Modeling the Kinetics of Acylation of Insulin using a Recursive Method for Solving the Systems of Coupled Differential Equations

Bartosz A. Grzybowski, Janelle R. Anderson, Ian Colton, Scott T. Brittain, Eugene I. Shakhnovich, and George M. Whitesides
Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT   This paper describes a theoretical method for solving systems of coupled differential equations that describe the kinetics of complicated reaction networks in which a molecule having multiple reaction sites reacts irreversibly with multiple equivalents of a ligand (reagent). The members of the network differ in the number of equivalents of reagent that have reacted, and in the patterns of sites of reaction. A recursive algorithm generates series, asymptotic, and average solutions describing this kinetic scheme. This method was validated by successfully simulating the experimental data for the kinetics of acylation of insulin.

INTRODUCTION

Even moderately complex kinetic systems can have complicated solutions. In fact, it is only for a limited number of systems (e.g., unimolecular systems, Scheme 1 a) that a closed-form solutions exist (Berberan-Santos and Martinho, 1990; Rodiguin and Rodiguina, 1960). In other instances, one can either approximate the answer by an infinite series, or integrate the kinetic equations numerically. The series solutions often diverge, and numerical integration may be time-consuming for complicated systems.

In this paper, we present a recursive method that allows us to find series, asymptotic, and approximate functional solutions for a kinetic system consisting of an arbitrary number of starting materials, intermediates, and products (X_i), and a reagent R that is consumed irreversibly in the sequence of conversions among them. Upon reaction with R, an intermediate X_i converts irreversibly to X_j with a characteristic rate constant k_{ij}. Rate constants are positive, if they correspond to production of a given species (intermediate or product), and negative for its consumption. This process is assumed to be first-order in both the reactant X_i and the reagent, and second-order overall. Moreover, each of the X_j can be a product of reaction of more than one X_i and can itself convert to more than one species. The reagent itself, aside from reacting with the starting material and the intermediates (the loss of the reagent due to reaction with X_i is described by rate constant k_{ri}), can disappear (e.g., by reaction with solvent or buffer, or by thermal decomposition), following first-order kinetics with rate constant k_R. Such a network of reactions can be represented in matrix notation as in Scheme 1 b. The important difference between this system and the unimolecular one (Scheme 1 a), is that now the equations are coupled by the $R(t)$ function (that

is, the concentration of the reagent R at time t). Nonlinear systems of equations of this type do not have a known closed-form solution. We will refer to this kinetic scheme as a reagent-coupled unimolecular system (RCUS).

An RCUS provides a general model for the kinetics of irreversible reactions of molecules having multiple recognition sites with complementary ligands. Such networks of reactions are common in biology: recations of multiple ligands with a protein (Imai, 1983; Matthews and van Holde

a.

$$\frac{d}{dt}\begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_n(t) \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} & \cdot & k_{1n} \\ k_{21} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ k_{n1} & \cdot & \cdot & k_{nn} \end{pmatrix} \cdot \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_n(t) \end{pmatrix}$$

b.

$$\frac{d}{dt}\begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_n(t) \\ R(t) \end{pmatrix} = R(t) \cdot \begin{pmatrix} k_{11} & 0 & \cdot & 0 & 0 \\ k_{21} & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & 0 & 0 \\ k_{n1} & \cdot & \cdot & k_{nn} & 0 \\ k_{r1} & k_{r2} & \cdot & \cdot & k_R \end{pmatrix} \cdot \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_n(t) \\ 1 \end{pmatrix}$$

c.

$$\frac{d}{dt}\begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_n(t) \\ R(t) \end{pmatrix} = R(t) \cdot \begin{pmatrix} k_{11} & k_{12} & \cdot & k_{1n} & 0 \\ k_{21} & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 \\ k_{n1} & \cdot & \cdot & k_{nn} & 0 \\ k_{r1} & k_{r2} & \cdot & \cdot & k_R \end{pmatrix} \cdot \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_n(t) \\ 1 \end{pmatrix}$$

SCHEME 1   The system of equations in (*A*) describes the kinetics of a unimolecular system; it has an analytical solution. In (*B*), the equations of the system are coupled by the $R(t)$ term (the concentration of the reagent). The system is nonlinear and cannot be solved analytically. Because the reactions are assumed irreversible, the entries in the matrix above the diagonal are all zero. The recursive method developed in the paper can also solve the more general case (*C*) which, however, is physically unrealistic.

1990; Mattevi, 1996; Monaco et al., 1978; Perutz, 1989) and humoral immune response (many antibodies attaching to an antigen) (Tizard, 1984) are two prominent examples. An RCUS can also be used to describe kinetics of protein–ligand interactions in certain analytical techniques, such as affinity capillary electrophoresis (ACE) (Chu et al., 1994; Kuhr and Monnig, 1992; Liu et al., 1994). Here, we use our theoretical model to simulate the data obtained by capillary electrophoresis (Gao et al., 1995) from the formation of a charge ladder (Gao et al., 1994, 1996; Gao and Whitesides, 1997) of insulin by acylation.

Insulin is a protein (MW 5700) that is made up of two chains: α, consisting of 21 amino acid residues, and β, 30 residues long. Insulin has three amino groups: the α-chain N-terminal glycine residue, the β-chain N-terminal phenylalanine and a lysine group located at position 29 of the β-chain. Each of these amino groups can be acylated. Thus, acylation gives rise to 7 possible acylation products (Scheme 2) in addition to native insulin (denoted N): insulin acylated at F (F), acylated at G (G), acylated at K (K), acylated at both F and G (FG), at both F and K (FK), at both G and K (GK) and, finally, with all three amino groups acylated (FGK). We studied acylation reactions of insulin with two acylating reagents: 1) the N-hydroxysuccinimide ester of 5-carboxyfluorescein (F-NHS, Figure 1 A), and 2) fluorescamine (FL, Figure 1 B) (Stein et al., 1973). We analyzed the results using our recursive method for solving the RCUS describing this system, and compared them to the results obtained by numerical integration of kinetic equations.

