Werner[7] in which a crystalline product, containing three moles of acetone per mole of sodium iodide, is prepared. Distillation of this compound and repeated distillation of the product was followed by drying over calcium chloride in a moisture-free dry box. All sampling operations and preparation of solutions were carried out in the dry box which was flushed with nitrogen during the operation. Alternatively, the commercial acetone was repeatedly fractionated and the fraction boiling at 56.1° collected. The products prepared by both methods gave the same boiling point and refractive index, *i.e.*, $n^{20}D$ 1.3590. The reported refractive index is 1.3591.[8] The acetone purified by these procedures did not decolorize dilute permanganate solutions. It was found necessary to transfer the solution into the sample cell in a dry box because, in some cases, filling the cells in the laboratory air introduced traces of water. The sample cells were fixed thickness sodium chloride cells (0.2 mm.) which were not affected by acetone containing small amounts of water. All measurements were carried out with cells of the same thickness and the spectra were measured on the Perkin–Elmer Model 21 recording spectrophotometer, with sodium chloride optics. Wave length calibrations were made using the known absorption bands of water vapor and carbon dioxide.

Results

Figure 1, curve I, gives the absorption spectrum of an acetone sample partially purified as described above. The presence of water in the sample is indicated by the absorption in the 3600 cm.$^{-1}$ region. Known volumes of distilled water were added to this sample by means of calibrated micropipets and the increased absorption in this region is illustrated by curves II and III, which were taken with samples to which increasing amounts of water had been added. The Beers' law plot was obtained from the absorbance at 3600 cm.$^{-1}$ plotted against the concentration of water added, in parts per thousand by volume, to purified acetone. The highest per cent. transmittance obtained in the case of the purest acetone prepared was of the order of 90%. In the range 0–20 p.p.t. of water added to this sample the curve is not linear but shows positive deviations from linearity and levels off at higher water concentrations (13–20 p.p.t. added water). Extrapolation of the smooth curve to zero absorbance gave an estimate of the amount of water present in the purified sample, assuming the 3600 cm.$^{-1}$ absorption of the purified sample was due to traces of water. This procedure gave approximately 2.5 p.p.t. by volume, or about 0.2% by weight. In the initial portion of the curve, to approximately 10 p.p.t. of added water, the curve gave satisfactory results for the estimation of water added to the sample. For five different concentrations in this range the results obtained gave an average deviation from the known amount of water added of 0.08%. However, at higher concentrations the absorbance of the sample is less sensitive to changes in water concentration.

It should be mentioned that a variety of salts (strontium bromide, strontium sulfate and others), when added in excess to an acetone–water solution of known concentration, gave a measurable decrease in the per cent. absorption at 3600 cm.$^{-1}$ but not at 3400 cm.$^{-1}$. This decrease, probably due to adsorption of water on the powder, became greater when the surface area of the solid increased. However, the variations were too small to permit a reliable estimate of apparent surface areas.

Acknowledgment.—The author gratefully acknowledges the financial support given by Research Corporation in support of this work.

CHICAGO, ILLINOIS

(7) K. Shipsey and E. A. Werner, *J. Chem. Soc.*, **103**, 1255 (1913).

(8) N. A. Lange, "Handbook of Chemistry," Handbook Publishers, Inc., Sandusky, Ohio, 1949.

[CONTRIBUTION FROM THE DEPARTMENT OF BIOCHEMISTRY AND THE VIRUS LABORATORY, UNIVERSITY OF CALIFORNIA, BERKELEY]

A Study of the Kinetics of the Enzymatic Digestion of Deoxyribonucleic Acid[1]

BY V. N. SCHUMAKER, E. G. RICHARDS AND H. K. SCHACHMAN

RECEIVED JANUARY 17, 1956

In a recent paper from this Laboratory dealing with structural models for deoxyribonucleic acid (DNA) with particular emphasis on the two strand model proposed recently by Watson and Crick it was pointed out that a detailed study of the kinetics of the enzymatic digestion of DNA might yield information about different models. For example a one strand structure would be split at each attack by the enzyme. A two strand structure would require at least two attacks and therefore the efficiency of splitting would increase with the square of the number of attacks. A two strand structure with some "preformed gaps," a model proposed in the above mentioned paper, would degrade in a manner intermediate between a single strand model and a continuous two strand model. This communication presents a detailed theory for the degradation of such long chain macromolecules. From the distribution of fragments formed upon random hydrolysis the decrease in the viscosity of such a solution is calculated. The calculation is generalized to include solutions of polymers each molecule of which contains a definite number of chains bound side by side. Certain parts of this treatment are general for molecules polydisperse with respect to both molecular weight and shape. This communication also presents new data on the viscosity and sedimentation changes as a function of the number of enzymatic attacks on the DNA macromolecule. The number of enzymatic attacks is determined in a "*p*H-stat" by measuring the hydrogen ion liberation resulting from the cleavage of the phosphodiester bonds. The theory developed may be applied to the experimental data, and thus the single strand model for the DNA molecule is ruled out. The best agreement is obtained with a doubly stranded model. Moreover the number of preformed gaps in the structure as determined by this study is less than one in about 3000 nucleotides.

