U.S.A. 66, 651.
Elliot, G. E., Lowy, J., & Worthington, C. R. (1963) *J. Mol. Biol. 6,* 295.
Habeeb, A. T. S. A. (1966) *Anal. Biochem. 14,* 328.
Hunter, M. J., & Ludwig, M. L. (1962) *J. Am. Chem. Soc. 84,* 3491.
Huxley, H. E. (1969) *Science 164,* 1356.
Huxley, H. E., & Brown, W. (1967) *J. Mol. Biol. 30,* 383.
Josephs, R., & Harrington, W. F. (1966) *Biochemistry 5,* 3473.
Lowey, S., Slayter, H. S., Weeds, A. G., & Baker, H. (1969) *J. Mol. Biol. 42,* 1.
Mendelson, R. A., & Cheung, P. (1976) *Science 194,* 190.
Mihalyi, E. (1950) *Enzymologia 14,* 224.
Morimoto, K., & Harrington, W. F. (1974) *J. Mol. Biol. 83,* 83.
Squire, J. M. (1972) *J. Mol. Biol. 72,* 125.
Sutoh, K., & Harrington, W. F. (1977) *Biochemistry 16,* 2441.
Thomas, D. D., Seidel, J. C., Gergely, J., & Hyde, J. S. (1975) *J. Supramol. Struc. 3,* 376.
Tregear, R. T., & Squire, J. M. (1973) *J. Mol. Biol. 77,* 279.
Wang, K., & Richards, F. (1974) *J. Biol. Chem. 249,* 8005.
Weber, K., & Osborn, M. (1969) *J. Biol. Chem. 244,* 4406.
Weeds, A. G., & Pope, B. (1977) *J. Mol. Biol. 111,* 129.
Weeds, A. G., & Taylor, R. S. (1975) *Nature (London) 257,* 54.

Conformational States of a Hydrophobic Protein. The Coat Protein of fd Bacteriophage†

Yasuhiko Nozaki, Jacqueline A. Reynolds, and Charles Tanford*

ABSTRACT: The coat protein of fd bacteriophage has a short polypeptide chain of only 50 amino acid residues, containing a highly hydrophobic segment of 19 amino acids that is entirely devoid of ionic or other strongly polar amino acids. In the viral particle the protein exists as a closely packed array of α helices. It can be transformed to a monomeric randomly coiled polypeptide in very concentrated ($\geq$7.3 M) guanidinium chloride. In anionic detergents or phospholipids the protein is dimeric, with a mixed conformation ("50% α"), the hydrophobic segment having a β structure, whereas the two ends are predominantly α helical. In guanidinium chloride at concentrations of 6 M or less, and under other conditions in the absence of an anionic detergent or phospholipid, the protein forms an intractable polymer, with a β-type conformation. If the protein is succinylated an oligomeric form of this structure (speculatively thought to be a soluble variety of a "β barrel") can be obtained as a metastable state. The 50% α conformation, the β oligomer, and the random coil can be interconverted reversibly, but formation of the β polymer appears to be irreversible.

The coat protein of fd bacteriophage has a single polypeptide chain of 50 amino acid residues, with a remarkable amino acid sequence (Nakashima & Konigsberg, 1974), shown in Figure 1. The polypeptide contains an uninterrupted sequence of 19 uncharged and predominantly hydrophobic residues (positions 21–39), 12 of them having side chains that can gain between 1300 and 3400 cal/mol apiece if they can be removed from contact with water (Nozaki & Tanford, 1971). The threonine residue at position 36 is the only one in this sequence that can be considered significantly hydrophilic. This segment of polypeptide chain must have an exceptionally strong propensity to form structures in which contact with an aqueous medium is avoided, and the behavior of the coat protein in various solvent systems might therefore be expected to be anomalous when compared to water-soluble polypeptides with a more even distribution of hydrophilic and hydrophobic residues along the chain.

The only other known polypeptide chain with a comparable amino acid sequence is that of the transmembrane MN-glycoprotein of the human erythrocyte (Tomita & Marchesi, 1975). It has a hydrophobic segment that is 23 residues long, but there are three hydrophilic residues within it, two threonines and one serine. Another membrane-bound protein, microsomal cytochrome b_5, has two eight-residue hydrophobic segments in its membrane-embedded region, but they are separated from each other by 13 intervening residues (Corcoran & Strittmatter, 1977). Sequence data for other intrinsic membrane proteins are not available, but it is considered likely that many of them may contain polypeptide segments of this type. Studies of the fd coat protein may therefore be of general interest in relation to the solution behavior of membrane proteins. The fd virus itself does not possess a membrane and the coat protein in the virus is not associated with lipid or any other amphiphilic substance. The protein does, however, reside in the phospholipid membrane of its host bacterium, *Escherichia coli,* during the process of viral biosynthesis and assembly (Smilowitz et al., 1972; Webster & Cashman, 1973); i.e., it is capable of existing as a membrane protein.