RECURSIVE METHOD FOR SOLVING RCUS

The procedure starts with solving a simplified version of the problem, without $R(t)$ dependence; this simplification gives a unimolecular system, such as that in Scheme 1 a. The eigenvalues p_m of the matrix of rate constants $\mathbf{K}$ are obtained from Eq. 1, in which $\mathbf{U}$ is the unit matrix. For a given root p_m, a particular solution is given by

$$\det(\mathbf{B} - p_m\mathbf{U}) = 0 \tag{1}$$

$$x_i(t) = A_{im} \cdot \exp(p_m \cdot t), \quad i = 1, \ldots, n \tag{2}$$

$$\forall_{h=1\ldots n} \sum_{i=1}^{n} (k_{ih} - \delta_{ih}p_m) \cdot A_{im} = 0 \tag{3}$$

$$X_j(t) = \sum_{i=1}^{n} A_{ji} \cdot \exp(p_i \cdot t) \tag{4}$$

$$\frac{d}{dt} \exp\left(\int_0^t R(t') \, dt' \right) = R(t)\exp\left(\int_0^t R(t') \, dt' \right) \tag{5}$$

Eq. 2, and the values of coefficients A_{im} are found from Eq. 3 in which δ is the delta function (Berberan-Santos and Martinho, 1990). The general solution can be written as a linear combination of particular solutions (Eq. 4). Eq. 4 will be useful in solving the original problem with $R(t)$ dependence. Note that the following identity holds (Eq. 5).

We postulate the solution to the RCUS in the form of Eq. 6. This equation is, indeed, a solution to the RCUS, as can

SCHEME 2 The kinetic scheme for the acylation of insulin. Insulin can be acylated by an acylating reagent R on three different residues (F, G, K). Each conversion is first order both in the substrate and in the reagent. In addition, the reagent can hydrolyze (the hydrolyzed product is inactive in the acylation reactions). This kinetic scheme can be described mathematically by an RCUS set of equations (Scheme 1 B). Assuming the rate constants of the acylation reactions are independent of the extent and pattern of the acylation of the protein (no cooperativity), the number of independent rate constants can be reduced to four. In this case, $k_1 = k_6 = k_8 = k_{12}$, $k_2 = k_4 = k_9 = k_{11}$, and $k_3 = k_5 = k_7 = k_{10}$.
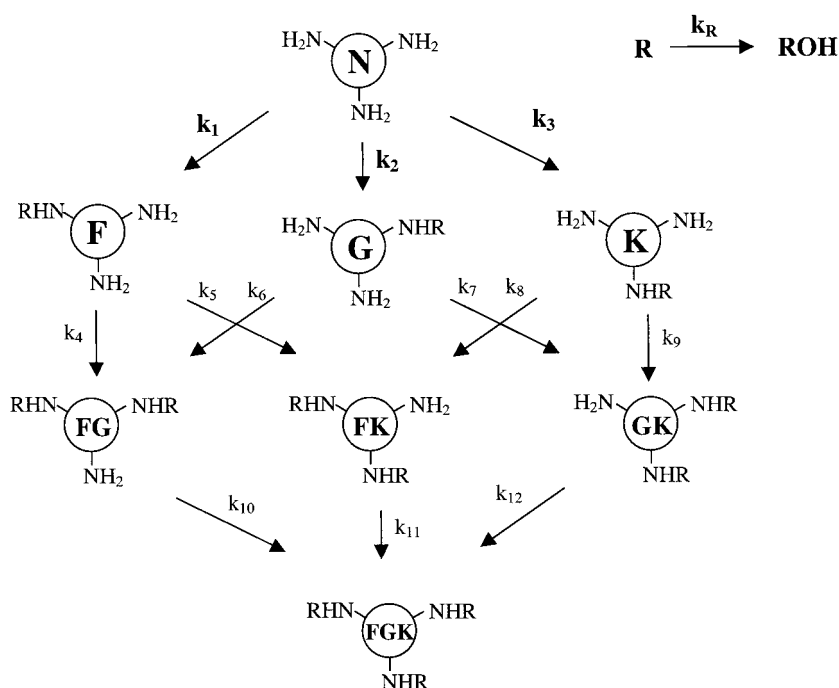
easily be proven (Eq. 7). The solution given by Eq. 6 is still incomplete, because it is written in terms of $R(t)$, which is unknown.

$$\tilde{X}_j(t) = \sum_{i=1}^{n} A_{ji} \cdot \exp\left(p_i \cdot \int_0^t R(t') \, dt'\right) \quad (6)$$

$$\frac{d}{dt} \tilde{X}_j(t) = \sum_{i=1}^{n} A_{ji} \cdot p_i \cdot R(t) \exp\left(p_i \cdot \int_0^t R(t') \, dt'\right) \quad (7)$$

$$= R(t) \cdot \left\{\sum_{i=1}^{n} (\tilde{X}_j(t) \cdot k_{ji})\right\}.$$

We solve this problem recursively, substituting the $\tilde{X}_i(t)$ solutions into the equation for $R(t)$ (Eqs. 8 and 9). Using Eq. 5, and introducing $S(t) = \int_0^t R(t') \, dt'$ and $B_m = C_m/p_m$, we can rewrite Eq. 9 in a slightly modified form (Eq. 10).

$$\frac{dR(t)}{dt} = \left\{\sum_{i=1}^{n} k_{ri} \cdot \tilde{X}_i(t) \cdot R(t)\right\} + k_R \cdot R(t) \quad (8)$$

$$\frac{dR(t)}{dt} = \sum_{m=1}^{n} \left\{C_m \cdot R(t) \cdot \exp\left(p_m \int_0^t R(t') \, dt'\right)\right\} + k_R \cdot R(t); \quad (9)$$

$$C_m = \sum_{i=1}^{n} k_{ri} \cdot A_{im}$$

$$\cdot \frac{d}{dt}\left\{R(t) - \sum_{m=1}^{n} B_m \cdot \exp(p_m \cdot S(t)) - k_R \cdot S(t)\right\} = 0. \quad (10)$$

This expression can be further simplified to Eq. 11 by making use of the boundary conditions at time $t = 0$. Eqs. 9 and 11 will be used to find series, asymptotic, and approximate solutions for $R(t)$ (note that if $R(t)$ is known, the entire RCUS problem is solved through Eq. 6).

$$R(t) - \sum_{m=1}^{n} B_m \cdot \exp(p_m \cdot S(t)) - k_R \cdot S(t) = R(0) - \sum_{m=1}^{n} B_m \quad (11)$$

The procedure for obtaining the series solution valid for small values of time is quite lengthy, and we relegate it to the Appendix. To investigate the asymptotic behavior of $R(t)$, we first note that $\lim_{t\to\infty} R(t) = 0$, and that $S(t)$ converges in infinity, i.e., $\lim_{t\to\infty} S(t) = S_\infty$. With these conditions, Eq. 11 for $t \to \infty$ simplifies to Eq. 12, which can

easily be solved numerically for S_∞.

