resulting in stabilization. In the absence of inhibitors, the free energy loss corresponding to formation of local bonds far outweighs the electrostatic work and the polymerization is unlimited. In the system studied here, the hexamethylene glycol interferes with the formation of local bonds,[29] so that the local and electrostatic contributions are of the same order of magnitude, and stabilization occurs at a rather low value of $n$. The degree of distribution about the mean value of $n$ would be determined by the sharpness of the minimum in the free energy of association.

Although the value of $n$ will no doubt depend on the composition of the solvent, and under normal physiological conditions there may be no stabilization of this sort at all, nevertheless the fundamental geometry of polymerization is probably unaffected by the artificial conditions under which these experiments are carried out; so that a detailed study of the geometry and structure of our intermediate polymers can elucidate the mechanism of conversion of fibrinogen to fibrin in nature.

**Acknowledgments.**—We are indebted to Professor J. W. Williams for use of the Svedberg oil turbine ultracentrifuge, and to Miss Carol Braatz for technical assistance.

(29) J. D. Ferry and S. Shulman, THIS JOURNAL, **71**, 3198 (1949).

MADISON, WISCONSIN

---

[CONTRIBUTION FROM THE GATES AND CRELLIN LABORATORIES OF CHEMISTRY,[1] CALIFORNIA INSTITUTE OF TECHNOLOGY]

# A Theory of Antibody–Antigen Reactions. I. Theory for Reactions of Multivalent Antigen with Bivalent and Univalent Antibody[2]

BY RICHARD J. GOLDBERG[3]

RECEIVED JULY 18, 1952

The most probable distribution of species is calculated for a system composed of univalent and bivalent antibody and $f$-valent antigen. This distribution is used as the basis for a theory of antibody–antigen reactions. The critical extent of reaction, which has been described previously by Flory and Stockmayer, is the point at which the system changes from one composed chiefly of small aggregates into one composed chiefly of relatively few exceedingly large aggregates. This point is interpreted to be the point at which precipitation commences in certain antibody–antigen systems. Since the fraction of reacted antibody sites and the fraction of reacted antigen sites cannot individually exceed unity, the ratio of bivalent antibody to $f$-valent antigen in the system must lie between specified limits in order for the system to attain the critical point. These limits are functions of the antigen valence and the fraction of antibody sites belonging to bivalent antibody. This theory, consequently, suggests a mechanism by which inhibition to precipitation is achieved. Limits may also be computed for points other than the critical point. Expressions for various antibody–antigen ratios are derived and are compared with the Heidelberger–Kendall equation, which is easily obtainable from the theory presented here. The aggregation as a function of the extent of reaction is compared to that described for a system containing $f$-valent antigen and univalent antibody only. Contrary to a previous belief, the possibility of large aggregate formation has a predominant effect on the system at small extents of reaction. Previously obtained experimental values for (1) the positions of the inhibition zones, and (2) the antibody–antigen ratios of the precipitates, of four antibody–antigen systems are compared with the values calculated by the theory. Two of the systems contain horse antibody and two contain rabbit antibody. Good agreement is obtained. The theory appears to have withstood a difficult set of tests. The calculations indicate that the apparent lack of antibody-excess inhibition in the systems containing rabbit antibody results chiefly from the relatively small solubility of the rabbit antibody. Results of Rh agglutination tests are in qualitative agreement with the theory in regard to the variation of the positions of inhibition zones with the combining power of the antigen.

## Part A. Introduction

For a long time it has been attractive to consider antibody–antigen reactions as involving the combination of specific sites by which very large aggregates are attained. To require the existence of aggregates of this kind, one must necessarily assume that the antibody and antigen molecules responsible for the size of the aggregate are multivalent with respect to each other. If one is to require further that the antibody–antigen molecular ratio of these very large aggregates be variable and no less than unity, then he must consider antibody molecules to be bivalent and antigen molecules to be greater than bivalent. It should be noted that the existence of univalent antibody molecules in the system is still permitted. They cannot, however, be responsible for the specific growth of the aggregate to a size involving more than one antigen molecule, since wherever they occur they end chains which might otherwise have grown longer than they are.

A general theory presenting the features of a system involving reactions between bivalent antibody molecules and multivalent antigen molecules has not yet been achieved. It is the purpose here to present such a theory, which, it is hoped, will create a greater understanding of antibody–antigen reactions through the interpretation of serological data.

Although theoretical treatments have been developed in the past, they have not been sufficiently general to predict the common characteristics of the precipitin reaction. Variable antibody–antigen ratios of the precipitate, which depend on the preparation of the system, are of fundamental importance to a good theory. Inhibition to precipitation in regions of antigen excess and also antibody excess should be accounted for without relying on artificial assumptions of solubility. One should attempt to describe the relative amounts of precipitate corresponding to the composition of the system. One should be able to explain the relative differences of the system arising from the manner in which composition is varied. The quantitative function of blocking antibody molecules has yet to be described for the common systems.

(1) Contribution number 1668.

(2) This work was supported by a grant from the U. S. Public Health Service.

(3) Department of Chemistry, University of Wisconsin, Madison, Wisconsin.

Boyd argues that there is no need to postulate the existence of bivalent antibody molecules if the concept of univalent antibody molecules satisfies the experimental data.[4] In the $\beta$-procedure the passage of the mass of the precipitate through a maximum clearly indicates inhibition. Boyd supposes that the addition of the first few antibody molecules to the antigen molecule causes the polar groups of both kinds to be covered up, thus making the aggregate less and less soluble. Beyond the point corresponding to most favorable precipitation he argues that addition of more antibody molecules to the remaining sites on the antigen molecule causes an increase in solubility as a result of the increase of polar groups on the aggregate. He, therefore, concludes that univalent antibody molecules cause precipitation to go though a maximum. It seems that the polar groups on the last antibody molecule to go onto the antigen molecule would be covered up for the same reason, at least, as the ones on the first antibody molecule were; and, therefore, his argument is unsound. Haurowitz, on the other hand, believes that antibody molecules produced in the horse are probably multivalent because of the solubility of the antibody–antigen precipitate in excess antibody.[5] Since antibody molecules produced in the rabbit do not appear to have this kind of action, Haurowitz feels that there is a fundamental immunological difference between these two kinds of antibody molecules. The theory presented here does not require such a difference.

## Part B. A Theory for Reactions of Multivalent Antigen Molecules with Bivalent and Univalent Antibody Molecules

The theory presented here uses the most probable distribution as a basis for determining the features of antibody–antigen reactions. Flory was the first to employ the statistical method for finding the distribution of species in polymer systems.[6] Using this method he studied molecular size distributions of branched-chain polymers.[7-9] Stockmayer, shortly thereafter, generalized Flory's method using the same basic assumptions. He obtained the most probable distribution of molecular sizes for certain types of branched-chain polymers.[10] It is this method of Stockmayer which is used in the following presentation. Consequently, the two assumptions used by Flory and Stockmayer also characterize this work. In this theory it is assumed that intra-aggregate reactions yielding cyclical structures cannot occur. This simplifying assumption fixes the number of bonds in an aggregate of a given size. This number of bonds is one less than the total number of antibody and antigen molecules of which the aggregate is composed. It is next assumed that any unreacted site is as reactive as any other site regardless of the size or shape of the aggregate to which it is attached. The use of these assumptions permits us to carry through the mathematical operations indicated in the treatment presented here. Without these assumptions the mathematical apparatus becomes unwieldy. The former assumption is not strictly correct for the general case. For without the occurrence of intra-aggregate reactions, steric hindrances would be too great to construct a large aggregate while maintaining equal availability of all reactive sites on the aggregate, as is demanded by the latter assumption. Only a small percentage of the reactions have to be intra-aggregate reactions to overcome this difficulty. Consequently, the error in the results need not be large. As one approaches the antigen or antibody excess region this error should continue to diminish, since the aggregates become smaller and smaller. The second assumption appears reasonable in view of the fact that the reactive sites are probably sufficiently far apart so as not to influence one another. Therefore, all bond energies are required to be equal, and only entropy effects control the mode of combination.

An antibody site is the reactive area on an antibody molecule which permits combination of the latter with an antigen molecule at one of its reactive areas. The antigen site is defined in an analogous way. An aggregate is defined as a group of antibody and antigen molecules, any two of which are connected by only one chain consisting of alternating antibody and antigen molecules bound to each other by their respective reaction sites, provided these two molecules are not bound together by their reaction sites (see Appendix). Therefore, if a bond in a single aggregate is broken the antibody and antigen molecules on either side of the bond are in no way connected to those on the other side. Two aggregates exist. An antibody–antigen reaction involves the combination of one antibody site with one antigen site in the formation of a bond. An aggregate consisting of two molecules must be composed of one antibody molecule and one antigen molecule with one bond between them. Furthermore, there can be only one bond holding any antibody molecule to an antigen molecule.