Considerable interest has been focussed on the helical model proposed by Watson and Crick[2] for the structure of deoxyribonucleic acid (DNA), and many lines of evidence support the view that DNA, as commonly isolated, is a hydrogen-bonded, intertwined, two strand structure. X-Ray diffraction data and chemical analysis furnish the principal sources of evidence for this model. Furthermore, the hydrodynamic behavior, as revealed by sedi-

(1) This work was supported by grants from the National Science Foundation, Lederle Laboratories, and The Rockefeller Foundation.

(2) J. D. Watson and F. H. C. Crick, *Nature*, **171**, 737 (1953).

mentation,[3] viscosity[4] and streaming birefringence studies,[5] and the structure factor, from light scattering measurements,[6] show that in solution the molecules are highly extended and fairly rigid, perhaps in the form of curved or bent, thread-like particles. The "anomalous" titration behavior and the changes in ultraviolet absorption upon degradation are consistent with the view that DNA is a highly organized structure with the purine and pyrimidine bases hydrogen bonded to one another and arranged in a regular fashion.

Another approach to the determination of the structure of DNA has involved the simultaneous measurement of a number of physical-chemical properties during the early stages of the enzymatic degradation of DNA.[7] From these studies it was concluded that the enzymatic attack involved a splitting of elongated, thread-like structures into a number of shorter pieces by the random hydrolysis of phosphodiester bonds in the backbone chains. This was accomplished during the very early stages without disrupting the hydrogen bonding between the purine and pyrimidine rings. However, as the fragments approached a molecular weight of about 50,000, the ultraviolet absorption began to increase. This was interpreted to mean that the hydrogen bonding between the bases was not sufficiently strong to hold these short chains together at room temperature. Thus no direct action by the enzyme need be evoked to account for the rupture of hydrogen bonds and the separation of the chains. In a discussion of various models for DNA, it was suggested that a study of the kinetics of enzyme action would discriminate among a one strand structure, a two strand structure and the proposed interrupted two strand structure. The preliminary results showed that while there did exist a short "lag" period between the addition of the enzyme and the decrease in the viscosity, still only about five bonds per molecule need be broken to produce a significant change in the viscosity. This result seemed to cast doubt upon a one strand structure but no differentiation was made between a continuous *versus* an interrupted two strand model.

In this communication is presented a detailed statistical theory for the degradation of macromolecules. This theory shows that differentiation can be made readily among one, two or three strand structures as long as the degrading agent attacks at random only one strand at a time. Moreover, it is possible to determine the number of preformed gaps in a macromolecule and the minimum number of secondary or cross-linking bonds which are required to hold the strands together. Also this communication presents new data on the viscosity and sedimentation changes as a function of the number of enzymatic attacks on the macromolecules. These results are then used with the theory in a discussion of different models for the structure of DNA. It should be emphasized that we are studying the structure of the macromolecule as it exists in solution which, of course, may be different from the DNA as it exists in the cell.

(3) A. R. Peacocke and H. K. Schachman, *Biochim. Biophys. Acta*, **15**, 198 (1954).

(4) J. Pouyet, *Compt. rend. Acad. sci. (Paris)*, **234**, 152 (1952).

(5) H. Schwander and R. Cerf, *Helv. Chim. Acta*, **34**, 436 (1951).

(6) M. E. Reichmann, S. A. Rice, C. A. Thomas and P. Doty, THIS JOURNAL, **76**, 3047 (1954).

(7) C. A. Dekker and H. K. Schachman, *Proc. Natl. Acad. Sci. (U. S.)*, **40**, 894 (1954).

Theory

Let us consider a solution of homogeneous long chain macromolecules each of which is composed of m monomer units linked together by bonds susceptible to attack by an enzyme. If the enzyme breaks bonds at random, then a distribution of molecules of varying size results. Consequently the viscosity of the solution decreases as the enzymatic degradation proceeds. For high polymers the intrinsic viscosity, $[\eta]$, is proportional to M^α where M is the molecular weight and α is a number varying from $^1/_2$ for random coils to 2 for long rods. The viscosity increments caused by the different species are additive at low concentrations, and we may write

$$\left[\frac{\eta_{sp}^{obs}}{c}\right]_{c=0} = \frac{A\sum c_i M_i^\alpha}{c} \qquad (1)$$

where the summation is carried out over all of the species, i, and where A is a constant characteristic of a given polymer and solvent system. The symbol c is the concentration and η_{sp}^{obs} is the observed specific viscosity.