$$-\sum_{m=1}^{n} B_m \cdot \exp(p_m \cdot S_\infty - k_R S_\infty) = R(0) - \sum_{m=1}^{n} G_m. \quad (12)$$

To find the asymptotic solution at infinity, we first write S_∞ as a sum of $S(t)$ and a function $y(t)$ defined by $y(t) = \int_t^\infty R(t') \, dt'$, and substitute for $S(t)$ in Eq. 9 to obtain Eq. 13. Because $y(t) \to 0$ for large times, we have $\exp(-p_m y(t)) \cong 0$, and the asymptotic solution at infinity can be written as in Eq. 14. By a similar method, the asymptote for $t \to 0$ is found: in this case $S(t) \cong 0$ and $\exp(-p_m S(t)) \cong 1$, leading to Eq. 15.

$$\frac{dR(t)}{dt} = \sum_{m=1}^{n} \{C_m \cdot R(t) \cdot \exp(p_m S_\infty) \cdot \exp(-p_m \cdot y(t))\} + k_R \cdot R(t) \quad (13)$$

$$R_{t\to\infty}(t) = R(0) \cdot \exp\left[\left(k_R + \sum_{m=1}^{n} \{C_m \cdot \exp(p_m S_\infty)\}\right) \cdot t\right] \quad (14)$$

$$R_{t\to 0}(t) = R(0) \cdot \exp\left[\left(k_R + \sum_{m=1}^{n} C_m\right) \cdot t\right] \quad (15)$$

If the difference between limiting exponential asymptotes is small, it is reasonable to assume that $R(t)$ may be approximated by an exponential for all times (average exponential, R_{avex}). With this assumption, we impose the condition that its integral $\int_0^\infty R_{avex}(t) \, dt$ is equal to the integral of an exact $R(t)$ over this range, i.e., to $S(t)$. Using a boundary condition at $t = 0$, Eq. 16 is obtained:

$$R_{avex} = R(0) \cdot \exp\left(-\frac{R(0)}{S_\infty} t\right). \quad (16)$$

KINETICS OF ACYLATION OF INSULIN

We used the mathematical model developed in the previous section to study the kinetics of acylation of insulin (Scheme 2). Partial acylation of the amino groups of insulin results in a set of derivatives that can be resolved into eight peaks by capillary electrophoresis based on differences in their values of electrophoretic mobility at pH 6.8. These eight peaks have all been assigned (Gao et al., 1995). We denote either of the two acylating agents used in the experiments (the NHS ester of 5-carboxyfluorescein [1] and flourescamine [2]), as R (Fig 1). The acylation reactions are irreversible, first-order in both the reagent and an insulin derivative, and second-order overall. Moreover, we assume no cooperativity among acylation reactions; that is, we assume there are only three rate constants, so that, for example, the rate constant for conversion of N to F is the same as for con-

version of GK to FGK. It has been reported (Gilson and Honig, 1987), that for the ionic strength ~ 0.2 (as in our experiments) and for amino groups separated by ~ 10 Å, the changes in the pKa of an amino group upon acylation of other amino groups of the same protein are ~ 0.1. For the amino groups of insulin, the changes in pKa are expected to be even smaller, because the distances between these groups are larger ($N_{Phe} - N_{Gly}$, 21 Å; $N_{Phe} - N_{Leu}$, 25 Å; and $N_{Gly} - N_{Leu}$, 13 Å) than the distances between the amino groups studied by Gilson and Honig. Noncooperativity assumption might be erroneous for proteins, in which the distances between the acylated amino groups are smaller than ~ 10 Å. Therefore, using Bronsted's equation, one can conservatively estimate that, by assuming no cooperativity, one introduces an error of at most $\sim 10\%$ in the values of rate constants. The nocooperativity assumption also greatly reduces the computational complexity of the problem (with full cooperativity, there would be twelve independent rate constants for insulin, instead of three).

With these simplifying assumptions, the kinetic equations describing the system can be written in the form of an RCUS as in Eq. 17 (the rate constants k_1, k_2, k_3, and k_R are positive numbers, and the minus sign denotes consumption of a given species).

solutions for $R(t)$ were calculated; they are presented in Table 2. In the next section, we will examine how accurately these equations fit the reality of the acylation of insulin.

RESULTS AND DISCUSSION

General properties of solutions

We begin this section with the discussion of general properties of series, asymptotic, and average exponential solutions for the kinetic scheme of the acylation of insulin. The model parameters (i.e., the observed rate constants k_1, k_2, k_3, k_R, and the initial concentrations $R(0)$ and $N(0)$) are chosen such that they illustrate the important characteristics of the solutions. Figure 2 A shows an example of a series solution for $R(t)$ compared with the results of numerical integration of kinetic equations ($k_1:k_2:k_3:k_R = 4:3:2:0$ and $N(0) = R(0)$). With a fourth-order expansion, the series follows the numerically integrated solution down to $\sim 0.4R(0)$ and then diverges to infinity. Expansions to orders higher than four do not markedly improve the quality of the solution. They influence, however, the way in which the series diverges: for even-order expansions $\Pi_{t\to\infty} R(t) = -\infty$. For small and

$$\frac{d}{dt} \begin{pmatrix} N(t) \\ F(t) \\ G(t) \\ K(t) \\ FG(t) \\ FK(t) \\ GK(t) \\ FGK(t) \\ R(t) \end{pmatrix} = R(t) \cdot \begin{pmatrix} -k_1-k_2-k_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ k_1 & -k_2-k_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ k_2 & 0 & -k_1-k_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ k_3 & 0 & 0 & -k_1-k_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_2 & k_1 & 0 & -k_3 & 0 & 0 & 0 & 0 \\ 0 & k_3 & 0 & k_1 & 0 & k_2 & 0 & 0 & 0 \\ 0 & 0 & k_3 & k_2 & 0 & 0 & -k_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_3 & k_2 & k_1 & 0 & 0 \\ -k_1-k_2-k_3 & -k_2-k_3 & -k_1-k_3 & -k_1-k_2 & -k_3 & -k_2 & -k_1 & 0 & -k_R \end{pmatrix} \cdot \begin{pmatrix} N(t) \\ F(t) \\ G(t) \\ K(t) \\ FG(t) \\ FK(t) \\ GK(t) \\ FGK(t) \\ 1 \end{pmatrix} \quad (17)$$