The following terminology will be used throughout the discussion.

$G$ = number of antigen molecules in the system

$A$ = number of antibody molecules in the system with two reactive sites (bivalent antibody)

$D$ = number of antibody molecules in the system with one reactive site (univalent or blocking antibody)

$M$ = number of aggregates in the system plus the number of free antibody and antigen molecules

$f$ = number of effective reaction sites on each antigen molecule ($f$-valent antigen)

$m_{ijk}$ = number of aggregates each of which is composed of $i$ bivalent antibody molecules, $j$ univalent antibody molecules and $k$ antigen molecules

$W_{ijk}$ = number of ways to construct a single $i,j,k$-aggregate containing no cyclic structures from $i$-given bivalent antibody molecules, $j$-given univalent antibody molecules, and $k$-given antigen molecules

$q$ = number of free antibody sites on an aggregate

$p$ = fraction of antigen sites in the system which have reacted; it is also called the extent of reaction

$\rho$ = fraction of antibody sites in the system which belong to bivalent antibody molecules

$r$ = $fG/2A$

$s$ = $fG/D$

$M_G$ = molecular weight of the antigen

$M_A$ = molecular weight of the bivalent antibody

$M_D$ = molecular weight of the univalent antibody

For a particular fraction of antigen sites reacted there are many distributions of species available from which to choose. We choose the one which

(4) W. Boyd, *J. Exp. Med.*, **74**, 369 (1941).

(5) F. Haurowitz, "Chemistry and Biology of Proteins," Academic Press, Inc., New York, N. Y., 1950.

(6) P. J. Flory, THIS JOURNAL, **58**, 1877 (1936).

(7) P. J. Flory, *ibid.*, **63**, 3083 (1941).

(8) P. J. Flory, *ibid.*, **63**, 3091 (1941).

(9) P. J. Flory, *ibid.*, **63**, 3096 (1941).

(10) W. H. Stockmayer, *J. Chem. Phys.*, **11**, 45 (1943).

corresponds to a maximum entropy. Equations 1 to 18 describe the mathematical operations involved. This distribution has a very high probability of occurrence, while all others (perceptibly different ones) have exceedingly small probabilities of occurrence. We state, therefore, that the system follows the most probable path, or path of maximum entropy to equilibrium. Under these circumstances, *all reactions are reversible. The theory does not require bonds to be fixed once they are formed.*

The total number of ways to form the number of aggregates $m_{ijk}$, for all appropriate $i$, $j$ and $k$ values out of the $A$, $D$ and $G$ molecules is

$$\Omega = G!A!D! \prod_{i,j,k} \left[ \left( \frac{W_{i,j,k}}{i!j!k!} \right)^{m_{ijk}} \frac{1}{m_{ijk}!} \right] \qquad (1)$$

In order to find the most probable distribution, that is, the set of the numbers $m_{ijk}$, corresponding to the maximum value of $\Omega$, one must set the derivative of $\Omega$ with respect to the variables $m_{ijk}$, equal to zero for constant $A$, $D$, $G$ and $M$ which are expressed by

$$\sum_{i,j,k} i m_{ijk} = A; \quad \sum_{i,j,k} j m_{ijk} = D, \quad \sum_{i,j,k} k m_{ijk} = G \qquad (2)$$

and

$$\sum_{i,j,k} m_{ijk} = M \qquad (3)$$

All values of $k$ up to and including the number of antigen molecules in the system are included in the sums and products of equations 1, 2 and 3. The values of $i$ and $j$ permitted for a given $k$ can be found as follows. The number of bivalent antibody molecules in an aggregate containing $k$ antigen molecules is the number required to hold the latter together, $k - 1$, plus the number of the former of which only one site is used, $q$. Furthermore, the number of bivalent antibody molecules of which only one site is used, $q$, plus the number of univalent antibody molecules, $j$, cannot exceed the number of antigen sites available to them, $fk - 2k + 2$. Hence, the numbers of bivalent and univalent antibody molecules which may be used to form an aggregate containing $k$ antigen molecules are given by

$$i = k - 1 + q \qquad (4)$$
$$0 \leqslant q + j \leqslant fk - 2k + 2$$

In the sums and products which follow, these relations give the limit unless otherwise specified.

The condition given by equation 3 implies a constant reacted fraction of antigen sites $p$. This can be shown by expressing $p$ in terms of $m_{ijk}$ as

$$p = \frac{1}{fG} \sum_{i,j,k} [2(k - 1) + q + j] m_{ijk} = \frac{A + D + G - M}{fG} \qquad (5)$$

since the number of reacted antigen sites in an $i$, $j$, $k$-aggregate is $2(k - 1) + q + j$.

The differentiation of equation 1 is performed with the help of Stirling's approximation yielding

$$\sum_{i,j,k} \left( \log \frac{W_{ijk}}{i!j!k!} - \log m_{ijk} \right) \mathrm{d}m_{ijk} = 0 \qquad (6)$$

On account of the restrictions imposed by equations 2 and 3, the number of independent variables $m_{ijk}$, is reduced by four. Hence, four of the differentials, $\mathrm{d}m_{ijk}$, are functions of the remaining ones and can be eliminated with the use of equations 2 and 3 in differential form. This can be accomplished by adding the following equations to equation 6 and choosing the four constants $\zeta$, $\eta$, $\xi$ and $B$, known as Lagrangean undetermined multipliers, so that the coefficients of four of the differentials vanish.[11]

(11) L. Page, "Introduction to Theoretical Physics," 2nd Ed., D. Van Nostrand Co., Inc., New York, N. Y., 1935.

$$\begin{aligned} \log \zeta \sum_{i,j,k} i \mathrm{d}m_{ijk} &= 0 \\ \log \eta \sum_{i,j,k} j \mathrm{d}m_{ijk} &= 0 \\ \log \xi \sum_{i,j,k} k \mathrm{d}m_{ijk} &= 0 \\ \log B \sum_{i,j,k} \mathrm{d}m_{ijk} &= 0 \end{aligned} \qquad (7)$$

Since all of the remaining differentials $\mathrm{d}m_{ijk}$ are independent, their coefficients can be made to vanish separately. Therefore, the most probable distribution becomes

$$m_{ijk} = \frac{W_{ijk}}{i!j!k!} \zeta^i \eta^j \xi^k B \qquad (8)$$

The constants $\zeta$, $\eta$, $\xi$ and $B$ can be evaluated from the four simultaneous equations given by equations 2, 3 and 8. To sum these expressions in equations 2 and 3 $W_{ijk}$ is needed. It is shown in the Appendix that

$$W_{ijk} = f^k 2^i \frac{(fk - k)!}{(fk - 2k + 2 - q - j)!} \frac{i!}{q!} \qquad (9)$$

If the running index $i$ is replaced by $q$ according to equation 4 the summing can be accomplished in the following manner, equation 3 being used as a typical example.

$$M = \frac{B}{2\zeta} \sum_{k=0}^{\infty}{}' \frac{(f\xi 2\zeta)^k (fk - k)!}{(fk - 2k + 2)!k!} \times \sum_{q=0}^{fk-2k+2} \frac{(2\zeta)^q (fk - 2k + 2)!}{(fk - 2k + 2 - q)!q!} \times \sum_{j=0}^{fk-2k+2-q} \frac{\eta^j (fk - 2k + 2 - q)!}{(fk - 2k + 2 - q - j)!j!} \qquad (10)$$

Extension of the maximum value of $k$ to infinity for the purpose of summation involves negligible error. The above limits for the $q$ and $j$ sums are obtained after interchanging summation procedures. After evaluating the term for $k$ zero, which yields $\zeta + \eta$, the sums over $j$ and $q$ are accomplished with the use of the binomial theorem.

$$\sum_{q=0}^{fk-2k+2} \frac{(2\zeta)^q (fk - 2k + 2)!}{(fk - 2k + 2 - q)!q!} \times \sum_{j=0}^{fk-2k+2-q} \frac{\eta^j (fk - 2k + 2 - q)!}{(fk - 2k + 2 - q - j)!j!}$$
$$= (1 + \eta)^{fk-2k+2} \sum_{q=0}^{fk-2k+2} \left( \frac{2\zeta}{1 + \eta} \right)^q \frac{(fk - 2k + 2)!}{(fk - 2k + 2 - q)!q!}$$
$$= (1 + \eta + 2\zeta)^{fk-2k+2} \qquad (11)$$

Equation 10 then becomes

$$M/B = \zeta + \eta + \frac{(1 + \eta + 2\zeta)^2}{2\zeta} \sum_{k=1}^{\infty} \frac{y^k (fk - k)!}{(fk - 2k + 2)!k!} \qquad (12)$$
$$y = f\xi 2\zeta (1 + \eta + 2\zeta)^{f-2}$$

