the utility of these correlation equations. (a) Log $P$, an important term in all the equations except that for *S. hemolyticus*, had a positive coefficient and no equation which included a $(\log P)^2$ term was significant. This indicated that unlimited increases in the lipophilicity of amide derivatives should result in unlimited increases in antibacterial activity. This is inconsistent with past experience. (b) As many as eight of the 44 derivatives had to be omitted (Table I) in order to obtain the best correlations. (c) There was some collinearity between log $P$ and $E_s$ for the compounds studied in several of the bacterial systems. In an attempt to improve the usefulness of the correlation equations, further studies were carried out.

Molar refractivity[14] (MR) was added as a dependent variable, but it did not prove useful due to its high degree of collinearity ($r^2$ = 0.86) with the log $P$ term for the substituents in the study.

Inspection of the data shows that, in general, activity increases with increasing substitution of the amide nitrogen (di > mono > unsubstituted).[1] Additionally, it is known that amide partition coefficients are often not calculable as strictly linear combinations of $\pi$ values for N-substituents.[10] For these reasons, a study of the effect of an additional parameter, $D$, was undertaken. A value of $D$ = 1.0 was assigned to all N,N-disubstituted amides and $D$ = 0.0 to the mono- and unsubstituted compounds. The single parameter equations in $D$ and the best equations using log $P$, $\sigma^*$, $E_s$, and $D$ as variables are given in Table VIII.

The addition of the dummy parameter significantly altered the form of the optimum correlation equations (Tables VII and VIII) and effectively negated, or at least reduced, many of the concerns stated above relative to the equations in Table VII. Inclusion of this parameter gave best equations which were parabolic rather than linear in log $P$. The greatest number of compounds which had to be omitted from any series in order to obtain good correlations dropped from eight to three of the 44 studied. Inspection of the compounds which had to be dropped (Table I) shows that many of the same compounds (**2, 6, 12, 28, 31, 38, 48**) gave correlation problems in more than one bacterial system. This could be due to several factors, among which are poor approximations for the calculated constants and uncertain biological data due to possible metabolic transformations of some of the derivatives.†

The collinearity among the various parameters studied is seen in Table IX for one of the bacterial test systems.

