absorption spectra of synthetic peptides, have presented evidence that peptide bonds could act as couplers between N- and C-terminal chromophore groups in a peptide chain. From this evidence they concluded that peptide bonds might have an effect on side-chain groups.

Acknowledgment

The author wishes to acknowledge the excellent technical assistance of Mrs. Judith Schmidt.

References

Goldfarb, A. R., and Hoffman, E. (1959), *Arch. Biochem. 81*, 493.

Okuyama, T., and Satake, K. (1960), *J. Biochem. (Tokyo) 47*, 454.

Satake, K., Okuyama, T., Ohashi, M., and Shinoda, T. (1960), *J. Biochem. (Tokyo) 47*, 654.

Heterogeneity of Amino Groups in Proteins. I. Human Serum Albumin*

A. R. Goldfarb

ABSTRACT: The reaction of human serum albumin with trinitrobenzenesulfonic acid has been studied in the pH range 6.98–8.40 at 16°. The reaction does not follow a simple second- or pseudo-first-order kinetics. A method is described for analyzing the data on the assumption that several classes of reacting amino groups exist, each having a specific velocity constant. This method of analysis reveals that human serum albumin contains 19 reactive ϵ-amino groups divided into three sets containing 14, 3.7, and 1.5 amino groups, respectively.

The structure of proteins depends, to a great extent, on the covalent bonds, hydrogen bonds, ionic bonds, apolar bonds, and van der Waals interactions. All of these are variously affected by the constituent amino acids and their associated functional groups. The proximity and interaction of these groups ultimately determine the shape and "fine structure" of the protein as well as the ability of the protein to function as an agent in biological systems (*e.g.*, as enzymes or as ion binding molecules). Conversely, the reactivity of a functional group is altered by its presence in a protein macromolecule. For example, studies of the titration of ionic groups and of the ultraviolet absorption characteristics of proteins (Foster, 1960) have shown that "masked" groups occur; *i.e.*, by various tests the number of groups that could be determined were fewer than the total number which were known to be present from amino acid analyses. "Masking" results either from the side chain being buried in the molecule or from some interaction with proximal groups or some combination of both. The older literature implied that masking was an "all or none" phenomenon. More recent evidence has demonstrated gradations of masking of the functional groups in native proteins which can be influenced by pH, temperature, ionic strength, and ionic composition (Shinoda, 1965; Klee and Richards, 1957; Scatchard *et al.*, 1957).

It should be relatively simple to determine whether a functional group in proteins is composed of subclasses by studying the kinetics of the reaction of the group with a specific reagent. If all of the unmasked groups are equally reactive the kinetic study should confirm this by revealing a simple second-order or pseudo-first-order curve. For example, M. Schlamowitz (unpublished observations, 1966), in an analysis of the data of Li (1945) on the iodination of pepsin, found that the reaction followed a simple second-order equation in which all ten iodinatable groups were equally reactive. On the other hand Klee and Richards (1957) reported that the reaction of ribonuclease with *O*-methylisourea did not give a pseudo-first-order curve and Boyer (1954) made a similar observation concerning the reaction of proteins with sulfhydryl reagents.

It had been reported by Okuyama and Satake (1960) that trinitrobenzenesulfonic acid (TNBS)[1] was a specific reagent for only the amino group of amino acids and peptides. Since TNBS is water soluble and reacts rapidly with amino groups under mild conditions it was chosen as a suitable reagent to investigate the reaction heterogeneity of the amino groups in proteins.

* From the Biochemical Research Laboratories of the Division of Research, Sinai Hospital, Detroit, Michigan. *Received February 4, 1966; revised June 13, 1966.* Supported in part by Grant AI-085386 from the National Institutes of Health.

[1] Abbreviations: TNBS, 2,4,6-trinitrobenzenesulfonic acid; HSA, human serum albumin; ϵ, absorbance.

To the best of our knowledge, no method has been reported in the literature for analyzing a curve that resulted from several simultaneous reactions. The present paper describes such a method. By the use of this analysis it is shown that human serum albumin at 16° contains three classes of reactive amino groups and a fourth class which reacts extremely slowly if at all.

Materials and Methods

Crystallized human serum albumin (Pentex, Inc.) was occasionally found to contain a dialyzable impurity which reacted with TNBS. Exhaustive dialysis at 4° removed this impurity and stock solutions of dialyzed HSA were prepared each week. The concentration of HSA was determined from the absorbance at 280 mμ assuming ϵ (0.1%) = 0.531. The molecular weight was assumed to be 69,000.

