assumed to be 1.5, the pK_a of compound *C*. On the other hand, the pK_a of protonated alprazolam was determined from the initial absorbance at 310 nm after a methanolic solution was diluted to a given pH (Fig. 6); the pK_a^{AH} was 2.40. Note that at 310 nm the opened-ring compound or its analogue *C* absorbed very little UV energy (Fig. 2) and the pH dependence shown on Fig. 6 should reflect the protonation at N-4. From the values of equilibrium pK_a, pK_a^{AH}, and pK_a^{BH2}, the fractional concentration of important ionic species involved in the $A \rightleftharpoons B$ equilibrium was calculated (Fig. 7).

The concentration of the opened-ring compound was also estimated experimentally. An alprazolam solution in methanol was first acidified with HCl and left to attain equilibrium. The mixture was then neutralized with an equivalent amount of hydroxide ion and immediately back-titrated. The titration end point should represent the amount of the opened-ring compound which was generated from the acid treatment. A critical assumption was that during the neutralization and the titration no significant $A \rightleftharpoons B$ reaction occurred. Since the titration end point occurs at a pH betweeen 4 and 6, alprazolam was never protonated in the titration. Three determinations of the total opened-ring compound by this rapid titration technique are shown on Fig. 7 (open circles). Since the titration system contained as much as 10% (v/v) methanol, the total concentration of the opened-ring compound recovered at a given pH was expected to be much lower than in the absence of methanol. In this particular instance, not only should the activity of water have been lower but the pK_a^{AH} was also expected to decrease significantly to result in a displacement of pH profile for the fractional concentration of opened-ring compound toward lower pH.

REFERENCES

(1) M. L. Bender, *Chem. Rev.*, **60**, 53 (1960).

(2) W. P. Jencks, *Prog. Phys. Org. Chem.*, **2**, 63 (1964).

(3) S. Johnson, *Adv. Phys. Org. Chem.*, **5**, 237 (1964).

(4) R. B. Martin, *J. Phys. Chem.*, **68**, 1369 (1964).

(5) T. C. Bruice and S. Benkovic, "Bioorganic Mechanisms," vol. 1, Benjamin, New York, N.Y., 1966, pp. 27–118.

(6) E. G. Sander and W. P. Jencks, *J. Am. Chem. Soc.*, **90**, 6154 (1968).

(7) J. Hine and R. W. Redding, *J. Org. Chem.*, **35**, 2769 (1970).

(8) M. J. Cho and I. H. Pitman, *J. Am. Chem. Soc.*, **96**, 1843 (1974).

(9) D. D. Perrin, *Adv. Heterocycl. Chem.*, **4**, 43 (1965).

(10) L. H. Sternbach, E. Reeder, and G. A. Archer, *J. Org. Chem.*, **28**, 2456 (1963).

(11) J. T. Carstensen, K. S. E. Su, P. Maddrell, J. B. Johnson, and H. N. Newark, *Bull. Parenter. Drug Assoc.*, **25**, 193 (1971).

(12) W. Mayer, S. Erbe, and R. Voight, *Pharmazie*, **27**, 32 (1972).

(13) W. Mayer, S. Erbe, G. Wolf, and R. Voight, *ibid.*, **29**, 700 (1974).

(14) H. V. Maulding, J. P. Nazareno, J. E. Pearseen, and A. F. Michaelis, *J. Pharm. Sci.*, **64**, 278 (1975).

(15) W. W. Han, G. J. Yakatan, and D. D. Maness, *ibid.*, **65**, 1198 (1976).

(16) *Idem.*, **66**, 573 (1977).

(17) *Idem.*, **66**, 795 (1977).

(18) C. H. Hassall, S. W. Holmes, W. H. Johnson, A. Krohn, C. E. Smithen, and W. A. Thomas, *Experientia*, **33**, 1492 (1977).

(19) M. Nakano, N. Inotsume, N. Kohri, and T. Arita, *Int. J. Pharm.*, **3**, 195 (1979).

(20) N. Inotsume and M. Nakano, *J. Pharm. Sci.*, **69**, 1331 (1980).

(21) N. Inotsume and M. Nakano, *Chem. Pharm. Bull.*, **28**, 2536 (1980).

(22) D. D. Perrin, *Aust. J. Chem.*, **16**, 572 (1963).

(23) A. Albert and E. P. Serjeant, "Ionization Constants," Wiley, New York, N.Y., 1962.

(24) W. P. Jencks, "Catalysis in Chemistry and Enzymology," McGraw-Hill, New York, N.Y., 1969, pp. 463–477.

(25) J. B. Hester, Jr., D. J. Duchamp, and C. G. Chidester, *Tetrahedron Lett.*, **1971**, 1609.

