REVIEWS OF MODERN PHYSICS VOLUME 31, NUMBER 1 JANUARY, 1959

12
Configurations of Biologically Important Macromolecules in Solution

PAUL DOTY

Department of Chemistry, Harvard University, Cambridge 38, Massachusetts

ALMOST all biologically important macromolecules are particular chemical and structural modifications of three basic, polymeric-chain structures. These polymeric chains are the polysaccharides, the polypeptides, and the polynucleotides. They are better known in their more particular forms; for example, cellulose and amylose in the first instance, fibrous and globular proteins in the second, and deoxyribonucleic and ribonucleic acid in the third.

A polysaccharide chain such as cellulose exists as a completely extended chain in the crystalline, fibrous form and as such has a unique configuration. In solution, however, the limited rotational freedom permitted at the juncture of each repeating unit results in the chain being flexible and, because of internal Brownian motion, it undergoes continuous, worm-like changes in configuration. This state is referred to as a randomly coiled configuration. Such a configuration is devoid of any fixed relationship between pairs of residues, and hence is said to have no secondary structure. Since the fixed configuration in the crystalline form and the randomly coiled configuration in solution are typical of the situation found in most synthetic polymers as well, these have been widely studied and are well understood. Consequently, they are not further dealt with here.

The polypeptides and polynucleotides offer a sharp contrast in configurational properties to the polysaccharides and all other known polymeric chains. This uniqueness lies in the ability of these two types of chains to form hydrogen-bonded helical configurations consisting on one, two, or three chains which are stable in aqueous solution and hence in the cellular environment. Each of these helical configurations is unique and is equivalent to one-dimensional crystallites in that they consist of a periodic arrangement of the repeating-chain units along the helical axis. In this way unique, stereospecific relations in the individual macromolecule can be maintained while the macromolecule itself moves about in solution. The randomly coiled configuration exhibited in solution by all other macromolecules does not have this property. Consequently, it is evident that the preservation of unique configurations by polypeptides and polynucleotides in solution offers a basis of biological specificity.

These unique configurations of individual macromolecules are similar to the crystalline state in several respects, in addition to their having a one-dimensional periodic order. Most important is the implication that they will have a melting point; that is, a temperature will exist at which the supporting hydrogen-bonded structure will undergo a transition to the equivalent of the liquid state. For the macromolecule, the liquid state is simply the randomly coiled configuration already described. Thus, it is not surprising that transitions are found in these macromolecules that can have a sharpness approaching a phase transition. These are generally known as helix-coil transitions, and their study in the present case is of interest because their location reveals the relative stability of the unique configurations upon which one's attention becomes focused.

In the last few years, the study of these helical structures has greatly benefited from the possibility of making pure polypeptides and pure polynucleotides—pure in the sense that the repeating units are identical rather than differentiated as they are in the naturally occurring counterparts, proteins and nucleic acids. As a consequence, the properties of the purely helical forms could be carefully studied and the means of detecting such forms in the naturally occurring materials were thereby greatly sharpened.

It is against this background[1] that some studies of the configurations of polypeptides and their relation to protein structure are examined briefly, and following that a similar look is taken at the corresponding situation in polynucleotides and nucleic acids. The configurations have been established in solution by the use of the methods outlined earlier (Doty, p. 61). In addition, optical-rotatory dispersion and ultraviolet spectroscopy have been found widely useful in detecting the configuration established by the physical methods.

POLYPEPTIDES AND PROTEINS

In 1951, shortly after the proposal of the α-helix by Pauling and Corey,[2,3] Perutz[4] showed by means of x-ray diffraction that a few fibrous proteins in crystalline form did contain this configuration in unspecified amounts. However, it was not possible to extend this method to globular proteins.

In 1953, E. R. Blout and the author initiated a program of synthesis and characterization of synthetic polypeptides: this had as one of its aims the direct testing of whether or not polypeptides could take up unique configurations such as the α-helix in solution. In 1954, we found that poly-γ-benzyl-L-glutamate could exist in two configurations, the α-helix and the solvated, randomly coiled chain, depending on the solvent. Moreover, we found that the two forms showed a substantial difference in specific rotation similar in sign and magnitude

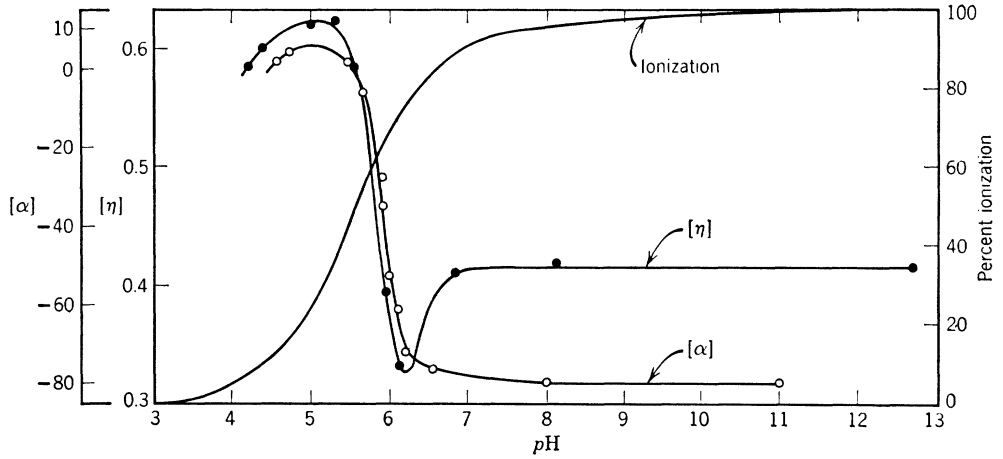

FIG. 1. The helix-coil transition in poly-L-glutamic acid.

to the difference between native and denatured proteins.[5] This fitted nicely with the suggestion made in 1955 by Cohen[6] that a specific main-chain configuration may be the cause of the difference in optical rotation between native and denatured states of proteins.

In 1956, more-detailed evidence was presented in support of the assignment of configuration in solution,[7] and it was shown that a sharp transition could be observed between the two configurations.[8] More to the point, however, was the investigation of the rotatory dispersion (i.e., the wavelength dependence of specific rotation), because this showed that, whereas the dispersion for the randomly coiled form was of the simple (or Drude) type found in most organic compounds containing asymmetric carbon, the dispersion of the α-helical form was anomalous.[8] At the same time, poly-L-glutamic acid was investigated in aqueous solution, and found to be helical in acid solution and randomly coiled in neutral and alkaline solutions.[9,10] This provided additional data on a system that more closely resembled proteins in aqueous solution.