Trinitrobenzenesulfonic acid (picryl sulfonate, Nutritional Biochemicals) was recrystallized once from 5 M HCl. The reaction was carried out in 0.1 M sodium phosphate buffer.

The reaction was run at 16 ± 0.1° in a stoppered 10-mm silica cell in a Bausch and Lomb 505 spectrophotometer. The absorbance at 345 mμ was continuously recorded on a Sargent Model SR recorder linked to the output of the spectrophotometer.

Calculations

The following is the derivation of the equations, and the method, which were used to calculate the velocity constant and the number of amino groups that had reacted for each constant. The general equation for a second-order reaction is given in eq 1

$$kt = \frac{1}{[T] - n_i[P]} \ln \left(\frac{[T] - x_i}{n_i[P] - x_i} \frac{n_i[P]}{[T]} \right) \quad (1)$$

where [T] = molar concentration of TNBS; [P] = molar concentration of protein; n_i = number of amino groups of the ith kind per mole of protein; x_i = molar concentration of ith kind of amino groups reacted at time t; k_i = velocity constant for ith subclass; and t = time.

When the reaction is run in a large excess of TNBS, $[T] \gg n_i[P] > x_i$ and eq 1 reduces to eq 2

$$k_i t = \frac{1}{[T]} \ln \frac{n_i[P]}{n_i[P] - x_i} = -\frac{1}{[T]} \ln \left(1 - \frac{x_i}{n_i[P]} \right) \quad (2)$$

Letting ϵ_i be the molar absorbance for the ith subclass of trinitrophenylamino groups, it can be shown that

$$-k_i t = \frac{1}{[T]} \ln \left(1 - \frac{a_i}{A_i} \right) \quad (3)$$

where $a_i = x_i \epsilon_i$ = absorbance due to the ith subclass at time t, $A_i = n_i[P]\epsilon_i$ = absorbance at 100% reaction due to the ith subclass.

If, for this discussion, three simultaneous reactions are assumed for which the maximum absorbances (100% reaction) are $A_1 = n_1[P]\epsilon_1$, $A_2 = n_2[P]\epsilon_2$, and $A_3 = n_3[P]\epsilon_3$, then the total absorbance at infinite time is given by $A_L = A_1 + A_2 + A_3$. At values of t when reactions 1 and 2 are essentially complete their contributions to the absorbance is $A_1 + A_2$ and $a_t = a_3 + A_2 + A_1$, where a_t is the observed absorbance at time t.

In this region, substituting the value of a_3 into eq 3

$$-k_3 t = \frac{1}{[T]} \ln \left((1 - \frac{a_t - A_1 - A_2}{A_3} \right) = \frac{1}{[T]} \ln \left(\frac{A_3 + A_2 + A_1 - a_t}{A_3} \right) = \frac{1}{[T]} \ln (A_L - a_t) - \frac{1}{[T]} \ln A_3 \quad (4)$$

Since three constant values are unknown, (k_3, A_L, and A_3) the solution of three simultaneous equations gives the values for these unknowns. This calculation is simplified as follows.

If in eq 4 readings are taken at times t_α, t_β, and t_γ such that $t_\gamma - t_\beta = t_\beta - t_\alpha$ and k_3 is constant in this region, then

$$-k_3(t_\gamma - t_\beta) = \frac{1}{[T]} \ln \frac{A_L - a_\gamma}{A_L - a_\beta} \quad (5a)$$

$$-k(t_\beta - t_\alpha) = \frac{1}{[T]} \ln \frac{A_L - a_\beta}{A_L - a_\alpha} \quad (5b)$$

since $t_\gamma - t_\beta = t_\beta - t_\alpha$, it can be shown that

$$\frac{A_L - a_\gamma}{A_L - a_\beta} = \frac{A_L - a_\beta}{A_L - a_\alpha} \quad (6)$$

$$A_L = \frac{a_\alpha a_\gamma - a^2_\beta}{(a_\alpha + a_\gamma) - 2a_\beta} \quad (7)$$

From eq 5a or 5b, k_3 can be determined, and then A_3 from eq 4 using the values of A_L and k_3. Since $A_3 = n_3[P]\epsilon_3$, n_3 is calculated from the known values of ϵ_3 and [P]. In practice, the values of A_L, A_3, and k_3 were calculated for many sets of points, starting at the largest t, and these results were averaged.