If the probability that a bond in any molecule has been split by the enzyme is p and the probability that it has not been broken is $q = 1 - p$, then the number of molecules $L(m)$ still containing m monomer units is

$$L(m) = Nq^{m-1} \qquad (2)$$

where N is the initial number of molecules. Similarly $L(i)_t$ is the number of molecules containing i monomer units including the terminal units

$$L(i)_t = 2Nq^{i-1}p \qquad (2a)$$

There can also be molecules containing i monomer units none of which are originally terminal units and this can be written as $L(i)_{\bar{t}}$

$$L(i)_{\bar{t}} = Nq^{i-1}p^2(m - i - 1) \qquad (2b)$$

The sum of 2, 2a and 2b gives the total number of molecules containing i monomer units. In equation 1, M_i and c_i are proportional to i and $iL(i)$, respectively. Substitution of these values in equation 1 and dividing by the initial intrinsic viscosity $[\eta]_0$ yields

$$R = \frac{[\eta]}{[\eta]_0} = \frac{\sum_i L(i)_t i^{1+\alpha} + \sum_i L(i)_{\bar{t}} i^{1+\alpha} + L(m)m^{1+\alpha}}{m^{1+\alpha}} \qquad (3)$$

where R is defined as the ratio of the observed viscosity at any time relative to the initial intrinsic viscosity. The summations can be replaced by integrals from 0 to m. The integrals may be evaluated directly for integral values of α, and closed expressions are obtained. After integration the approximation $\ln q = -p$ is made, and for the case for $\alpha = 1$, we obtain,[8] where $S = mp$

$$R = \frac{2}{S^2}(e^{-S} + S - 1) \qquad (4)$$

(8) Equation 4 is identical to that derived by A. Charlesby for the change in molecular weight as a function of time, *Proc. Roy. Soc. (London)*, **A224**, 120 (1954). Our approach is much different from that used by Charlesby and can be generalized for any value of α for a homogeneous distribution.

When α has values different from unity we obtain

$$R = e^{-S}(1 + S) - \frac{(\alpha - S)}{S^{1+\alpha}} \int_0^S e^{-z} x^{1+\alpha}\, dx \quad (5)$$

subject[9] to the condition $S<20$. Values of the integral in equation 5 have been tabulated.[10] Equation 5 may also be written

$$R = 1 - \frac{\alpha}{\alpha + 2} S + \frac{(\alpha + \alpha^2)S^2}{2!(\alpha + 2)(\alpha + 3)} + \cdots + \frac{(\alpha + \alpha^2)S^n(-1)^n}{n!(\alpha + n)(\alpha + n + 1)} + \cdots (6)$$

For computational purposes only terms including S, S^2 and S^3 need be used for the initial stages of degradation, *i.e.*, $S<1$.

It should be noted that equation 6 applies to the degradation of a solution of macromolecules which were initially homogeneous with regard both to shape and to molecular weight. This equation will now be generalized for any initial distribution of molecular weights and shapes.[11] For small values of S we can ignore terms containing higher powers and write for each species of weight, m, and shape, j

$$R_{mj} = 1 - \frac{\alpha_j}{\alpha_j + 2} ms \quad (7)$$

where s is the number of scissions per monomer unit, that is, $ms = S$. Summing over the distribution of shapes and molecular weights where

$$R^{obs} = \frac{\sum_{mj} R_{mj}(\eta^0_{sp})_{mj}}{\eta^0_{sp}} = \frac{\sum_{mj} 1 - \left(\frac{\alpha_j}{\alpha_j + 2}\right) ms\, (\eta^0_{sp})_{mj}}{\eta^0_{sp}} \quad (8)$$

$(\eta^0_{sp})_{mj}$ is the initial specific viscosity of the mj component and η^0_{sp} is the total initial specific viscosity. An averaging process may now be carried out

$$R^{obs} = 1 - \left(\overline{\frac{\alpha_j}{\alpha_j + 2}}\, m\right) s \frac{\sum_{mj} (\eta^0_{sp})_{mj}}{\eta^0_{sp}} \quad (9)$$

Since $\sum_{mj}(\eta^0_{sp})_{mj} = \eta^0_{sp}$ we can write

$$R^{obs} = 1 - K_1 s \quad (10)$$

where K_1 lies between the minimum and maximum values of $\alpha_j/(\alpha_j+2)m$. It is of interest to note that from values of R^{obs}, s and α, it is possible to calculate $\bar{m}$, the $\alpha + 2$ average degree of polymerization for any initial distribution of molecular weights.

Thus far we have been considering a single strand polymer in which there would be mp scissions in the time, t, where m is the number of monomer units in the macromolecule ($m>>1$) and p is the probability that a given bond connecting any two monomer units is broken in the time, t. For a polymer composed of two strands, both strands must be broken if a scission is to occur. This scission will occur if the breaks in each strand are opposite to one another or if the breaks are staggered only several units apart as long as the interstrand bonding is not sufficiently strong to maintain the integrity of the molecule. Let β equal the maximum number of units by which the breaks can be staggered without preventing scission of the molecule. We can now write for the total number of scissions per molecule, S

$$S = \frac{m}{2} (2\beta + 1)p^2 \quad (11)$$

where $m/2$ is the number of units in each strand. In an analogous manner we shall write for a molecule composed of n strands[12]

$$S = \frac{m}{n} (2\beta + 1)^{n-1} p^n \quad (12)$$

or

$$s = \frac{S}{m} = \frac{(2\beta + 1)^{n-1}}{n} p^n \quad (12a)$$

For the initial stage of the degradation we can combine equations 10 and 12a to give

$$\log_{10}(1 - R^{obs}) = n \log_{10} p + \text{constant} \quad (13)$$

Thus the number of strands can be evaluated from data showing the decrease in viscosity as bonds are being broken early in the reaction. When the enzymatic activity is constant during the initial stage of the reaction,[13] p will be proportional to the time, t, and equation 13 may be written

$$\log_{10}(1 - R^{obs}) = n \log_{10} t + \text{constant} \quad (13a)$$

Since it has been suggested[7] that the DNA macromolecule might be composed of two strands each of which is interrupted at many points, it is of interest to consider the statistics of such a model. If mg is the number of preformed gaps in the macromolecule, the number of scissions becomes

$$S = \frac{m}{2} (2\beta + 1)p^2 + 2\left(\frac{m}{2}\right)(2\beta + 1)pg \quad (14)$$