**Table III.** Rifamycin Amides. Physicochemical Parameters

| Compd | $R_1$ | $R_2$ | $\Sigma E_s$ | Log $P$ | $\Sigma\sigma^*$ |
|---|---|---|---|---|---|
| **1** | H | H | 2.48 | −0.68 | 0.98 |
| 2 | H | Me | 1.24 | −0.18 | 0.49 |
| 3 | H | Et | 1.17 | 0.32 | 0.39 |
| 4 | H | *n*-Pr | 0.88 | 0.82 | 0.38 |
| 5 | H | *i*-Pr | 0.77 | 0.62 | 0.30 |
| 6 | H | *t*-Bu | −0.30 | 0.85 | 0.19 |
| 8 | H | Phenyl | −1.34 | 1.45 | 1.09 |
| 9 | H | *p*-Cl-Ph | −1.34 | 2.15 | 1.09 |
| 10 | H | *p*-Br-Ph | −1.34 | 2.31 | 1.09 |
| 11 | H | *p*-I-Ph | −1.34 | 2.57 | 1.09 |
| 12 | H | CH(OH)Me | 1.34 | −0.84 | 0.94 |
| 13 | Me | Me | 0.00 | 0.32 | 0.00 |
| 14 | Et | Et | −0.14 | 1.32 | −0.20 |
| 15 | *n*-Pr | *n*-Pr | −0.72 | 2.32 | −0.23 |
| 16 | *n*-Bu | *n*-Bu | −0.78 | 3.32 | −0.26 |
| 17 | *i*-Bu | *i*-Bu | −1.86 | 2.92 | −0.25 |
| 18 | *n*-Phenyl | *n*-Pentyl | −0.80 | 4.32 | −0.26 |
| 19 | Allyl | Allyl | −0.42 | 1.72 | 0.32 |
| 20 | Benzyl | Benzyl | −0.76 | 4.70 | 0.43 |
| 21 | Me | Et | −0.07 | 0.82 | −0.10 |
| 22 | Me | Pr | −0.36 | 1.32 | −0.11 |
| 23 | Me | *i*-Pr | −0.47 | 1.12 | −0.19 |
| 24 | Me | *n*-Bu | −0.39 | 1.82 | −0.13 |
| 25 | Me | *t*-Bu | −1.54 | 1.35 | −0.30 |
| 26 | Et | *n*-Pr | −0.43 | 1.82 | −0.22 |
| 27 | Et | *n*-Bu | −0.45 | 2.32 | −0.23 |
| 28 | *n*-Pr | Propynyl | −0.57 | 1.80 | 0.65 |
| 29 | Me | Cyclopentyl | −0.51 | 1.96 | −0.20 |
| 30 | Me | Cyclohexyl | −0.79 | 2.33 | −0.15 |
| 31 | Et | Phenyl | −2.65 | 2.45 | 0.50 |
| 32 | Me | Benzyl | −0.38 | 2.51 | 0.21 |
| 33 | Me | $-CH_2CH_2OH$ | −0.36 | −0.34 | 0.28 |
| 34 | Et | $-CH_2CH_2OH$ | −0.43 | 0.16 | 0.18 |
| 35 | Me | $-CH_2CH_2CN$ | −0.99 | −0.02 | 0.65 |
| 38 | $-CH_2CH_2$-Cl | $-CH_2CH_2Cl$ | −1.80 | 2.10 | 0.77 |
| 39 | $-(CH_2)_4-$ | | −0.14 | 0.92 | −0.20 |
| 40 | $-CH(CH_3)(CH_2)_2CH(CH_3)-$ | | −0.94 | 1.52 | −0.38 |
| 43 | $-(CH_2)_5-$ | | −0.43 | 1.32 | −0.22 |
| 44 | $-(CH_2)_2CH(CH_3)(CH_2)_2-$ | | −0.46 | 1.62 | −0.23 |
| 45 | $-(CH_2)_4CH(CH_3)-$ | | −0.83 | 1.62 | −0.30 |
| 46 | $-(CH_2)_6-$ | | −0.72 | 1.72 | −0.23 |
| 47 | $-(CH_2)_2O(CH_2)_2-$ | | −0.43 | 0.09 | 0.18 |
| 48 | $-(CH_2)_2OCH_2CH(CH_3)-$ | | −0.83 | 0.39 | 0.09 |
| 49 | $-CH(CH_3)CH_2OCH_2CH(CH_3)-$ | | −0.94 | 0.69 | 0.00 |

†Inclusion of the compounds designated as omitted (Table I) in the generation of eq 18, 20, 22, and 24 degraded the correlation statistics in all cases. In addition, a new equation with $-\sigma^*$ replacing $-(\log P)^2$ was generated which was statistically similar to the degraded equation containing the variables shown in eq 18, 20, 22 and 24. These new equations were poorly predictive for the two compounds prepared to test the correlation equations.

**Table IV.** Rifamycin Hydrazides. Physicochemical Parameters

| Compd | $R_1$ | $R_2$ | $R_3$ | $\Sigma E_s$ | Log $P$ | $\Sigma\sigma^*$ |
|---|---|---|---|---|---|---|
| 50 | H | H | H | 3.72 | −1.93 | 1.47 |
| 51 | H | H | Ph | −0.10 | 0.21 | 1.58 |
| 52 | Me | Me | Me | 0.00 | −0.43 | 0.00 |
| 53 | Me | Et | Et | −0.14 | 0.57 | −0.20 |
| 54 | Me | Pr | Pr | −0.72 | 1.57 | −0.23 |
| 55 | Me | Bu | Bu | −0.78 | 2.57 | −0.26 |
| 56 | Et | Me | Me | −0.07 | 0.07 | −0.10 |
| 57 | Et | Et | Et | −0.21 | 1.07 | −0.30 |
| 58 | Et | Pr | Pr | −0.79 | 2.07 | −0.33 |
| 59 | Et | Bu | Bu | −0.85 | 3.07 | −0.36 |
| 60 | Pr | Me | Me | −0.36 | 0.57 | −0.12 |
| 61 | Pr | Et | Et | −0.50 | 1.57 | −0.32 |
| 62 | Pr | Pr | Pr | −1.08 | 2.57 | −0.34 |
| 63 | Pr | Bu | Bu | −1.14 | 3.57 | −0.38 |
| 64 | Bu | Me | Me | −0.39 | 1.07 | −0.13 |
| 65 | Bu | Et | Et | −0.53 | 2.07 | −0.33 |
| 66 | Bu | Pr | Pr | −1.11 | 3.07 | −0.36 |
| 67 | Pentyl | Me | Me | −0.40 | 1.57 | −0.13 |
| 68 | Me | $-(CH_2)_5-$ | | −0.43 | 0.67 | −0.22 |
| 69 | Et | $-(CH_2)_5-$ | | −0.50 | 1.17 | −0.32 |
| 70 | Pr | $-(CH_2)_5-$ | | −0.79 | 1.67 | −0.34 |
| 71 | Bu | $-(CH_2)_5-$ | | −0.82 | 2.17 | −0.35 |
| 72 | Me | $-(CH_2)_2O(CH_2)_2-$ | | −0.43 | −1.26 | 0.18 |
| 73 | Et | $-(CH_2)_2O(CH_2)_2-$ | | −0.50 | −0.76 | 0.08 |
| 74 | Pr | $-(CH_2)_2O(CH_2)_2-$ | | −0.79 | −0.26 | 0.06 |
| 75 | Bu | $-(CH_2)_2O(CH_2)_2-$ | | −0.82 | 0.24 | 0.05 |