Previous work has established that the coat protein exists in the native virus as a closely packed layer of helices, with essentially all of the polypeptide backbone in an α-helical conformation (Marvin et al., 1974; Marvin & Wachtel, 1975; Nozaki et al., 1976). The virus may be disrupted by use of deoxycholate, and the purified protein in this detergent has a quite different conformation, judged by circular dichroism measurements to contain about 50% α helix and about 30% β

†From the Whitehead Medical Research Institute and the Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710. *Received October 4, 1977.* This work was supported by Research Grants AM-04576 from the National Institutes of Health and PCM7615240 from the National Science Foundation. Charles Tanford is a Research Career Awardee of the National Institutes of Health.

0006-2960/78/0417-1239$01.00/0

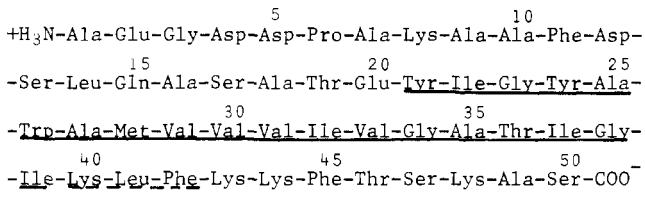

FIGURE 1: The amino acids of fd coat protein (Nakashima & Konigsberg, 1974). The hydrophobic segment is underlined. Residues 40–42 may be able to participate in the formation of an ordered hydrophobic domain because the length of the lysyl side chain would permit the amino group to extend beyond the next two residues.

structure (Makino et al., 1975; Nozaki et al., 1976). This conformation, which we shall call the "50% α" conformation, is retained when the coat protein is transferred from deoxycholate to certain other detergents, to lysophospholipid micelles, or to diacylphospholipid vesicles, and we have inferred from this that this is also the conformation of the protein when it is bound to the host bacterial membrane (Nozaki et al., 1976). This conclusion has been disputed (Williams & Dunker, 1977) because the protein can be incorporated into phospholipid membranes in other states as well, suggesting that the conformational state of the protein in detergents or lipids may depend on the history of treatment. One of the results of the present work will be to show that the 50% α conformation is in fact the stable conformation, independent of history, at least in anionic detergents.

One unusual property of the coat protein already observed before the present study was begun was its existence in an associated state in sodium dodecyl sulfate or in 6 M guanidinium chloride (GdmCl),[1] solvents that normally dissociate oligomeric proteins to their constituent polypeptide chains. The conformation in sodium dodecyl sulfate is the 50% α conformation also observed in deoxycholate, and the molecular weight in both detergents is about 10 000 corresponding to a dimer of the 50-residue polypeptide (Knippers & Hoffmann-Berling, 1966; Makino et al., 1975). In 6 M GdmCl the protein is highly aggregated, with a circular dichroism spectrum characteristic of a β-sheet structure. Ikehara et al. (1975) have, however, indicated that the coat protein adopts a random coil conformation in GdmCl solutions at concentrations above 7 M, but did not measure the molecular weight. The present study confirms their observation and shows in addition that the conversion to a random coil is accompanied by dissociation to the monomeric state.

Materials and Methods

Materials. The preparations of fd phage and of purified coat protein in deoxycholate have been described previously (Makino et al., 1975). Sodium deoxycholate (Fisher Scientific Co.) was purified as described by Makino et al. (1973). Sodium dodecyl sulfate (BDH Chemicals, Poole, England) was used without purification, but all samples were checked for homogeneity of the dodecanol moiety by gas chromatography. The nonionic detergent dodecyl octaethylene glycol monoether was a product of Nikko Chemicals, Co., and was obtained from Kouyoh Trading Co., Tokyo, Japan. Guanidinium chloride (GdmCl) from three sources was used: the Ultra Pure reagent obtained from Schwarz/Mann was found to satisfy the criteria for purity (Wong et al., 1971) and was used without purification; GdmCl from J. T. Baker Co. was recrystallized before use; guanidinium carbonate (Eastman) was converted to GdmCl and purified as described by Nozaki (1972). Recrystallized succinic anhydride was a gift from Dr. R. L. Hill. Beef insulin was obtained from Eli Lilly, and 6× recrystallized hen's egg lysozyme was from Miles Laboratories.