For degradation of two strand molecules containing preformed gaps, it is necessary to retain the term in S^2 in equation 6 and the averaging process leads to

$$R^{obs} = 1 - \left(\overline{\frac{\alpha}{\alpha + 2}}\, m\right) s + \left(\overline{\frac{(\alpha + \alpha^2)m^2}{2!(\alpha + 2)(\alpha + 3)}}\right) s^2 - \cdots \quad (15)$$

$$= 1 - K_1 s + K_2 s^2 - \cdots \quad (15a)$$

Combination of equations 14 and 15a gives

$$1 - R^{obs} = [K_1(2\beta + 1)g]p + \left[K_1 \frac{2\beta + 1}{2} - K_2(2\beta + 1)^2 g^2\right] p^2 + \cdots \quad (16)$$

To use equation 16, $(1 - R^{obs})/p$ should be plotted against p. The initial slope and the intercept of this plot are given by

$$(\text{Slope}) = K_1 \frac{2\beta + 1}{2} - K_2(2\beta + 1)^2 g^2 \quad (17)$$

and

$$(\text{Intercept}) = K_1(2\beta + 1)g \quad (17a)$$

(9) A solution of the same general form as equation 5 may be obtained without restriction to the condition that $S < 20$.

(10) "Tables of the Incomplete Γ-Function," Edited by Karl Pearson, Cambridge University Press, London, 1946.

(11) For the particular case in which the initial distribution of molecular weights is a Poisson distribution, the appropriate equations are given by A. Charlesby.[8]

(12) For our purposes, equation 12a may be replaced by the more general expression $s = Bp^n$, where B is independent of p.

(13) For the digestion of DNA, it was found that the enzymatic activity was constant until about 260 bonds were broken per weight average molecule.

These two simultaneous equations may be solved for g and $(2\beta+1)$

$$g = 1/2 \frac{\text{Intercept}}{\text{Slope} + (K_2/K_1^2)(\text{Intercept})^2} \quad (18)$$

$$2\beta + 1 = \frac{(\text{Slope}) + (K_2/K_1^2)(\text{Intercept})^2}{K_1} \quad (18a)$$

Now the coefficient of the $(\text{Intercept})^2$ term, K_2/K_1^2, is approximately equal[14] to

$$2\frac{(\alpha+1)(\alpha+2)}{(\alpha+3)}$$

which has a value less than or equal to 5 for $2 > \alpha > 1/2$. Hence if $5\,(\text{Intercept})^2 << (\text{Slope})$ it follows that

$$g = 1/2 \frac{(\text{Intercept})}{(\text{Slope})} \quad (19a)$$

$$(2\beta + 1) = \frac{(\text{Slope})}{K_1} \quad (19b)$$

Thus g may be evaluated without the necessity of making assumptions as to the average shape or molecular weight. To obtain β, on the other hand, requires that these quantities be estimated.

For the study of the degradation of DNA, which is described below, certain of the equations just derived are of special importance. These are equations 13a, 16, 19a and 19b. From equation 13a, the number of strands in the molecule, n, can be estimated. Equation 16 is the general expression relating the viscosity of the solution to the number of bonds broken, and will be used later. Finally, the frequency of preformed gaps is given by equation 19a, while 19b is used to determine the maximum number of interstrand bonds which may occur between interruptions on opposite chains without a scission of the molecule at that site.

Materials and Methods

Several different preparations of purified thymus DNA were used in these studies and the authors would like to express their thanks to Dr. S. Katz of this Laboratory for a sample prepared by the procedure of Signer and Schwander and to Dr. N. S. Simmons of the University of California at Los Angeles for a sample prepared through the use of sodium xylene sulfonate. The pancreatic deoxyribonuclease (D-Nase) was a commercial preparation from Worthington Enzyme Laboratories.

Viscosity measurements were made at 25 ± 0.002° in a three bulb viscometer at average shear gradients of 43, 92 and 145 sec.$^{-1}$. All data were plotted as specific viscosity against shear gradient, and the data were extrapolated in a linear fashion to zero shear gradient. The values of η_{sp}/c obtained by this method are independent of concentration as long as low concentrations of DNA are used[6] and are assumed to be equal to the intrinsic viscosity. For the DNA preparations used in this work the intrinsic viscosity was 50 ± 1 (g./100 ml.)$^{-1}$ in good agreement with the values reported by Reichmann, *et al.*,[6] Pouyet,[4] Sadron,[15] and Conway and Butler.[16]

For measurement of the number of bonds broken per unit time, experiments were conducted at *p*H 7.60 in an unbuffered solution containing 0.2 *M* sodium chloride and 0.03 *M* magnesium sulfate. The *p*H was maintained constant in a "*p*H-stat," described by Neilands and Cannon,[17] which yields a continuous recording of the amount of alkali automatically added to the reaction vessel when the amplified signal indicates departures from the pre-set *p*H value. The authors would like to thank Dr. Neilands for making his instrument available to us.