**Table V.** Calculation of $E_s$ Values for Cyclic Amides

| Compd | Ring[a] | Calcn method | $\Sigma E_s$ |
|---|---|---|---|
| 39 | Pyrrolidine | 2 × Et | −0.14 |
| 40 | 2,5-Dimethylpyrrolidine | 2 × *i*-Pr | −0.94 |
| 43 | Piperidine | Et + Pr | −0.43 |
| 44 | 4-Methylpiperidine | Bu + Et | −0.46 |
| 45 | 2-Methylpiperidine | Pr + *i*-Pr | −0.83 |
| 46 | Azepine | 2 × Pr | −0.72 |
| 47 | Morpholine | Pr + Et | −0.43 |
| 48 | 3-Methylmorpholine | Pr + *i*-Pr | −0.83 |
| 49 | 3,5-Dimethylmorpholine | 2 × *i*-Pr | −0.54 |

[a] The N atom of the ring is the amide nitrogen.

**Table VI.** Equations Generated for Rifamycin Amides in *M. aureus*

| Equations | $n$ | $r$ | $s$ | Eq no. |
|---|---|---|---|---|
| Log $1/C$ = 7.48 (±0.26) − 0.42 (±0.25) $E_s$ | 41 | 0.481 | 0.700 | 5 |
| Log $1/C$ = 7.10 (±0.34) + 0.39 (±0.18) log $P$ | 41 | 0.574 | 0.653 | 6 |
| Log $1/C$ = 7.90 (±0.18) − 1.29 (±0.38) $\sigma^*$ | 41 | 0.742 | 0.535 | 7 |
| Log $1/C$ = 7.13 (±0.34) − 0.21 (±0.27) $E_s$ + 0.30 (±0.21) log $P$ | 41 | 0.607 | 0.642 | 8 |
| Log $1/C$ = 7.68 (±0.15) − 0.41 (±0.14) $E_s$ − 1.27 (±0.28) $\sigma^*$ | 41 | 0.874 | 0.392 | 9 |
| Log $1/C$ = 7.85 (±0.21) − 1.54 (±0.76) $\sigma^*$ + 0.36 (±0.94) $(\sigma^*)^2$ | 41 | 0.746 | 0.538 | 10 |
| Log $1/C$ = 6.94 (±0.39) + 0.69 (±0.42) log $P$ − 0.08 (±0.10) $(\log P)^2$ | 41 | 0.610 | 0.640 | 11 |
| Log $1/C$ = 7.38 (±0.21) + 0.33 (±0.11) log $P$ − 1.18 (±0.27) $\sigma^*$ | 41 | 0.882 | 0.380 | 12 |
| Log $1/C$ = 7.41 (±0.18) − 0.25 (±0.15) $E_s$ + 0.22 (±0.11) log $P$ − 1.20 (±0.24) $\sigma^*$ | 41 | 0.915 | 0.331 | 13 |

Molar refractivity (MR) is very highly correlated with log $P$ and is strongly correlated with $E_s$. The problem of multiple collinearity is a difficult one[15] which is often quite serious in bio- or medicinal chemical QSAR. The problem can be largely circumvented by careful selection of the substituents to be used in drug modification.[16] The dummy parameter, $D$, which corresponds to mono- or disubstitution on the amide nitrogen, is strongly correlated with $\sigma^*$. This correlation between $\sigma^*$ and $D$ is further confirmed by the fact that $\sigma^*$ is the most important single variable in non-$D$-containing equations, while $D$ is the most important single variable in equations containing that term.

