Stochastic Aspects of Enzyme Inhibition*

Hiroshi Asai† and Manuel F. Morales‡

ABSTRACT: The "activity" (enzymatic, antigenic, and the like) of a macromolecule can often be modified by reacting some of its groups with suitable reagents. In the usual experiment, activity and the number of reagent molecules attached (per macromolecule or per fragment) are measured; however, whether attachment is to relevant or irrelevant sites and whether reaction of more than one site is necessary is unknown. For this reason the relationship, activity versus extent of reaction, must be interpreted taking into account certain stochastic factors. W. J. Ray and D. E. Koshland (*J. Biol. Chem. 236*, 1973 [1961]) first clearly recognized this problem. In this paper the problem is formulated, and solved for some plausible models; also, a few literature cases susceptible to analysis are discussed. The present theory treats not only the case wherein groups respond independently to an inexhaustible concentration of reagent, but also cases in which reagent amount is finite and those in which the reaction of some groups is conditional on the reaction of others.

Identification of the active site of an enzyme is one of the most important problems in contemporary protein chemistry. For this purpose, chemical modification of enzymes has been extensively used. However, conclusions derived from such experiments are frequently unreliable because (1) generally it is very difficult to find a reagent which reacts with only a single type of amino acid residue, (2) usually there are too many amino acids with comparable reactivities involved and not involved in the active site, and (3) modification of an amino acid residue which is not related directly to the active site may cause a change in tertiary or secondary structure of the active site and a concomitant loss in enzyme activity.

On the other hand, there are no other good methods for identifying the active site of a high molecular weight enzyme which cannot be crystallized, because it is then hopeless to determine its three-dimensional structure and often even its primary amino acid sequence. Therefore, in spite of the difficulties mentioned, chemical modification has been and still is the best generally applicable method for studying the active site. In order to get reliable conclusions about the active site from the experiments on chemical modification, however, some strategy of experiments and quantitative analysis of data must be used.

In this paper we will present a rather general statistical theory of the correlation between probability of chemical modification of enzyme and the corresponding loss in activity. The probabilities of chemical modification are usually determined experimentally by amino acid analysis or peptide analysis. Such an approach was first developed by Ray and Koshland (1961). However, their theory is limited to the following two cases: (1) when the reaction rate of chemical modification can be determined directly as a function of incubation time because the reaction is very slow, and during reaction the amount of unreacted reagent is much higher than that of reacted reagent, and (2) photooxidation, wherein the above-mentioned condition of reaction is spontaneously satisfied. The present theory is of a more general nature, and therefore can also be applied to fast chemical modification. Such an extension of the theory is practically very important.

The assumption underlying the theory in section (*A*) is that an amino acid residue in an enzyme is modified by the reagent independently of other residues; in other words, chemical modification is stochastically independent. Therefore this theory cannot be applied to the case when chemical modification of a special amino acid residue of the enzyme changes the reactivity of some other residues through a change in secondary or tertiary structure of the enzyme. A completely general treatment of such conditional cases is too complicated. In section (*B*), however, a few special cases in which chemical modification causes structural change of enzyme are solved analytically, and experimentally applicable expressions are obtained.

Theory

We shall be concerned with the *irreversible* reaction between a reagent and an enzyme. We shall consider the case wherein the receptor groups (e.g., amino acid residues of a particular kind or peptide fragments) can be identified by their primary sequence in the enzyme or by their positions on chromatography. For a given extent of reaction, i.e., for a given amount of

* From the Cardiovascular Research Institute, University of California, San Francisco. *Received December 17, 1964.* Supported by grants from the National Science Foundation and the American Heart Association.

† U.S. Public Health Service Postdoctoral Trainee (grant HTS-5251) and Travel Awardee of the International Cardiology Foundation. Present address: Institute of Molecular Biology, Faculty of Science, Nagoya University, Nagoya, Japan.

‡ Career Investigator, American Heart Association.

reagent used up, we are interested in knowing how the bound reagent is distributed among the various receptor groups. If the activity of the enzyme is always expressible as a function of the states of *particular* receptor groups, we can say that these particular groups are involved in the active site. The problem is to find this function and these groups from experimental results. We shall at first encounter a chemical kinetic problem, then a combinatorial problem.

(*A*) *Stochastically Independent Reaction.* Let us suppose that each receptor group constitutes one reactive site which reacts with reagent *independently* of other groups. The velocity constant characterizing the bimolecular reaction between the reagent and each one of n groups, say the ith, is k_i. The molar concentration of reacted ith group is x_i. The kinetic aspect of the problem is then described by the equations:

$$dx_i(t)/dt = k_i[N - x_i(t)][R_o - T(t)] \qquad (i = 1, 2, \ldots, n) \quad (1)$$

and

$$T(t) = \sum_{i=1}^{n} x_i(t) \qquad (2)$$

where N is the molar concentration of macromolecule, R_o is the initial molar concentration of reagent, and $T(t)$ is the total molar concentration of the reacted reagent. Dividing both sides of equation (1) by N and introducing the notation, $a_i = x_i(t)/N$, we obtain

$$da_i/dt = k_i(1 - a_i)[R_o - T(t)] \qquad (1)'$$

and

$$T(t) = \sum_{i=1}^{n} Na_i \qquad (2)'$$

In equations (1)′ and (2)′ we may think of a_i as the *probability* that the ith group of an enzyme has been reacted. This concept of probability will prove very useful. Distribution and reaction probabilities of receptor groups are determined experimentally by some method of chemical analysis, e.g., chromatography, paper electrophoresis, and the like. It is very convenient to use isotopically labeled reagents for chemical modification, because the reaction probabilities can then be determined very easily.