The ultracentrifuge studies were performed with a Spinco Model E analytical ultracentrifuge especially fitted with an ultraviolet absorption optical system. The absorption patterns were then analyzed with the Spinco Model R analytrol, a recording photodensitometer which gives the concentration as a function of position in the cell. Very low concentrations of DNA may be used for these experiments, and the distribution of sedimentation coefficients for the undegraded sample is found to be independent of concentration in agreement with the results of Shooter and Butler.[18] A detailed report in this method will be presented elsewhere.[19]

In preparing the DNA solutions, the solid dry DNA was placed directly in 0.2 *M* salt solutions. The enzyme solution was stored at 4° at a concentration of 100 γ/ml. and dilutions of it were made immediately prior to use in the kinetic studies. As a measure of enzyme activity, the enzyme was added to an unbuffered solution of DNA (0.003 g./100 ml.) contained in the water jacketed titration cell of the "*p*H-stat." The *p*H was maintained constant between *p*H 7.59 and 7.60, and a stream of nitrogen was continuously blown over the solution. Sodium hydroxide, specially protected against carbon dioxide, was used as a reagent in the titrimeter, and the recorder plotted the volume of reagent added to the reaction mixture as a function of time. The initial slopes of such curves, which were straight, gave a measure of the number of phosphodiester bonds split per unit time for the enzyme concentration used in the experiment. From experiments of this type it was found that the initial activity was proportional to the enzyme concentration.

For the kinetic studies in the viscometer and the ultracentrifuge the reaction mixture contained 0.01 *M* phosphate buffer at *p*H 7.60 in addition to 0.2 *M* sodium chloride. The *p*H of these solutions decreased only to 7.45 during the digestion. To determine the number of bonds broken per unit time comparable solutions of DNA *without* buffer and with enzyme concentration about 100 times greater were studied in the "*p*H-stat." These studies provide the data for the bonds split at the lower enzyme concentrations used in the viscosity studies.

Results and Discussion

In Fig. 1 are shown the viscosity data obtained at different times during the enzymatic digestion. The data are reproducible; four experiments using two different DNA preparations gave essentially identical results. The viscosity data are plotted in Fig. 2 as $\log_{10}[(1 - R^{obs})\ 1000]$ against $\log_{10} t$ in accordance with equation 13a. The slope of the straight line so obtained is, according to the theory, equal to n, the number of strands possessed by each DNA macromolecule. The value observed is, however, non-integral and equal to 1.47.

This non-integral value of n is readily explainable if a doubly stranded structure is postulated for the DNA molecule, as will be shown below. Therefore, let us first investigate the one strand[20] and multistrand structures in order to see if they too may be used to fit the experimental data.

The characteristics of a one strand model which distinguish it from all multistrand models is that the one strand structure scissions at each enzymatic

(14) The actual value of K_2/K_1^2 depends upon the distribution of molecular weights and shapes in the solution. We have assumed a value of about 4 for the ratio of the $Z + 1$ average molecular weight to the Z average molecular weight. For the experiments presented here the magnitude of the slope is such that the ratio $(K_2/K_1)^2$ could be much greater than the large value assumed and still not affect equation 19a.

(15) C. Sadron, *Progr. Biophys. Chem.*, **3**, 237 (1953).

(16) B. E. Conway and J. A. V. Butler, *J. Polymer Sci.*, **12**, 199 (1954).

(17) J. B. Neilands and M. D. Cannon, *Anal. Chem.*, **27**, 29 (1955).

(18) K. V. Shooter and J. A. V. Butler, *Nature*, **175**, 500 (1955).

(19) V. N. Schumaker and H. K. Schachman, in preparation.

(20) A doubly stranded molecule tied together at one end could be considered to be a singly stranded molecule. In terms of enzymatic degradation, however, the tying together of the strands at one end would constitute a trivial effect and the kinetics would be like those of a two strand model.

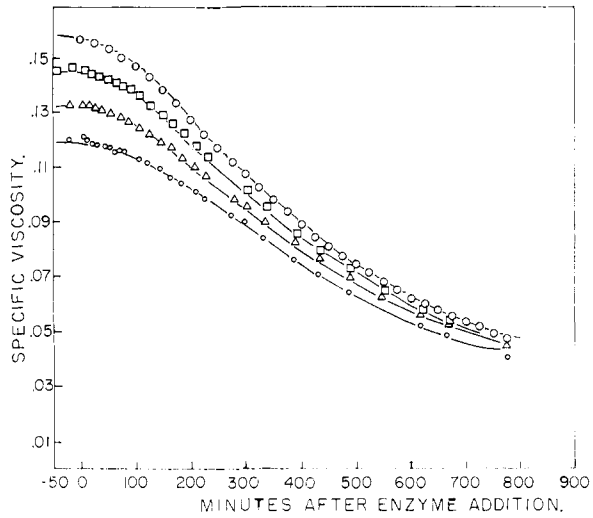

Fig. 1.—Viscosity of DNA solution undergoing enzymatic digestion: □, lower bulb shear gradient 43 sec.$^{-1}$; Δ, middle bulb shear gradient 92 sec.$^{-1}$; o, upper bulb shear gradient 145 sec.$^{-1}$; O, extrapolated shear gradient 0 sec.$^{-1}$.