The sign of the coefficient with $\sigma^*$ is always negative which means that activity increases with increasing electron density on the amide nitrogen. Since $D$ and $\sigma^*$ are so highly collinear, $D$ may also, to a considerable extent, reflect an increase in the electron density on nitrogen. This may represent some special hydrogen bonding ability.

The most useful correlations appear to be derived from the equations containing a $(\log P)^2$ term (eq 18, 20, 22, 24). These equations indicate that activity is a parabolic function of log $P$ and predict the following ideal lipophilic values (log $P_0$): *S. faecalis*, 3.68 (3.0–5.4); *M. aureus*, 3.94 (2.8–13); and *S. hemolyticus*, 2.37 (1.9–3.2). It was not possible to place 95% confidence limits[17] on log $P_0$ (6.97) for *B. subtilis*. Ideal lipophilic character can be seen to vary considerably with the type of microorganism. Barbaro, *et al.*, performed thin-layer chromatography on rifamycin B, ri-

**Table VII.** Rifamycin B Amide Correlation Equations

| System | Equation | $n$ | $r$ | $s$ | Eq no. |
|---|---|---|---|---|---|
| *M. aureus* | Log $1/C$ = 7.41 (±0.18) − 0.25 (±0.15) $E_s$ + 0.22 (±0.11) log $P$ − 1.20 (±0.24) σ* | 41 | 0.915 | 0.331 | 13 |
| *B. subtilis* | Log $1/C$ = 5.74 (±0.14) − 0.15 (±0.10) $E_s$ + 0.27 (±0.08) log $P$ − 0.62 (±0.17) σ* | 39 | 0.924 | 0.239 | 14 |
| *S. faecalis* | Log $1/C$ = 6.38 (±0.16) − 0.22 (±0.13) $E_s$ + 0.31 (±0.10) log $P$ − 0.78 (±0.22) σ* | 41 | 0.912 | 0.300 | 15 |
| *S. hemolyticus* | Log $1/C$ = 7.83 (±0.10) − 0.93 (±0.20) σ* | 36 | 0.858 | 0.276 | 16 |

**Table VIII.** Equations Utilizing a Disubstituted Amide Dummy Parameter

| | $n$ | $r$ | $s$ | Eq no. |
|---|---|---|---|---|
| *M. aureus* | | | | |
| Log $1/C$ = 6.63 (±0.28) + 1.46 (±0.33) $D$ | 42 | 0.820 | 0.426 | 17 |
| Log $1/C$ = 6.20 (±0.26) + 0.51 (±0.23) log $P$ − 0.06 (±0.06) (log $P$)$^2$ + 1.30 (±0.24) $D$ | 42 | 0.920 | 0.324 | 18 |
| *B. subtilis* | | | | |
| Log $1/C$ = 5.42 (±0.28) + 0.96 (±0.32) $D$ | 41 | 0.690 | 0.456 | 19 |
| Log $1/C$ = 5.04 (±0.19) + 0.41 (±0.16) log $P$ − 0.03 (±0.04) (log $P$)$^2$ + 0.76 (±0.20) $D$ | 41 | 0.909 | 0.270 | 20 |
| *S. faecalis* | | | | |
| Log $1/C$ = 6.05 (±0.38) + 1.00 (±0.43) $D$ | 42 | 0.599 | 0.569 | 21 |
| Log $1/C$ = 5.43 (±0.23) + 0.74 (±0.19) log $P$ − 0.10 (±0.05) (log $P$)$^2$ + 0.79 (±0.23) $D$ | 42 | 0.915 | 0.294 | 22 |
| *S. hemolyticus* | | | | |
| Log $1/C$ = 6.74 (±0.28) + 1.23 (±0.31) $D$ | 41 | 0.786 | 0.393 | 23 |
| Log $1/C$ = 6.40 (±0.28) + 0.48 (±0.22) log $P$ − 0.10 (±0.05) (log $P$)$^2$ + 1.18 (±0.26) $D$ | 41 | 0.866 | 0.325 | 24 |