Before considering the solution of equations (1)′ and (2)′ we note that if R_o is in such great excess that $T(t)$ is always negligible relative to R_o, or if the agency which modifies the group is not a chemical substance but, say, a constant flux of radiation, then,

$$e^{u_i} = 1 - a_i = e^{-k_i R_o t} \qquad (3)$$

This is the solution obtained by Ray and Koshland (1961). In this case, if the probability, $1 - a_i$, that the ith group has not reacted is determined experimentally at various time intervals, it should be found that $u_i = \log(1 - a_i)$ is proportional to time. Such experiments may be made when the velocity constant k_i is relatively small. Another point to note is that in some instances $T(t)$ is directly measurable, so that $T(t)$ may be considered to be a known function, and

$$1 - a_i = \exp\left\{-k_i\left[R_o t - \int_0^t T(t)dt\right]\right\} \qquad (4)$$

Equations of the form (1)′ can be combined to give

$$\frac{da_1}{k_1(1 - a_1)} = \frac{da_2}{k_2(1 - a_2)} = \cdots = \frac{da_i}{k_i(1 - a_i)} = \cdots = [R_o - T(t)]dt$$

from which it follows that

$$\frac{u_i}{u_1} = \frac{\log(1 - a_i)}{\log(1 - a_1)} = \frac{k_i}{k_1} \qquad (i = 2, 3, \cdots, n) \quad (5)$$

Although a_i is a function of R_o and time t, equation (5) does not contain R_o and t explicitly. Therefore, all k_i can be determined by the use of this equation even though they may be too large to be determined directly from equation (3). The ratio of u_i to u_1 gives the velocity constant (k_i/k_1) of the ith group relative to a particular group (e.g., the 1st group). Since k_i/k_1 is constant, independent of R_o and t, u_i is a linear function of u_1, with slope k_i/k_1. Preservation of this proportionality verifies the assumption of stochastic independence of reaction.

By substituting (5) into equation (1)′, we obtain

$$da_1/dt = k_1(1 - a_1)(R_o - nN + N\sum_{i=1}^{n}(1 - a_1)^{k_i/k_1}) \quad (6)$$

The analytical integration of equation (6) in general will not be easy. If we are considering the situation in which all the reagent that is going to react has reacted at a sufficient time of incubation of reaction solution, then equation (6), with $da_1/dt = 0$, constitutes an algebraic equation for finding a_1 (and, with equation [5], for finding all a_i). If the right-hand side of equation (6) is set at zero, then obviously one solution for a_1 is $a_1 = 1$. This corresponds to the case wherein R_o has exceeded the total number of receptor sites and the reaction has gone to completion. This is therefore a case of no interest, and we pursue the search for a solution for which $0 < a_1 < 1$ in the equation

$$\sum_i (1 - a_i) = \sum_i (1 - a_1)^{k_i/k_1} = n - R_o/N \qquad (R_o < nN;\ t = \infty) \quad (7)$$

Therefore, especially in the case of large k_i, a_i must be measured after a sufficient time of incubation of reac-

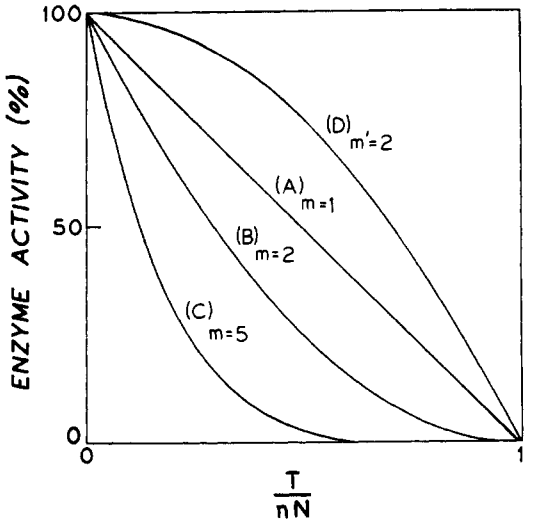

FIGURE 1: Relation between enzyme activity (%) and the probability (T/nN) that an arbitrary group has reacted. Here m and m' are the numbers of groups involved in the active site directly and indirectly, respectively. For detail, see equations (9) and (11), and the text.

tion solution under the condition that the initial concentration of reagent R_0 does not exceed the total concentration of receptor sites nN.

Solution of equations (5) and (6) yields all the probabilities, a_i, in terms of quantities which at least in principle are experimentally measurable. If a_i is the probability that the ith group is reacted, and b_i ($= 1 - a_i$) is the probability that the ith group is not reacted, then each term in the expansion of the "generating function," $\prod_{i=1}^{i=n} (a_i + b_i)$, gives the probability that on a molecule chosen at random all groups contributing "a's" to the term are reacted, and all groups contributing "b's" to the term are unreacted. This probability can also be considered as the fraction of all the molecules having the stated distribution of reacted and unreacted groups; from this point of view the solution is a mixture of diverse molecular species. Now, to specify the active site is to specify that for a molecule to be active certain groups must be reacted and certain others must not be reacted. Given this criterion, one may decide which of the diverse molecular species satisfy the criterion. Adding together the probabilities of the acceptable species then gives A, the fraction of the macromolecules which are active (i.e., the normalized "activity"). If one knows the a_i's (either by calculating from knowledge of concentrations and velocity constants or from direct determination on, say, a "fingerprint") then a comparison of the predicted A with the observed A is a check on whether the active site was correctly specified in the model. Several illustrative examples follow.