Notice that, in our insulin example, the rate constants are the observed ones, because only deprotonated neutral amino groups react with the acylating reagent. The values of p_m, B_m, and C_m coefficients (defined in the previous section) for this system of equations are given in Table 1. Using these coefficients, the series, asymptotic, and average exponential

comparable initial concentrations of N and R, and when the rate of decomposition of R is comparable with the rate constants of the interconversion reactions (k_1, k_2, k_3), the series solution assumes a very simple form. Because in such case, $k_i N(0) \ll k_R$, $i = 1, 2, 3$, we can approximate $R(t)$ by a single exponential (Eq. 18). Obviously, the concentrations
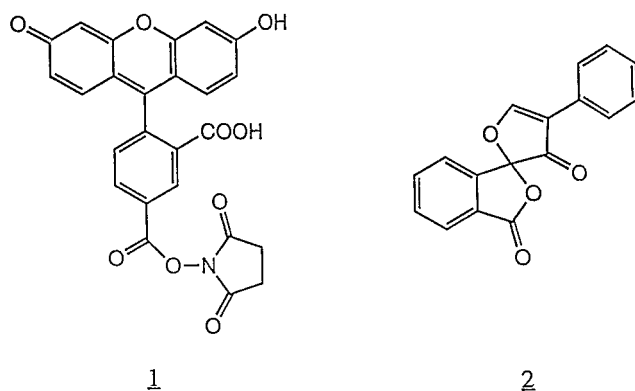
FIGURE 1 Molecular structures of the acylating reagents used in the experiments: (*1*) the NHS ester of 5-carboxyfluorescein and (*2*) fluorescamine.

of acylated derivatives of insulin can be well approximated by series.

$$R(t) = R(0) - ((k_1 + k_{21} + k_3)N(0) + k_R)R(0)t$$

$$+ \left(-\frac{k_1^2 + k_2^2 + k_3^2}{2}N(0)R^2(0) \right.$$

$$\left. + \frac{((k_1 + k_2 + k_3)N(0) + k_R)^2 \cdot R(0)}{2} \right)$$

$$t^2 + \cdots \approx R(0) - k_R R(0)t + \frac{k_R^2 R(0)}{2}t^2 - \cdots$$

$$= R(0)\exp(-k_R t) \qquad (18)$$

solutions only in the region, where the series solution for $R(t)$ is exact (Fig. 2 *C* and *D*).

The average exponential solutions are shown in Fig. 2, *B–D*. They were obtained by first calculating S_∞ from Eq. 12 using the bisection method (Press et al., 1996), and then using this value in Eq. 16. The maximum deviation of these solutions from the results of numerical integration does not exceed 6%. The asymptotic solutions for small and large times approximate the numerical curves even better than average exponentials: the differences for small/large times,

TABLE 1 Values of p_m, C_m, and B_m coefficients for the model describing the kinetics of acylation of insulin

m	p_m	C_m	B_m
1	$-k_1 - k_2 - k_3$	0	0
2	$-k_2 - k_3$	0	0
3	$-k_1 - k_3$	0	0
4	$-k_1 - k_2$	0	0
5	$-k_3$	$-k_3 \cdot N(0)$	$N(0)$
6	$-k_2$	$-k_2 \cdot N(0)$	$N(0)$
7	$-k_1$	$-k_1 \cdot N(0)$	$N(0)$
8	0	0	0

For definitions, see the Theory section.

respectively, are less than 1% of the numerically calculated values.

Kinetics of acylation of insulin

Our study of the kinetics of acylation of insulin is an example of a problem where the values of the rate constants describing the kinetics of a complex reaction network are unknown and must be decoded from available experimental data (concentrations of intermediates at different times and/or for different initial concentrations of reagents). The procedure for finding the values of rate constants is conceptually similar to finding a global minimum of a function of many variables. First, a trial set of rate constants $\{k\}_{trial}$ is chosen, and the concentrations of intermediates are calculated by substituting the trial set into the kinetic equations and integrating them. Second, the calculated concentrations of intermediates are compared to the experimental ones, and the difference between these two sets is quantified by some function $\Delta(\{k\}_{trial})$. In our study, Δ was defined as a square difference between simulated (x) and experimental (y) data, i.e., $\sum_{i=1}^{n} (x_i - y_i)^2$, where n is the number of data. The rate constants are then changed according to some algorithm, (such as, e.g., grid search or Monte Carlo), and the procedure is repeated. This cycle continues until the optimal set of rate constants $\{k\}_{optimal}$ that minimizes Δ is found.

This procedure for finding the optimal rate constants involves large numbers of calculations: even in a simple case of only 10 kinetic equations and four rate constants, for a grid search with ~ 100 possible values for each rate constant, and for $\sim 10^3$–10^4 numerical integration steps for each trial set $\{k\}_{trial}$, the total number of computer substitutions to be made is on the order of 10^{13}. Our method of solving an RCUS, makes it possible to reduce this number significantly. Specifically, because the method does not require numerical integration of kinetic equations, the CPU time needed for finding optimal rate constants is shorter by roughly the factor of the number of numerical integration steps ($\sim 10^3$–10^4). In the remaining part of this section, we will illustrate the use of our theoretical method in two experiments involving acylation of insulin (Fig. 3).

In the first experiment, fluorescein (*1*) was used as the acylating agent. The reaction of fluorescein with insulin was monitored as a function of time (Fig. 4). Two methods were independently used to find the optimal values of rate constants. In the first approach, the grid search in the space of rate constants (100 possible values for each rate constant) was performed, and the kinetic equations were integrated for each trial set $\{k\}_{trial}$ (time step 0.01 s). The procedure gave the ratio of the values of rate constants $k_1:k_2 k_3:k_R = 1.07:0.69:0.57:1.00$, and the standard deviation per peak σ (defined as $\sigma = \sqrt{\Delta/n}$) between the simulated and the experimental data was calculated to be 0.0093. The total time required for calculations (Pentium 233 MHz processor) was ~ 47 h. In the alternative approach based on our theo-

TABLE 2 Functional forms of various solutions for the *R(t)* function for the kinetic scheme of acylation of insulin