**Table IX.** Correlation Coefficients ($r^2$) between Regression Parameters for Rifamycin B Amides in the *S. faecalis* System

| | $E_s$ | Log $P$ | σ* | MR | $D$ |
|---|---|---|---|---|---|
| $E_s$ | 1.00 | 0.29 | 0.01 | 0.46 | 0.15 |
| Log $P$ | | 1.00 | 0.02 | 0.86 | 0.05 |
| σ* | | | 1.00 | 0.01 | 0.55 |
| MR | | | | 1.00 | 0.07 |
| $D$ | | | | | 1.00 |

famycin SV, and four related compounds of this series.[18] They also obtained a correlation which showed a parabolic dependence of *S. aureus* activity on the log $P$ related chromatographic parameter $R_M$.

The fact that equations of the type $\pi + \pi^2 + D$ correlate both mono- and diamides implies that with each type of amide there is a relatively simple parabolic dependence of activity on log $P$. $D$ is the correction factor which merges the two parabolas. To a certain extent this can be confirmed by factoring the data into two sets and comparing regression equations for the two sets. Unfortunately, the much smaller sets of monoamides do not constitute as well balanced a series as the diamides. While *S. aureus*, *S. faecalis*, and *B. subtilis* give roughly similar parabolas for mono- and disubstituted amides, *S. hemolyticus* does not. The correlation in this system is quite poor with the monoamides and from every point of view, this organism gives the poorest correlations.

At this point in the study, two statistically similar sets of correlation equations had been developed for the rifamycin B amides (Tables VII and VIII). However, the two sets predicted activity as a different function of the partition coefficient (linear and parabolic). In an attempt to determine which equation set had the most predictive utility as a possible guide for future synthetic modifications, two additional derivatives were designed, prepared, and tested. (These compounds were prepared and tested by Gruppo Lepetit, Milan, Italy, under the direction of Dr. G. C. Lancini.) Both equation sets make the same qualitative prediction, *i.e.*, that *N,N*-dioctyl and *N,N*-dicyclohexyl rifamycin B amide derivatives should be active in the bacterial systems studied. However, they predict different quantitative results. A comparison of the predictive ability of the equations using linear and parabolic log $P$ functions is given in Table X. Both derivatives were found to have very good ac-

**Table X.** Comparison of Predictive Ability of Equations for the Activity of *N,N*-Dioctyl- and *N,N*-Dicyclohexylamides of Rifamycin B

| | | | | | $D$[a] | | |
|---|---|---|---|---|---|---|---|
| System | Eq | $n$ | $r$ | $s$ | Di-$C_8$ | Di-$C_6$ | SD[b] |
| *M. aureus* | 13 | 41 | 0.915 | 0.331 | −2.66 | −0.94 | 0.79 |
| *M. aureus* | 18 | 42 | 0.920 | 0.324 | −0.87 | −0.31 | 0.80 |
| *S. faecalis* | 15 | 42 | 0.903 | 0.318 | −3.16 | −1.12 | 0.71 |
| *S. faecalis* | 22 | 42 | 0.915 | 0.294 | −0.36 | −0.35 | 0.71 |
| *S. hemolyticus* | 16 | 36 | 0.856 | 0.278 | −0.50 | −0.51 | 0.53 |
| *S. hemolyticus* | 24 | 41 | 0.866 | 0.325 | 1.93 | −0.17 | 0.63 |

[a] Deviation of predicted and experimentally determined −log $C$ values. [b] Standard deviation for $n$ compounds in the data set.

**Table XI.** Rifamycin B Hydrazide Correlation Equations

| System | Equation | $n$ | $r$ | $s$ | Eq no. |
|---|---|---|---|---|---|
| *S. faecalis* | Log $1/C$ = 7.47 (±0.08) − 0.96 (±0.15) $\sigma^*$ | 24 | 0.943 | 0.178 | 25 |
| *M. aureus* | Log $1/C$ = 8.42 (±0.08) − 0.96 (±0.16) $\sigma^*$ | 24 | 0.937 | 0.189 | 26 |
| *S. hemolyticus* | Log $1/C$ = 8.38 (±0.09) − 0.96 (±0.16) $\sigma^*$ − 0.65 (±0.22) $A$ | 23 | 0.929 | 0.190 | 27 |
| *B. subtilis* | Log $1/C$ = 6.74 (±0.14) − 0.81 (±0.25) $\sigma^*$ + 0.14 (±0.10) log $P$ | 24 | 0.936 | 0.222 | 28 |
| *M. tuberculosis* | Log $1/C$ = 6.58 (±0.09) + 0.40 (±0.16) $\sigma^*$ + 0.17 (±0.06) log $P$ + 0.44 (±0.17) $A$ | 24 | 0.872 | 0.144 | 29 |

tivity. The equations with $(\log P)^2$ terms, however, did a significantly better job of quantitatively predicting the activity in the *M. aureus* and *S. faecalis* systems than did the equations with linear log $P$ terms. The two new compounds were not tested against *B. subtilis*.