TYPE 1. If reaction of any one of the first m residues in an enzyme abolishes activity, and reaction of other residues is without effect, then the normalized (to unity) average activity (A) from many enzyme molecules is

$$A = \prod_{i=1}^{m} b_i \prod_{i=m+1}^{n} (a_i + b_i) = \prod_{i=1}^{m} (1 - a_i) \qquad (8)$$

This situation arises when all groups which participate in the active site are directly involved in enzyme action. When the situation obtains, a plot of log A against a particularly selected log $(1 - a_i)$ is straight.

If in fact all $a_i = a$, then the combination of equations (8) and (2)′ leads to

$$A = (1 - T/nN)^m \qquad (9)$$

Equation (9) indicates that a plot of log A against log $[1 - (T/nN)]$ would yield the value of m, i.e., the number of groups in the active site. In Figure 1, the activities at various values of m are plotted against T/nN. When $m = 1$, the activity decreases linearly with increasing T and it becomes 50% at $T/nN = \frac{1}{2}$. If m becomes large, the activity decreases rapidly with reaction of residues. Such plots are useful when detailed information about the reaction probabilities is lacking.

TYPE 2. The first m residues are involved in the activity, but in order to destroy the activity all of these residues must be reacted. Then in the expansion of the generating function it follows that the term $\prod_{i=1}^{m} a_i$ is the probability that all the groups 1 to m are reacted; therefore $1 - \prod_{i=1}^{m} a_i$ is the probability that not all groups 1 to m are reacted. Therefore

$$A = (1 - \prod_{i=1}^{m} a_i) \qquad (10)$$

This situation may arise when modification of the groups 1 to m causes indirect inactivation of the active site through destruction of its secondary or tertiary structure, and the active site itself is not modified by the reagent used.

If again we assume that all $a_i = a$, then the combination of equations (10) and (2)′ leads to

$$A = 1 - (T/nN)^{m'} \qquad (11)$$

The activity when $m' = 2$ is plotted in Figure 1.

TYPE 3. If an enzyme has two regions (sites) at which a single substrate molecule binds and if destruction of the first site causes concomitant loss of activity, and modification of the second site activates the enzyme f-fold, then the activity of the enzyme is given by

$$A = (1 - P_1)[(1 - P_2) + fP_2] \qquad (12)$$

where P_1 is the probability that the first site is destroyed, and P_2 is the probability of reaction of the

second site. Mathematical expressions for P_1 and P_2 depend, of course, on the types of these sites. An example of this situation might be myosin ATPase modified by an SH reagent (see Rainford *et al.*, 1964).

(*B*) *Conditional Reaction of Groups.* In the previous section (*A*), we have presented a simple kinetic theory of reaction of receptor groups in enzymes, useful for identifying enzymatically active sites, based on the assumption that each group reacts with reagent independently of reactions of other groups. This assumption may be realistic if all receptor groups are located on the surface of the enzyme and if the secondary and tertiary structures are not changed in the course of reactions of the groups. However, chemical modification sometimes causes structural change in the enzyme, or causes steric hindrance to the reaction of vicinal groups. To treat such cases we must develop a more general kinetic equation of chemical modification. Unfortunately, a completely general case of chemical modification is too complex to solve, and only some simplified special cases of contingency will be solved analytically to obtain experimentally applicable formulas.

Let us denote by E the probability that no group of the enzyme molecule has reacted, by E_i the probability that the ith group has reacted (but no other groups), by E_{ij} the probability that at first the ith group and then the jth group have reacted (but no other groups), by E_{ijk} the probability that at first the ith, then the jth, and finally the kth group has reacted (but no other groups), and so on. Then, understanding that in any term no two indices can be the same, we have

$$1 = E + \sum_i E_i + \sum_i \sum_j E_{ij} + \sum_i \sum_j \sum_k E_{ijk} + \cdots \quad (13)$$

and

$$a_i = E_i + \sum_j (E_{ij} + E_{ji}) + \sum_j \sum_k (E_{ijk} + E_{jik} + E_{jki}) + \cdots \quad (14)$$

Taking into account equation (14), we may rewrite and rearrange equation (13) to:

$$E = 1 - \{a_i + \sum_j E_j + \sum_j \sum_k E_{jk} + \sum_j \sum_k \sum_l E_{jkl} + \cdots\} \quad (15)$$

where it is understood that in the terms of the summations of equation (15) no two indices can be the same and no index can be i. To obtain the rate of growth of a_i we must multiply the probability of occurrence of each species in which the ith group is unfilled by a velocity constant characteristic of that species; thus, if k_i is the velocity constant of reacting the ith group when no other groups are reacted, k_{ji} is the velocity constant of filling the ith group when only the jth group is reacted, k_{jki} is that when only the jth and kth are filled, and so on, then

$$da_i/dt = \{k_i E + \sum_j k_{ji} E_j + \sum_j \sum_k k_{jki} E_{jk} + \cdots\}\{R_o - T(t)\} \quad (16)$$

But substituting from (15) into (16), we obtain

$$da_i/dt = \{k_i(1 - a_i) + \sum_j (k_{ji} - k_i) E_j + \sum_j \sum_k (k_{jki} - k_i) E_{jk} + \cdots\}\{R_o - T(t)\} \quad (17)$$

Note that if all velocity constants are equal, equation (17) reduces to equation (1)′. Also, note that the first bracketed factor of equation (17) is a factor proportional to e_i', the probability that those conditions requisite for the reaction of the ith group have been met, but that the group has not yet reacted. This meaning will be made clear by some illustrations.