The sum over $k$ in equation 12 as well as the corresponding ones in equation 2 can be expressed in the following way.

$$S_i \equiv \sum_{k=1}^{\infty} k^i y^k \frac{(fk - k)!}{(fk - 2k + 2)!k!}; \quad i = 0, 1 \qquad (13)$$

Stockmayer has summed this expression for $i$ zero, one and two.[10] He obtained the results

$$\begin{aligned} S_0 &\equiv \frac{\alpha(1 - \alpha f/2)}{(1 - \alpha)^2 f} \\ S_1 &\equiv \frac{\alpha}{(1 - \alpha)^2 f} \\ S_2 &\equiv \frac{\alpha(1 + \alpha)}{f(1 - \alpha)^2 [1 - \alpha(f - 1)]} \\ y &= \alpha(1 - \alpha)^{f-2} \end{aligned} \qquad (14)$$

Consequently, equations 2 and 3 yield

$$M/B = \zeta + \eta + \frac{(1+\eta+2\zeta)^2}{2\zeta}\frac{\alpha(1-\alpha f/2)}{(1-\alpha)^2 f}$$

$$G/B = \frac{(1+\eta+2\zeta)^2}{2\zeta}\frac{\alpha}{(1-\alpha)^2 f} \qquad (15)$$

$$D/B = \eta + \eta\frac{1+\eta+2\zeta}{2\zeta}\frac{\alpha(f-2)}{(1-\alpha)^2 f} + \eta\frac{(1+\eta+2\zeta)}{\zeta}\frac{\alpha(1-\alpha f/2)}{(1-\alpha)^2 f}$$

$$A/B = \zeta + \frac{(1+\eta+2\zeta)}{2\zeta}[1+\eta+2\zeta(f-1)]\frac{\alpha}{(1-\alpha)^2 f} - (1+\eta-2\zeta)\frac{1+\eta+2\zeta}{2\zeta}\frac{\alpha(1-\alpha f/2)}{(1-\alpha)^2 f}$$

The Lagrangean undetermined multipliers are found rather tediously to be

$$\zeta = \frac{\rho p(1-pfr)}{2(1-p)}$$

$$\eta = (1-p)\frac{p}{1-p}$$

$$\xi = \frac{\alpha(1-p)^{f-1}}{f(\rho p-\alpha)} \qquad (16)$$

$$B = \frac{fG(1-p)(1-\rho pr)}{\rho pr}$$

in which

$$\alpha = \rho^2 p^2 r \qquad (17)$$

$\alpha$ is the probability that an antigen site has reacted with a bivalent antibody molecule, the other site of which has also reacted.

With the use of equations 16 and 17, the distribution given in equation 8 becomes

$$m_{ijk} = fG\frac{(jk-k)!}{(fk-2k+2-q-j)!k!q!j!}r^{k-1}\rho^{k+i-1}\times p^{k+i-j-1}(1-p)^{fk-k-i-j+1}(1-\rho pr)^{i-k+1}(1-\rho)$$
$$i = k-1+q \qquad (18)$$
$$0 \leq q+j \leq fk-2k+2$$

Therefore, the number of every kind of aggregate in the system including the free antibody and antigen molecules can be determined if the composition of the system, valence of the antigen, and extent of reaction are known. The distribution reduces to one for a system consisting of bivalent antibody molecules and $f$-valent antigen molecules alone, if $\rho$ and $j$ are given the values unity and zero, respectively.

The terms in the sum of equation 13 have important properties. These properties will be explained on the basis of their physical implications. From equation 14, one finds that $y$ has a maximum value given by

$$\alpha_c = \frac{1}{f-1}$$
$$y_c = \frac{(f-2)^{f-2}}{(f-1)^{f-1}} \qquad (19)$$

The point at which this maximum occurs will hereafter be designated the critical point and indicated with the subscript c. The most probable distribution $m_{ijk}$, was obtained for a fixed extent of reaction, or a fixed value of $y$. Once $m_{ijk}$ was obtained, however, $y$ could take on many values each corresponding to the system for a definite value of $p$. Therefore, as the antibody–antigen reactions proceed, $p$ becomes progressively larger. The value of $y$ increases in a corresponding fashion up to the critical point where it passes through a maximum. It is convenient to evaluate the total number of aggregates containing $k$ antigen molecules, $m_k$, in order to understand the nature of the system at the critical point. This can be done by summing the distribution $m_{ijk}$, over all allowed values of $i$ and $j$.

$$m_k = \sum_{i,j} m_{ijk} = fG\frac{(fk-k)!}{(fk-2k+2)!k!}\times \alpha^{k-1}(1-\alpha)^{fk-2k+2} \qquad (20)$$

Then the rate of change of $m_k$ with respect to $\alpha$ is found to be

$$\left(\frac{\partial m_k}{\partial \alpha}\right)_k = fG\frac{(fk-k)!}{(fk-2k+2)!k!}\alpha^{k-2}(1-\alpha)^{fk-2k+1}\times [k[1-\alpha(f-1)]-(1+\alpha)] \qquad (21)$$

Equation 21 shows that the number of $k$-aggregates for $k$ unity decreases from the very start of the reaction. The numbers of aggregates for all other values of $k$ increase for sufficiently small values of $\alpha$. As the reactions proceed, that is, for a somewhat larger value for $\alpha$, $m_2$ begins to decrease, later $m_3$ begins to decrease, and so on. In other words, the aggregates continue to build up into larger aggregates as the reactions proceed. Just preceding the critical point all $m_k$ except those for the largest $k$ values are decreasing. Finally, at the critical point and beyond, $(\partial m_k/\partial\alpha)_k$ is negative for all values of $k$. This means, of course, that all sizes of aggregates are disappearing at the critical point. In a real system this cannot be true, however, since the very largest of the aggregates must be growing in size. The reason for this difficulty is that the sum over all finite values of $k$ was replaced by the sum extending the $k$ values to infinity. This implies that aggregates can be infinite and for these, $(\partial m_k/\partial\alpha)_k$ would not be negative. So, although the physical picture is clear, there is this difficulty with the model. This can be avoided to some extent by discussing the relative magnitudes of the rates of disappearance of the aggregates. With the use of Stirling's approximation equation 21 becomes

$$\left(\frac{\partial m_k}{\partial\alpha}\right)_k = G\left(\frac{y}{y_c}\right)^k\frac{1}{k^{5/2}}\times\frac{(1-\alpha)}{\alpha^2}\times [k[1-\alpha(f-1)]-(1+\alpha)]\frac{fe^2(f-1)^{1/2}}{[2\pi(f-2)^5]^{1/2}} \qquad (22)$$
$$k >> 1$$

It is obvious that at the critical point the rate of disappearance of the very largest aggregates is negligible compared to the rate of disappearance of relatively small aggregates. The difference beyond the critical point is even greater, since $y$ is a maximum at $y_c$. Therefore, all aggregates are growing into a few exceedingly large ones. The bulk of the system is in these few. Equation 20 yields with the use of Stirling's approximation

$$m_k = G\left(\frac{y}{y_c}\right)^k\frac{1}{k^{5/2}}\times\frac{(1-\alpha)^2}{\alpha}\frac{fe^2(f-1)^{1/2}}{[2\pi(f-2)^5]^{1/2}} \qquad (23)$$
$$k >> 1$$

The changes in the numbers of aggregates which occur in the region of the critical point are relatively little for small aggregates, while they are tremendous for large aggregates. The critical point is, therefore, characterized by the fact that the system at this point is changing from one composed chiefly of small aggregates into one composed of relatively few exceedingly large aggregates.

It should be mentioned that on account of the $k^{5/2}$ in the denominator of equation 23, $S_0$ and $S_1$ of equation 14 can be used at the critical point, but $S_2$ cannot since it becomes infinite at that point, that is, the corresponding series diverges.