Several points of caution should be noted. First, if the interval $t_\gamma - t_\beta$ is too small, the value $a_\alpha + a_\gamma - 2a_\beta$ in eq 7 becomes very small so that small errors in any absorbance may cause large errors in calculating A_L. It is evident, therefore, that as large intervals of $t_\beta - t_\alpha$ as possible should be taken. Second, errors in k_3 will arise from errors in calculating A_L. And third, this is a method of approximations.

Substituting the values of A_3 and k_3, obtained as above, in the equation $a_3 = A_3(1 - e^{-[k_3 t]}$ gives the

TABLE I: Experimental Data.

	Calcd Constants							Blanks	
pH	n_t[a]	k_3	n_3[a]	k_2	n_2[a]	k_1	n_1[a]	Exptl	Calcd
6.96	18.6	0.96	14.9	13.39	3.6	76.3	0.8	0.020	0.024
7.00	18.4	1.18	13.9	15.76	4.1	108.4	1.4	0.040	0.057
7.00	18.9	1.17	14.0	11.64	3.6	43.1	1.40	0.011	0.017
8.00	19.6	3.57	14.0	29.7	3.4	104.9	1.3	0.011	0.016
8.03	20.3	5.60	13.8	53.3	4.7	335	1.7	0.050	0.060
8.35	19.1	9.70	14.2	75.2	2.5	264	1.9	0.095	0.127
8.40	19.4	6.70	13.0	51.1	3.9	271	1.8	0.020	0.047
Mean	19.2 ±0.23[b]		14.0 ±0.30[b]		3.7 ±0.22[b]		1.5 ±0.133[b]		

[a] The values of n are calculated on the basis of $\epsilon = 1.15 \times 10^4$ (Goldfarb, 1966). [b] Standard error.

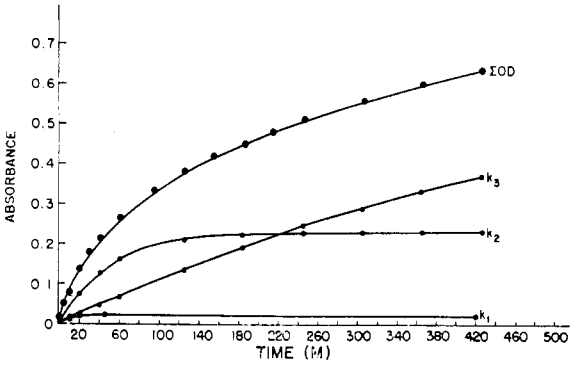

FIGURE 1: Kinetics for the reaction between TNBS (1254×10^{-6} M) and HSA (4.92×10^{-6} M), 0.1 M phosphate buffer, pH 7.0, at 16°. The solid lines are the calculated curves for $k_1 = 108.4$, $k_2 = 11.6$, and $k_3 = 1.2$ l. mole^{-1} min^{-1}. The points correspond to the experimental observations.

values of a_3 for various times due to k_3. When a_3 is subtracted from a_t, the values of $a_1 + a_2$ result. The residual absorbances, $a_1 + a_2$, are treated as above to give A_2 and k_2 and this process is repeated to give A_1 and k_1.

Results

The results of a representative run are given in Figure 1. The data could not be fitted to a simple second- or pseudo-first-order curve. When analyzed as described above, it was found possible to fit the data to a curve which was the sum of three simultaneous reactions with velocity constants $k_1 = 108$, $k_2 = 15.8$, and $k_3 = 1.18$ l. mole^{-1} min^{-1}. The theoretical curves were calculated for the individual k values, and the curve marked ΣOD was the sum of these curves. It is seen that the experimental points were all slightly above the theoretical curve. When the ΣOD values were subtracted from the experimental absorbance at any given time, constant differences were obtained which were fairly close to a blank that occurred at zero time (about 15 sec). This blank was found in all of the runs (Table I). The absolute values of the blanks increased with increasing concentrations of the protein in different runs.

Experimental results are summarized in Table I. In general, there appeared to be no change in the total number of amino groups reacted ($n_t = 19.2$/mole of HSA) within the pH range studied. Nor did there appear to be any more than random variations in the number that were distributed in the subsets ($n_1 = 1.5$, $n_2 = 3.7$, $n_3 = 14.0$). The summation of the values of n_1, n_2, and n_3 agreed with n_t within about 5%.