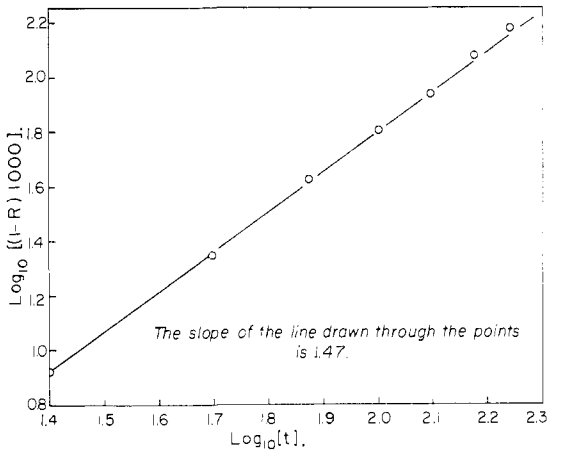

Fig. 2.—Determination of the number of strands for a multistrand model for the DNA molecule in solution.

attack. In these experiments the rate of enzymatic attack is constant and therefore, if a one strand model is correct, the number of scissions should be directly proportional to the time. However, a value of n greater than unity indicates that the efficiency of scissioning increases with time. Further emphasizing the inadequacy of the one strand model are the curves of the distribution of sedimentation coefficients shown in Fig. 3. After over 250 bonds are broken per molecule, the sedimentation coefficients have fallen by less than 50% although if the one strand model is correct the molecule should break into 250 fragments. Hence the one strand model is inconsistent with the experimental data.

The three strand model is now to be examined to see if it may account for the observed pattern of the enzymatic degradation. If there existed a very large number of preformed gaps in a three strand model, then the kinetics of degradation would be similar to those which would be observed for a doubly stranded model. Hence we are not able to differentiate between an "interrupted" three strand structure and a continuous two strand model. Moreover, four, five, six, etc., strand structures might be postulated, but here the preformed gaps would have to come in pairs, triplets, quadruplets, etc., respectively.

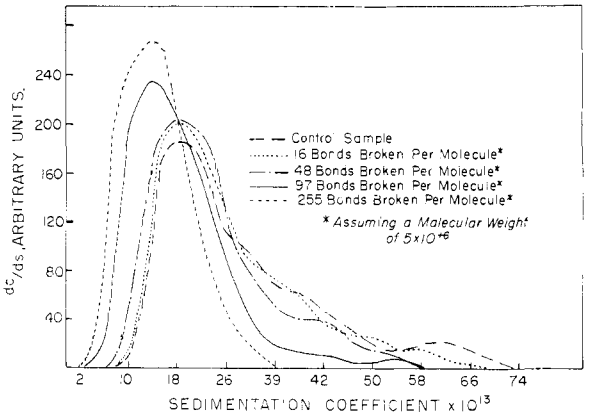

Fig. 3.—Distribution of sedimentation coefficients during enzymatic degradation of DNA.

It seems to us that the several *ad hoc* assumptions which are required if the four, five, six, etc., strand models are to be evoked are sufficiently unjustified to exclude these models from further consideration. The singly stranded structure has been ruled out because the observed degradation is very much slower than would be expected from such a model and also because there seems to be no reasonable way in which such a structure could be made to fit the observed pattern of the viscosity decrease.

Since the X-ray data point so strongly toward a two strand model let us assume, for the remainder of this discussion, that DNA is adequately represented in solution by a doubly stranded structure and determine whether such a structure is consistent with the value $n = 1.47$. For this purpose it is of interest to investigate the following problems: (1) Are the two strands continuous, or are they frequently interrupted along the length of the macromolecule? (2) Could the solution be "contaminated" by a large fraction of single strands? (3)

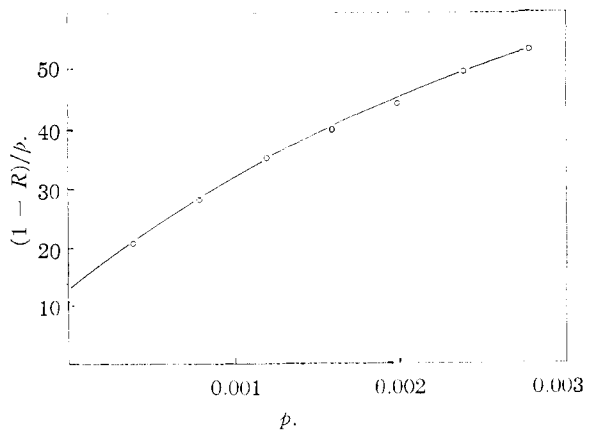

Fig. 4.—Plot of $(1 - R)/p$ *vs.* p for the enzymatic digestion of DNA.