	Functional Form
S_∞	$-N(0) \cdot \{\exp(-k_1 \cdot S_\infty) + \exp(-k_2 \cdot S_\infty) + \exp(-k_3 \cdot S_\infty)\} - k_R \cdot S_\infty = R(0) + k_1 + k_2 + k_3$ Solve numerically for S_∞
Series solution	$R(t) = R(0) + H_1 \cdot R(0)t + \left(H_2 \cdot R^2(0) + \dfrac{H_1^2 \cdot R(0)}{2}\right)t^2 + (H_3 \cdot R^3(0) + \frac{4}{3} \cdot H_2 \cdot H_1 \cdot R^2(0) + \frac{1}{6}H_1^3 \cdot R(0))t^3 + \cdots$ where $\quad H_i = \delta_{1,i} - \dfrac{k_1^i + k_2^i + k_3^i}{k!}$
Asymptote t → 0	$R_{t\to 0}(t) = R(0) \cdot \exp[(-N(0) \cdot (k_1 + k_2 + k_3) - k_R)t]$
Asymptote t → ∞	$R_{t\to\infty}(t) = R(0) \cdot \exp[(-k_R - N(0)\sum_{m=1}^{3}k_m\exp(-k_m S_\infty))t]$
Average Exponential	$R(t) = R(0) \cdot \exp\left(-\dfrac{R(0)}{S_\infty} \cdot t\right)$

retical method, the numerical integration was not used. Instead, for each $\{k\}_{trial}$, the average-exponential solution was found, and the concentrations of the intermediates were calculated from it. This procedure gave the same values of

the optimal rate constants. The standard deviation per peak between the simulated and the experimental data was $\sigma = 0.0094$, and the difference between this simulation and that based on numerical integration was calculated to be $\sigma =$
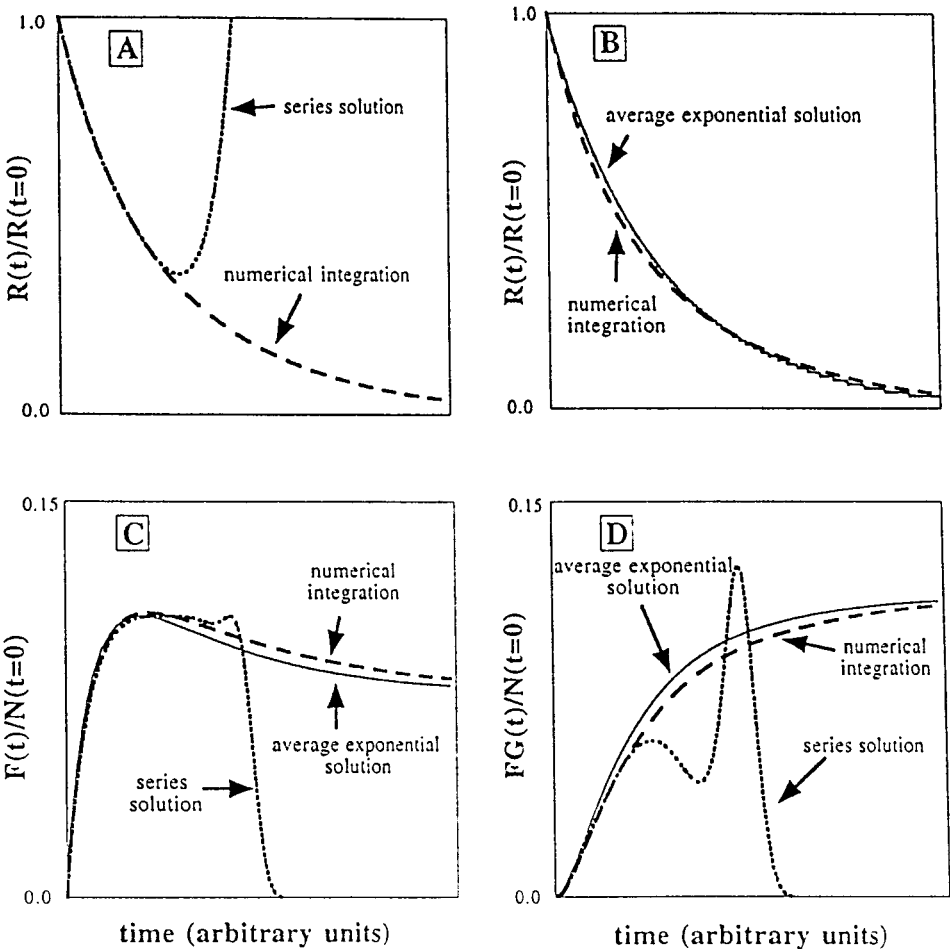

FIGURE 2 Performance of various methods for finding solutions for $R(t)$ and two intermediates, F(t) and FG(t), in the acylation of insulin. The model rate constants were (*A* and *B*) $k_1{:}k_2{:}k_R = 4{:}3{:}2{:}0$ and $N(0) = R(0)$ and (*C* and *D*) $k_1{:}k_2{:}k_3{:}k_R = 4{:}3{:}2{:}0.5$ and $N(0) = 0.4R(0)$.
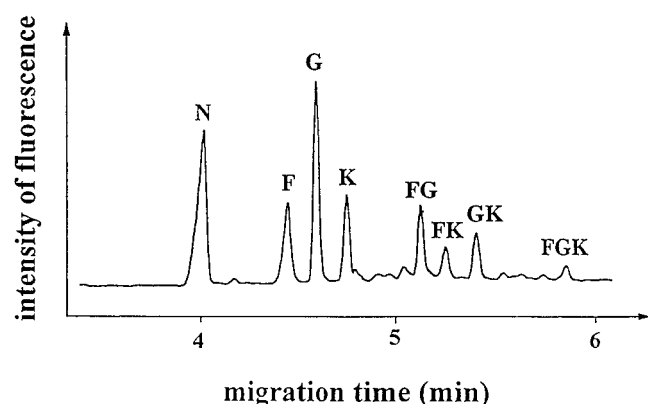
FIGURE 3   Example of an electropherogram for the reaction of insulin (1.5 μl 260 μM buffer solution) with flourescamine (4 μl 100 mM buffer solution). The products of the reaction were analyzed by capillary electrophoresis in a capillary of 50 μm (internal diameter) and 77 cm (length). A voltage of 30 kV was applied to the ends of the capillary.