(26) T. A. Scahill, Ph.D. thesis, University of Kentucky, 1981.

(27) J. B. Hester, Jr., C. G. Chidester, and J. Szmuszkovicz, *J. Org. Chem.*, **44**, 2688 (1979).

(28) J. Barrett, W. F. Smyth, and I. E. Davidson, *J. Pharm. Pharmacol.*, **25**, 387 (1973).

(29) J. March, "Advanced Organic Chemistry: Reactions, Mechanisms, and Structure," McGraw-Hill, New York, N.Y., 1968, pp. 228–230, and references cited therein.

ACKNOWLEDGMENTS

A part of the text was presented at the Higuchi Symposium, Orlando, Fla., November 1981.

M. J. Cho thanks Dr. B. D. Anderson and Dr. W. C. Krueger for helpful discussions.

Factorial Designs in Pharmaceutical Stability Studies

SANFORD BOLTON

Received October 26, 1981, from the *College of Pharmacy and Allied Health Sciences, St. John's University, Jamaica, NY 11439.* Accepted for publication May 27, 1982.

ABSTRACT □ An approach to analyzing and interpreting kinetic data from stability studies using factorial designs is presented. This may be useful for screening purposes or as an aid in identifying significant effects in complex systems. A typical 2^n factorial experiment is discussed, and methods of variance estimation and statistical testing are presented. An example of simulated data is used to demonstrate how typical results may be analyzed, as well as the potential and limitations of this design in interpretation and construction of kinetic models.

Keyphrases □ Factorial designs—in pharmaceutical stability studies, kinetic models, statistical analysis □ Stability studies, pharmaceutical—factorial designs, kinetic models, statistical analysis

Factorial designs are extremely useful in a wide variety of experimental situations, and applications of these designs to pharmaceutical problems have appeared in the recent literature (1–3). Factorial designs applied to stability studies of pharmaceuticals can be used for screening purposes or to help interpret complex systems. This paper deals with an approach to the design and statistical analysis of such experiments.

BACKGROUND

A factorial experiment considers the effects of various *factors* (*e.g.*, temperature, pH, drug concentration, buffer concentration) at several *levels* (*e.g.*, 2 pHs, one high pH and one low pH) where results of all combinations of the factor levels are observed. Modifications of the complete factorial design may be used in situations where it is not convenient or possible to do all of the combinations or trials (4). For this presentation, only experiments with all factors at two levels, a 2^n factorial design (where n is the number of factors, the effects of which are to be investigated), will be considered. The *main effect* is the difference in response (*e.g.*, rate constant) caused by the change in level of a factor (pH,

0022-3549/83/0400-0362$01.00/0

for example) averaged over all levels of the other factors. This has meaning in a practical sense if the effect of pH is not dependent on the levels of the other factors. If the effect of pH is dependent on the level of another factor (buffer, for example) an interaction between pH and buffer is said to exist. In this case a description of the effect of the factor, pH, would not be complete without consideration of the buffer level. This concept may be extended to higher order interactions. Thus, if an AB interaction exists and is dependent on the level of, say, factor C, the level of C should be specified when describing the AB interaction. This latter situation describes a 3-factor ABC, interaction.

A 2^3 factorial experiment with factors A, B, and C each at two levels consists of the following eight trials using the usual notation: (1), a, b, ab, c, ac, bc, abc. In this case (1) refers to all factors at their low level, a refers to the experiment with factor A at the high level and B and C at low levels, *etc.* The main effect of A is computed as $\frac{1}{4}$ [{a + ab + ac + abc} − {(1) + b + c + bc}]; the interaction AB is $\frac{1}{4}$ [{(1) + ab + c + abc} − {a + b + ac + bc}]; *etc.* Thus, the results of all experiments are used to calculate each main effect and interaction.

A relatively simple method of calculating the effects, is described by Davies (4).

THEORETICAL

The usual approach to kinetic studies is to examine, one at a time, the factors that are thought to affect the reaction rate. In many cases this approach is sufficient, since either some fundamental relationship is to be examined, or the effect of the various factors may be fairly well understood. One disadvantage of this experimental method is that, when present, interactions of factors may not be observed. Usually, little or no extra effort is required to uncover all possible effects by use of the factorial design. Factorial experiments are particularly advantageous in situations in which (*a*) the effects of several factors (and their interactions) are to be determined simultaneously in the absence of prior information, (*b*) a relatively small preliminary screening experiment is desired to obtain an estimate of the magnitude of effects of various variables, (*c*) the effects of factors are determined in a complex experimental situation (*e.g.*, many simultaneously varied factors), and (*d*) the experimental error (*e.g.*, assay) is relatively large. Statistical techniques may then be applied to test the significance of the main effects and interactions.