The conditions for inhibition of the attainment of the critical point can be obtained from the relations

$$p \leq 1$$
$$\rho pr \leq 1 \qquad (24)$$

These inequalities express the impossibility of having the fraction of antigen sites reacted and the fraction of antibody sites reacted exceed unity. Furthermore, from equation 17, $p_c$ is found to be

$$p_c = \left(\frac{\alpha_c}{r\rho^2}\right)^{1/2} = \frac{1}{\rho}\left[\frac{2}{f(f-1)}\times\frac{A}{G}\right]^{1/2} \qquad (25)$$

The extent of reaction at which the material passes into the form of very large aggregates is dependent on the valence of the antigen and the composition of the system. Equations 19, 24 and 25 yield the interesting result

$$\frac{f}{2(f-1)} \leq \frac{A}{G} \leq \frac{f(f-1)}{2}\rho^2 \qquad (26)$$

for the attainment of $p_c$. If the system is prepared in such a manner that the bivalent antibody–antigen ratio lies outside the limits given by equation 26, then $p_0$ can never be reached. With the same kind of argument limits can be found for any other value of $\alpha$. If the attainment of $p_0$ is required for precipitation to occur, then equation 26 predicts that the regions outside the above limits are the antigen-excess and antibody-excess inhibition zones, in which precipitation does not occur. It further predicts that the beginning of the inhibition zone of antibody excess but not antigen excess is altered by altering the value of $\rho$ in the sense that an increase in the amount of univalent antibody in the system decreases the range of antibody–antigen ratios over which precipitation occurs. The univalent antibody acts as an inhibitor. Equation 26 also predicts that in a system of bivalent antibody and bivalent antigen there is only one antibody–antigen ratio, namely, unity, for which the critical point can be reached. It, therefore, gives theoretical grounds for the interesting experimental fact that it is difficult to obtain precipitation in a system of this kind.[12] Figure 1 illustrates how these limiting antibody–antigen ratios are affected by the valence of the antigen for $\rho$ values of unity and one-half. The differences between the corresponding ordinate values for the upper and lower limits give the range of ratios for which the critical point can be attained. This range increases as the valence of the antigen increases. As the antigen-excess or antibody-excess region is approached the aggregates become smaller and smaller. When they become sufficiently small to be soluble, inhibition exists. Consequently, a system which has a more soluble antibody than another can attain antibody-excess inhibition with less antibody excess than the other. The unsymmetrical combining powers of antibody and antigen, however, make the antigen-excess inhibition zone less sensitive to solubility differences, particularly when these differences stem from the antibody molecules involved. Since the horse antibody used in the past is more soluble than rabbit antibody,[4] this argument can largely explain the antibody-excess inhibition in systems employing horse antibody and the apparent lack of such inhibition in systems employing rabbit antibody. Of course, if a very soluble antigen is used with rabbit antibody, or if the valence of the antigen is decreased, antibody-excess inhibition should be more easily attainable. Pauling, *et al.*, have found such inhibition in a system containing simple bihaptenic antigens.[13]

Another interesting but not surprising feature of these reactions is given by equation 20. The number of aggregates $m_k$, each of which has the same number of antigen molecules $k$, is independent of the amount of antibody in the system at the critical point. This is also true for any other value of $\alpha$. It can be seen there that each system of a set with the same constituents having a different value of $A$ but the same value of $G$ and $\alpha$, has the same number of aggregates $m_k$, for each $k$. The differences lie only in the numbers of antibody molecules in these aggregates occupying positions other than between two antigen molecules. These differences can be determined from the average numbers of bivalent and univalent antibody molecules, $\bar{\imath}_k$ and $\bar{\jmath}_k$, respectively, in a $k$-aggregate. They are

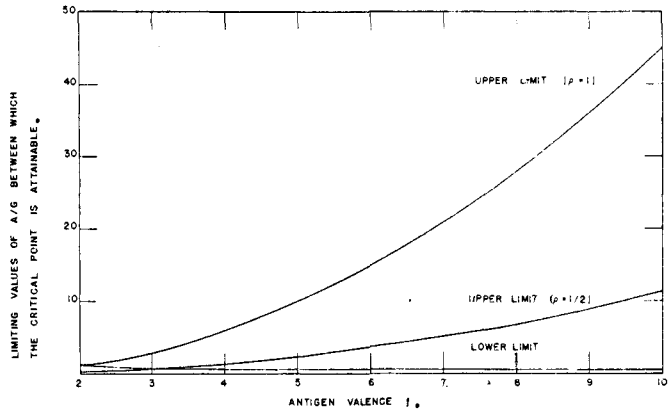


Fig. 1.—The effect of antigen valence on the critical point limits. These limits are defined by (lower limit), $f/2(f-1) \leq A/G \leq f(f-1)\rho^2/2$, (upper limit).

$$\bar{\imath}_k = \frac{\sum_{i,j} i m_{ijk}}{\sum_{i,j} m_{ijk}} = k - 1 + (fk - 2k + 2)\frac{\rho p - \alpha}{1 - \alpha}$$

$$\bar{\jmath}_k = \frac{\sum_{i,j} j m_{ijk}}{\sum_{i,j} m_{ijk}} = (fk - 2k + 2)p\frac{1 - \rho}{1 - \alpha} \qquad (27)$$

As the reactions proceed $\bar{\imath}_k$ and $\bar{\jmath}_k$ both increase. They depend on the extent of the reaction $p$, which is different for each of the systems in the set considered above. The average number of antibody molecules of both kinds found in a $k$-aggregate is the same number as there would be for $\bar{\imath}_k$ alone if there were no univalent antibody present. In other words, the univalent antibody acts like bivalent antibody of which only one site is used. This means that one can determine the correct average total antibody in a $k$-aggregate by using a hypothetical system which contains no univalent antibody. In this hypothetical system, however, the $k$-aggregate may not be the one of interest since it is formed more easily than that in the real system. That is to say, in order to attain the same value of $\alpha$ in a system with $\rho$ unity as in one with $\rho$ less than unity, $p$ need have a correspondingly smaller value in the former than in the latter since $\rho$ and $p$ are inversely proportional to one another.

The average fraction of the free sites on a $k$-aggregate belonging to antibody molecules can be obtained from equation 27. This fraction is independent of $k$, and it is, therefore, the same for all aggregates no matter how many antigen molecules are in them. It can be used to determine the effect of composition on the probability of combination of two aggregates. When this fraction is unity, or when it vanishes, the probability for combination vanishes and equation 26 is deduced.

The average antibody–antigen ratios $\bar{\imath}/k$ and $\bar{\jmath}/k$ of all aggregates containing $k(>> 1)$ antigen molecules are also obtainable from equation 27. These ratios are also independent of $k$. Therefore, in a given system for a particular extent of reaction the average antibody–antigen ratio is the same for all large aggregates. These ratios increase as the extent of the reaction increases and attain their maximum values when $p$ has its maximum value, over most of the range of composition of interest. This value is $f - 1$ for $(\bar{\imath} + \bar{\jmath})/k$ and also for $\bar{\imath}/k$ if $\rho$ is unity.

Equation 4 can also be used to calculate certain ratios.

$$(i/k)_{max} = 1 + \frac{q_{max} - 1}{k} = f - 1 + 1/k$$

$$(j/k)_{max} = f - 2 + \frac{2 - q_{min}}{k} = f - 2 + 2/k \qquad (28)$$

$$q_{max} = fk - 2k + 2;\ q_{min} = 0$$

(12) W. Boyd, *J. Exp. Med.*, **75**, 407 (1942).

(13) L. Pauling, D. Pressman and D. Campbell, THIS JOURNAL, **66**, 330 (1944).

These ratios are exact for all values of $k$ and can be used to determine the valence of the antigen $f$, from experimental data. An equivalence ratio can be defined from equation 4 as

$$(i/k)_e = 1 + (\eta_e - 1)/k = f/2 \qquad (29)$$
$$\eta_e = \eta_{max}/2$$

Therefore, from equation 31

$$(i/k)_{max} = 2(i/k)_e - 1 + 1/k \qquad (30)$$

which for large values of $k$ strongly resembles Pauling's expression which relates the antibody-antigen molecular ratio of a precipitate in the antibody-excess region to the corresponding ratio in the equivalence zone.[14]

At the critical point equations 27 reduce to

$$(\bar{i}/k)_c = \left[\frac{2(f-1)}{f} \times \frac{A}{G}\right]^{1/2}$$
$$(\bar{j}/k)_c = \left(\frac{1}{\rho} - 1\right)\left[\frac{2(f-1)}{f} \times \frac{A}{G}\right]^{1/2} \qquad (31)$$
$$[(\bar{i}+\bar{j})/k]_c = 1/\rho\left[\frac{2(f-1)}{f} \times \frac{A}{G}\right]^{1/2}$$
$$k >> 1$$

These ratios for $i$, $j$, $k$-aggregates are independent of concentration and increase as the bivalent antibody–antigen ratio for the system increases. Variations in the number of univalent antibody molecules present do not affect the top ratio. This appears reasonable since a $k$-aggregate at the critical point would have required a particular number of bivalent antibody molecules to form it.