Let us imagine that the receptor sites of the protein are of two kinds. Each group of the first kind is "buried" when the protein is native, and its velocity constant of reaction with reagent under these circumstances is zero. Each group of the second kind reacts independently with reagent; if a particular pattern of reacted sites is achieved among the sites of the second kind, the protein undergoes a conformational change, and every group of the first kind becomes free to react, each with its characteristic velocity constant. The object is to derive the expected relation between a for any group of the first kind, say, the ith, and groups of the second kind involved in the particular pattern requisite for the conformational change, say, the sth group, the tth group, and so on.

Referring to equation (17), we see that the typical velocity constant is $k_{\alpha,\beta,\ldots\omega i}$, in which the numbers, $\alpha, \beta, \ldots \omega$ specify the sites reacted prior to the proposed reaction of the ith group. If the requisite distribution does not appear in $\alpha, \beta, \ldots \omega$ we set that velocity constant equal to zero, and if it does appear we set it equal to some value characteristic of the ith group, say, k_i'. For example, if the requisite distribution is simply "reaction of the sth group," then only those constants whose subscript contains s will survive; if it is "reaction of the sth group and the tth group," then only those constants whose subscript contains both s and t will survive; if it is "reaction of the sth group, or the tth group, or both," then a constant will survive if its subscript contains either s or t or both, and so on. Since all surviving constants will have the same value (k_i'), they will each factor out, and equation (17) will always be expressible as

$$da_i/dt = k_i' e_i' \{R_o - T(t)\} \quad (18)$$

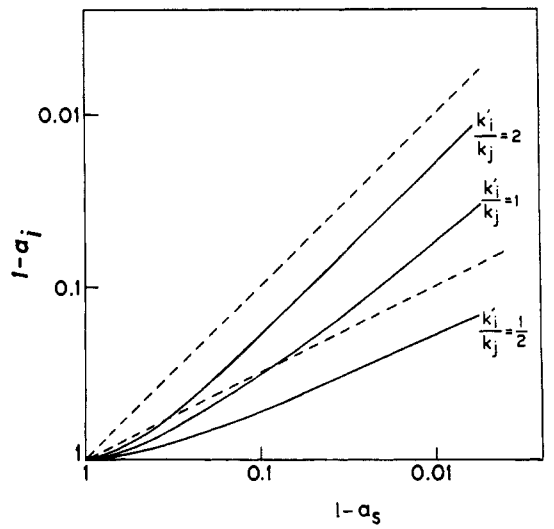

FIGURE 2: Logarithmic relation between the probability, $(1 - a_i)$, that the ith group of the first class has not reacted and the probability, $(1 - a_s)$, that the sth group of the second class has not reacted. It is assumed that the groups in the first class react with reagent only after the sth group of the second class has reacted and caused a change of enzyme structure. See equation (23).

where e_i' will be a function of those E's whose subscripts also contain the requisite distribution. The form of equation (18), compared to that of equation (1)', suggests that e_i' is the probability that, in a molecule chosen at random, the requisite distribution has been achieved, but the ith group has not yet reacted. Therefore, if we call e_i the probability that the requisite distribution has been achieved (regardless of whether the ith group has reacted), it must be that

$$e_i = e_i' + a_i \tag{19}$$

Recalling that groups of the second kind react independently, it is clear that for the three illustrations given e_i is, respectively, a_s, or $a_s a_t$, or $1 - (1 - a_s)(1 - a_t)$. In other words, e_i is a function specified by the hypothesis and numerically calculable if the a's of critical groups of the second kind are experimentally measurable. Therefore, using equation (19) we may rewrite (18) as

$$da_i/dt = k_i'(e_i - a_i)\{R_o - T(t)\} \tag{20}$$

where e_i is to be considered some known function of a_s, a_t, and so on. Applying equation (1)' to one of the critical groups, say, the sth, we may divide equation (20) by equation (1)', rearrange, and obtain

$$\frac{da_i}{da_s} + \frac{r}{1 - a_s}a_i = \frac{re_i}{1 - a_s} \tag{21}$$

where $r \equiv k_i'/k_s$. Changing to the variable, $u_s = \log (1 - a_s)$, and integrating,

$$a_i(u_s) = -re^{ru}\int_0^{u_s} e_i e^{-ru}\, du_s \tag{22}$$

since $u_s = 1$ when $a_i = 0$. Now, let us consider the simple cases already mentioned. Suppose $e_s = a_s \equiv 1 - e^{u_s}$, then, from equation (22) we find that

$$e^{u_i} \equiv (1 - a_i) = \frac{1}{1 - r}\{e^{ru_s} - re^{u_s}\} \tag{23}$$

whence

$$\delta \equiv \frac{du_i}{du_s} = r\frac{e^{ru_s} - e^{u_s}}{e^{ru_s} - re^{u_s}} \tag{24}$$