In typical kinetic studies, effects are usually additive. Thus, the following might describe a kinetic model:

$$k_{obs} = k_b[B] + k_c[C] + \ldots$$

in which [B] and [C] might represent hydroxide ion and buffer concentration, respectively. One objective is to estimate the rate constants, k_b, k_c, *etc.*

Since the usual statistical analysis is based on additivity of effects, it is important to consider carefully any analyses based on a model that also shows multiplicative relationships. For example, if ionic strength were a factor, the model could be of the following form:

$$k_{obs} = f(A)\{k_b[B] + k_c[C] + \ldots\}$$

in which $f(A)$ is a function of ionic strength.

If interaction of the multiplicative factor A (*e.g.*, ionic strength) with the additive factors B and C is present[1], the usual statistical analysis will be difficult, if not impossible, to interpret. In these cases, it is recommended that two separate analyses be performed, one at each level of A. Then, the effects of B and C can be estimated under the experimental conditions, *i.e.*, A constant at either the low or high level.

If interactions with A are absent, then the usual ANOVA of the rate constants (suitably weighted as described below) would, in general, provide a valid statistical test for the significance of the BC interaction only. The effects of the additive factors will be confounded with the multiplicative effects, and the effects of the additive factors should be evaluated separately at each level of A.

The factorial analysis of ln k (if CV is small and the k values are suitably weighted) will result in approximately valid statistical tests for A, AB, AC, and ABC under the null hypothesis that AB = AC = 0 (A is assumed not to interact with the additive factors B and C). In the absence of these interactions, the antilog of the A effect approximates the multiplicative effect of A (if CV is small).

Otherwise, the usual assumptions for ANOVA are considered to hold, as discussed later in this paper.

Because of the above implications, some knowledge of the functional relationships of the factors is helpful to come to meaningful conclusions. This knowledge may come from inspection of the data or from prior experience.

Statistical Considerations—In addition to the additivity concept discussed above, other assumptions inherent in the usual analysis of factorial experiments are that the errors are normally and independently distributed, and the variance (σ^2) is the same for each observation (homoscedasticity). In general, this latter assumption does not hold for experimentally determined rate constants, and the treatment of the data needs special consideration.

The variation or error in these studies arises from several factors, among which are assay error, other manipulative errors, and, possibly, the fact that the theoretical functional relationships such as linearity are not exactly satisfied. Measurement of the variance may be checked at several points during the study. Thus, it is useful during the assay development to check the assay variation at several concentrations of intact drug. Often the standard deviation is found to be proportional to the concentration of drug, *i.e.*, CV is constant; ($CV = S/C$, where S is the standard deviation and C is the concentration.) If the CV is constant and not too large, the logarithms of the assayed concentrations will have approximately equal variance at different concentration levels of drug (see *Appendix*). In this case the slope of the ordinary least-squares fit of the line, ln C *versus* time (first-order plot), is an unbiased estimate of the rate constant, k. The remainder of the discussion will be based on the assumption that the CV is constant and the reactions follow first-order kinetics. Certain modifications of the following analysis may be necessary if the kinetics are other than first order (see *Appendix*).

The variance of concern is the variation of ln C, σ^2 (which is approximately constant). Two estimates of σ^2 are available: (*a*) an estimate of the variance of ln C from replicates of assays of intact drug and (*b*) the variance estimate obtained from the least-squares line fit. If the error is due only to assay-related variation, these two estimates should show good agreement. In general, there will be sources other than assay variation contributing to the experimental error, and the line fitting should give a more realistic estimate of the variance. Another check on the error estimate could be obtained from replicate runs. (A run consists of determining the concentration as a function of time at specified levels of the factors.) Each set of replicates will contribute $n - 1$ degrees of freedom (df) to the error estimate in which n is the number of replicates. The details of this calculation for replicate runs (*i.e.*, repeat assessments of k with factor levels the same) are described in the *Appendix*.

Other studies[2] have shown excellent agreement of the variance estimate using replicates and line fitting. The line fitting approach has distinct advantages compared with replicates, since less runs are needed and more degrees of freedom for error are available.