Is there a large change in the flexibility of the DNA macromolecule at the site of an enzymatic attack? (4) Does the enzyme attack bonds at random or does it occasionally break both chains simultaneously in a non-random fashion?

Each of these possibilities may be satisfactorily evaluated by assuming that it alone accounts for the observed value of n being less than 2. Hence the maximum number of preformed gaps will be obtained by assuming that the deviation from double strand kinetics is due only to the presence of these interruptions. Therefore when $(1 - R^{obs})/p$ is plotted against p as is suggested by equation 19a, one half of the ratio of the intercept to the initial slope is equal to the maximum number of preformed gaps per nucleotide. The value calculated in this manner for the data presented in Fig. 4 is one preformed gap for every 3000 nucleotides.[21] It is interesting that this small number of preformed gaps is sufficient to account for the deviation from double strand kinetics. Moreover, that such a small value may be determined emphasizes the sensitivity of the viscometer as a tool for following the kinetics of the degradation of such macromolecules.

The second, third and fourth problems mentioned above may all be investigated through the use of the same mathematical formulation. In order to maximize the "contamination," flexibility increase, or non-random enzymatic attack it is necessary to set the number of preformed gaps equal to zero. When this is done, equation 16 becomes

$$1 - R^{obs} = K_1 \frac{(2\beta + 1)}{2} p^2 + \cdots \qquad (20)$$

No first power term appears in this equation and consequently a plot of $(1 - R^{obs})/p$ *vs.* p should go through the origin. That the experimental results do not give a plot going through the origin is shown in Fig. 4. Thus some first power term must be added to equation 20 and it is this term which can be interpreted according to the possibilities mentioned above.

Let us now investigate the possibility that the first power term is due to the contamination of the solution by a large fraction of single strands. The viscosity contribution of a singly stranded molecule will initially decrease an amount which is directly proportional to the number of enzymatic attacks. Therefore, if the justification for the first power term is to be found in the assumption that the solution contains singly as well as doubly stranded molecules, then the coefficient of this term is equal to $f\alpha m/(\alpha+2)$, where f is the fraction of the initial viscosity of the solution which is contributed by the single strands. Making reasonable assumptions for the single strands that $\alpha = 0.8$ and $m = 8000$, we calculate that $f = 0.0057$. The order of magnitude of f is, of course, the information of interest, and we see that the "contamination" of the solution by single strands contributes only a small fraction of the total viscosity of the solution. Hence if the viscosity of the single strand is only 5% of the viscosity of a double strand of the same contour length, then the "contamination" of the solution by single strands is only about 5%. It is important to realize that if one of the other possibilities actually explains the departure from double strand kinetics the actual contamination is less than this value.

In an analogous manner we may examine the possibility that at the site of each enzymatic attack there is created a point of additional flexibility which decreases the viscosity contribution of the molecule. This effect certainly would vary directly with the number of enzymatic attacks during the initial stage of the reaction, and the maximum value for the effect is obtained by attributing the coefficient of the first power term to the additional flexibility alone. In this case, the intercept of $(1 - R^{obs})/p$ plotted against p will be equal to $m\delta$, where δ is the average fractional decrease in the viscosity contribution of the molecule caused by the first enzymatic attack. Assuming a value of 16,000 for m, and 13.3 for the intercept, we calculate that δ is less than one tenth of one per cent. Therefore, the molecules change their initial shapes very little during the first part of the degradation. It is quite interesting to note that such a slight increase in flexibility is capable of explaining the deviations from double strand kinetics.

The next problem to be investigated is whether the enzyme always acts in a random fashion or if it may occasionally break both strands in a simultaneous, non-random attack. Such a non-random attack will cause the molecule to scission, and the number of such scissions will be proportional to the number of enzymatic attacks. The intercept of the $(1 - R^{obs})/p$ against p plot is then equal to yK_1, where y is the fraction of attacks which result in this type of non-random scission. Assuming that $\alpha = 1$ and $m = 16{,}000$, y is calculated to be 0.0024. This means that non-random scissions occur less often than one time in each 400 attacks. This again is a maximum value and would be less if any of the other effects contributed significantly to the decrease in viscosity.

There is another piece of information which this study affords us. So far we have been considering largely the intercept of the $(1 - R^{obs})/p$ against p plot. From the initial slope of this plot the quantity β may be calculated. This quantity is the maximum number of nucleotides which may occur between two breaks in opposite strands and yet have a scission occur at that point on the molecule. From equation 19b, it is seen that in order to calculate β, the constant K_1 must be evaluated. Since α lies somewhere between one-half and two while m is most likely between 10,000 and 20,000, we can calculate minimum and maximum values

(21) The possibility that the enzyme may be partially inhibited in the region of previously formed interruptions is readily considered. In this case, let the percentage inhibition in this region be denoted by the quantity, ω. Neglecting very small terms, it is now possible to show that both terms on the left-hand side of equation 14 are multiplied by ω. This equation is now combined with equation 15a. Again we may solve for the initial slope and the intercept of the $(1 - R^{obs})/p$ against p plot. Two equations analogous to 17 and 17a result. Solving these simultaneously for the number of gaps, g, we find that the quantity ω cancels out of the final expression, and equation 18 is obtained. Hence the number of gaps determined in this manner is theoretically independent of this type of enzymatic inhibition. Moreover, since the quantity ω may be larger than 1, the number of gaps so determined is also independent of an enhancement of enzymatic activity in the region of previously formed interruptions.

for the constant K_1. Combining these values with the slope according to equation 19b we find that the value of β should lie between 0.5 and 5.