0.00038. The computer time required for calculations was only ~50 min.

In the second experiment, fluorescamine (*2*) was used as the acylating reagent. This time, the mixture of underivatized protein and the acylating reagent was allowed to react until stable concentrations of all the derivatives were reached (when the acylating agent disappeared, and the reactions came to a halt; ~2 h). The experiment was repeated for 6 different initial amounts of fluorescamine (Fig. 5). As before, two theoretical approaches were taken to fit to the experimental data. Using the grid search method (100 possible values for every rate constant) with numerical integration (5000 steps for each set of data), the ratios of the rate constants were found to be $k_1:k_2:k_3:k_R = 2.08:2.48:0.88:1.00$. The calculations took ~39 h. In the approach based on our theoretical method, numerical integration was replaced by calculating a long-time asymptotic solution, giving the same values of the optimal rate constants. This time, however, the computer time required was much shorter (~10 min). The difference between the two simulated sets was $\sigma = 0.00027$, and that between the experimental dataset and the set simulated using long-term asymptotic solution (without numerical integration) was $\sigma = 0.028$.

To assess the reasonableness of our results, the relative magnitudes of the observed rate constants obtained by our method for the acylation of insulin were compared with those estimated from the Bronsted equation. An observed rate constant k_i, $i = 1, 2, 3$ can be related to the real rate constant for the acylation reaction by a simple relation $k_i = \theta_i k_{i,true}$, in which θ_i, expresses the fraction of a respective amino group in unionized form, $\theta_i = [RNH_2]_i/([RNH_2]_i + [RNH_3^+]_i) = (10^{pK_a-pH} + 1)^{-1}$. Using the Bronsted equation

$\log(k_{i,real}) = \beta pK_a$, in which β is a parameter that correlates basicity with nucleophilicity and is assumed to be 0.8 (Jencks and Carriuolo, 1960), we arrive at Eq. 19 relating the pK_a of an amino group with the observed rate constant of the acylation of this group. The pK_a of Phe($i = 1$) and Gly($i = 2$) groups of insulin have

$$K_I = \frac{10^{pK_a}}{10^{pK_a-pH}+1}, \qquad (19)$$

previously been experimentally determined from the pH dependence of the electrophoretic mobilities, and the extent of acylation of these groups with acetic anhydride (Gao et al., 1995) to be 8.4 and 7.1, respectively. Using these pK_a values, and a value of 10.0 for Lys($i = 3$) (Matthews and van Holde, 1990), and the value of pH = 6.8 for the acylation reaction, the relative magnitudes of the observed rate constants are $k_1:k_2:k_3 = 2.5$ (±0.7):2.0 (±0.4):1 (±0.25). The deviations in the rate constants were estimated assuming uncertainty in the values of pK_a of 0.5. The relative magnitudes of the observed rate constants in our experiments fall into these uncertainty limits.

CONCLUSIONS

In this paper, we presented a recursive method for solving reagent-coupled unimolecular systems: series, asymptotic, and average solutions were found. We validated the theory using experimental data for the acylation of insulin. Because the algorithms presented in this study are of general nature (but limited to irreversible reactions), they can be used to study kinetics of more complicated systems, in which molecules with multiple recognition (or reaction) sites interact with multiple equivalents of reagent. The calculations performed using the method are fast, because it does not require numerical integration of kinetic equations. Although, in this study, we assumed noncooperativity of acylation reactions, this assumption is not necessary: the method can solve an RCUS of an arbitrary dimension. When the rate constants are to be found from the experimental data, our method can accelerate the calculations, but cannot circumvent the problems associated with common search algorithms (e.g., Monte Carlo): for large sets of independent rate constants, finding the global minimum (best fit) becomes exceedingly complicated and, if the uncertainties in the experimental data are large, might yield meaningless solutions.

EXPERIMENTAL

Equipment

A Beckman P/ACE system 5500 capillary electrophoresis system was used in the experiments. The capillary tubing (Polymicro Technologie, Inc., Phoenix, AZ) was of un-

FIGURE 4 Schematic electropherograms of acylated derivatives of insulin for different values of time. In this study, fluorescein was used as the acylating agent. The hatched bars correspond to experimental data, and the solid bars were generated using an average-exponential method.
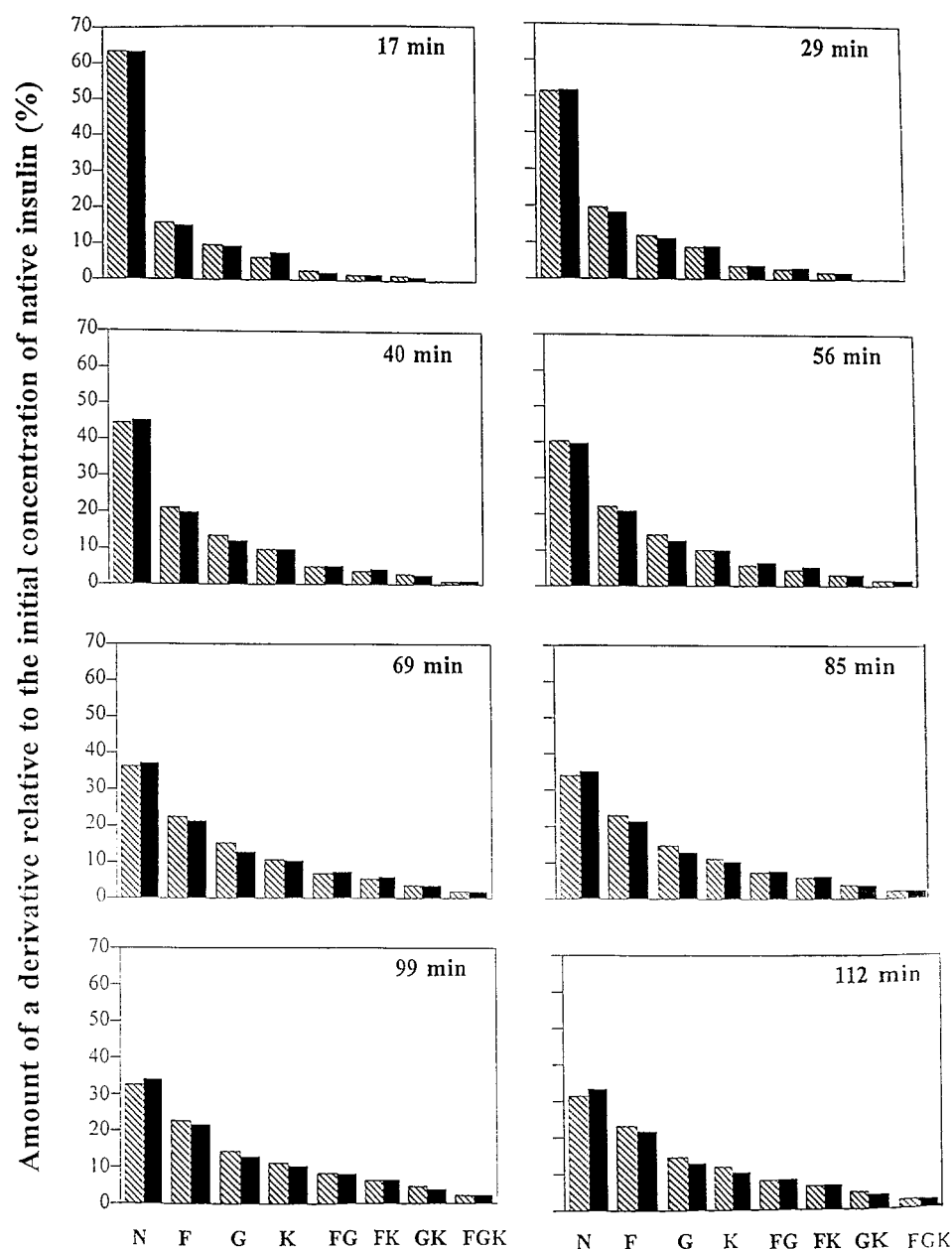
coated fused silica with an internal diameter of 50 μm. The samples (8 nl) were introduced into the capillary by a positive pressure injection. A voltage of 30 kV was applied to the capillary. The detection of the derivatized protein was done by monitoring the UV absorbance at 214 nm.