An independent estimate of the variance as described is particularly important in this type of study. Higher order interaction terms (third order or higher) are often assumed to be nonexistent and are used to obtain estimates of the error. This approach has the disadvantage of yielding few degrees of freedom for error and possibly making false assumptions about the existence of such interactions. Most importantly, the computation of an independent variance estimate enables us to calculate the significance of the effects and all interactions in a relatively uncomplicated manner without prior consideration of the magnitude of interaction terms. As previously mentioned, the usual analysis of factorial designs assumes the variance of the observations to be equal. Under conditions usually present in kinetic studies this will very rarely be the case. The variance of a slope (the rate constant in this case) is $\sigma^2/\Sigma(t - \bar{t})^2$, in which t is the time at which the sample is assayed. Since $\Sigma(t - \bar{t})^2$ will, in general, be different for the different runs, the situation is one of variance inequality. The variance of ln k is $\sim\sigma^2/k^2\Sigma(t - \bar{t})^2$, if σ^2/k^2 is small (see *Appendix*). Interestingly, if, for each run, the same number of equally spaced time intervals are used that go to the same point of decomposition (say, 1 half-life), $k^2\Sigma(t - \bar{t})^2$ will be constant, and the variance of the ln k values will be approximately equal. This situation could be closely realized in practice if the approximate *magnitude of the* rates were known in advance. Then the usual factorial analysis of the ln k values could be used with the variance estimated as σ^2/K, with $K = k^2\Sigma(t - \bar{t})^2$. In the more realistic case, the variances are expected to differ, and, thus, one must approach the analysis differently. Initially, the main effects and interactions may be computed in the usual way (4). This is

[1] A is defined as not interacting if the response at high A is a constant multiple of the response at low a for all combinations of the additive factors, *i.e.*, ab/b = ac/c = abc/bc = a/(1).

[2] To be reported in a subsequent publication; S. Bolton, personal data.

Table I—Results of Simulation Study for 2^3 Factorial (run in duplicate) Untransformed Values

Combination	Average Rate Constant	Average ln Rate Constant	$\frac{1}{\Sigma\Sigma(t-\bar{t})^2}$	$\frac{1}{k^2\Sigma(\Sigma t-\bar{t})^2}$
(1)	1.87	0.63	0.77	0.22
a	3.77	1.38	5.00	0.34
b	3.10	1.16	2.92	0.30
ab	7.04	1.99	40.00	0.79
c	3.07	1.20	5.00	0.50
ac	5.35	1.72	40.00	1.36
bc	3.67	1.40	8.62	0.60
abc	8.04	2.11	31.25	0.47
Average = 4.49			Total 133.56	4.58

an appropriate unweighted average of the 2^n trials. The variance of these effects is estimated as described below, and the usual t test is applied by forming the ratio:

$$t = \frac{|\text{effect}|}{\sqrt{\text{Variance (effect)}}}$$

If t exceeds the tabulated α% value with appropriate degrees of freedom, the effect is significant. Nonsignificance does not necessarily mean that the effect is not present, but rather that it cannot be dissociated from error. This is important because most factors probably will affect the rate to some extent. If the effect is so small that it cannot be dissociated from error, it would probably not be of interest from a practical standpoint. The variance estimates of an effect may be calculated as follows: a main effect or interaction is the change in the rate constant due to changes in the levels of a factor or combination of factors appropriately averaged over all runs:

$$\hat{\sigma}^2 \text{ effect} = \frac{\hat{\sigma}^2\Sigma 1/\Sigma(t-\bar{t})^2}{2^{2n-2}} \qquad \text{(Eq. 1)}$$

in which $\hat{\sigma}^2$ is the estimated variance and $1/\Sigma(t-\bar{t})^2$ is summed over all 2^n trials or runs.

For ln rate constants:

$$\hat{\sigma}^2 \text{ effect} = \frac{\hat{\sigma}^2\Sigma 1/[k^2\Sigma(t-\bar{t})^2]}{2^{2n-2}} \qquad \text{(Eq. 2)}$$

In an experiment where many factors are being investigated, it may be convenient to look at segments of the data in which some factors are kept constant, while others are allowed to vary. The same statistical analysis as above may be used, but the summation in Eqs. 1 and 2 will include only those experiments that are appropriate to that segment of the experiment.

Example of Analysis—To illustrate the analysis described above, a hypothetical experiment was simulated with three factors, each at two levels, conforming to the following model:

$$k_{obs} = 10^{k_A\sqrt{A}}(k_B[B] + k_c[C])$$

The levels of A, B, and C are as follows:

Factor	Low Level	High Level
A	0	0.09
B	10^{-3}	2×10^{-3}
C	0.1	0.2

k_a, k_b, and k_c were assumed to be equal to 1, 1000, and 10, respectively.

A constant 0.2 CV, using random normal deviates, was imposed upon concentration readings at arbitrary time intervals, and the rate constants were calculated. (Duplicate runs were made for each of the eight combinations comprising the factorial, a total of 16 runs.) It should be mentioned that this error is relatively large for kinetic work; if anything, one should expect better results with real data. Pertinent results are shown in Table I.

The average rate constants and average ln rate constants were calculated as a weighted average of the respective duplicates with the weights being equal to $\Sigma(t-\bar{t})^2$ and $k^2\Sigma(t-\bar{t})^2$, respectively. The last two columns in Table I represent the weighting factors for computing the variances of the effects for rate constants and ln rate constants respectively, as shown in Eqs. 1 and 2; here $\Sigma[\Sigma(t-\bar{t})^2]$ is the sum from the two replicates.