The small value for β obtained in this manner is not in agreement with the interpretation given a previous observation[7] that as the degradation products approach a molecular weight of 100,000, the ultraviolet absorption begins to increase and when the molecular weight reaches 50,000 the increase reaches 80% of its final value. This was interpreted to mean that the hydrogen bonding between the bases was not sufficiently strong at room temperature to hold these short chains together. If this explanation is correct, it is possible to show that the value of β must be large. This may be done in the following manner: first, if the molecule is to dissociate into single strands, then

$$2\beta \cong 1/p \qquad (21)$$

Secondly, assuming that the initial molecular weight of the DNA molecules averages 5 million, then after 100 scissions have occurred the molecular weight will drop to 50,000. Thus it follows from equation 11 that

$$\beta p^2 = 1/160 \qquad (22)$$

where the approximation $(2\beta+1) \cong 2\beta$ has been made. Combining equations 21 and 22 gives the large value for β of 40.

In summary, various models for DNA have been examined, and it has been shown that the theory for the kinetics of the degradation of a doubly stranded model essentially fits the experimental data. The correspondence between theory and experiment is not exact, however, since the efficiency of the enzyme in causing a decrease in the viscosity contribution of the macromolecules is greater than the theory predicts. The discrepancy may be explained in a variety of ways. If the preparation of DNA contained some molecules composed of only single strands one would obtain kinetics like those found experimentally, but it can be shown that such DNA molecules, if present at all, could constitute no more than 5% of the total. Alternatively, we might propose that enzyme can attack both strands simultaneously, but such attacks could occur no more than 1 out of every 400 times. This more rapid decrease in viscosity over that expected from the theory could also be attributed to increase in flexibility of the macromolecules. Such increase in flexibility due to each enzymatic attack on the backbone structure could cause a viscosity decrease of no more than one tenth of one per cent. Finally the difference between theory and experiment can be explained if there were pre-existing gaps in the original macromolecules as in the model proposed by Dekker and Schachman. If such gaps were present, the theory shows that there could be no more than one in about 3000 nucleotides. This is far less than the number of gaps suggested by them on the basis mainly of the number of end groups per macromolecule as obtained from titration studies. From the present studies it appears therefore that the DNA molecules are formed of two continuous strands. It should be pointed out, however, that the existence of preformed gaps occurring in regions of the macromolecule which are not susceptible to enzymatic attack could not be detected by the approach outlined above. Finally, it is shown that the minimum number of nucleotides which must be present between two gaps to prevent the molecule from scissioning is no more than six. This is a much lower number than that inferred previously from the changes in optical density as a function of molecular weight.[22]

(22) Since the submission of this manuscript, a paper has appeared by C. A. Thomas, THIS JOURNAL, **78**, 1861 (1956), in which the enzymatic degradation of DNA is followed by the technique of light scattering. Similar conclusions to some of those presented in this paper were obtained by assuming that the distribution of molecular weights in the original sample was Poisson. Too few experimental measurements during the early stages of degradation were made, however, to allow Thomas' data to be analyzed in the manner we have described.

BERKELEY, CALIFORNIA

[CONTRIBUTION FROM THE DEPARTMENT OF CHEMISTRY, PENNSYLVANIA STATE UNIVERSITY]

Acetal and Ketal Hydrolysis Rates in Moderately Concentrated Perchloric Acid Solutions Containing 50% Dioxane[1]

BY MAURICE M. KREEVOY[2]

RECEIVED DECEMBER 2, 1955

The rates of hydrolysis of four acetals and ketals and the ratio of protonated to unprotonated 2-nitro-4-chloroaniline have been measured in moderately concentrated perchloric acid solutions containing 50% by volume of dioxane. The quantities $\log\{k_1/(HClO_4)\}$ and $\log\{(BH^+)/(B)(HClO_4)\}$ were empirically found to be linear functions of the concentration of perchloric acid with the slopes varying only slightly. Some possible implications of these relations are discussed briefly.

In the course of a study[3] of the relative rates of hydrolysis of acetals and ketals it was necessary to study a number of rates as a function of acid concentration in fairly concentrated perchloric acid solutions containing 50% dioxane by volume. (The exact method of making up these solutions is described in the Experimental section. Pure dioxane–water mixtures made up as described contained 49.6% dioxane and 50.4% water by weight.) Because the data bear on the interesting problem of the acidity of strong acid solutions in mixed sol-

(1) The work herein reported was carried out on Project NR055-328 between the Office of Naval Research and the Pennsylvania State University. Reproduction in whole or in part is permitted for any purpose of the United State Government.

(2) Department of Chemistry, University of Minnesota, Minneapolis, Minn.

(3) M. M. Kreevoy and R. W. Taft, Jr., THIS JOURNAL, **77**, 5590 (1955); other papers forthcoming.