It is perhaps useful to examine the case of poly-L-glutamic acid in some detail. Measurement of its intrinsic viscosity, sedimentation constant, and molecular weight[10] at *p*H values below *p*H 5 (where the carboxyl group was largely non-ionized) showed that it had the rod-like form and mass-to-length ratio of the α-helix. As the *p*H was raised, however, substantial changes in all of the physical properties of this polypeptide were observed[9] (Fig. 1), and the interpretation of these changes indicated that the polypeptide has gone over to a random-coil form. This can be considered as a melting process brought about by the electrostatic repulsion. That is, owing to the ionization of the carboxyl groups resulting in turn from the increase in *p*H, there is imposed upon the original helical structure an outer helix of negative charges; the repulsion between these charges is more than the hydrogen-bonding frame can withstand; and the α-helix structure breaks apart. One can observe this transition by measuring changes in the viscosity, for example, corresponding to the change in the shape of the molecule. However, the most generally useful of all of the changes accompanying this transition is the change in optical rotation. In this case, one observes a change of optical rotation from +4 to −80°C accompanying the helix-coil transition. With this nearly instantaneous means of detecting the configuration, it is easy to show that this is a reversible transition.

It is an old observation that the specific rotation of proteins falls* upon denaturation.[11] During the last decade, much specific support has been given to this generalization by quantitative studies, particularly by Kauzmann.[12] These studies show that the specific rotation, $[\alpha]_D$, for the majority of proteins lies between −30 and −60°C and that these values fall to approximately −100°C upon denaturation. Now, if the values of $[\alpha]_D$ observed for the helix and coil forms of poly-L-glutamic acid are adjusted to the mean residue weight of proteins (151 for sodium glutamate to 115 for proteins), the result is that a value of about +5°C would be expected for a protein in the completely helical form and about −105°C for a protein in the completely random-coil form. Thus, the observed specific rotation of proteins is compatible with 40 to 70% of their residues being in the α-helical configuration, and the specific rotation of denatured proteins corresponds to that of the randomly coiled configuration as previously surmised.[13] Naturally, this proposal requires extensive testing, and much of this is under way or has been done.

Since it is well known that proteins denature upon heating, one would expect the helical form to break up as the temperature is raised. The data in Fig. 2 show that this does occur if one uses the optical-rotation

* In this article, changes in specific rotation are always described with reference to the absolute value, not the magnitude as is customary. This is necessary since we now have both positive and negative values of specific rotation to discuss in contrast to earlier times when the only measured values of proteins were negative.

method to gain a measure of helix content.[9] At *p*H 4.1 and temperatures below 40°C, the optical rotation is nearly constant. Above 40°C, the optical rotation begins to decrease, showing the start of the helix-coil transition. It melts out nearly all of the way at *p*H 4.6. One can pick up the rest of the transition by going to *p*H 5.0 where it melts out completely.

It is possible that the stability of the α-helix in poly-L-glutamic acid resides not so much in the hydrogen-bonding framework of the peptide units as in the pairwise interaction of carboxyl groups. This is the kind of hydrogen bonding that is always pointed out as being quite strong, and, if one makes models of the α-helix configuration of polyglutamic acid, one sees that uncharged glutamic residues can pair very nicely on the surface of the helix. Therefore, it is of great interest to see if other polypeptides that do not have this possibility also show the same phenomena.

Poly-L-lysine has an amino group on the side chain. This amino group is in the charged form ($-NH_3^+$) at *p*H values below 9.5, and becomes uncharged ($-NH_2$) above *p*H 10.5. Thus, one would expect that, if the helix is only marginally stabile, it would exist only at high *p*H values and would melt out as one lowered the *p*H. Current experiments show just this.[14] Again, there is about a 90°C change in optical rotation upon passing through the helix-coil transition. Physicochemical investigations were used to show that it exists as a pure α-helix at high *p*H.

If one examines the state of charge of the poly-L-lysine molecule when the helix begins to melt out in a quantitative fashion, one sees that this helix is considerably weaker than the poly-L-glutamic-acid helix. However, the difference is not enough to prevent the poly-L-lysine helix from being stable in aqueous solution. From this and a few other examples, it can be concluded that the α-helical configuration is stable in aqueous solution in the absence of pronounced electrostatic repulsions or other disrupting influences.

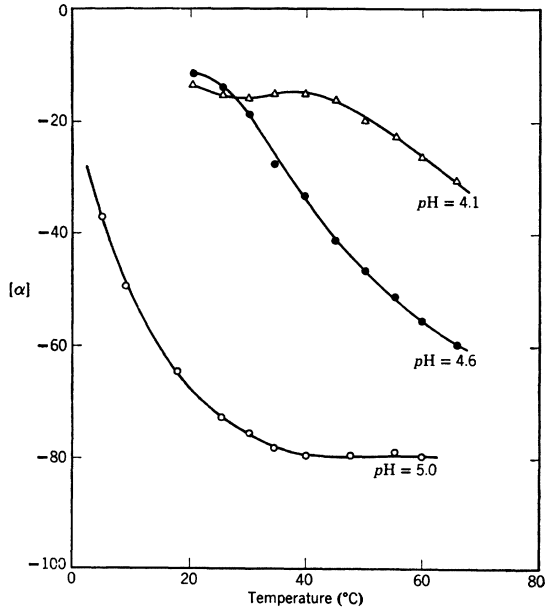

FIG. 2. Optical rotation of poly-L-glutamic acid as a function of temperature as several values of *p*H.

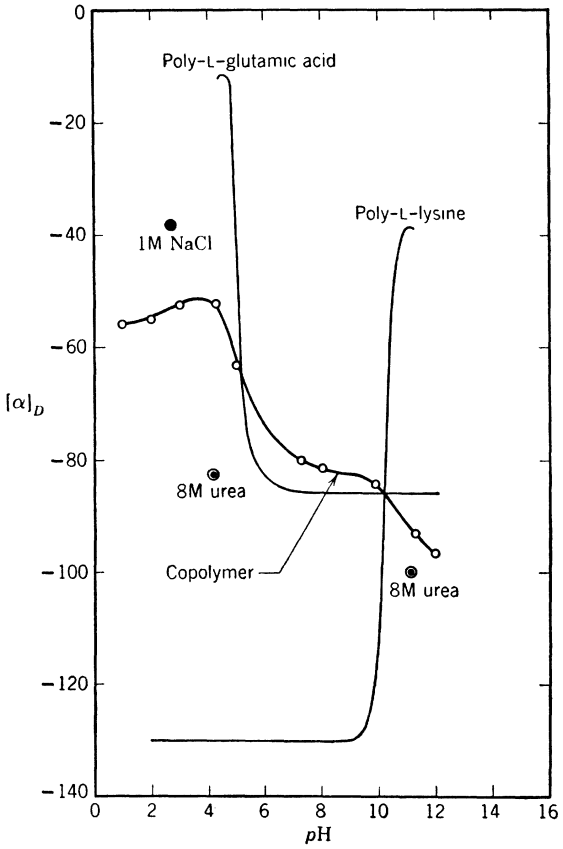

FIG. 3. The specific rotation as a function of *p*H for copoly-L-lysine-L-glutamic acid.