The structure–activity properties of the *S. hemolyticus* system appear to obey a different type of relationship. Neither the equation with $D$ (eq 24) nor without $D$ (eq 16) gave as good a correlation of the *S. hemolyticus* data as was found for the other bacterial systems. Equation 16 showed that only $\sigma^*$ was significant, while the comparable equations for the other bacterial systems also contain $E_s$ and log $P$ terms. The equation in $D$ (eq 24) does well in predicting the activity of the *N,N*-dicyclohexylamide but fails badly for the dioctyl analog. The reason for this inconsistency is not apparent. The equation in $\sigma^*$ alone (eq 16) predicts the activity of both new derivatives quite well. As seen from the negative values of the deviations in Table X, the equations tend to overpredict activity.

As mentioned earlier in the Methods section, compounds **7**, **36**, **37**, **41**, and **42** in the original study[1] were omitted from this investigation because of difficulties in the estimation of steric and electronic constants. When it became evident that equations which did not contain these variables (eq 18, 20, 22, 24) correlated the data for the other derivatives, log $P$ values were calculated for these five compounds and the data were reexamined with these materials included. Compounds **7**, **41**, and **42** were well correlated while **36** and **37** were poorly correlated by the four equations in log $P$, $(\log P)^2$, and $D$. It can be noted that the two compounds which did not correlate contain additional aliphatic amine groups which might be protonated.

**Rifamycin B Hydrazides.** Data for 26 hydrazides (compounds **50–75**) were available for study.[1] Once again $\sigma^*$, $E_s$, and log $P$ were used as calculated variables. Because of the approximations used in the calculation of the cyclic hydrazide derivatives, a dummy parameter, $A$, was also employed. It was given a value of 1.0 for all hydrazides containing the piperidine ring (**68–71**) and 0.0 for the other 22 derivatives. The application of the dummy parameter, $A$, to the morpholine analogs **72–75** provided no advantage in the development of a better equation.[‡] The best statistically significant equations are shown in Table XI.

The correlation equations for the hydrazides are considerably different from those derived for the amides (Tables VII and VIII). The single parameter $\sigma^*$ accounts for most of the variance in log $1/C$. On closer consideration of the facts, this may not be true in a fundamental sense. From Table XII it is seen that $\sigma^*$ and log $P$ are strongly collinear. Moreover, since for each organism the coefficient with $\sigma^*$ is negative, an increase in electron density on nitrogen increases activity. As pointed out above for the amides, the main function of increased electron density on nitrogen may simply be to increase hydrophilic character through some kind of conformational effect. The more electronegative hydrazine function appears to make this the dominant effect for the hydrazides. The data in two bacterial systems (eq 25 and 26) are correlated by $\sigma^*$ alone. The dummy parameter $A$ helped produce the most significant equation in the *S. hemolyticus* system. The negative coefficient of $A$ indicates the deleterious effect of the piperidine function. The dummy parameter also helped produce a reasonable hydrazide correlation in the *M. tuberculosis* system which was not possible for the amides. Emphasizing the difference between *M. tuberculosis* and the other bacterial systems, however, are the positive coefficients for $\sigma^*$ and $A$ (eq 29). A small but significant ($F_{1,21} = 10.0$) improvement occurred with the addition of a log $P$ term for *B. subtilis*.

It is difficult to draw any firm conclusions about the hydrazide QSAR. The large amount of collinearity between $E_s$ and log $P$, $\sigma^*$ and $E_s$, and log $P$ and $\sigma^*$ seriously obscures the details of the QSAR.