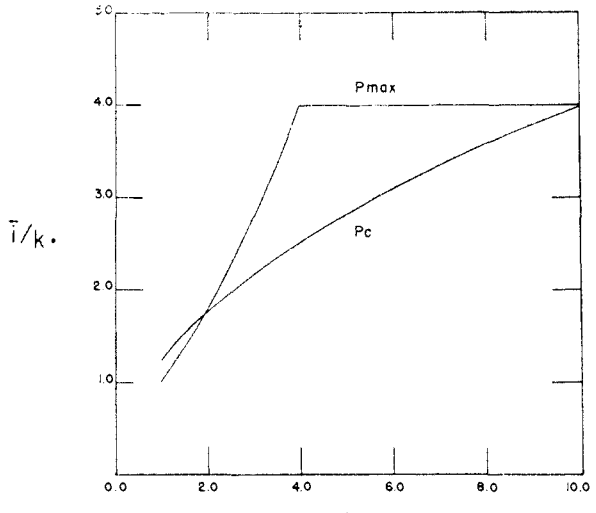


Fig. 2.—The relation between $\bar{i}/k$ and $A/G$ is shown for the critical extent of reaction $p_c$, and the maximum extent of reaction $p_{max}$. $\bar{i}/k$ is the average antibody–antigen ratio of all aggregates containing $k$ antigen molecules. $A/G$ is the antibody–antigen ratio of the entire system. The antibody molecules referred to here are bivalent. A valence of five is assumed for the antigen.

If an expression for the theoretical maximum value of $p$ is substituted into equations 27 the corresponding ratios designated with the subscript max can be obtained. The maximum value of $p$ is, from equation 5, obviously where

$$p_{max} = \frac{A + G + D - M_{min}}{fG} \qquad (32)$$
$$= \frac{1}{f}\left(\frac{1}{\rho} - \frac{1}{2}\right) + \frac{1}{f}\left(1 - \frac{M_{min}}{G}\right)$$

(14) L. Pauling, THIS JOURNAL, **62**, 2643 (1940).

$M_{min}$ is the lowest possible value of $M$ which can be calculated for the system. For example, when $\rho$ is unity

$$M_{min}/G = 0;\ 1 \leqslant A/G \leqslant f - 1$$
$$M_{min}/G = A/G - (f-1);\ A/G \geqslant f - 1 \qquad (33)$$
$$M_{min}/G = 1 - A/G;\ A/G \leqslant 1$$

When univalent antibody is not present in the system, it is found that

$$(\bar{i}/k)_{max} = \frac{(f-3)(A/G)^2 + 2(f-1)(A/G) - (f-1)}{2(f-1)(A/G) - (A/G)^2 - 1};$$
$$1 \leqslant A/G \leqslant f - 1$$
$$(\bar{i}/k)_{max} = f - 1;\ A/G \geqslant f - 1 \qquad (34)$$

The relation between $\bar{i}/k$ and $A/G$ is shown for the critical extent of reaction $p_c$, and the maximum extent of reaction $p_{max}$, in Fig. 2.

The Heidelberger–Kendall equation, which expresses the total amount of antibody combined in terms of the composition of the system and the valence of the antigen, is in the notation here

$$A_b = fG - f^2G^2/4A \qquad (35)$$

where $A_b$ is the total number of bivalent antibody molecules bound in one form or another.[15,16] It is usually written in terms of grams rather than numbers of molecules and $A_b/G$ is assumed to be the antibody–antigen ratio of the precipitate. The distribution $m_{ijk}$ can be used quite simply to obtain their result.

$$A_b = A - m_{(0)} = fG\rho p - \rho^2p^2f^2G^2/4A$$
$$D_b = D - m_{(0)} = fGpD/(2A + D) \qquad (36)$$

If univalent antibody is present, equation 35 cannot be obtained since $p$ cannot have a value greater than unity. If $\rho p$ is taken to be unity, however, the top one of the equations 36 reduces to the Heidelberger–Kendall equation. We might emphasize here that there has been no assumption of consecutive, non-reversible reactions. $p$ can be unity only for the region of extreme antibody excess, namely, where the antibody–antigen ratio of the system is no less than the maximum ratio $f - 1$, of an $i$, $k$-aggregate. This is readily shown from equations 33 and 34. Even in this region of extreme antibody excess the attainment of $p_{max}$ is certainly questionable.

Equations 36 at the critical point reduce to

$$(A_b/G)_c = \left[\frac{2f}{f-1} \times \frac{A}{G}\right]^{1/2} - \frac{f}{2(f-1)}$$
$$(D_b/G)_c = \left[\frac{f}{2(f-1)} \times \frac{D^2}{AG}\right]^{1/2} \qquad (37)$$

The top equation indicates that the total number of bound bivalent antibody molecules is not dependent on the number of univalent antibody molecules present. This should be expected in the light of equation 34. In a linear plot of $(A_b/G)_c$ against $G$, as is generally the case with experimental results in the antibody excess region, equations 37 predict that the negative slope of the curve should increase as the corresponding values of $G$ decrease. Figure 3 illustrates this point well with the curve labeled $G$. Therefore, on this basis one can expect increasing deviation from the Heidelberger–Kendall equation with decreasing amounts of $G$ (as antibody excess increases). This observation has been verified experimentally.[17] Equations 37 are not required to obtain this result. Any other set of values for $p$ in equations 36 will give the same effect, provided $\alpha$ is assumed to be independent of composition, an assumption which is not unreasonable in view of the interpretation of $\alpha$. The appropriate dilutions of antigen in an experiment of this kind should be determined, therefore, from the fact that $A_b/G$ depends on $G$ in an inverse manner.

(15) M. Heidelberger and F. E. Kendall, *J. Exp. Med.*, **61**, 563 (1935).

(16) F. E. Kendall, *Ann. N. Y. Acad. Sci.*, **43**, 85 (1942).

(17) W. Boyd, "Fundamentals of Immunology," 2nd Ed., Interscience Publishers, Inc., New York, N. Y., 1947, p. 257.

Although there is a theoretical basis for the use of equation 37 over the entire range of precipitation in contrast to the Heidelberger–Kendall equation, nevertheless the quantity $A_b/G$ is not the one of interest since, obviously, considerable antigen as well as bound antibody is not precipitated in the region of antigen excess. The expressions which would be more appropriate, however, are those for $\bar{i}/k$ or $(\bar{i}+\bar{j})/k$ given by equations 27, 31 and 34, since they determine the antibody–antigen ratios for aggregates. They do not require the number of free antigen molecules in the system to be negligible. Consequently, they may be used for the entire region of precipitation for either the $\alpha$- or $\beta$-titration procedure.

It is of interest here to discuss the distribution of species for a system containing $f$-valent antigen and univalent antibody molecules only. In such a system aggregates containing more than one antigen molecule cannot exist, and the distribution is described by $m_{ojk}$. Obviously, $k$ can only have the value zero or unity. If it has the former value than $j$ must take the value unity. If it has the latter value $j$ can vary from zero to $f$. Therefore, equation 18 reduces to

$$m_{ojk} = fG\left[\frac{(f-1)!}{(f-j)!j!}\right]^k p^{j+k-1} s^{k-1}(1-p)^{fk-k-j+1}(1-sp)^{1-k} \quad (38)$$

$$k = 0, 1$$

$$j = kl + 1$$

$$-1 \leqslant l \leqslant f-1$$

If for thermodynamic equilibrium values of $m_{ojk}$ and $p$, $K_j$ is defined by

$$K_j = \frac{m_{oj1}}{m_{100}\, m_{oj-1,1}} \quad (39)$$

then with the use of equation 38, one obtains the familiar results[18]

$$K_j = \frac{K_1}{f} \times \frac{f-j+1}{j}$$

$$\frac{K_1}{f} = \frac{p}{m_{010}(1-p)} \quad (40)$$

and

$$\frac{\sum_{j=1}^{f} j m_{oj1}}{\sum_{j=1}^{f} m_{oj1}} = \frac{K_1 m_{010}(1+K_1 m_{010}/f)^{f-1}}{(1+K_1 m_{100}/f)^f - 1} \quad (41)$$

Equation 41 gives the equilibrium ratio of the number of bound antibody molecules to the number of bound antigen molecules. It is clear that the equilibrium distribution is a special case of equation 38.

A comparison of two systems with the distributions $m_{iok}$ (all $i$ and $k$ with $j$ zero) and $m_{ojk}$ (all $j$ and $k$ with $i$ zero), respectively, can be made after computing their average molecular weights. The number average molecular weight $<M>_n$ and the weight average molecular weight $<M>_w$ are defined by

$$<M>_n = \sum_{i,j,k} M_{ijk} f_{ijk}^{(n)};\ f_{ijk}^{(n)} = m_{ijk}/\sum_{i,j,k} m_{ijk}$$

$$<M>_w = \sum_{i,j,k} M_{ijk} f_{ijk}^{(w)};\ f_{ijk}^{(w)} = M_{ijk} m_{ijk}/\sum_{i,j,k} M_{ijk} m_{ijk}$$

$$M_{ijk} = iM_A + jM_D + kM_G \quad (42)$$

(18) M. F. Morales, J. Botts and T. L. Hill, THIS JOURNAL, **70**, 2339 (1948).