An interesting point associated with the nature of these helix-coil transitions is the question of whether or not they have an "all-or-none" character. That is, at the midpoint of the transition are half of the molecules in each configuration or are the residues in each molecule partitioned between these two forms? One of several answers indicating that the latter was the case was obtained in the following way. The intrinsic viscosity increases much more strongly with molecular weight in the helical form than in the random-coiled form. Thus, there is one molecular weight for which the viscosity is the same in both configurations. By choosing a sample of this molecular weight and carrying it through the transition, the intrinsic viscosity would be expected to remain unchanged if the transition were of the all-or-none type. This was not observed in several cases.[10,14,15] Consequently, the transition must be viewed as one that proceeds via intermediate states in which individual molecules have interspersed helical and nonhelical regions.

It is of some interest to consider the behavior of a copolymer composed of equal amounts of L-glutamic acid and L-lysine. The optical rotation, which again is taken as a measure of helix constant for this copolymer, is shown in Fig. 3.[16] These results can be compared with those previously reported for poly-L-glutamic acid and poly-L-lysine which are also shown. The copolymer exhibits an intermediate behavior here; that is, it is about 50% helical at acid *p*H, and not at all helical at alkaline *p*H. This is interpreted as resulting from the difference in intrinsic stability of the glutamic-acid and lysine residues in supporting the helical structure. When the lysine residues are uncharged, there appears to be no helix, because the lysine residues cannot compensate for the electrical repulsion arising from the 50% negative charges that exist here owing to the ionized glutamic-acid residues. At neutral *p*H, one finds about 15 or 20% helix.

ROTATORY DISPERSION AND HELICAL CONTENT OF POLYPEPTIDES AND PROTEINS

Although one could proceed to estimate the fraction of residues (helical content) in the helical configuration from the specific rotation of proteins in the manner indicated in the foregoing, one would be failing to take advantage of another aspect of the situation that, upon proper analysis, offers an independent means of estimating the helical content and in addition offers a much more perceptive view of the problem. Soon after the first observations of the helix-coil transition in polypeptides, the rotatory dispersion (wavelength dependence of the specific rotation) of the two forms was measured[8] and found to be simple Drude dispersion for the random-coil form and a complex dispersion for the helical form. The complex dispersion could be fitted in a manner expected for coupled oscillators and originally suggested by Moffitt.[17] This gave rise to a very considerable theoretical interest in the problem of the rotatory dispersion of helical macromolecules,[18–22] but it has not been possible to take advantage of much of this because of the limited number of parameters that can be uniquely determined from the experimental data.

From the practical point of view, it is sufficient to say that the following equation was found adequate to express the observed data[23–25]:

$$[m'] = \frac{3}{n^2+2}\frac{M_0}{100}[\alpha] = (a_0^R + a_0^H)\left(\frac{\lambda_0^2}{\lambda^2-\lambda_0^2}\right) + b_0\left(\frac{\lambda_0^2}{\lambda^2-\lambda_0^2}\right). \quad (1)$$

Here, the specific rotation is multiplied by two factors that eliminate the trivial effect of residue weight, M_0, and refractive index, n. The corrected specific rotation, called the effective residue rotation, $[m']$, is then assumed to be made up of two components. The first of these is a Drude term but with two parts to its coefficient, one depending on the intrinsic residue rotation (asymmetric carbon), a_0^R, and the second is owing to the effect of the helical configuration, a_0^H. The second term is the anomalous one, being the square of a Drude term, and its coefficient is wholly owing to the effects of the helical configuration on the rotatory dispersion.

This equation is of the same form as that proposed by Moffitt and shown to be applicable to the first dispersion data obtained on helical polypeptides,[8,9,20] but it now seems proper to consider it an empirical proposal for two reasons. The equation itself is of the form expected for coupled oscillators. Secondly, Moffitt concentrated his attention on the second term, recognizing the complexities that may be associated with the first since, in any rigorous view, the first term must instead be replaced by several Drude terms each having a different λ_0. But if this last point is admitted, then the analysis of any available dispersion data is impossible, because there is not enough precision to allow the independent assignment of values to more than the three constants contained in Eq. (1), i.e., $(a_0^R+a_0^H)$, b_0, and λ_0. Consequently, in order to proceed, it was necessary to assume that λ_0 had the same value in both terms. This compromise, added to the later finding that the absolute value of b_0 cannot be computed with enough accuracy to be useful, forces us to accept Eq. (1) as empirical. It is important to make this point because, when it is recognized that Eq. (1) is not firmly supported by theory, it is more likely that it will be used with the caution that an empirical relation deserves.

The Fitts-Kirkwood theory[18,19] cannot be expressed in simple functional form, but since it consists of four different contributions, it is clear that these cannot possibly be resolved from experimental dispersion data since again too many parameters would have to be independently evaluated. However, if one takes the computed results of Fitts and Kirkwood,[22] one finds that they can be quite well fitted by Eq. (1). This observation emphasizes that the procedure that we have adopted in using Eq. (1) is not in conflict with the Fitts-Kirkwood theory.

Examine the adequacy of Eq. (1) in representing the dispersion data of polypeptides in the helical form. By dividing Eq. (1) by $\lambda_0^2/(\lambda^2-\lambda_0^2)$, it is clear that dispersion data obeying this relation should yield linear plots when $[m'](\lambda^2-\lambda_0^2)/\lambda_0^2$ is plotted against $\lambda_0^2/(\lambda^2-\lambda_0^2)$, provided the appropriate value of λ_0 has been selected. In this plot, the intercept of the straight line would then provide the value of $(a_0^R+a_0^H)$, which may be denoted a_0 for convenience, and the slope would provide the value of b_0. Since in this type of plot an empirical relation is being tested, it will be necessary to have specific rotation measurements at a number of wavelengths, because the absence of curvature in the plot must be demonstrated in each application.

In selecting the appropriate value of λ_0, a process of trial and error must be followed until that particular

value which linearizes the dispersion data is found. In the fitting of the first dispersion data, a value of 212 mμ was found for poly-γ-benzyl-L-glutamate and poly-L-glutamic acid. Since that time, the value has been confirmed on both of these polypeptides[10,26] and found to apply to at least four other helical polypeptides, as well as to several copolypeptides.[27] Dispersion data plotted in this manner are shown in Fig. 4 for three samples of poly-γ-benzyl-L-glutamate.[26] One sample had an average chain length of 4 units and gave a horizontal plot indicative of no helical content. Another had an average chain length of 10 residues and gave an intermediate slope and intercept, which, in comparison with the high molecular-weight sample, indicates about 60% helical content.

In all of the cases referred to in the foregoing (poly-γ-benzyl-L-glutamate, poly-L-glutamic acid, poly-carbobenzoxy-L-lysine, poly-L-lysine, and the copoly-L-lysine-L-glutamic acid), the values of b_0 obtained from the slopes of what one might call the "coupled oscillator" plot have clustered about -630. Thus, there is good reason to accept this value with an uncertainty of no more than $\pm 10\%$ as a constant characterizing in an empirical way the anomalous dispersion of helical polypeptides of L-residues. In some cases, we have found somewhat lower values, but these have in each case been with polypeptides that had not been proved to be in the completely helical form.