Thus, regardless of the value of r, if u_s tends to zero ($a_s \to 0$), $\delta \to 0$; on the other hand, if u_s becomes a large negative number ($a_s \to 1$), then the behavior of δ depends on r, for if r is large e^{ru_s} will be negligible relative to e^{u_s} and $\delta \to 1$, while if r is small e^{u_s} will be negligible relative to e^{ru_s} and $\delta \to r$. This diagnostic behavior of δ is illustrated in Figure 2. If $e_i = a_s a_t$, then applying equation (1)' separately to a_s and a_t we may express one probability in terms of the other, and find, for example, that $e_i = a_s a_t = a_s - a_s(1 - a_s)^p$, where $p = k_t/k_s$. Substituting this expression for e_t into equation (22) we obtain

$$\begin{aligned} e^{u_i} = 1 &+ \left\{\frac{1}{1 - r} + \frac{r}{p - r} - \frac{r}{1 + p - r}\right\}e^{ru_s} \\ &- \left\{\frac{r}{1 - r}\right\}e^{u_s} - \left\{\frac{r}{p - r}\right\}e^{pu_s} \\ &+ \left\{\frac{r}{1 + p - r}\right\}e^{(1+p)u_s} \end{aligned} \tag{25}$$

If $e_i = 1 - (1 - a_s)(1 - a_t) = 1 - (1 - a_s)^{1+p}$, then, on substitution into equation (22), we obtain

$$e^{u_i} = \left\{\frac{1 + p}{1 + p - r}\right\}e^{ru_s} - \left\{\frac{r}{1 + p - r}\right\}e^{(1+p)u_s} \tag{26}$$

Another solvable case is the following: We suppose two kinds of receptor sites. Sites of the first kind are such that as soon as one of them reacts all other sites of this kind become unreactive; sites of the second kind undergo independent reaction, in conformity with equation (1)'. If we apply equation (17) to the ith site of the first kind, we note certain simplifications in summated terms of the type $(k_{\alpha\beta\ldots\omega i} - k_i)E_{\alpha\beta\ldots\omega}$. If the reaction sequence, $\alpha\beta \ldots \omega$, involves sites of the second kind only, then the two velocity constants are equal, and the term is zero. If two or more elements in $\alpha\beta \ldots \omega$ refer to sites of the first kind, then $E_{\alpha\beta\ldots\omega} = 0$, since hypothesis forbids reaction of more than one such site.

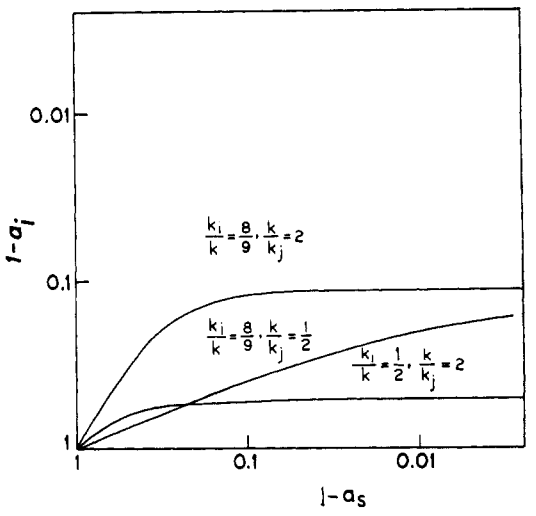

FIGURE 3: Logarithmic relation between the probability, $(1 - a_i)$, that the ith group of the first class has not reacted and the probability, $(1 - a_s)$, that the sth group of the second class has not reacted. It is assumed that a group of the second class reacts with reagent independently of all other groups and the reaction of one of the groups of the first class inhibits the reaction of all other groups of the first class (equation 29).

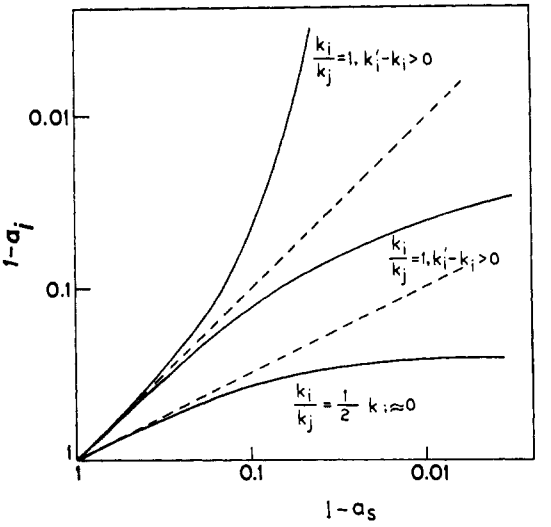

FIGURE 4: Logarithmic relation between the probability, $(1 - a_i)$, that the ith group of the first class has not reacted, and the probability, $(1 - a_s)$, that the sth group of the second class has not reacted. It is assumed that on reaction of the sth group the ith group changes its reactivity from k_i to k_i'. This is a generalization of Figure 4 (equation 32).

If $\alpha\beta \ldots \omega$ consists of the λth site of the first kind ($\lambda \neq i$), and a complementary set, $C(\lambda)$, all of whose elements refer to sites of the second kind, then $k_{\lambda,C(\lambda)i} = 0$, since once λ has reacted i may not, but a term, $-k_i E_{\lambda,C(\lambda)}$, survives. If we group together all terms thus involving λ, we find their sum to be $-k_i \sum_{C(\lambda)} E_{\lambda,C(\lambda)}$, where the summation is over all possible distributions of reacted sites of the second kind; therefore, $\sum_{C(\lambda)} E_{\lambda,C(\lambda)} = a\lambda$, and the sum is $-k_i a_\lambda$. The summations of equation (17) thus yield $-k_i \sum_\lambda a_\lambda$, where λ is, in turn, each site of the first kind except the ith site. Combining this last sum with the term $k_i(1 - a_i)$, we obtain finally the special case of equation (17) applicable to sites of the first kind, viz.,