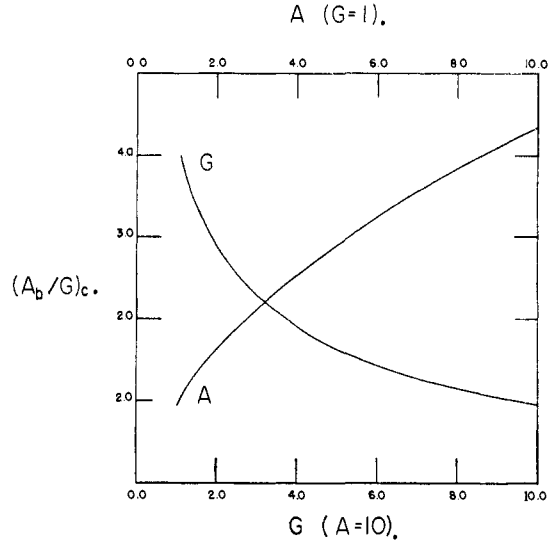


Fig. 3.—Variations in $(A_b/G)_c$ with increasing amounts of $A$ and $G$ are shown for the critical extent of reaction. $(A_b/G)_c$ is the ratio of bound bivalent antibody to total antigen at the critical point. The curve labeled A refers to the upper abscissa which represents additions of bivalent antibody to the system. The curve labeled G refers to the lower abscissa which represents additions of antigen to the system. A valence of five is assumed for the antigen.

We shall designate the system containing univalent antibody by the superscript one and the system containing bivalent antibody by the superscript two. Consequently, equations 14, 18, 38 and 42 lead to

$$<M>_n^{(1)} = \frac{DM_D + GM_G}{D + G - fGp}$$

$$<M>_n^{(2)} = \frac{AM_A + GM_G}{A + G - fGp} \quad (43)$$

and also to

$$[<M>_w^{(1)} - <M>_{w0}^{(1)}]\, W^{(1)} = 2M_D M_G fGp + M_D^2 Gf(f-1)p^2$$

$$[<M>_w^{(2)} - <M>_{w0}^{(2)}]W^{(2)} = \frac{2AM_A^2(f-1)\alpha + GM_G^2 f\alpha + 2M_A M_G fGp}{1-(f-)\alpha}$$

$$W = \sum_{i,j,k} M_{ijk} m_{ijk} \quad (44)$$

The symbol $<M>_{wo}$ is the initial weight average molecular weight of the system (that is, for $p$ zero). Differences between the two systems are manifested in a graph of $[<M>_w - <M>_{wo}]W$ plotted against the extent of reaction in Fig. 4. We may conclude from it that there are significant differences in aggregation between the two systems, even at small extents of reaction. Therefore, in the system containing bivalent antibody the reactions yielding aggregates with only one antigen cannot be separated in time from the reactions causing the formation of larger aggregates. This conclusion tends to render as unrealistic the basis for Hershey's theory.[19–21]

(19) A. D. Hershey, *J. Immunol.*, **42**, 455 (1941).
(20) A. D. Hershey, *ibid.*, **42**, 485 (1941).
(21) A. D. Hershey, *ibid.*, **42**, 515 (1941).

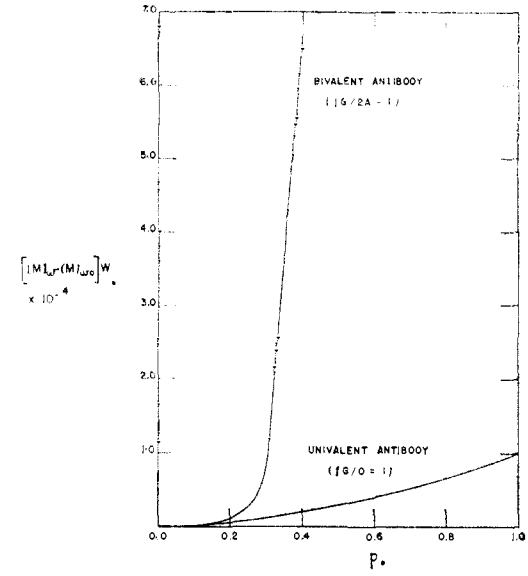


Fig. 4.—The effect of the extent of the reaction $p$, on a function of the weight average molecular weight, $[<M>_w - <M>_{wo}]\,W$, for a system containing bivalent antibody and one containing univalent antibody. $<M>_{wo}$ is the weight-average molecular weight of the system for $p$ zero. $W$ is the mass of the system.

## Part C. Comparison of the Theory with Experiment

In this section we shall examine experimental data from the literature on antibody–antigen systems with regard to the theory presented here. Two of these systems contain horse antibody and two contain rabbit antibody. They were chosen for two reasons: we wish to compare horse antibody systems with rabbit antibody systems, and secondly, the data in the literature describing them appear to be more complete than those describing other systems. For purposes of treating these systems theoretically we assume that all antibody molecules are bivalent. The antibody–antigen ratio of the precipitate, $R$, is compared with values of $i/k$ for the maximum extent of reaction and for the critical extent of reaction. If the critical point is the point at which precipitation is initiated, then the value of the antibody–antigen ratio of the precipitate should lie between the corresponding values of $(\bar{i}/k)_{max}$ and $(\bar{i}/k)_c$. According to equations 31 and 34, the composition of the system and the valence of the antigen must be known in order to evaluate the last two ratios.

The experimental data used here were obtained from the usual kind of titration experiment which can be described briefly as follows. A series of test-tubes is set up, each with a known amount of antigen nitrogen. The amount of antigen varies logarithmically with the tube number. The same amount of antiserum is added to each tube. The system in each tube differs from the others, therefore, only in the amount of antigen present, and in a definite way. Systems differing only in this way will be referred to as a set. The number of antigen molecules $G$, in each system was determined from the amount of antigen nitrogen added and the amount of nitrogen per antigen molecule (the nitrogen factor). The total nitrogen in the precipitate of each tube is plotted against the tube number. The curve invariably has a maximum. Although the amount of antigen in the system is determined directly, the amount of antibody is not. The latter value is generally obtained as the difference between the total nitrogen in the precipitate and the known antigen nitrogen in the system which corresponds to some point in the vicinity of the maximum of the curve just described. The location of this point is decided upon by specific ring tests on the supernatants from all the tubes. The supernatants which give no ring with either antibody or antigen solution are said to belong to systems which lie in the equivalence region, and it is in this region where the above point is chosen.

The theory, however, predicts that there is still uncombined antibody in the system after the maximum extent of reaction has been reached. As a second approximation, therefore, we consider the total amount of antibody in the system to be that usually taken plus the calculated amount of uncombined antibody. The latter amount is calculated by a method of successive approximations using the equation

$$m_{100} = A(1 - rp_{max})^2 = (A/4)(1 - G/A)^2 \quad (45)$$

Equation 45 follows from equations 18, 32 and 33. The trial value of $A$ used is that value usually taken as total antibody. The uncombined antibody $m_{100}$ is calculated and added to the trial value of $A$. This sum is used as a new value of $A$ to calculate another $m_{100}$. The process is continued until a constant value of $m_{100}$ is obtained. The valence of the antigen is determined from the experimental antibody–antigen ratio of the precipitate or soluble complex corresponding to the greatest antibody excess attainable, $R_{max}$, by substituting the latter for $(\bar{i}/k)_{max}$ in equation 28.

$$f = R_{max} + 1;\ k >> 1 \quad (46a)$$

$$f = R_{max};\ k = 1 \quad (46b)$$

The molecular weights used in the calculations are 170,000 for all antibody, 40,000 for egg albumin and 70,000 for serum albumin and diphtheria toxin. All nitrogen factors were taken to be 6.25 except that for the horse antibody for which 6.95 was used.[22] The system egg albumin–rabbit anti egg albumin is described in Table I. The data have been taken from the paper by Heidelberger and Kendall.[23] They yield for the total antibody nitrogen in the system 0.752 mg. N. We have altered this value to 0.811 mg. N. A value of five has been taken for the valence of egg albumin according to equation 46a. This equation was used here because $R_{max}$ corresponds to exceedingly large aggregates ($k >> 1$). The data of Kabat and Heidelberger have been used to obtain Table II, which describes the system horse serum albumin–rabbit antiserum albumin.[24] They yield 0.875 mg. N for the total antibody nitrogen in the system. We have altered this value to 0.998 mg. N. According to equation 46a the valence of the

(22) A. M. Pappenheimer, H. P. Lundgren and J. W. Williams, *J. Exp. Med.*, **71**, 247 (1940).

(23) M. Heidelberger and F. E. Kendall, *ibid.*, **62**, 697 (1935).

(24) E. A. Kabat and M. Heidelberger, *ibid.*, **66**, 229 (1937).

serum albumin is seven. The last two systems to be described here contain horse antibody. The data on them have been obtained by Pappenheimer.[25] We have used the value 1.18 mg. N for the total antibody nitrogen, rather than the observed value 1.10 mg. N in the case of the egg albumin–horse anti egg albumin system shown in Table III. The valence of egg albumin has already been given. The diphtheria toxin–horse antitoxin system appearing in Table IV has an antigen valence of seven according to equation 46b and the data on soluble complexes in antibody-excess inhibition.[22] We altered the observed antibody nitrogen value from 0.476 mg. N to 0.516 mg. N.