Turning finally to the coefficient a_0, we do not expect to find constant values for different polypeptides, because a_0 consists of two components, one of which (a_0^R) depends on the intrinsic rotatory power of the individual amino-acid residue. It is sufficient to say that, in the cases of the five polypeptides mentioned in the previous paragraph, the values of a_0 gave an average of zero. Since in the helical form a_0 is found to equal 650, this must be approximately the value of a_0^H itself.

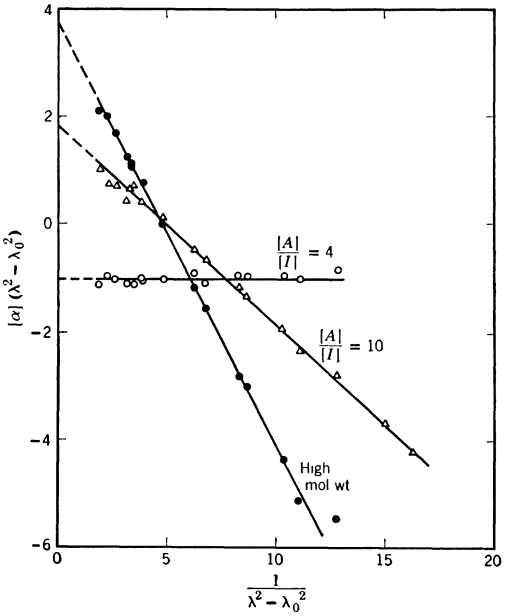

FIG. 4. Coupled oscillator plot of rotatory-dispersion data for poly-γ-benzyl-L-glutamate of various molecular weights. $[A]/[I]$ indicates degree of polymerization (number average).

One is now in a better position to examine the extent to which the work on polypeptides enables one to understand the optical activity and rotatory dispersion of proteins. Actually, one is forced to proceed in a very simple manner. The polypeptide studies have given quantitative information about the rotatory dispersion of two configurations of polypeptide chains. One can, therefore, only ask if the rotatory dispersion of proteins can be accounted for by a linear combination of these two characteristic rotatory dispersions. It is important to recognize that, if this is the case, the residues not occurring in the α-helical configuration need not be in a randomly coiled configuration. It is necessary only that they not be in any kind of periodic arrangement, because this is the condition that the rotatory dispersion be of the kind observed for the randomly coiled configuration, subject of course to "solvent effects" on the intrinsic residue rotation, i.e., a_0^R.

The data summarized in terms of a_0^H and b_0 values can be translated directly into the contribution that the α-helical configuration would make to $[\alpha]_D$. It is $+117°$ for the values of $a_0^H=650$, and $b_0=-630$ if the mean residue weight is taken as 115. The value of a_0^H contributes 84% to this result and, hence, it is the uncertainty of this term that affects most the predicted value. Taking the limits of a_0^H to be 550 and 750 would make the helix contribution to $[\alpha]_D$ 17° higher or lower, respectively. Thus, the prediction is made that for proteins the purely helical configuration and the completely denatured form should be separated by 100° to 135° in specific rotation. The polypeptide studies, adjusted to mean residue weights on 115 and aqueous solutions, establish the lower end of this increment—that is, the denatured protein—at about 110° and the high end—the pure helix—at about +5°. This is nearly the same as that obtained earlier from the measurements on poly-L-glutamic acid.

In terms of the model proposed here, the rotatory dispersion of proteins in aqueous solution should be given by converting Eq. (1) into the following form[28]:

$$[\alpha]=1.39\left[(\Sigma\, a_0^R+fa_0^H)\left(\frac{\lambda^2}{\lambda^2-\lambda_0^2}\right)+fb_0\left(\frac{\lambda^2}{\lambda^2-\lambda_0^2}\right)\right],$$

where f denotes the fraction of residues in helical form and a_0^R the sum over the a_0^R values characteristic of each residue in the protein. The model is not subject to a considerably more rigorous test than in the comparison dealing only with $[\alpha]_D$. This can be carried out in the following way. Rotatory-dispersion measurements are made on the protein both in aqueous solution and in the completely denatured state. For the former set of data, $[\alpha]1.39(\lambda-\lambda)$ is plotted against $(\lambda-\lambda)^{-1}$. The result should be a straight line yielding a slope equal to fb_0

TABLE I. Excess right-handed helical contents of (f) of various proteins in water.

	$-b_0/630$	$a_0^H/650$
Tropomyosin	0.88	0.87
Insulin	0.38	0.57
Bovine serum albumin	0.46	0.58
Ovalbumin	0.31	0.50
Lysozyme	0.29	0.39
Pepsin	0.31	0.26
Histone	0.20	0.30
Ribonuclease	0.16	0.17
Globin (H)	0.15	0.09

and an intercept equal to ($\sum a_0^R + f a_0^H$). With the value of b_0 set at -630, the value of f, the fraction of residues in the helical configuration, is obtained at once.

The rotatory dispersion in the denaturated state should yield at once the value of a_0^R, since $f=0$ in this case. Indeed, only a measurement of the specific rotation at one wavelength is needed if it is known that $\lambda_0=212$ mμ, as appears to be the case in $8M$ urea. With a_0^R known, the value of $f a_0^H$ can be obtained from the intercept evaluated in the plot described above. By taking the value of a_0^H as $+650$, an independent estimate of f can be made.

This kind of analysis has been applied to a number of proteins.[28] In each case, a linear plot was obtained. The test of the model consists then in seeing if the two different estimates of f are reasonable, that is, lie between 0 and 1, and are in agreement. A selection of this data is shown in Table I. It is seen that the values of f are indeed reasonable ones and that the agreement is fairly good.

The reasonably good performance of our model in this test is not, of course, proof that the model is a faithful description of the secondary structure of the proteins to which it has been applied. However, it does appear to represent an advance which deserves further and more-incisive testing. At present, it greatly increases our confidence in the existence of regions having the α-helical structure in proteins, it sharpens our views on protein denaturation, and it provides a framework of reference in which studies relating protein structure and function can be at least tentatively interpreted.