Reagents

All chemicals were of analytical grade and were used without further purification. Bovine pancreatic insulin was purchased from Sigma (St. Louis, MO). 5-carboxyfluorescein succimidyl ester was purchased from Molecular Probes

(Eugene, OR), and fluorescamine was purchased from Aldrich (Milwaukee, WI). Stock solutions of insulin (1.5 mg/ml, 260 μM) were prepared by dissolving the protein in pH 12 water and then diluting this solution with an equal volume of potassium phosphate buffer (100 mM, pH 6.8).

Procedures

1) Acylation with Fluorescamine. Different volumes of 100 mM buffer solution of Fluorescamine (2, 4, 6, 8, 10, and 12 μl) were added to six samples of 1.5 μl insulin buffer solution (260 μM). The mixtures were allowed to react until
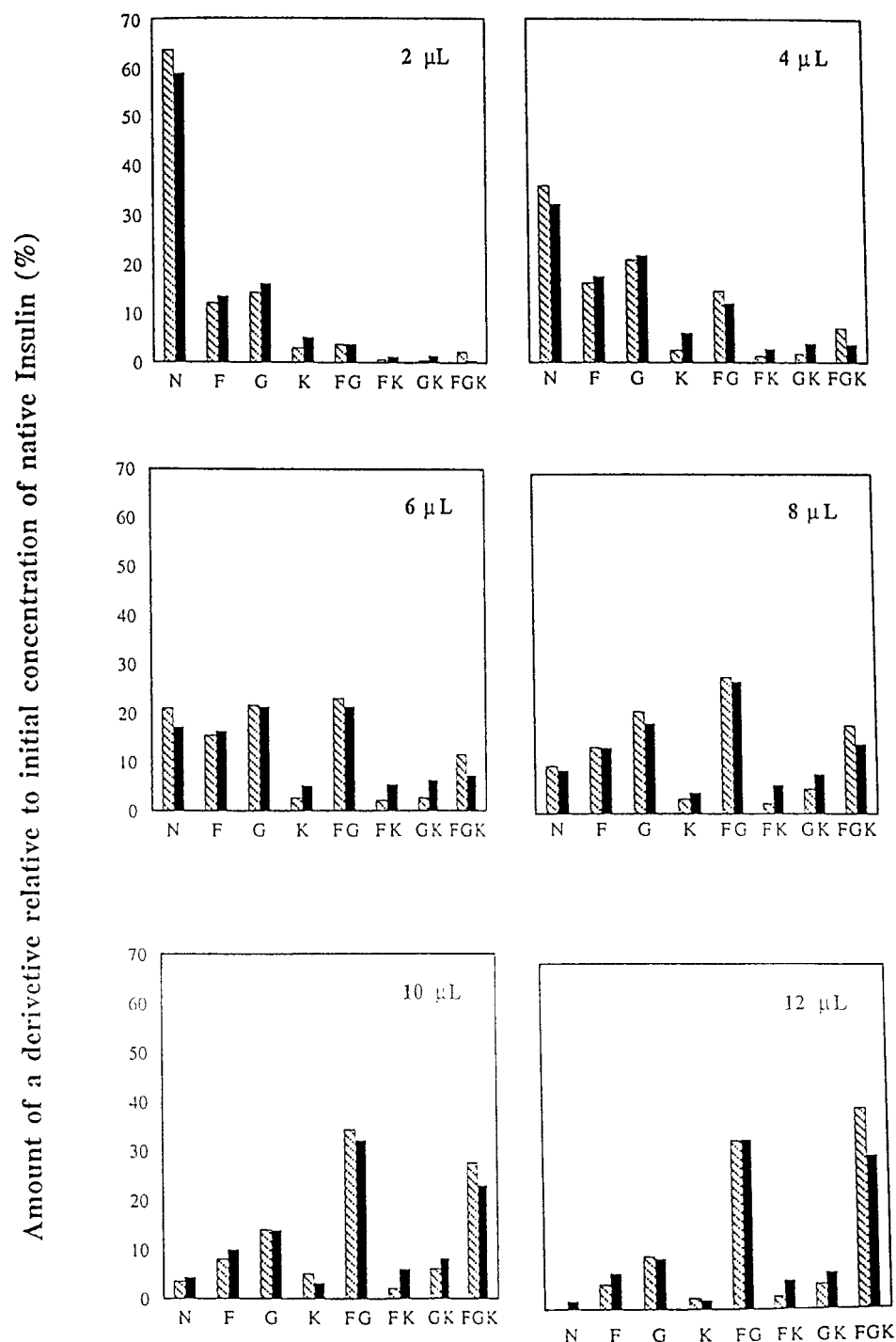
FIGURE 5 Set of schematic electropherograms of acylated derivatives of insulin for different initial amounts of the acylating agent (fluorescamine). The hatched bars correspond to experimental data, and the solid bars were generated using long-time asymptotic method.

stable concentrations of intermediates were reached (~2 h), and were subsequently analyzed by capillary electrophoresis. 2) Acylation with Fluorescein. Buffer solution of Fluorescein (7.5 μl, 100 mM) was mixed with 500 μl of buffer insulin solution (260 μl). The reaction was quenched at specified times (up to 112 min) with fivefold excess of hydroxylamine. The samples were analyzed by capillary electrophoresis.