TABLE I

THEORETICAL AND EXPERIMENTAL VALUES OF PRECIPITATE RATIOS FOR EGG ALBUMIN–RABBIT ANTI EGG ALBUMIN[23]

| $A/G$ | $R$ | $(i/k)_{max}$ | $(i/k)_e$ |
|---|---|---|---|
| 21 | 3.9 | 4.0 | ... |
| 13 | 3.5 | 4.0 | ... |
| 7.6 | 3.3 | 4.0 | 3.5 |
| 4.8 | 2.9 | 4.0 | 2.8 |
| 3.8 | 2.7 | 3.7 | 2.5 |
| 2.9 | 2.4 | 2.6 | 2.2 |
| 2.6 | 2.3 | 2.3 | 2.0 |
| 2.3 | 2.1 | 2.1 | 1.9 |

TABLE II

THEORETICAL AND EXPERIMENTAL VALUES OF PRECIPITATE RATIOS FOR HORSE SERUM ALBUMIN–RABBIT ANTISERUM ALBUMIN[24]

| $A/G$ | $R$ | $(i/k)_{max}$ | $(i/k)_e$ |
|---|---|---|---|
| 17 | 6.0 | 6.0 | 5.3 |
| 11 | 5.6 | 6.0 | 4.3 |
| 8.2 | 5.1 | 6.0 | 3.8 |
| 5.1 | 4.0 | 4.7 | 3.0 |
| 4.4 | 3.6 | 3.8 | 2.8 |
| 4.1 | 3.3 | 3.5 | 2.7 |
| 3.6 | 3.1 | 3.1 | 2.5 |
| 3.3 | 2.9 | 2.8 | 2.4 |

TABLE III

THEORETICAL AND EXPERIMENTAL VALUES OF PRECIPITATE RATIOS FOR EGG ALBUMIN–HORSE ANTI EGG ALBUMIN[25]

| $A/G$ | $R$ | $(i/k)_{max}$ | $(i/k)_e$ |
|---|---|---|---|
| 3.1 | 2.4[a] | 2.8 | 2.2 |
| 2.6 | 2.3 | 2.3 | 2.0 |
| 2.2 | 2.0 | 2.0 | 1.9 |
| 2.1 | 1.9 | 1.9 | 1.8 |
| 1.8 | 1.6 | 1.7 | 1.7 |

[a] Precipitation was probably incomplete. The value of $(i/k)_{max}$ corresponds to 88% of the antigen precipitating.

TABLE IV

THEORETICAL AND EXPERIMENTAL VALUES OF PRECIPITATE RATIOS FOR DIPHTHERIA TOXIN–HORSE ANTITOXIN[25]

| $A/G$ | $R$ | $(i/k)_{max}$ | $(i/k)_e$ |
|---|---|---|---|
| 2.9 | 2.7 | 2.5 | 2.2 |
| 2.6 | 2.4 | 2.2 | 2.1 |
| 2.3 | 2.1 | 2.0 | 2.0 |
| 1.7 | 1.6 | 1.6 | 1.7 |
| 1.3 | 1.2 | 1.2 | 1.5 |

The values determined for the valences of the antigens in the four systems just described can be substituted into equation 26 to furnish us with theoretical limits imposed on the antibody–antigen ratio for attaining the critical point. Table V shows the comparison of these limits with those beyond which precipitation did not occur. It should be noted that precipitation can occur in both of these rabbit antibody systems for compositions at which the critical point is not attainable. The horse antibody systems show no such behavior. In fact, for these systems the precipitation limits are more restricting than the critical point limits. The theory is not to be interpreted, however, as requiring the precipitation limits to coincide with the critical point limits. If the attainment of the critical point is required for precipitation to occur, then the theory implies that precipitation cannot occur outside these limits. It does not predict what the limits will be. From the point of view of this theory the limits for the horse antibody systems may be as restricting as they are either on account of equilibrium requirements or the presence of univalent antibody. It is clear, however, that precipitation can occur before the critical point is reached in the rabbit antibody systems studied here. One can interpret this to mean that the rabbit antibody is more insoluble than the horse antibody, as has already been mentioned.

TABLE V

A COMPARISON OF INHIBITION ZONE LIMITS AND CRITICAL POINT LIMITS

| System | Experimental limits beyond which precipitation does not occur | Theoretical limits beyond which the critical point is not attainable |
|---|---|---|
| Egg albumin–horse anti egg albumin | $1+ \leqslant A/G \leqslant 4+$ | $5/8 \leqslant A/G \leqslant 10$ |
| Diphtheria toxin–horse antitoxin | $1 \leqslant A/G \leqslant 5$ | $7/12 \leqslant A/G \leqslant 21$ |
| Egg albumin–rabbit anti egg albumin | $1+ \leqslant A/G \leqslant 21$ | $5/8 \leqslant A/G \leqslant 10$ |
| Horse serum albumin–rabbitt antiserum albumin | $? \leqslant A/G \leqslant \sim 41$ | $7/12 \leqslant A/G \leqslant 21$ |

Figures 5 and 6 present an interesting experiment which might be mentioned briefly. They represent a kind of three dimensional diagram of an Rh agglutination test.[26] The abscissas give the ratio of inhibiting to agglutinating antibody molecules, a variable not ordinarily available to the experimenter. The ordinates give the usual antiserum dilutions. The amount of agglutination is expressed by the different kinds of crosshatching. In these figures the prozone becomes larger with increasing amounts of inhibiting antibody molecules until, finally, complete inhibition exists. This trend is not apparent when the relative amount of agglutinating serum present is small (the upper values of the ordinates). This is just the effect predicted by equation 26 and Fig. 1. Figures 5 and 6 correspond to red blood cells of different origins. In performing these experiments, Sturgeon noted that the red blood cells used in preparing the tests for Fig. 6 had considerably more combining power than those corresponding to

(25) A. M. Pappenheimer, *J. Exp. Med.*, **71**, 263 (1940).

(26) Personal communication with P. Sturgeon.

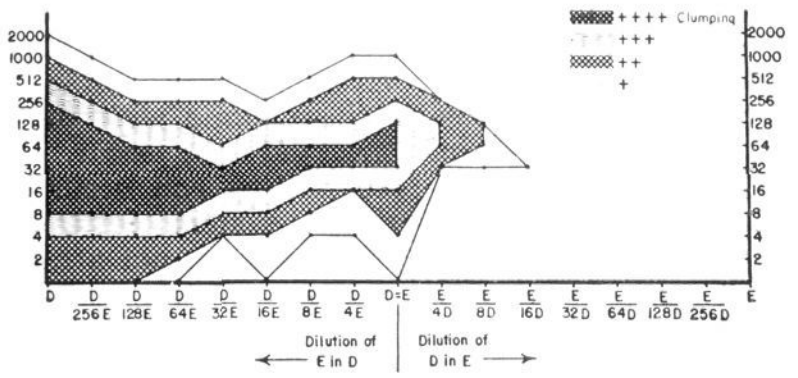


Fig. 5.—The effect of inhibiting antibody on Rh agglutination. D represents the serum containing agglutinating antibody. E represents the serum containing inhibiting antibody. Antiserum dilutions are indicated on the ordinate. The red blood cells used in this test were considered by Sturgeon to have the usual combining power.

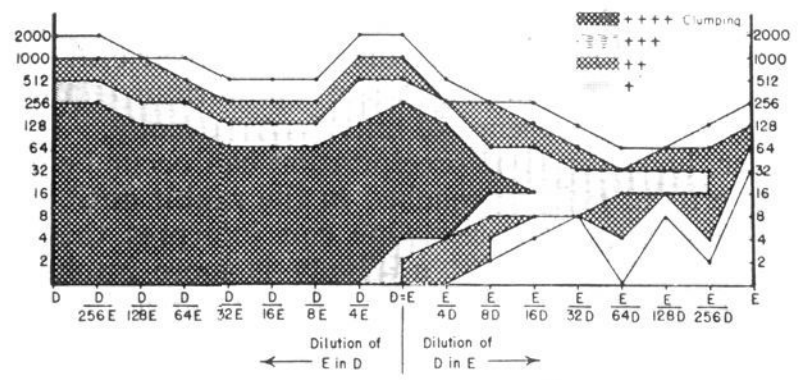


Fig. 6.—The effect of inhibiting antibody on Rh agglutination. D represents the serum containing agglutinating antibody. E represents the serum containing inhibiting antibody. Antiserum dilutions are indicated on the ordinate. The red blood cells used in this test were considered by Sturgeon to have an abnormal combining power.