Having observed the striking dependence of the configuration of polypeptides on solvent, we were led to wonder if the intermediate development of the helical structure of proteins suggested by the foregoing experiments could not be increased by altering the solvent. For this, a solvent that was miscible with water, of comparable polarity and cohesive energy density, but with less hydrogen-bonding capacity, was needed. Our search indicated that 2-chloroethanol was well suited, and it was found that the addition of this to aqueous solutions increased the helical content as measured by rotatory dispersion in nearly every case. Some examples are shown in Table II for proteins dissolved in chloroethanol.[28]

Thus, the expectation that the helical content of proteins can be increased by lowering the hydrogen-bonding capacity of the solvent is confirmed. This behavior is precisely the opposite of denaturation. In the few cases that we have studied in detail (e.g., ribonuclease and insulin), the configurational changes on going from water to chloroethanol and back to water have been completely reversible. Thus, it would appear that the helical content of proteins is a result of balance between the intramolecular hydrogen bonding of the protein and the hydrogen-bonding capacity of the solvent. When water is the solvent, the point of balance is in many cases near 50%, and can be shifted in either direction by proper alteration of the solvent.

It is of little interest to note that insulin was the only protein in the list whose helical content was not substantially increased in chloroethanol over water. Since the model of insulin proposed by Lindley has about 20% of the residues in the left-handed helical configuration, owing to restrictions imposed by the cystine bridges, it appears quite possible that insulin has approximately this configuration both in aqueous and in chloroethanol solutions. The residues in the left-handed helix cancel the effect of an equal number in the right and thus, even when the helical content is nearly 100%, only about half of this amount registers with the rotatory-dispersion method. This anomaly is removed when it is recognized that the method measures only the excess right-handed helical content, as indicated in the titles of the tables. Nevertheless, this is the sort of detail that emerges only on making additional measurements. Obviously, there is now a great need of a completely independent method to measure total helical content so that the rotatory-dispersion methods can be checked in regularly behaving cases, and the amount of left-handed helical configuration evaluated in those proteins where cystine bridges may prevent the normal development of right-handed helical configurations.

As a final illustration of the relation between rotatory dispersion and protein structure, no better example can be quoted than the current work of Kay and Bailey on *Pinna* tropomyosin.[29] Kay's study of this molecule by physical methods shows it to have the character of thin, rigid rods having the dimensions of an α-helix of the proper molecular weight. Rotatory-dispersion measurements have now shown it to behave precisely like an α-helix with values of a_0^H and b_0 equal to those used

TABLE II. Excess right-handed helical contents of various proteins in chloroethanol.

Tropomyosin	110
Insulin	45
Bovine serum albumin	75
Ovalbumin	85
Lysozyme	63
Pepsin	44
Histone	72
Ribonuclease	67
Globin (H)	74

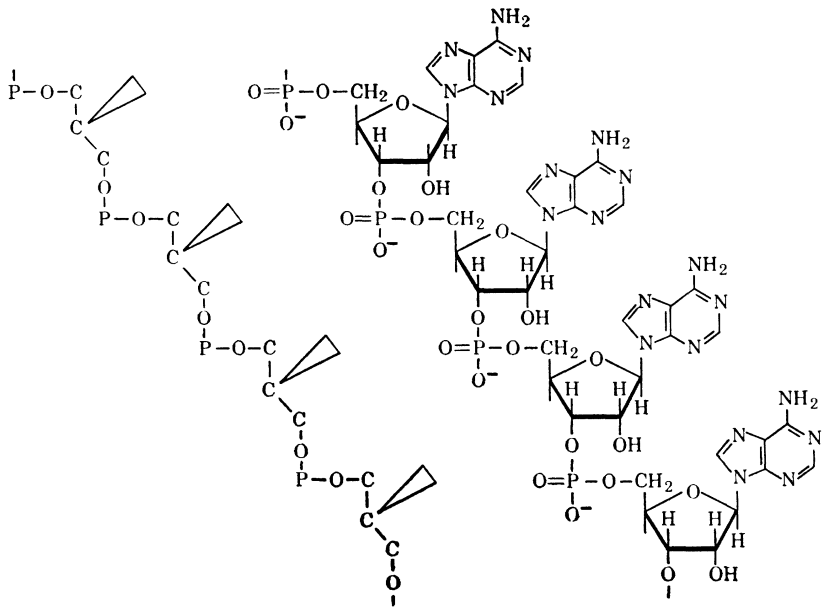

FIG. 5. The chemical structure of polyriboadenylic acid.

here.[30] Consequently, one has here the perfect example of the extreme of the scale of protein structure provided by our model. However, the work that is needed to establish this model as a framework of reference for protein structure in general remains to be done for the most part. It is certainly to be expected that numerous exceptions will be found, but there is even now reason to hope that the concept of describing the secondary structure of proteins by a partitioning of the residues between helical and nonhelical regions will play a useful role in the advance of knowledge of protein structure.

POLYNUCLEOTIDES AND NUCLEIC ACIDS

From the viewpoint of chemical structure, nucleic acids are known to consist of chains of repeating units such as are shown in Fig. 5. The backbone is a phosphoester polymer with six chain atoms per repeating unit. The chain shown in Fig. 5 is of the ribonucleic-acid type. That of deoxyribonucleic acid differs in that the oxygen at the 2-position on the ribose ring is absent. Generally, four different monomeric units, differing in the heterocyclic rings attached to the ribose group, are found in a given nucleic acid. In ribonucleic acid (RNA), these four groups are adenine, uracil, guanine, and cytosine. In deoxyribonucleic acid (DNA), thymine (5-methyl uracil) replaces uracil. DNA is found in cell nuclei as the principal component of chromosomes, and plays the central role in carrying and passing on the genetic endowment of the cell. RNA occurs principally at the sites of protein synthesis in the cytoplasm, and is intimately involved with that process.

Deoxyribonucleic Acid

The nucleic-acid chain is a very flexible one and, in the absence of secondary structure, it would have the configuration of a random coil. About 10 years ago, it became evident that this could certainly not be the case for DNA. With increasing refinement in the study of DNA in solution, it became clear that this molecule was the most extended ever examined.[31] Light-scattering studies and, more recently, viscosity and sedimentation measurements, show that typical DNA samples have average molecular weights in the range of 5 to 10 million and occupy volumes in solution about one-half micron in diameter. Ordinarily, polyelectrolytes are characterized by their molecular size being very dependent on ionic strength. This was not the case for DNA. Thus, it appeared to be not only very greatly extended, but stiff rather than flexible, as well.

These observations, as well as a number of others, became understandable with the structural proposal made by Watson and Crick[32] in 1953 on the basis of the x-ray diffraction studies made by Wilkins and his collaborators.[33] This structure consisted of two antiparallel DNA chains united through hydrogen bonds connecting the heterocyclic rings (usually called bases). Only two types of pairings were permitted, those between adenine and thymine and those between guanine and cytosine as shown in Fig. 6. The resulting structure is a two-stranded helix about 20 A in diameter. With two residues every 3.4 A, a DNA molecule with a molecular weight of 10 million would have a length of 50 000 A (5 μ). Such a long, thin structure would have a slight flexibility, presumably enough to account for several gentle folds that would reduce its maximum extent by about tenfold to agree with the molecular size found in solution. (Technically, it is correct to speak of this molecule as randomly coiled, but the degree of coiling is minute as compared with the single-chain coils to which this description is usually applied.)