Coat Protein in Concentrated GdmCl. When starting from intact phage fd stock solutions were diluted with 0.02 M Tris buffer (pH 8.2) to a final concentration below 1.5 mg/mL. Crystalline GdmCl was added to the desired final concentration: 7.35 M and 8.22 M correspond, respectively, to solutions containing 60 and 66% GdmCl by weight. The resulting solutions were chromatographed on Sepharose 4B or Sephacryl S200 to separate DNA and the A protein of the phage from the coat protein. Coat protein solutions in GdmCl were also obtained starting with the purified protein in 10 or 30 mM deoxycholate. Crystalline GdmCl was added to the protein-containing fractions from the preparative column used for purification (Makino et al., 1975) to the desired final concentration.

Transfer of Coat Protein from GdmCl to Sodium Dodecyl Sulfate. The precipitation of guanidinium dodecyl sulfate can be avoided if a temperature of 75 °C is used. Powdered sodium dodecyl sulfate was added to a solution of coat protein in 7.35 M GdmCl at that temperature and stirred briefly until a homogeneous (slightly opaque) solution was obtained. The solution was dialyzed for 45 min against 50 mM sodium dodecyl sulfate at 75 °C, and the dialysis was continued for 1 h at 40 °C and several hours at 30 °C. The dialysate was replaced with fresh solution at the appropriate temperature at intervals. In some experiments the sodium dodecyl sulfate concentration was reduced to 10 mM, without effect.

Succinylation of Coat Protein. The coat protein was succinylated in concentrated GdmCl or in deoxycholate solution (Klapper & Klotz, 1972). A 50-fold molar excess of succinic anhydride was used and the pH was maintained between pH 8 and 9 by addition of Tris crystals in small amounts. Low molecular weight impurities were removed by dialysis or chromatography. The product was not analyzed to determine the final extent of reaction, but the conditions used were designed to ensure complete reaction of all amino groups.

Circular Dichroism. Circular dichroism measurements were carried out using a Jobin Yvon Dichrograph Mark III. Path lengths of 0.02 to 0.10 cm and protein concentrations of 0.05 to 1.0 mg/mL were used.

Interpretation of Circular Dichroism Data. We have used the ellipticity at 208 nm as a measure of α-helix content, as recommended by Greenfield & Fasman (1969). Other published procedures would give similar results. The recommended procedures of Greenfield & Fasman (1969) and Chen et al. (1974) for the determination of the content of β structure have proved to be not self-consistent in our hands, and we have preferred to use the ellipticity near 202 nm if measurements to that wavelength are feasible. The α-helix contribution here is small, the β structure has a large positive ellipticity, and randomly coiled polypeptides (and, by inference, unordered regions of folded proteins) have a large negative ellipticity. In concentrated GdmCl solutions, where it is difficult to obtain reliable data below 210 nm, two kinds of spectra were obtained, one of which was unmistakably that of an essentially pure random coil, so that no further analysis was needed. The other spectrum resembled that of a pure β polypeptide in the 210–240 nm range and we used the ellipticity at 217 nm given by Greenfield & Fasman (1969) as a measure of the percent β structure present, assuming the α-helix content to be zero. The observed ellipticities at the lowest wavelength attained (210 nm) were consistent with this assumption, tending to be too *positive* where they deviate at all; i.e., the deviation was in

[1] Abbreviation used is: GdmCl, guanidinium chloride.

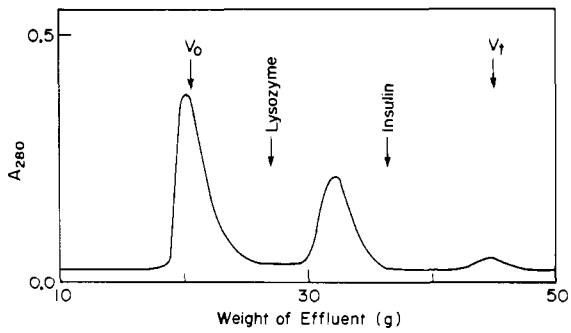

FIGURE 2: Gel chromatography of a solution of fd phage in 7.35 M GdmCl on Sephacryl 200 SF. The first peak consists almost entirely of viral DNA and was used as void volume marker (V_0). Other arrows indicate the elution positions of reduced and carboxymethylated lysozyme and insulin and of the total volume marker (V_t), for which β-mercaptoethanol was used.

the opposite direction from that expected for a contribution from an α-helical structure.