The variance calculated from the line fitting (16 plots of ln C *versus* time) was 0.039 with 44 *df*. The variance calculated from the duplicate runs (using rate constants) was 0.060 with 8 *df*. This latter calculation

Table II—Effects Based on Data of Simulation Study (Untransformed Values)

Source	Average Effect
A	3.12
B	1.95
AB	1.03
C	1.09
AC	0.20
BC	−0.30
ABC	0.01

is described in the *Appendix*. The theoretical value should be ~0.04 ($CV^2 = 0.2^2$) and the results are within expectation, especially in view of the approximations involved. In this case the more dependable value of 0.039 determined from the least-squares line fitting for the variance will be used.

If the usual analysis on the untransformed rate constants is performed, one may expect to obtain estimates only for the BC interaction because of the multiplicative effects of A in the model. Table II shows the average effects based on the rate constants. The variance of an effect, calculated according to Eq. 1 is:

$$\frac{1}{16}\left[\frac{1}{\Sigma[\Sigma(t-\bar{t})^2]}\right]\hat{\sigma}^2$$

where $\hat{\sigma}^2 = 0.039$ and the summation includes all runs. The square root of the variance yields the standard deviation which, in this case, is 0.57 $[\sqrt{(0.039)(133.56)/16}]$. The BC interaction (−0.30) is not significant at the 5% level (Fig. 1). Keeping the multiplicative factor (A) constant, the results can be examined separately at both levels of A. With A at the low level, Eq. 1 is used where the summation includes readings (1), b, c, and bc. The *SD* of an effect is now 0.206. The results show that the main effects, B and C (0.92 and 0.89, respectively) are significant and the BC interaction (−0.32) is not. Similarly, an analysis with A at the high level yields 0.535 *SD*. As before, both main effects, B and C, are significant (2.98 and 1.29) while, again, the interaction (−0.29) is not (Table III).

Analysis of the log rate constants gives information on the effect of A and its interactions. The calculated effects are listed in Table IV. The standard deviation of an effect is 0.106 (Eq. 2), $\sqrt{0.039(4.58)/16}$.

The main effect of A is significant, and in the absence of interaction the antilog estimates its effect. The antilog of 0.70 is 2.01, which is very close to the known effect, 2. The analysis of these results also shows that all interactions with A are not significant. The results for B, C, and BC are difficult to interpret, since the additive and multiplicative properties are now confused with these effects.

Once the significant effects have been identified, it is of interest to estimate the rate constants associated with the different factors, which may be hydrogen ion concentration, buffer concentration, ionic strength, *etc.*, in a real example. Also, it would be most useful to construct a kinetic model that can be used to predict stability under various conditions (different levels of the factors). The fitting of such a model, however, would be rather precarious in the case of a design with factors at only two levels. In this situation, responses must be assumed to be linear functions

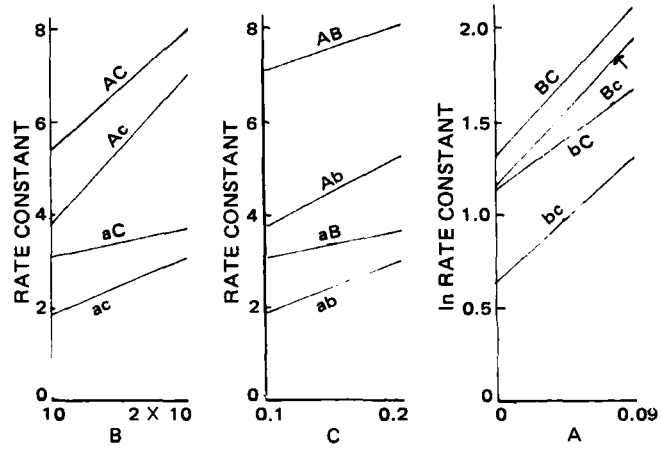

Figure 1—*Plot showing lack of interaction of factors. Capital letters are factors at high levels; lower case letters are factors at low levels. For example, aC is factor A at low level (0) with factor C at high level (0.2).*

Table III—Effects at Low and High Level of A

A = 0		A = 0.09	
Source	Average Effect	Source	Average Effect
B	0.92	B	2.98
C	0.89	C	1.29
BC	−0.32	BC	−0.29

of the factor levels. If a curved response exists, a 2^n design will not be sufficient to fit a proper model.