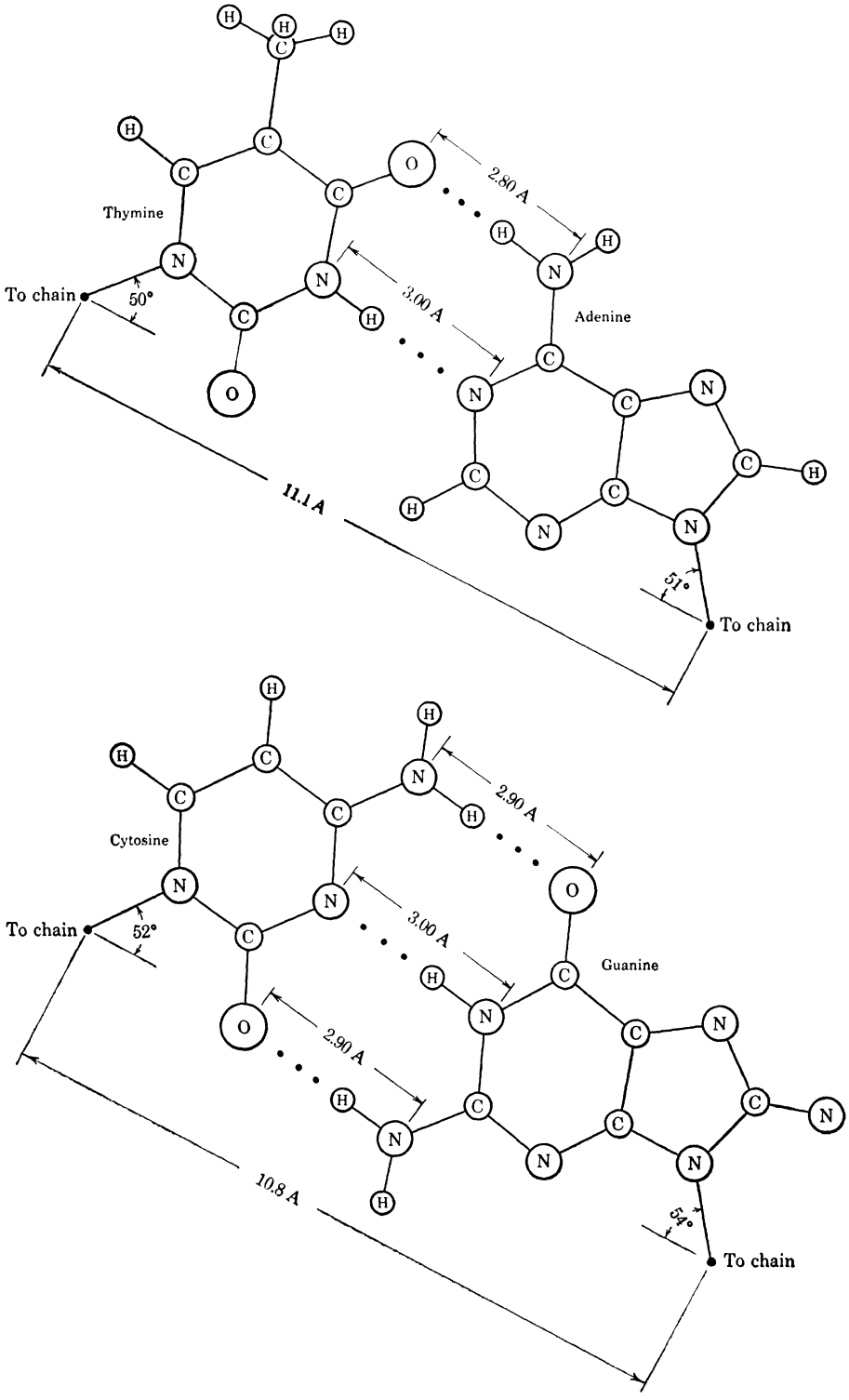

FIG. 6. Detail of the hydrogen-bonded base pairs occurring in DNA.

This periodic, secondary structure undergoes a cooperative melting out upon raising the temperature. The transition can be seen in terms of the viscosity, since the collapse of the structure is accompanied by a twenty-fold decrease in intrinsic viscosity. This is illustrated in Fig. 7.[34] The transition in this case is essentially irreversible. The reason for this is the low probability of the bases to find their original partners: without this the

perfect arrangement in the original DNA cannot be recovered. This helix-coil transition can be induced by other means, for example by raising or lowering the *p*H, and it can be followed by other means, such as changes in optical rotation. However, the most sensitive indicator of structure has turned out to be the ultraviolet spectrum. DNA exhibits a broad maximum at about 2600 A, and it has been widely observed that the maximum of this absorption is substantially depressed relative to that found for the corresponding monomeric units (nucleotides). Thus, hydrolysis which brings about such a conversion from DNA in the helical form to individual nucleotides is accompanied by approximately a 40% increase in the optical density or extinction at 2600 A. This suppression of absorbance in the DNA is known as hypochromicity. Now, it is found also that, if the helical form is converted to the randomly coiled, denatured form by raising the temperature or by lowering the *p*H, a rise of nearly 40% in the extinction coefficient also occurs. Thus, electronic states of the base groups are substantially different, depending on whether or not they are held in a hydrogen-bonded, helically arranged configuration. It is not yet clear if the hypochromicity is a result of the hydrogen bonding or of the stacking of the base groups one on top of another.

Polynucleotides

In 1955, Ochoa and Manago discovered a new enzyme which could bring about the polymerization of nucleoside diphosphates to form polymers having the chemical structure of ribonucleic acid. The study of pure polynucleotides which this has made possible has greatly increased knowledge of the basic properties which are combined in naturally occurring RNA itself. The problem that can be directly attacked with these new polyribonucleotides is the following. In 1956, Donohue[35] showed that the two specific base pairs used in the Watson-Crick structure of DNA are only two out of a large number of possible pairs. Thus, the availability of these polymers has made possible the investigation of which pairs are actually stable in aqueous solution. Warner[36] and Rich[37] quickly showed that polyadenylic acid (Poly A) and polyurdylic acid (Poly U) combined to form a double-stranded helix when their solutions were mixed. A number of other pairs have now been found, and in addition several triple-stranded helical complexes have been demonstrated. These are discussed elsewhere by Rich (p. 191).

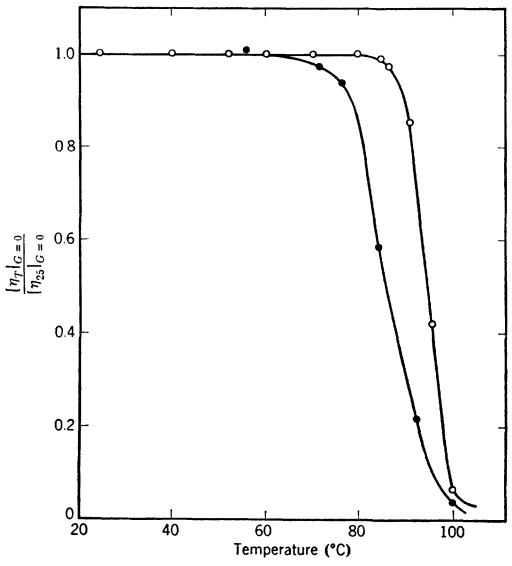

FIG. 7. The intrinsic viscosity (measured at 25°C) of DNA as a function of the temperature to which it has been heated for one hour.