**Table XII.** Correlation ($r^2$) between Regression Parameters for Rifamycin B Hydrazides in the *S. faecalis* System

| | $E_s$ | Log $P$ | $\sigma^*$ | $A$ |
|---|---|---|---|---|
| $E_s$ | 1.00 | 0.47 | 0.52 | 0.02 |
| Log $P$ | | 1.00 | 0.42 | 0.00 |
| $\sigma^*$ | | | 1.00 | 0.04 |
| $A$ | | | | 1.00 |

## Conclusions

Our results not only substantiate and quantitate the qualitative conclusions of Sensi, *et al.*, but also they bring to light the limiting value of hydrophobic character as defined by log $P_{oct}$. It can be concluded, at least for *in vitro* conditions, that a combination of two amides, one with log $P$ of about 3.0 and one with log $P$ of 5.0 or higher, would be effective against a mixture of all four microorganisms studied.

It is very important to consider the collinearity of structural parameters when deciding which analogs to prepare. This can be illustrated by the results of this investigation. Statistically acceptable correlation equations were obtained for both the rifamycin B amides and hydrazides. The predictive usefulness of the hydrazide equations is severely limited, however, by the significant collinearity among the correlation parameters which, in turn, are dependent upon the functional groups chosen for study. A high degree of collinearity precludes a clear separation of the roles of the various substituents in the QSAR.

While the amide and hydrazide functions are undoubtedly not the intrinsic pharmacophores in the rifamycin B antibiotics, it can be seen that these groups greatly affect drug potency, probably by changes in transport properties. It can also be stated that the antibacterial activity of molecules with the structural complexity of the rifamycins can be quantitatively correlated with structural parameters.

**Acknowledgment.** The authors wish to thank Dr. Harry B. Wood, Jr., and Mrs. Nancita Lomax for their help dur-

‡ Inclusion of compounds designated as omitted (Table II) in the generation of eq 25–29 degraded the correlation statistics and produced no significant new equations of altered format.

ing this study. They also wish to thank Dr. G. C. Lancini, Gruppo Lepetit, Milan, Italy, and Dr. C. W. Hinman, Dow Chemical Co., Midland, Mich., for the synthesis and testing of the dioctyl and dicyclohexyl rifamycin amide derivatives.

### References and Notes

(1) P. Sensi, N. Maggi, R. Balotta, S. Furesz, R. Pallanza, and V. Arioli, *J. Med. Chem.*, 7, 596 (1964).
(2) C. Hansch in "Drug Design," Vol. I, E. J. Ariens, Ed., Academic Press, New York, N.Y., 1971, p 271.
(3) M. S. Tute, *Advan. Drug Res.*, 6, 1 (1971).
(4) C. Daniel and F. S. Wood, "Fitting Equations to Data," Wiley-Interscience, New York, N.Y., 1971.
(5) G. Pelizza, G. C. Lancini, G. C. Allievi, and G. G. Gallo, *Farmaco. Ed. Sci.*, 28, 298 (1973).
(6) R. C. Gallo, J. W. Abrell, M. S. Robert, S. S. Yang, and R. G. Smith, *J. Nat. Cancer Inst.*, 48, 1185 (1972).
(7) W. J. Haggerty, "Assay of Experimental Medicinal Chemicals and Their Pharmaceutical Formulations," Contract NO1-CM-33722, Progress Report No. 17, Nov 1973.
(8) A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, 71, 525 (1971).
(9) C. Hansch, J. E. Quinlan, and G. L. Lawrence, *J. Org. Chem.*, 33, 347 (1968).
(10) C. Hansch, unpublished data.
(11) R. W. Taft, "Steric Effects in Organic Chemistry," Wiley, New York, N.Y., 1956, p 559.
(12) I. V. Talvik and V. A. Palm, *Reakts. Sposobnost Org. Soedin.*, 8, 445 (1971).
(13) H. K. Hall, Jr., *J. Amer. Chem. Soc.*, 79, 5441 (1957).
(14) C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien, *J. Med. Chem.*, 16, 1207 (1973).
(15) (a) D. E. Farrar and R. R. Glauber, *Rev. Econ. Stat.*, 49, 92 (1967); (b) Y. Haitovsky, *ibid.*, 51, 486 (1969).
(16) C. Hansch, S. H. Unger, and A. B. Forsythe, *J. Med. Chem.*, 16, 1217 (1973).
(17) C. Hansch, A. R. Steward, S. M. Anderson, and D. L. Bentley, *J. Med. Chem.*, 11, 1 (1968).
(18) A. M. Barbaro, M. C. Guerra, and G. L. Biagi, *Bull. Soc. Ital. Biol. Sper.*, 47, 556 (1971).