The rate constant associated with a factor is the slope of the observed rate constant *versus* factor level plot (Fig. 2). For example, in the case of factor B, at the low level of A (the multiplicative factor), the main effect is 0.92. The term k_B can be determined as follows: The average result at the high level of B (2×10^{-3}) is 3.39 ({3.10 + 3.67}/2), and the average result at the low level of B is 2.47 ({1.87 + 3.07}/2). The rate constant, k_B, is $(3.39 - 2.47)/(2 \times 10^{-3} - 1 \times 10^{-3}) = 920$. If interactions or multiplicative effects are not present, the rate constants can be simply calculated. If interactions exist, kinetic models may not be simply and clearly described in a single equation, and it may be preferable to describe the system at each level of the interacting factors. If a multiplicative factor is present, one may wish to include this as part of the final model. In this example the observed rate constant can be expressed as:

$$k_{obs} = 10^{k_A\sqrt{A}}(k_B[B] + k_C[C]) \qquad \text{(Eq. 3)}$$

Since interactions are nonsignificant in this example, it is convenient to express the kinetics in terms of Eq. 3.

To determine k_A, proceed as described, but use the main effect of A, expressed as $\log_{10}$:

$$k_a = \frac{0.70/2.303}{\sqrt{0.09} - \sqrt{0}} = 1.01 \qquad \text{(Eq. 4)}$$

One way to estimate k_B to satisfy Eq. 3 is to initially estimate the main effect of B, taking into account the multiplicative effect of A. From Table II, it is seen that the main effect of B is 1.95, which is the average of the main effects at low A and high A ({0.92 + 2.98}/2), as shown in Table III. Since the result of increasing the level of A is to multiply the effect of B by 2.01 (antilog of 0.70), the effect of B can be estimated: 2.01 (B $\text{effect}_{\text{low A}}$) + B $\text{effect}_{\text{low A}} \simeq 3.90$ (Note: 2.01 (B $\text{effect}_{\text{low A}}$) = B $\text{effect}_{\text{high A}}$):

$$\text{B effect}_{\text{low A}} \cong \frac{3.90}{3.01} \cong 1.30 \qquad \text{(Eq. 5)}$$

$$k_B \cong \frac{1.30}{2 \times 10^{-3} - 1 \times 10^{-3}} \cong 1300 \qquad \text{(Eq. 6)}$$

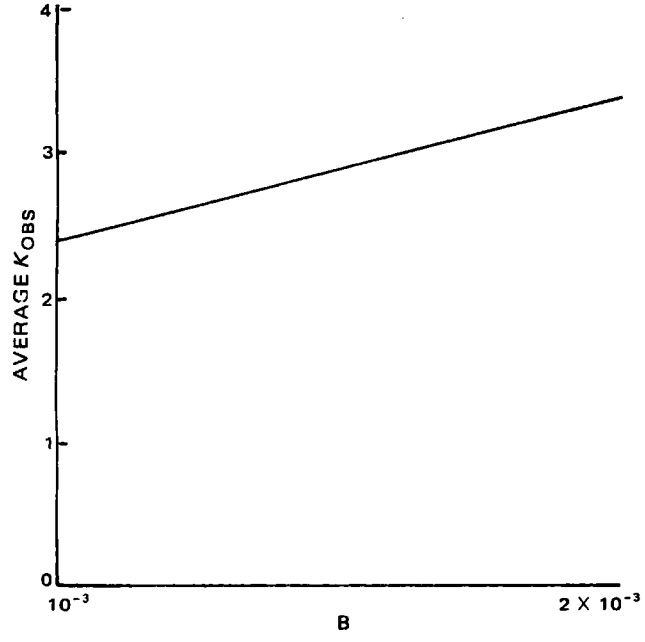

Figure 2—*Plot showing calculation of effect of factor B at low level of factor A, averaged over both levels of factor C. Slope = k_B = 920.*

Table IV—Factorial Analysis of ln Rate Constants

Source	Average Effect
A	0.70
B	0.43
AB	0.07
C	0.32
AC	−0.09
BC	−0.14
ABC	0.03

Similarly, $k_c = 7.24$, and the model (Eq. 3) can be estimated:

$$k_{obs} = 10^{\sqrt{A}}(1300[B] + 7.24[C]) \qquad \text{(Eq. 7)}$$

For example, this equation can be used to predict the rate constant with the combination, abc (A = 0.09, B = 2×10^{-3}, C = 0.2):

$$k_{obs} = 10^{\sqrt{0.09}}(1300[2 \times 10^{-3}] + 7.2[0.2]) = 8.10 \qquad \text{(Eq. 8)}$$

Without multiplicative effects, a model may also be constructed based on multiple regression techniques. In the case of a factorial design, it is convenient to use transformed (coded) values of the factor levels, where the low level equals −1 and the high level equals +1. The transformation is:

$$\frac{\text{Factor Level} - \dfrac{\text{Low Level} + \text{High Level}}{2}}{\dfrac{\text{High Level} - \text{Low Level}}{2}}$$