$$da_i/dt = k_i\{1 - \sum_j a_j\}\{R_o - T(t)\} \qquad (27)$$

In equation (27) the summation extends only over sites of the first kind (including the ith), so that the first bracketed factor is, as it should be, the probability that no site of the first kind has reacted. If we apply equation (27) to the mth and nth sites of the first kind and combine the results we find that $a_m = (k_m/k_n)a_n$. By using this relationship we may show that $k_i \sum_j a_j = a_i \sum_j k_j \equiv a_i S_k$. Making this substitution into equation (27), and combining it with equation (1)′ as applied to the sth site of the second kind, we obtain

$$da_i/(k_i - a_i S_k) = -du_s/k_s \qquad (28)$$

Integrating and rearranging, we obtain finally

$$1 - a_i = e^{u_i}$$

$$= 1 - (k_i/S_k)(1 - e^{(S_k/k_s)u_s}) \qquad (29)$$

An example of equation (29) appears in Figure 3.

Finally, we may consider the case that the ith site is accessible to reaction with velocity constant k_i, but that if a particular set of other sites, namely, $\alpha,\beta, \ldots \omega$, is achieved, then the ith site reacts with velocity constant $k_i' \equiv k_{\alpha\beta\ldots\omega i}$. If also $e_i' \equiv E_{\alpha,\beta,\ldots\omega}$, then equation (17) is

$$da_i/dt$$

$$= \{k_i(1 - a_i) + (k_i' - k_i)e_i'\}\{R_o - T(t)\} \qquad (30)$$

If we imagine that $e_i' = e_i'(a_s)$ (physically this means that reaction of the sites, $\alpha,\beta, \ldots \omega$ is related in an explicit way to reaction of the sth site), then

$$\frac{da_i}{k_i(1 - a_i) + (k_i' - k_i)e_i'} = -\frac{du_s}{k_s} \qquad (31)$$

and

$$e^{u_i} = 1 - a_i$$

$$= \frac{k_i' - k_i}{k_s} e^{(k_i/k_s)u_s} \int_0^{u_s} e_i' e^{-(k_i/k_s)u_s} du_s \qquad (32)$$

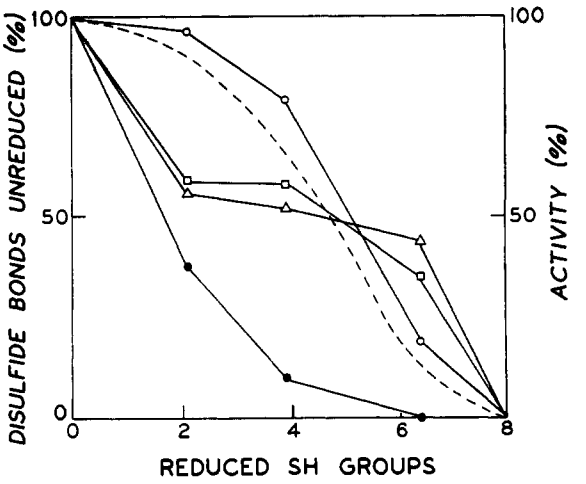

FIGURE 5: RNAase activity (%) and disulfide bonds unreduced (%) at various average values of reduced SH groups. Dotted line, RNAase activity; ●, unreduced 1–6 disulfide bond; ○, unreduced 4–5 disulfide bond; □, unreduced 3–7 disulfide bond; △, unreduced 2–8 disulfide bond.

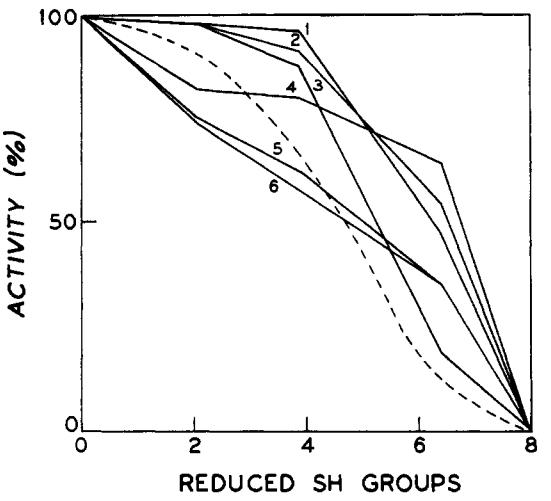

FIGURE 6: RNAase activity (%) for various average values of reduced SH groups. Activity was calculated by equation (10) on the assumption that only simultaneous reduction of two disulfide bonds out of four bonds destroys active structure of the enzyme. Curve 1, 4–5 and 3–7 bond; curve 2, 4–5 and 2–8 bond; curve 3, 4–5 and 1–6 bond; curve 4, 3–7 and 2–8 bond; curve 5, 3–7 and 1–6 bond; curve 6, 2–8 and 1–6 bond; dotted line, RNAase activity.

Alternatively, we may write (31) as,

$$\frac{du_i}{du_s} = \frac{k_i}{k_s} + \frac{k_i' - k_i}{k_s} e_i' e^{-u_i} \tag{33}$$

from which it follows that $\lim_{u_s \to 0} (du_i/du_s) = k_i/k_s$ (since $\lim_{u_s \to 0} e_i' = 0$). Furthermore, whether for small values of u_s the derivative du_i/du_s is larger or smaller than k_i/k_s will be governed by the sign of $k_i' - k_i$, i.e., by whether $k_i' > k_i$ or $k_i' < k_i$. If the latter, we may say that the reaction of the ith group is inhibited as a result of a structural change occurring when the sth group is reacted. An example of equation (32) is shown in Figure 4.