Fresco, Klemperer, and the author have been concerned with self-pairing, as typically exhibited by Poly A.[38,39] We first investigated the sedimentation and intrinsic viscosity of a series of such polymers of different molecular weight at neutral *p*H. The molecular-weight dependence of these properties was found to be given by the molecular weight to the 0.45 and 0.65 power, respectively. These are self-consistent values for random coils. The absence of a significant birefringence of flow and the rise in viscosity that was observed when the ionic strength was lowered forced us to conclude that the configuration was that of randomly coiled, single chains. However, upon lowering the *p*H, we noticed a sharp transition from one type of ultraviolet adsorption to another. The titration curve showed a similar abruptness that could occur only if some cooperative transition were taking place. At *p*H's below this transition, the solutions showed very marked negative birefringence of flow. Moreover, the molecular weights were found to be greatly increased.

These observations clearly indicated that a cooperative association was taking place, upon lowering the *p*H through a critical value. The nature of this was further clarified in the following way. Neutral solutions of a single polymer were made up at a series of different concentrations. When these were acidified and the sedimentation and viscosities determined, the results showed the regular behavior that can be seen on the right-hand side of Fig. 8. It is seen that the molecular weights, or rather particle weights, span a twentyfold range and that the sedimentation constants and viscosities vary in the manner expected for homologous polymers. The weight and size increase with the concentration at which they were formed. The respective slopes are again self-consistent and their numerical values, 0.36 and 0.92, are indicative of a more-extended chain structure than was present in the neutral solutions. The results already mentioned at neutral *p*H are seen at the left of Fig. 8.

This evidence on the acid-stable form indicates fairly clearly that the polyadenylic-acid molecules have joined together in a regular structure. We were then able to show that the amino group of adenine is unavailable for reaction with formaldehyde in the acid-stable complex,

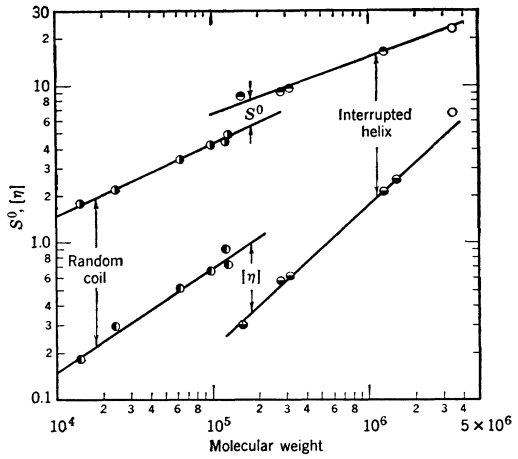

FIG. 8. The molecular-weight dependence of sedimentation constant and intrinsic viscosity for the two forms of polyadenylic acid.

but that it is highly reactive in the randomly coiled state. This showed that the pairing of the adenine groups was responsible for the association. When a model is built to satisfy this type of pairing, it is seen that it must resemble the double-helical configuration of DNA with the adenine bases nearly perpendicular to the axis. The structure requires that bases be 3.8 A apart. To test this point, J. Fresco[38] has taken x-ray photographs of solutions of Poly A (1 to 10%) above and below the transition. At low *p*H, he finds several very sharp rings, including one with a spacing of 3.8 A. At neutral *p*H, there are none. From all of this information, we conclude, therefore, that polyadenylic-acid molecules react as shown in Fig. 9. For the helix to form, the negative charge owing to the phosphate groups must be at least half-neutralized by the uptake of protons.

This behavior of Poly A is similar in many ways to other pairwise interactions that have been studied. Thus, we have nearly reached the point where all possible pairwise interactions have been cataloged. At present, it appears that, of the ten possible combinations between the four pure nucleotide polymers, 6 or 7 actually exist. The next problem is to list these pairwise interactions in order of their strength in analogy with the tables of bond energies one has in the case of covalent bonds. This information can be obtained by a careful study of the temperature at which the respective helix-coil transitions take place.

Ribonucleic Acid

Although DNA and RNA are so very similar in their chemical structure, their configurations are strikingly different. This can be said even though the configuration of RNA is only now beginning to become clear. I conclude by summarizing some investigations that have been made in this laboratory on RNA from calf-liver microsomal particles by Hall[40,41] and on RNA from tobacco mosaic virus by Boedtker.[42]

All examinations of RNA solutions have indicated that the intrinsic viscosity of such solutions is suprisingly low when compared with the molecular weight. This is quite the opposite of the case of DNA. Furthermore, RNA shows only a small birefringence of flow, and this is of the opposite sign of DNA. Thus, it is certain that the configuration of RNA in solution is not that of the extended double-stranded helix found in DNA.

An examination of the dependence of sedimentation and intrinsic viscosity on molecular weight shows that it is very close to the 0.5 power in both cases. This can be unambiguously interpreted as showing that the molecules are highly coiled and that the coils are contracted somewhat in comparison with the usual polymer chain. This result is surprising because the RNA chain carries negative charges on each repeating unit and, as a consequence, at the ionic strength (0.01 *M*) where these observations were made, it would be expected that the electrostatic repulsions would result in a very highly swollen polymer coil. If this had been the case, the intrinsic viscosities would have been much larger and the molecular-weight dependences far different. These considerations clearly suggest that there are very substantial intramolecular attractions which hold the molecules in contracted configurations.

The next problem lies in the identification of these points of internal attraction. That they indeed could be broken, and reversibly formed again, was shown in two ways.[40,41] By removing the salt from the solution, the viscosity was found to rise indicting that, by increasing the electrostatic repulsion sufficiently, the intramolecular bonds could be broken, allowing the coil to expand in the manner expected for a polyelectrolyte.

Similarly, it was found that the viscosity would rise with the temperature of the RNA solution. This again indicated the breakup of the intramolecular bonding.

It is known that RNA is hypochromic relative to its hydrolysate. Consequently, it was of interest to see if the optical density increased upon removing the salt or upon heating. Such was indeed found to be the case. A quantitative study showed that the increase in extinction was parallel with the increase in viscosity. Thus,

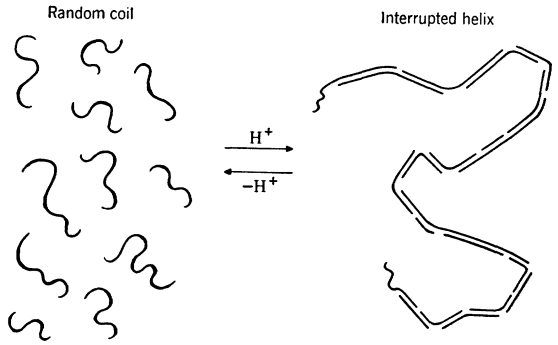

FIG. 9. Schematic illustration of the helix-coil transition in polyadenylic acid.

the base pairing through hydrogen bonds appears to be the cause of the contracted state of the RNA.