It had been shown (Goldfarb, 1966) for model compounds containing α- and ϵ-amino groups that the velocity constants were dependent on pH according to the general expression $\log k = a\text{pH} + b$. Therefore, the velocity constants for HSA were similarly plotted with the results shown in Figure 2. The equations of the lines in Figure 2 were calculated by the method of least squares. In order to evaluate the over-all discrepancies in this method, the k values were calculated from the least-square equations. The ratios k(exptl)/k(calcd), where k(calcd) was obtained from the least-square line, were determined, as well as their standard errors. The values of the ratios and standard errors of the ratios were 1.13 ± 0.17, 1.024 ± 0.078, and 1.020 ± 0.064, respectively, for k_1, k_2, and k_3. The average ratios for k_2 and k_3 were close to 1.0 but the ratio for k_1 indicated that a larger discrepancy existed in determining k_1. This was further confirmed by the relative values for the standard errors.

Discussion

The reproducibility of the results deserves some dis-

cussion. When duplicate runs were made with the same concentrations of reactants, identical curves were obtained. The duplicate experiments, the results of which are reported in Table I, differed from each other in respect to the concentrations of HSA and TNBS. The agreement between the values of n_i were satisfactory and showed relatively small random variations over the pH range investigated. On the other hand, the duplicate values of k_i were not so satisfactory. As can be seen in Figure 2, the deviation of the points from the least-square line increased in going from k_3 to k_2 to k_1. This is not surprising since the method has a "built in" cumulative error in the sense that any errors in the evaluation of A_L, k_3, and A_3 will accumulate into A_2, k_2, etc. It was concluded that the values of k_3 and k_2 were fairly reliable but the value for k_1 was of limited reliability.

It has been shown that the molar absorbance, corrected for the presence of 1 mole of sulfite, for model TNP-α-amino groups of peptides is about 1×10^4 and for the TNP-ϵ-amino groups of lysine is 1.15×10^4 (Goldfarb, 1966) and these figures were assumed in calculating n_3, n_2, and n_1. On this basis HSA was shown to contain 19.2 reactive amino groups of the total of 57.6 lysine ϵ-amino groups and 1 terminal α-amino group (Foster, 1960). These reactive groups consisted of three subsets, containing 14.0, 3.7, and 1.5 amino groups, each of which had a different velocity constant. These results are comparable, in two respects, with those obtained in the studies by Scatchard *et al.* (1957), Karush (1951), and Goodman (1958) on the anion binding properties of serum albumins. In the first respect all three authors reported that it was necessary to assume three subsets of binding sites in order to account for their data and this is in agreement with the finding of three subsets of amino groups in this report. In the second instance we have found that 14 reacting groups corresponded to the smallest velocity constant, 3.7 to an intermediate constant, and 1.5 groups to the largest velocity constant. This is parallel to the findings of Karush, Scatchard, and Goodman who found 15 to 20 sites that had the weakest affinity, four to eight sites of intermediate affinity, and one to two sites of greatest affinity, these being for a diverse set of reagents, *e.g.*, Cl^-, dodecyl sulfate, and fatty acid anions. A relation between binding affinity and velocity constant has been described by Branch and Calvin (1941) who showed that if an association is strong, a reaction is fast. Therefore, it would appear to be reasonable to assume a reaction mechanism which, at least partially, involves complex compound formation between TNBS and HSA. As will be described below, such a hypothesis will account qualitatively for the results.

It had been pointed out (Goldfarb, 1966), from a study of the pH dependence of the reaction of TNBS with glycine and α-acetyllysine, that k was proportional to $1/[H^+]$, from which it was inferred that TNBS reacted with the free amino group. On the other hand k was proportional to $1/[H^+]^a$ for di- and triglycine, where $a \equiv 0.5$, and this would not account for a