Analysis of Experimental Data and Discussion

Many experimental studies have recently been made on the loss of activity of enzyme induced by treatment with specific amino acid reagents, and important information about the active site in such enzymes has thus been obtained. For example, it has been suggested that cysteine or histidine residues take part in many enzymatic processes probably by making primary linkages with substrate molecules. However, almost all experimental data published are inadequate for positively identifying the amino acid residues involved in the active site because the probability of chemical modification of amino acid residues in the enzyme has not been considered.

In this section, using the foregoing theory, we will analyze several experimental data in the literature, and point out desirable strategy in future experiments of this type.

The modification of ribonuclease (RNAase) is perhaps the best example, since the primary structure of this enzyme has been determined and many studies have been made of its active site. If all four disulfide bonds of RNAase are reduced, its enzymatic activity is detroyed. At various stages of reduction of disulfide bonds, White *et al.* (1959) measured the loss of activity and the percentage reduction of each bond (Figure 5). The probability of reduction of the 1–6 disulfide bond is larger than that of the other three bonds, and the 4–5 disulfide bond has the smallest probability of reduction. From comparing the measured activity with the probabilities of reduction of the four disulfide bonds, it is concluded that simultaneous maintenance of two or more disulfide bonds among four bonds is not necessary for retention of activity, because, if it were necessary, the activity calculated from equation (8) with $m = 2$ would be much smaller than the measured one. If only one disulfide bond of four bonds is necessary for maintenance of the active structure, the 4–5 disulfide bond is the best candidate, because the probability of reduction of this bond follows closely the loss in activity (Figure 5). As an alternative assumption, if simultaneous reduction of two disulfide bonds of four bonds destroys the active structure, the 1–6 and 4–5 disulfide bonds are the best candidates, as shown in Figure 6 where the loss of activity was calculated for various cases by equation (10).

According to Venetianer and Straub (1963), and to

Goldberger *et al.* (1963), reduced RNAase recovers its activity by enzymatic disulfide bond formation, and this process occurs exponentially with time. In view of equation (3), such behavior would be expected if each pair of reduced cysteines is reoxidized independently of other pairs. Therefore the cited results suggest that only a particular disulfide bond is enough for maintenance of activity, provided that the bonds are formed by kinetic constants of about equal magnitude.

Therefore, from these experimental results, it is conceivable that either the 4–5 or 1–6 disulfide bond is necessary for maintenance of active structure in RNAase, but *simultaneous* preservation of these bonds is unnecessary. However, the assumption of independence of reduction or reoxidation of each disulfide bond seems to be wrong, because a plot of log percentage survived for one disulfide bond versus another did not give the straight lines expected from equation (5). Therefore exact definition of the reaction awaits more extensive studies.

It has been reported that lysine is one of the amino acid residues in the active site of RNAase (Klee and Richards, 1957). RNAase activity is decreased very sharply by guanidination of lysine residues, and the relation between the loss in activity and extent of guanidination resembles curve (B) or (C) in Figure 1. From this result, Klee and Richards inferred that there is one lysine which is involved in the active site and is guanidinated five times faster than the other nine residues. If their conclusion is true the essential lysine residue should be identifiable by amino acid analysis or by fingerprinting. In the absence of such identification, however, one cannot exclude the possibility that two or more lysine residues of the ten residues participate in the enzymatic activity. By photooxidation, Weil and Seibles (1955) showed that histidine is also one of the amino acids in the active site of RNAase. According to these authors, the photooxidation of one histidine out of four histidines in RNAase resulted in 76% inactivation. In this case, again, there are two possibilities: the responsible histidine is photooxidized faster than the other histidines, or two or more histidines are involved in the activity. Recently, by alkylation of histidine residues in RNAase by iodoacetate, it has been shown that two residues at the position 119 and 12 are responsible for enzymatic activity (Crestfield *et al.* 1963). In this case alkylation of either histidine-119 or histidine-12 prevents alkylation of the other and the ratio of the two inactive products remains always constant at about 8:1 even when the conditions for reaction, e.g., incubation time, ratio of iodoacetate to protein, and the like, are varied. Therefore this case is a good example to which to apply kinetic equation (29). The kinetic constant of alkylation of histidine-119 must be eight times larger than that of histidine-12.

Reagents of diverse reactivity "modify"[1] the nucleoside triphosphatase of myosin; among these are EDTA and sulfhydryl reagents. As regards SH reagents, it is known from amino acid analysis that myosin has upward of forty cysteinyl residues per mole, so proper consideration of statistical effects is imperative. Recently Rainford *et al.* (1964) interpreted statistically (using a variant of equation [12], i.e., assuming independently reacting sites) the curve, ATPase activity versus extent of reaction, taking care to wait until the reaction stopped ($da_i/dt = 0$; $0 < a_i < 1$); these authors achieved reasonable data fits assuming that reaction of one SH group per enzymatic site extinguished activity. Earlier, Sekine *et al.* (1962) also provided observations subject to the present analysis. In their work, myosin brought to maximal activation by EDTA was subjected to ^{14}C-labeled *N*-ethylmaleiimide, and the decay of activity with extent of label fixation by peptides of tryptic hydrolysate "fingerprint" was observed. A particular peptide ("no. 8") was predominantly labeled, and the relationship between activity and extent of labeling is shown in Figure 7. Noting that initially the activity falls linearly with the extent, these authors concluded that the reaction of one SH group per independent enzymatic site was sufficient to destroy activity. Therefore the Rainford *et al.* (1964) and the Sekine *et al.* (1962) results appear to be in harmony, but the latter invites further scrutiny from our point of view. The probability of reaction of cysteinyl groups in peptide 8, a_8, was approximately the same as the sum of the probabilities of reaction of all other cysteinyls, i.e.,