By referring to the extinction-coefficient changes accompanying helix-coil transitions in the study of polynucleotides, it is possible to estimate the fraction of base pairs that exist in RNA solutions at moderate ionic strength and at room temperature. This comes out to be the surprisingly large figure of about 50%. At present, this conclusion can only be tentative, but it does raise a very interesting question. The studies of polynucleotide interactions reveal the large number of different types of base pairing that are possible among the bases occurring in RNA. We now have the strong indication that such pairing does occur to a great extent. Obviously, the next problem to face is whether or not such extensive base pairing occurs in a random fashion, or whether there is sufficient organization in the RNA structure to justify describing it as secondary structure.

Finally, it is important to recognize that RNA occurs in the cytoplasm, not as a discrete substance, but in intimate association with protein in the microsomal particles. Unlike DNA, which can of itself be biologically active (as revealed in bacterial transformation), RNA performs its function as a complex with protein, and it is to the analysis of the configuration of both components in these particles that much future effort is certain to be directed.

BIBLIOGRAPHY

[1] A more detailed exposition may be found in P. Doty, Proc. Natl. Acad. Sci. U. S. **42**, 791 (1956).

[2] L. Pauling, R. B. Corey, and H. R. Branson, Proc. Natl. Acad. Sci. U. S. **37**, 205 (1951).

[3] L. Pauling and R. B. Corey, Proc. Natl. Acad. Sci. U. S. **37**, 235 (1951).

[4] M. F. Perutz, Nature **167**, 1053 (1951).

[5] P. Doty, A. M. Holtzer, J. H. Bradbury, and E. R. Blout, J. Am. Chem. Soc. **76**, 4493 (1954).

[6] C. Cohen, Nature **175**, 129 (1955).

[7] P. Doty, J. H. Bradbury, and A. M. Holtzer, J. Am. Chem. Soc. **78**, 947 (1956).

[8] J. T. Yang and P. Doty, J. Am. Chem. Soc. **78**, 498 (1956).

[9] P. Doty, A. Wada, J. T. Yang, and E. R. Blout, J. Polymer Sci. **23**, 851 (1957).

[10] A. Wada and P. Doty, J. Am. Chem. Soc. (to be published).

[11] See, e.g., P. Doty and E. P. Geiduschek in *The Proteins*, H. Neurath and K. Bailey, editors (Academic Press, Inc., New York, 1953), Vol. I, p. 393.

[12] See, e.g., R. B. Simpson and W. Kauzmann, J. Am. Chem. Soc. **75**, 5139 (1953).

[13] J. T. Yang and P. Doty, J. Am. Chem. Soc. **79**, 761 (1957).

[14] J. Applequist and P. Doty, Abstracts, Am. Chem. Soc. Meeting, San Francisco (April 7–11, 1958).

[15] J. T. Yang and P. Doty (unpublished results).

[16] P. Doty, K. Imahori, and E. Klemperer, Proc. Natl. Acad. Sci. U. S. **44**, 424 (1958).

[17] W. Moffitt, J. Chem. Phys. **25**, 467 (1956).

[18] D. D. Fitts and J. G. Kirkwood, Proc. Natl. Acad. Sci. U. S. **42**, 33 (1956).

[19] D. D. Fitts and J. G. Kirkwood, J. Am. Chem. Soc. **78**, 2650 (1956).

[20] W. Moffitt and J. T. Yang, Proc. Natl. Acad. Sci. U. S. **42**, 596 (1956).

[21] W. Moffitt, D. D. Fitts, and J. G. Kirkwood, Proc. Natl. Acad. Sci. U. S. **43**, 723 (1957).

[22] D. D. Fitts and J. G. Kirkwood, Proc. Natl. Acad. Sci. U. S. **43**, 1046 (1957).

[23] K. Imahori, E. Klemperer, and P. Doty, Abstracts, Am. Chem. Soc. Meeting, Miami, Florida (April 1–7, 1957).

[24] K. Imahori, E. Klemperer, and P. Doty, "Estimate of α-helical content of proteins," J. Am. Chem. Soc. (to be published).

[25] See W. Kauzmann [Ann. Rev. Phys. Chem. **8**, 413 (1957), reference 3], for a demonstration of the existence of a square term for coupled oscillators.

[26] J. C. Mitchell, A. E. Woodward, and P. Doty, J. Am. Chem. Soc. **79**, 3955 (1957).

[27] K. Imahori, P. Doty, and E. Blout (in preparation).

[28] K. Imahori, E. Klemperer, and P. Doty (in preparation).

[29] C. M. Kay and K. Bailey, Biochem. et. Biophys. Acta (to be published).

[30] C. M. Kay, Biochem. et Biophys. Acta **27**, 469 (1958).

[31] For a review, see for example: P. Doty, J. Cellular Comp. Physiol. **49**, Suppl. 1, 27 (1957).

[32] F. H. C. Crick and J. D. Watson, Proc. Roy. Soc. (London) **A233**, 80 (1954).

[33] M. H. F. Wilkins, W. E. Seeds, A. R. Stokes, and H. R. Wilson, Nature **172**, 759 (1953).

[34] S. A. Rice and P. Doty, J. Am. Chem. Soc. **79**, 3937 (1957).

[35] J. Donohue, Proc. Natl. Acad. Sci. U. S. **42**, 60 (1956).

[36] R. C. Warner, Ann. N. Y. Acad. Sci. **69**, 314 (1957).

[37] R. Rich, *The Chemical Basis of Heredity*, W. D. McElroy and B. Glass, editors (The Johns Hopkins Press, Baltimore, Maryland, 1957), p. 557.

[38] J. R. Fresco and P. Doty, J. Am. Chem. Soc. **79**, 3928 (1957).

[39] J. R. Fresco and P. Doty, "Polynucleotides II: The x-ray diffraction patterns of solutions of the coil and helical forms of polyriboadenylic acid," J. Molecular Biol. (to be published).

[40] B. Hall and P. Doty, "The configurational properties of ribonucleic acid isolated from microsomal particles of calf liver," *Symposium on Microsomal Particles*, American Biophysical Society (Pergamon Press, London, to be published).

[41] B. Hall and P. Doty, "The preparation and physical chemical properties of ribonucleic acid from microsomal particles," J. Molecular Biol. (to be published).

[42] H. Boedtker, "Some physical properties of infective ribonucleic acid isolated from tobacco mosaic virus," Biochem. et Biophys. Acta (to be published).