Fig. 1. They cause a tremendous decrease in the prozone and, hence, tend to counteract the effect of the inhibiting antibody molecules. The point at which the agglutination occurs at the upper end of the ordinate remains the same, however. Equation 26 and Fig. 1 are again in agreement. They demonstrate the insensitivity of the critical point composition to changes in the combining power of the antigen in the antigen-excess region (lower limit). They do show a very large effect at the other end. An increase in antigen combining power $f$, for a fixed composition, reduces the antibody-excess inhibition zone or prozone considerably. Therefore, the Sturgeon diagrams appear to be in good qualitative agreement with this theory. No quantitative study will be attempted, however, because the assumptions upon which the theory is based do not warrant it.

## Part D. Discussion

The most probable distribution has been calculated for a system containing univalent and bivalent antibody molecules and $f$-valent antigen molecules. The assumptions of equal reactivities and no intra-aggregate reactions in certain instances are not quite correct, but the use of them probably leads to only a few per cent error. The average ratios of antibody to antigen for the critical and maximum extents of reaction, described by equations 31 and 34, are obtained from the distribution of species. For these extents of reaction the aggregates are so large that their ratios probably describe the antibody–antigen ratios of the precipitates for the systems of interest here. Comparison of them with data on four antibody–antigen systems previously described in the literature yields good agreement without the use of any empirical approach. The data required for the calculations are the composition of the system and the valence of the antigen. Although we have adjusted the antibody content of the system to be consistent with the theoretical distribution, the amount of antibody used has never differed by more than about 10% from that obtained experimentally. The valence of the antigen has been computed from data giving the maximum combining power of the antigen. Since the critical point is the point at which the system changes from one containing chiefly small aggregates to one containing chiefly large aggregates, we can use it to compute limits for obtaining inhibition in systems whose aggregates are sufficiently soluble so as not to precipitate out until they become as large as those given by the critical point. The theory appears to stand up rather well under all the tests which have been applied. These tests represent quite rigid requirements.

Even though the theory has the disadvantage of considering only univalent and bivalent antibody, it has many advantages over its predecessors. It contains no arbitrary parameters for curve fitting purposes. The calculated antibody–antigen ratios agree with the experimental data available for all compositions of the system. They also predict ratios for which no experimental data are available (antigen excess side of the equivalence zone). In contrast to other theories they are valid for systems in which the composition is varied by varying the antibody content. Furthermore, no previous theory has furnished a method for determining the limits of inhibition in antigen excess and antibody excess.

A theory describing $g$-valent antibody and $f$-valent antigen will be the next task undertaken. It is hoped that there will be significantly distinguishing features between systems differing only in the valence of the antibody. The valence of the antibody manifests itself in the limits of inhibition for obtaining the critical point. The limits for this inhibition in a system containing $g$-valent antibody are found to be

$$\frac{f}{g(g-1)(f-1)} \leqslant \frac{A}{G} \leqslant \frac{f}{g}(f-1)(g-1) \quad (47)$$

The substitution of the value two for $g$ yields equation 26 if $\rho$ is unity. It should be noted that as the valence of the antibody increases, the upper limit increases. Such increases are not sufficient to explain the large deviations of experimental inhibition from theoretical inhibition of systems containing rabbit antibody (see Table V). They can, however, contribute to these deviations. Solubility properties are probably more important in this regard.

**Acknowledgment.**—The author wishes to express his gratitude to Professor John G. Kirkwood for his illuminating remarks about the theory presented here.

## Appendix. Evaluation of $W_{ijk}$

$W_{ijk}$ is defined as the number of ways in which $i$ bifunctional units (to be called $S_i$-units), $j$ unifunctional units (called $S_j$-units), and $k$ $f$-functional units (called $S_k$-units) can be formed into a single $i$, $j$, $k$-aggregate containing no cyclic structures. All units and all functional sites thereon are distinguishable. All sites on the $S_k$-units are equivalent. All sites on the $S_i$ and $S_j$-units are equivalent. Furthermore, sites on $S_i$ and $S_j$-units are permitted to react only with sites on $S_k$-units and *vice versa*.

This problem can be solved by the device invented by Mayer and Mayer[27] and adopted by Stockmayer in similar problems.[10] $S_k$-units are represented by mechanical frames containing $f$ holes. Indistinguishable bolts are required to hold the frames together, each bolt passing through a pair of holes belonging to different frames. Bolts are also required to fill all other holes. These, however, do not connect different frames with each other. Each of them has one free end.

The number of ways to bolt all the frames together into a so-called $k$-aggregate, containing no cyclic structures, is $W_k$. It should be noted that the insertion of $i$ $S_i$-units and $j$ $S_j$-units into the $k$-aggregate does not change the number of ways of forming the latter. $k - 1$ of the $S_i$-units must take the place of those bolts connecting two frames together. The rest of the $S_i$-units and the $S_j$-units must replace bolts which have one end free. The number of ways of inserting the $i$ $S_i$-units and $j$ $S_j$-units into the $k$-aggregate is defined as $R_{ijk}$. Therefore

$$W_{ijk} = W_k R_{ijk} \qquad \text{(A1)}$$

$W_k$ is determined in the following manner. Since a $k$-aggregate requires $k - 1$ bonds, $k - 1$ bolts are required for this purpose. Since bolts are required to fill all other holes, the total number of bolts used is then

$$fk - (k - 1) = fk - k + 1$$

Any one of the bolted arrangements can be dissociated into $k$ separate frames, each containing $f - 1$ holes occupied by bolts and one empty hole. There will be one free bolt left over. The bolt chosen as the free bolt uniquely determines the empty hole in each of the $k$ frames. Since there are $fk - k + 1$ bolts altogether, there are likewise $fk - k + 1$ different dissociated arrangements of the required kind which correspond to the same bolted arrangement. Now, if $P$ is the number of possible dissociated arrangements of this kind, and if $Q$ is the number of ways of bolting each dissociated arrangement together, the number of different bolted arrangements is

$$W_k = \frac{PQ}{fk - k + 1} \qquad \text{(A2)}$$

Since any one of the holes on the frames can be the empty one

$$p = f^k \qquad \text{(A3)}$$

To find $Q$, $k - 1$ indistinguishable washers are introduced, no more than one being placed on any bolt. The number of ways to choose $k - 1$ out of the $fk - k + 1$ bolts, on which to place washers, is

$$\frac{(fk - k + 1)!}{(fk - 2k + 2)!(k - 1)!}$$

Washered bolts are now inserted into holes in frames with which they are not already connected. The free bolt is kept for last. That is, the first washered bolt can select any one of $k - 1$ empty holes (excluding the one on its own frame). There are then $k - 2$ single frames and one double frame. In a like manner the second washered bolt can select any one of $k - 2$ empty holes. This process continues until only the one free bolt remains. If the free bolt has a washer there remain two structures, each with a hole, which must be bolted together. If it does not have a washer, there remains just one hole on one structure which must be filled. Therefore, washered bolts can be inserted in $(k - 1)!$ ways. This number of ways combined with the number of ways of assigning washers is

$$Q = \frac{(fk - k + 1)!}{(fk - 2k + 2)!} \qquad \text{(A4)}$$

Therefore, the substitution of equations A3 and A4 in A2 gives

$$W_k = \frac{f^k(fk - k)!}{(fk - 2k + 2)!} \qquad \text{(A5)}$$

This proof for $W_k$ is the same as that given by Stockmayer.[10]

$R_{ijk}$ can be obtained in the following manner. $k - 1$ of the $i$ $S_i$-units must be selected for the bonding positions now occupied by bolts. These can be selected in $i!/(i - k + 1)!$ ways. The remainder of the $S_i$-units, $i - k + 1$, and all the $S_j$-units must replace any of the $fk - 2k + 2$ bolts each of which has one end free. This selection can be accomplished in $(fk - 2k + 2)!/[fk - 2k + 2 - (i - k + 1) - j]!$ ways. Now since each of the $S_i$-units has two distinguishable functional sites, $R_{ijk}$ will contain the factor $2^i$. Therefore

$$R_{ijk} = 2^i \frac{(fk - 2k + 2)!}{(fk - k - i - j + 1)!} \times \frac{i!}{(i - k + 1)!} \qquad \text{(A6)}$$

Let the number of $S_i$-units of which only one functional site is used be defined by

$$q = i - k + 1 \qquad \text{(A7)}$$

With the use of equations A5, A6 and A7, equation A1 becomes

$$W_{ijk} = f^k 2^i \frac{(fk - k)!}{(fk - 2k + 2 - q - j)!} \times \frac{i!}{q!} \qquad \text{(A8)}$$

MADISON 6, WISCONSIN

(27) J. E. Mayer and M. G. Mayer, "Statistical Mechanics," John Wiley and Sons, Inc., New York, N. Y., 1940, p. 456 ff.