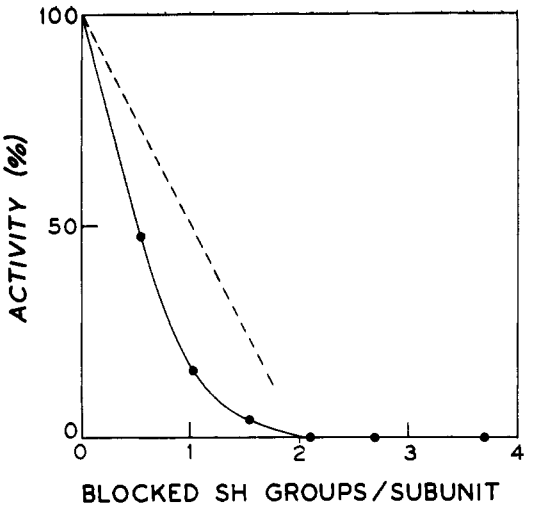

FIGURE 7: Relation between myosin ATPase activity (%) and the number of SH groups blocked by *N*-ethylmaleimide per subunit. Dotted line, the ATPase calculated by equation (26).

[1] Under "standard conditions," limited reaction of myosin with a modifying reagent elevates its ATPase activity to the level of its native ITPase activity; further reaction destroys ATPase activity. Limited reaction with reagent destroys ITPase activity. In a general way the effects of different reagents is additive; for example, ATPase may be maximally activated with one reagent and then destroyed by adding another.

$a_8 \cong \Sigma a_i\ (i \neq 8)$, so $2a_8 \cong T/N$, where T/N is the average number of moles bound per the weight which contains one enzymatic site. Therefore, according to the hypothesis of Sekine *et al.* (1962), when $T/N < 1$ the surviving fractional activity should be

$$\bar{V} = 1 - a_8 = 1 - (T/2N) \qquad (34)$$

However, equation (34) (dotted line of Figure 7) significantly overestimates the number of cysteinyls reacted for a given reduction in $\bar{V}$; furthermore, it predicts a linear fall in $\bar{V}$ with increasing T/N. Therefore further considerations are necessary. Label also appears in peptides 4, 6, 7, and 15; this labeling might arise from randomness in tryptic attack (in which case the label per cysteinyl should be the same as in 8, and the amino acid sequences in these peptides should "match up" with those in 8), or it may arise because, in fact, the model is wrong and more than one residue is involved in the loss of activity. Sekine *et al.* (1962) regard myosin as composed of three identical chains, each with one independent enzymatic site. If this *structural* proposal be correct, one may conceive of enzymatic sites cooperative between chains; for example, if there were an enzymatic site between two chains and the reaction of only the cysteinyls in no. 8 were necessary to destroy activity, then we should have $\bar{V} = [1 - (T/2N)]^2$, a relationship which actually fits the data of Figure 7 rather well. At any rate, either experimental or analytical alterations are necessary to reconcile the data with the authors' hypothesis.

Conclusions

In contemporary enzymology there is a dogma that the active center of an enzyme is small. Therefore, when a reagent affects activity its effect is supposed to be on just one group, i.e., remote effects tend to be ignored, and the existence of other similar groups tends to be dismissed on the assumption that the one group "at the site" is much more reactive than are "nonspecific" groups. This reasoning may be correct, but at the present time it cannot be advanced unequivocally. The "message" in this paper is that in many instances the activity-extent relationship can be well interpreted by models in which several groups are supposed sensitive to reagent. By recognizing the consequent ambiguity, better experimental design is possible, e.g., including substrate or analog protection experiments when the reagent reaction is studied. As already noted by Ray and Koshland (1961), the foregoing precautions apply as well to modification by reagents of other activities, e.g., immunological, hormonal, genetic, indeed to any chemical or physiological system which is potentially polyvalent to reagent.

Acknowledgement

We are indebted to Professor Fumio Oosawa of Nagoya University for valuable discussion of this work. We are also indebted to Professor Jean Botts of the University of California for a critical reading of the manuscript and for suggesting the shorter derivation of equation (17) used here.

References

Crestfield, A. M., Stein, W. H., and Moore, S. (1963), *J. Biol. Chem. 238*, 2413.

Goldberger, R. F., Epstein, C. J., and Anfinsen, C. B. (1963), *J. Biol. Chem. 238*, 628.

Klee, W. A., and Richards, R. M. (1957), *J. Biol. Chem. 229*, 489.

Rainford, P., Hotta, K., and Morales, M. (1964), *Biochemistry 3*, 1213.

Ray, W. J., and Koshland, D. E. (1961), *J. Biol. Chem. 236*, 1973.

Sekine, T., Barnett, L. M., and Kielley, W. W. (1962), *J. Biol. Chem. 237*, 2769.

Venetianer, P., and Straub, F. B. (1963), *Biochim. Biophys. Acta 67*, 166.

Weil, L., and Seibles, T. S. (1955), *Arch. Biochem. Biophys. 54*, 368.

White, F. H., Jr., and Anfinsen, C. B. (1959), *in* Symposium on Sulfur in Protein, Benesch, R., Benesch, R. E., Boyer, P. D., Klotz, I. M., Middlebrook, W. R., Szent-Gyorgyi, A. H., and Schwarz, D. R., eds., New York, Academic, p. 279.