For Factor B, the transformation is:

$$\frac{\text{Factor Level} - 1.5 \times 10^{-3}}{0.5 \times 10^{-3}}$$

For transformed values (+1 or −1), in general, the coefficients for factors in the model are the main effects divided by 2. With the multiplicative factor A equal to 2.01, the coefficients are equal to the main effects divided by (1 + 2.01) = 3.01. For example, the coefficient for factor B is: 1.95/3.01 = 0.65. When coded values are used, an intercept value is usually calculated as the average of the rate constants, 4.49 (Table I). With the multiplicative factor A in the model, one can calculate the intercept as the average of the rate constants at the low level of A as follows:

$$\frac{2.93 + 6.05}{3.01} = \frac{\text{Average at low A + average at high A}}{3.01} = 2.98 \qquad \textit{(Eq. 11)}$$

The final equation is:

$$k_{obs} = 10^{1.0\sqrt{A}}(2.98 + 0.65[B'] + 0.36[C']) \qquad \text{(Eq. 12)}$$

where B′ and C′ are the transformed values[3], *i.e.*, the levels are transformed to equal ±1.

APPENDIX

(*a*) Approximation to the Variance of ln *y*—Using a Taylor series expansion, it can be shown that variance (ln *y*) = $CV^2(1 + \frac{1}{2}CV^2 + \ldots)$. If the *CV* is small, variance (ln *y*) = $CV^2 = \sigma^2/y^2$. Obviously, the smaller the *CV*, the better the approximation.

This result may be used to calculate the approximate variance of ln rate constant, ln *k*. The estimated variance of a rate constant is $\hat{\sigma}^2/\Sigma(t - \bar{t})^2$. Thus, the approximate variance of ln *k* is $\hat{\sigma}^2_k/k^2 = \hat{\sigma}^2/k^2\Sigma(t - \bar{t})^2$.

(*b*) The Use of Weighting in Line Fitting and Variance Estimation—The ordinary least-squares line fitting procedure assumes equal variance at each point, C. If the *CV* is constant, the variance (σ^2) is not constant, but depends on the value of C. However, in view of the above discussion, if the *CV* is constant, ln C has approximately equal variance (CV^2) and the fitting of the line, ln C *versus* time, can be computed as usual. The residual variance estimates CV^2.

[3] Note that this equation is not exactly equivalent to the equation previously constructed due to differences in calculating the multiplicative effect. If the error and interactions are small, different ways of calculating the effect of A will be very close. Here, where the error is relatively large, the results are equivocal. For example, the effect of A can be determined as the antilog of the effect of A when analyzing ln rate constants (2.01); as the ratio of the average results at high A compared with the average at low A (2.01); or as the average, in this example, of the four ratios of the rate at high level with the rate at low level of A under constant conditions of the other factors, B and C, *i.e.*, the average of a/(1), ab/b, ac/c, and abc/bc (2.06).

If the variance is constant at each point C, a fit of ln C *versus* time would require weighting each point proportional to the reciprocal of the variance, approximately C^2 in the case of constant CV. With constant variance, a zero-order plot would require no weighting.

From similar considerations, an estimate of σ^2 (equal to CV^2 in the case of constant CV and a first-order reaction) can be obtained from replicate determinations of the rate constant using the formula:

$$\Sigma w(k - \bar{k}_w)^2/(n - 1)$$

Where $\bar{k}_w$ is the weighted mean, n is the number of replications and w is $\Sigma(t - \bar{t})^2$. (Term w is equal to $k^2\Sigma(t - \bar{t})^2$ in the case of ln k). A shortcut formula for $\Sigma w(k - \bar{k})^2$ is $\Sigma wk^2 - \Sigma w(\bar{k}w)^2$.

An example of some calculations for the combination ab follows. The duplicate determinations of the rate constants were 5.86 and 8.22, and $\Sigma(t - \bar{t})^2$ was 0.0125 for each determination:

(1) The weighted average, $\bar{k}_w$, = [(0.0125)5.86 + (0.0125)8.22]/(0.025 + 0.025) = 7.04 (since the weights are equal, the weighted average equals the ordinary arithmetic average).

(2) σ^2 (One degree of freedom) from the duplicates = $[\Sigma wk^2 - \Sigma w(\bar{k}_w)^2]/(n - 1)$ = [(0.0125)(5.86)^2 + (0.0125)(8.22)^2 − 0.025 (7.04)^2]/1 = 0.0348

(3) The weighted average of ln k = ln k_w = [(5.86)^2(0.0125) ln 5.86 + (8.22)^2(0.0125)ln 8.22]/[(5.86)^2(0.0125) + (8.22)^2(0.0125)] = 1.99

The approximations inherent in the method should be kept in mind, namely the transformation to logs and the fact that observed values are used rather than true values in the weighting. However, if the CV is not large, the reliability of any conclusions from this analysis should not be in doubt.