Molecular Weight by Sedimentation Equilibrium. A Beckman analytical ultracentrifuge with photoelectric scanner was used. The sample examined was succinylated in 3.4 M GdmCl. The partial specific volume was calculated on the basis of the amino acid composition (Cohn & Edsall, 1943). A small correction (less than 0.01 cm^3/g) was required to allow for succinylation, and was based on the assumption that all amino groups of the protein had reacted with the reagent. The possible error in molecular weight from this source and from preferential interactions in the GdmCl solution (Hade & Tanford, 1967; Lee & Timasheff, 1974) is less than 10%.

Molecular Weight of Randomly Coiled Polypeptide in Concentrated GdmCl. We have found in this and related studies that the chromatographic procedure of Fish et al. (1969) is applicable to the higher concentrations of GdmCl here employed, the only effect being that the flow rate on the column is decreased. In the one quantitative measurement reported here (which was repeated several times on different samples), reduced and alkylated lysozyme and reduced and alkylated insulin were used as standards. The mixture of the two insulin peptides is not separated on the column and was assigned an average molecular weight of 2900, as in the calibration procedure of Fish et al. (1969).

Other Methods. Protein concentrations were measured spectrophotometrically, using $A_{280}^{1cm} = 1.66$ for a solution containing 1 mg/mL (Nozaki et al., 1976). All measurements were made in 0.02 M Tris buffer (pH 8.2) at room temperature. Where slow changes over a period of days are reported, they refer to protein solutions stored at room temperature.

Results

Formation of Random Coil. Disruption of fd phage, at 7.35–8.2 M GdmCl, is a slow process, as was also observed by Ikehara et al. (1975). The time of incubation routinely employed was 2–4 days. The resulting solution was chromatographed on Sepharose 4B or Sephacryl S 200. Figure 2 shows a typical elution pattern. One peak elutes in the void volume and, on the basis of its absorption spectrum, was judged to consist of DNA with no significant amount of protein. The second peak is pure coat protein. The circular dichroism spectrum, shown by curve 1 in Figure 3, is the characteristic spectrum for a randomly coiled polypeptide chain. The molecular weight, estimated from the elution positions of several column experiments, is 5300 ± 200, essentially equal to the true molecular weight of 5245 calculated from the amino acid sequence. (When disruption of phage is carried out at phage concentrations above 1.5 mg/mL, complete conversion to the random coil form was not obtained. Detailed studies were not carried out.)

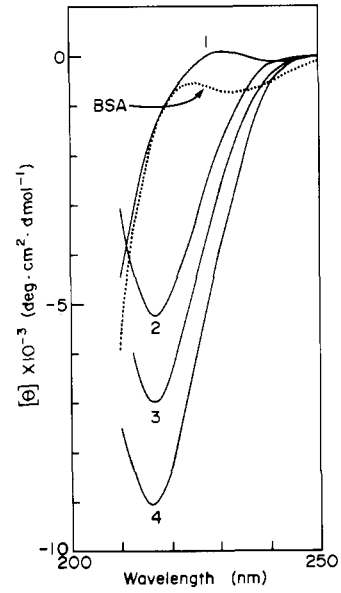

FIGURE 3: Circular dichroism spectrum of coat protein in the random coil and β conformations. (Curve 1) Random coil conformation in 7.35–8.2 M GdmCl. Succinylation of the protein did not change the spectrum. The spectrum of reduced and carboxymethylated bovine serum albumin (BSA) in 6 M GdmCl is shown for comparison. (Curve 2) β conformation in 7.35 M GdmCl, from leading chromatographic fraction of Figure 4. (Curve 3) β conformation in 6 M GdmCl. (Curve 4) β conformation obtained by diluting a solution of randomly coiled protein to 4 M GdmCl. When solid GdmCl was added to a final concentration of 7.4 M, the spectrum asymptotically approached the spectrum of curve 2 over a period of about a week.