# *ara*-Cytidine Acylates. Use of Drug Design Predictors in Structure–Activity Relationship Correlation[1,2]

W. J. Wechter,* M. A. Johnson, C. M. Hall, D. T. Warner, A. E. Berger,[3] A. H. Wenzel,[4] D. T. Gish, and G. L. Neil

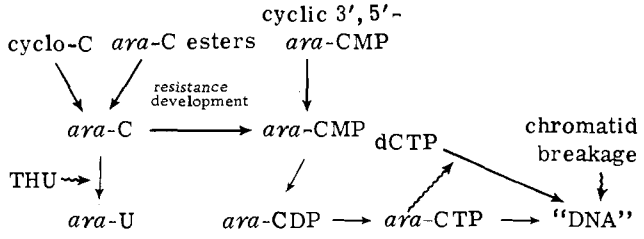
*Research Division, The Upjohn Company, Kalamazoo, Michigan 49001.* *Received September 4, 1974*

This manuscript is one of a series of investigations into modifying the pharmacologic properties of the antitumor, antiviral, and immunosuppressive nucleoside *ara*-cytidine (cytarabine, Cytosar). The present paper summarizes our studies on depot ester derivatives of the nucleoside. We are able to predict with reasonable accuracy the biological activity as measured by increased life span in the L1210 leukemic mouse from a combination of two predictor variables: (1) the solubility of the ester in water and (2) its rate of hydrolysis by the mixed esterase system of animal plasma. We have tried unsuccessfully to correlate enzymatic hydrolysis rates with an alkaline hydrolysis model. Calculated Hansch partition ($p$) values had a correlation of $r = 0.86$ with water solubility. These $p$ values had no additional predictive value. Based on our results, two esters were selected for clinical trial in cancer and rheumatoid arthritis.

Since the demonstration of the antitumor, antiviral, and immunosuppressive effects of the nucleoside *ara*-cytidine (Cytosar, cytarabine) (for leading references, see ref 1), we have attempted to modify the pharmacological properties of this unique nucleoside. We have tried to increase its potency, modify its catabolism to *ara*-uridine, develop depot and orally active forms, increase its specificity, obviate resistance development, modify its biological profile, and understand its mechanism of action. Some success has been achieved in reaching all of these objectives.[2] This paper deals exclusively with our development of a depot form of the drug employing *in vitro* correlates for the design of a drug for clinical application in cancer and rheumatoid arthritis, the latter to be effective as a locally administered (intraarticular) immunosuppressive agent in rheumatoid joints.

Early efforts by members of our group with synthetic dinucleoside phosphates containing *ara*-C led to minor increases in potency[5,6] and no apparent improvement in specificity. Synthesis of the potent cytidine aminohydrolase inhibitor tetrahydrouridine[7] provided a compound which, given in combination with *ara*-cytidine, afforded increased plasma half-life of the drug by blockage of the facile catabolic pathway *via* deamination to *ara*-uridine. The major impetus to the present work was the dramatically enhanced activity of the 5′-adamantoate ester[8] over *ara*-cytidine itself when administered as a single dose. This, in turn, led to the examination of a host of acylates at C-2′, C-3′, and C-5′. In three earlier publications[9-11] the synthesis and biological activity of the majority of the esters were reported. In this paper we attempt to analyze the structural requirements for activity for the purpose of the design and synthesis of additional esters employing our predictor correlates. We also prepared some diesters for evaluation. From an understanding of the biochemistry of *ara*-cytidine (Scheme I) and our rudimentary theses for drug design, we

**Scheme I**




felt that the four most important variables with which we would have to deal were (1) the deamination to the inactive *ara*-uridine esters; (2) the dissolution of the drug *in vivo*; (3) its transport and distribution; and (4) its *in vivo* hydrolysis to the active species, *ara*-cytidine, which necessarily must be the precursor of the active drug (see Scheme I) *ara*-CTP in all cases. Early work with the deaminase enzyme of human serum established that the esters were not substrates. Consequently, this factor could be discarded as a design parameter.[9] As an *in vitro* correlate of dissolution we chose, as an approximation, water solubility. The Hansch $p$ values were selected as the corresponding correlate of transport and distribution. Our most thoroughly in-