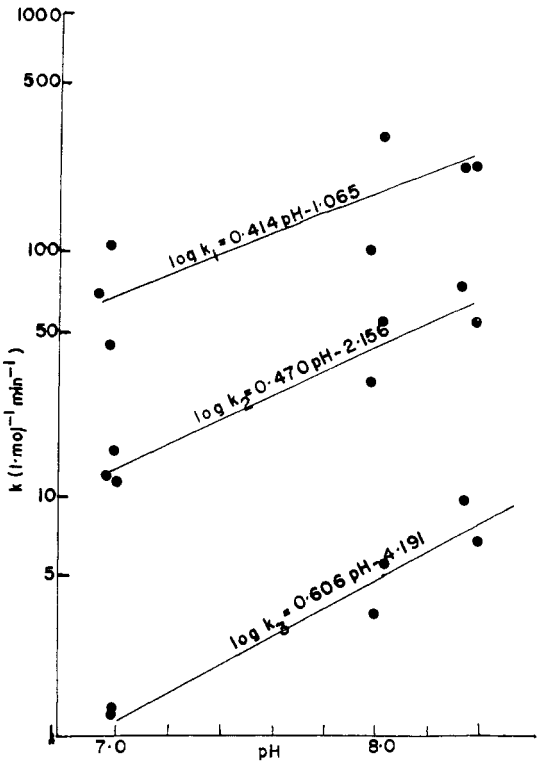

FIGURE 2: Graph of log k_1, log k_2, and log k_3 *vs.* pH.

similar simple reaction mechanism. Since most of the amino groups of HSA are the ϵ-amino groups of the lysine residues they should have had values of "a" near 1.0. However, in this report it has been shown that k_1, k_2, and k_3 for HSA are all proportional to $1/[H^+]^a$, where a has different values near 0.5, and the same problem of mechanism arises as with di- and triglycine. If, as suggested in the last paragraph, it is assumed that the rate-determining step depends on the concentration of a complex between TNBS and protein (T–P) and that the association constant is pH dependent, then in the reaction TNBS + P $\rightleftarrows$ T–P the association may be given as in eq 8

$$K_{\text{assocn}} = \frac{[\text{T–P}]}{[\text{T}][\text{P}]} f[\text{H}] \qquad (8)$$

where $f[\text{H}]$ is some function of the concentration of $[H^+]$. It follows then, that the rate equation takes the form

$$\frac{d[\text{TS}]}{dt} = -\frac{d[\text{T–P}]}{dt} = k[\text{T–P}] = kK_{\text{assocn}}[\text{T}][\text{P}]/f[\text{H}] \qquad (9)$$

and this would, at least qualitively, account for the presence of the term $1/[H^+]^a$ in the kinetic expression.

It is also desirable to account for the large differences between k_1, k_2, and k_3 at any given pH, as well as the fact that these are up to several hundred times greater than the velocity constant for the model compound,

α-acetyllysine. Differences in pK_a for each of the subsets of amino groups can be eliminated as a contributing factor since it would require unreasonably low values of pK_a (8–9) for such ϵ-amino groups. Conformational considerations could play a role in determining the difference between k_1, k_2, and k_3; *e.g.*, it is conceivable that one to two groups are on the outer surface of the protein,[2] three to four groups are near the surface, and 14 groups are "deeply" buried but approachable, while the remaining nonreactive groups are completely masked. Other environmental differences, *e.g.*, the presence of environmental charged groups and hydrophobic groups, could act cooperatively to make for complexes between TNBS anion and the reacting amino groups of varying stability and thus account for the differences between k_1, k_2, and k_3. It appears probable that both conformational and environmental factors affect the reactivity of the amino groups in proteins.

[2] The placement of the groups as "on the outer surface" or "buried" is given only in the operational sense that the spatial configuration allows for penetration of the reagent to the reacting group. It is quite possible that different sized reagents may give quantitatively different results.

Acknowledgment

The author wishes to acknowledge the excellent technical assistance of Mrs. Judith Schmidt.

References

Boyer, P. D. (1954), *J. Am. Chem. Soc. 76*, 433.

Branch, G. E., and Calvin, M. (1941), "The Theory of Organic Chemistry," New York, N. Y., Prentice Hall, p 385.

Foster, J. F. (1960), "The Plasma Proteins," Vol. 1, Putnam, F. W., Ed., New York, N. Y., Academic, pp 179–239.

Goldfarb, A. R. (1966), *Biochemistry 5*, 2570 (this issue; preceding paper).

Goodman, D. S. (1958), *J. Am. Chem. Soc. 80*, 3892.

Karush, F. (1951), *J. Am. Chem. Soc. 73*, 1246.

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

Li, C. H. (1945), *J. Am. Chem. Soc., 67*, 1065.

Okuyama, T., and Satake, T. (1960), *J. Biochem. (Tokyo) 47*, 454.

Satake, K., Okuyama, T., Ohashi, M. and Shinoda, T. (1960), *J. Biochem. Tokyo 47*, 654.

Scatchard, G., Coleman, J., and Shen, A. (1957), *J. Am. Chem. Soc. 79*, 12.

Shinoda, T. (1965), *Biochim. Biophys. Acta, 97*, 382.