Computational Methods

All the computer codes were implemented in C language. Numerical integration of kinetic equations was performed using a fourth-order Runge–Kutta method. Optimization of the values of rate constants was achieved using a grid search algorithm. The equation for S_∞ was solved using a bisection method.

APPENDIX

To derive a series solution for $R(t)$ for small values of times, we expand the exponentials in Eq. 11 in the power series

$$R(t) - \sum_{m=1}^{n} B_m \cdot \left\{ 1 + p_m \cdot S(t) + \frac{p_m^2}{2!} \cdot S^2(t) + \ldots \right\} - k_R \cdot S(t)$$

$$= R(0) - \sum_{m=1}^{n} B_m. \quad \text{(A1)}$$

Grouping the terms at equal powers of $S(t)$ and introducing

$$H_i = \sum_{m=1}^{n} \left\{ \frac{p_m^i \cdot B_m}{i!} \right\} + \delta_{1,i} \cdot k_R$$

leads to Eq. A2. Next, $S(t)$ is expanded in a power series: $R(t) = \sum_{z=0}^{\infty} f_z \cdot t^z$. From

$$R(t) - \sum_{i=1}^{\infty} H_i \cdot S^i(t) = R(0), \quad \text{(A2)}$$

the definition of $S(t)$, we obtain

$$S(t) = \int_0^\infty R(t') \, dt' = \int_0^\infty \left(\sum_{z=0}^{\infty} f_z \cdot t'^z \right) dt' = \sum_{z=0}^{\infty} \frac{f_z}{z+1} \cdot t^{z+1}.$$

$$\text{(A3)}$$

Substitution of A3 into A2 leads to

$$\sum_{z=0}^{\infty} f_z \cdot t^z - \sum_{i=1}^{\infty} H_i \left\{ \sum_{z=0}^{\infty} \frac{f_z}{z+1} \cdot t^{z+1} \right\} - R(0) = 0. \quad \text{(A4)}$$

After rearranging the terms and equating coefficients at all powers of t to zero, the series expansion coefficients f_z are obtained. The first four of them are given in A5:

$$f_0 = R(0) \quad \text{(A5)}$$

$$f_1 = H_1 \cdot R(0)$$

$$f_2 = H_2 \cdot R^2(0) + \frac{H_1^2 \cdot R(0)}{2}$$

$$f_3 = H_3 \cdot R^3(0) + \frac{4}{3} \cdot H_2 \cdot H_1 \cdot R^2(0) + \frac{1}{6} H_1^3 \cdot R(0)$$

REFERENCES

Berberan-Santos, M. N., and J. M. G. Martinho. 1990. The integration of kinetic rate equations by matrix methods. *J. Chem. Educ.* 67:375–379.

Chu, Y. H., W. J. Lees, A. Stassinopoulos, and C. T. Walsh. 1994. Using affinity capillary electrophoresis to determine binding stoichiometries of protein–ligand interactions. *Biochemistry.* 33:10616–10621.

Gao, J., F. A. Gomez, R. Haerter, and G. M. Whitesides. Determination of effective charge of a protein in solution by capillary electrophoresis. 1994. *Proc. Natl. Acad. Sci. USA.* 91:12027–12030.

Gao, J., M. Mrksich, F. A. Gomez, and G. M. Whitesides. 1995. Using capillary electrophoresis to follow the acylation of the amino groups of insulin and to estimate their basicities. *Anal. Chem.* 67:3093–3100.

Gao, J., M. Mammen, and G. M. Whitesides. 1996. Evaluating electrostatic contributions to binding with the use of protein charge ladders. *Science.* 272:535–537.

Gao, J., and G. M. Whitesides. 1997. Using protein charge ladders to estimate the effective charges and molecular weights of proteins in solution. *Anal. Chem.* 69:575–580.

Gilson, M. K., and B. H. Honig. 1987. Calculation of electrostatic potentials in an enzyme active site. *Nature.* 330:84–86.

Imai, K. J. 1983. The Monod–Wyman–Changeux allosteric model describes Haemoglobin oxygenation with only one adjustable parameter. *Mol. Biol.* 167:741–749.

Jencks, W. P., and J. Carrriuolo. 1960. Reactivity of nucleophilic reagents towards esters. *J. Am. Chem. Soc.* 82:1778–1786.

Kuhr, W. G., and C. A. Monnig. 1992. Capillary Electrophoresis. *Anal. Chem.* 64:389–407.

Liu, J., K. J. Volk, M. S. Lee, M. Pucci, S. Handwerger. 1994. Binding studies of vancomycin to the cytoplasmic peptidoglycan precursors by affinity capillary electrophoresis. *Anal. Chem.* 66:2412–2416.

Mathews, C. K., and K. E. van Holde. 1990. Biochemistry. Benjamin-Cummings, Redwood City, CA.

Mattevi, A., M. Rizzi, and M. Bolognesi. 1996. New structures of allosteric proteins revealing remarkable conformational changes. *Curr. Opin. Struc. Biol.* 6:824–829.

Monaco, H. L., J. L. Crawford, and W. N. Lipscomb. 1978. Three-dimensional structures of aspartate carbamoyltransferase from *Escherichia coli* and of its complex with cytidine triphosphate. *Proc. Nat. Acad. Sci. U.S.A.* 75:5276–5280.

Perutz, M. F. 1989. Mechanisms of Cooperativity and Allosteric Regulation in Proteins. *Quart. Rev. Biophys.* 22:139–236.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1996. Numerical Recepies in C. The Art of Scientific Computing. Cambridge University Press, Cambridge, MA.

Rodiguin, N. M., and E. N. Rodiguina. 1960. Consecutive Chemical Reactions. Van Nostrand, Princeton, NJ.

Stein, S., P. Bohlen, and S. Udenfriend. 1973. Amino acid analysis with fluorescamine at the picomole level. *Arch. Biochem. Biophys.* 155:203–212.

Tizard, I. 1984. Immunology, an Introduction. Saunders College Publishers, Philadelphia, PA.