REFERENCES

(1) J. A. Plaizier-Vercammen and R. E. De Neve, *J. Pharm. Sci.*, **69**, 1403 (1980).

(2) D. E. Fonner, Jr., J. R. Buck, and G. S. Banker, *ibid*, **59**, 1587 (1970).

(3) S. Dincer and S. Ozdurmus, *ibid*, **66**, 1070 (1977).

(4) O. L. Davies, "Design and Analysis of Industrial Experiments," Hafner, New York, N.Y., 1963.

Kinetics and Mechanism of Hydroxy Compound Cinnamoylation in Acetonitrile Catalyzed by *N*-Methylimidazole and 4-Dimethylaminopyridine

CHUKWUENWENIWE J. EBOKA and KENNETH A. CONNORS ˣ

Received March 25, 1982 from the *School of Pharmacy, University of Wisconsin, Madison, WI 53706.* Accepted for publication June 14, 1982.

Abstract □ The kinetics of reaction of the acylating agents *trans*-cinnamic anhydride and *trans*-cinnamoyl chloride with the hydroxy compounds *n*-propyl alcohol and water in the presence of *N*-methylimidazole and 4-dimethylaminopyridine were studied spectrophotometrically in acetonitrile solution at 25°. The acid chloride reacted *via* the intermediate formation of the *N*-acyl catalyst, which underwent general base-catalyzed reaction with the hydroxy compound. The anhydride did not form the *N*-acyl intermediate, but instead underwent direct general base catalysis. In the presence of water, all systems formed the *N*-acyl intermediate. The mechanistic route followed by the system was determined by the nucleophilicity of the catalyst, the ability of the leaving group, and the polarity of the solvent.

Keyphrases □ Cinnamoylation—hydroxy compounds in acetonitrile, catalyzed by *N*-methylimidazole and 4-dimethylaminopyridine, kinetics □ Kinetics—cinnamoylation of hydroxy compounds in acetonitrile □ 4-Dimethylaminopyridine catalysis—cinnamoylation of hydroxy compounds, kinetics □ *N*-Methylimidazole catalysis—cinnamoylation of hydroxy compounds, kinetics

Acylation is an important synthetic and analytical reaction. Pyridine is the classical acylation catalyst, but during the past decade more powerful catalysts have been introduced, most notably 4-dimethylaminopyridine, which has been used in synthesis (1–4) and analysis (5–8). More recently this laboratory introduced *N*-methylimidazole as an analytical acylation catalyst (9–13).

Acylation reactions are usually carried out in nonaqueous solvents. Although the mechanisms of acyl transfer in aqueous systems have been well studied (14–16), the nature of these reactions in nonhydroxylic solvents is not yet understood. Among the features that have been suggested as important in determining the kinetics and mechanisms of these reactions are the balance between nucleophilic and general base catalysis, the complex nature of rate equations (17–19), the possibility of kinetically significant ion-pair formation (3, 20, 21), competing reactions (22), and formation of molecular complexes (23). Some of these factors were addressed in a recent study on the kinetics of acetylation of alcohols by acetic anhydride and acetyl chloride, catalyzed by *N*-methylimidazole and 4-dimethylaminopyridine, in acetonitrile solution (24).

Since the cinnamoyl group, $C_6H_5CH{=}CHCO$, is a powerful UV chromophore, it is an interesting analytical acyl group (25, 26). An earlier study (27) reported the kinetics of hydrolysis of *trans*-cinnamic anhydride and of its reactions with some alcohols, catalyzed by pyridine, 4-dimethylaminopyridine, and *N*-methylimidazole, but the study was not designed to explore the detailed nature of the mechanism. In the present paper the reactions of *trans*-cinnamic anhydride and *trans*-cinnamoyl chloride with *n*-propyl alcohol and water, in acetonitrile solution, are described. The catalysts were *N*-methylimidazole and 4-dimethylaminopyridine; the reactions were studied by UV spectrophotometry.

EXPERIMENTAL

Materials—*trans*-Cinnamoyl chloride[1] was distilled under reduced pressure to give colorless crystals, mp 34–35° [lit. mp 35–36° (28)]. The molar absorptivity at 298 nm, in acetonitrile, was 2.42 × 10^4 liter/mole cm. *trans*-Cinnamic anhydride was synthesized as previously described (27), mp 136–137° [lit. mp 136° (29)]. Its molar absorptivity at 294 nm was 4.26 × 10^4 liter/mole cm. *N*-Methylimidazole[1] was distilled under

[1] Aldrich Chemical Co.

0022-3549/83/0400-0366$01.00/0