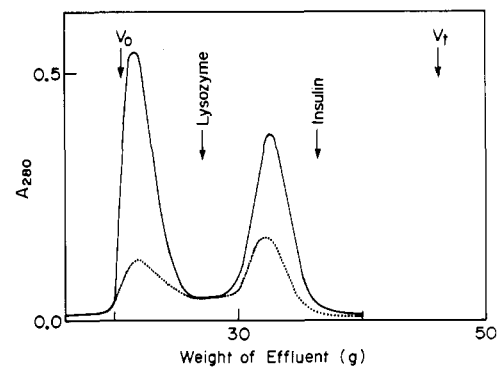

FIGURE 4: Gel chromatography on Sephacryl 200 SF of solutions of pure coat protein in 7.35 M GdmCl, obtained by adding GdmCl to solutions of the protein in deoxycholate. The amount of protein in the leading peak was variable, and two examples are shown. When succinylated protein is used, the fraction in the leading peak becomes extremely small. Positions of arrows are explained in the caption to Figure 2.

When GdmCl is added to purified coat protein in deoxycholate, where the initial conformation is the 50% α conformation, a transient cloudiness is observed. The solutions clear rapidly, however, and the circular dichroism reaches values near equilibrium in minutes, i.e., in much less time than is required when starting with intact phage. Chromatography of the resulting solution is illustrated by the typical example shown in Figure 4. Two peaks are again observed, despite the fact that one is starting here with pure coat protein. One peak elutes at the same position as the randomly coiled polypeptide obtained from intact phage and has the same circular dichroism spectrum (Figure 3, curve 1). The other, corre-

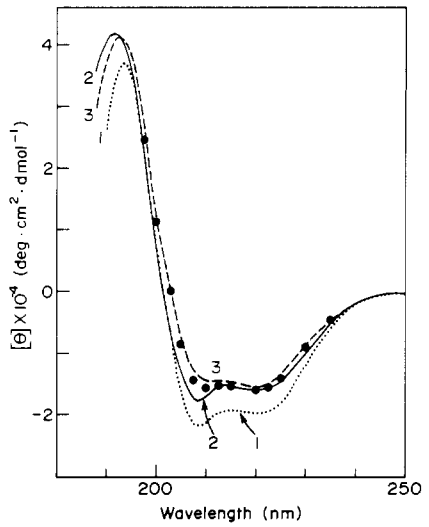

FIGURE 5: Circular dichroism spectra of the coat protein in the 50% α conformation. (Curve 1) Purified coat protein in 10 mM deoxycholate. (Curve 2) In 3 mM sodium dodecyl sulfate after initial solubilization in 30 mM deoxycholate. The filled circles represent a similar sample in 4 mM sodium dodecyl sulfate and are intended to illustrate the slight variability in the spectrum from one experiment to another. (Curve 3) In 50 mM sodium dodecyl sulfate after conversion from the randomly coiled state in 7.35 M GdmCl.

sponding to a larger particle size, has the circular dichroism spectrum given by curve 2 of Figure 3, which is essentially the spectrum of a polypeptide chain in the β conformation. The coat protein in this peak is an example of what we have called the "β-polymer" conformation, which is the stable form of the protein at GdmCl concentrations of $\leq$6 M. Its formation in the present experiment presumably occurs because the protein is exposed to GdmCl concentrations below the final concentration during dissolution of the added crystals. When GdmCl is added to intact virus, disruption of the virus is the slow step, and free protein is never exposed to GdmCl concentrations below the final concentration.

Chromatographed solutions containing randomly coiled coat protein, whether derived from intact phage or from coat protein in deoxycholate, could be stored for weeks without change. However, the polymerized protein in the leading peak of Figure 4 was likewise stable and did not revert to the random coil form on standing. Furthermore, when the two fractions of solutions derived from deoxycholate were not separated from each other, a very slow change in the circular dichroism spectrum in the direction of an increase in β structure occurred, suggesting that polymer already present can act as a nucleus for conversion of randomly coiled protein to the polymeric form. These experiments indicate that the polymeric form of the coat protein may be marginally more stable than the random coil even at the highest GdmCl concentrations employed.

50% α Conformation. When intact phage is dissolved in deoxycholate or sodium dodecyl sulfate (Makino et al., 1975), one obtains the coat protein in a conformation defined by the circular dichroism spectra shown in Figure 5. Analysis by the procedure given under Materials and Methods indicates an α-helix content of about 50% and about 30% β structure. The ellipticity at 192.5 nm is not used in this analysis: the predicted ellipticity (Greenfield & Fasman, 1969) at this wavelength for 50% α, 30% β, and 20% unordered structure is about +40 000, in agreement with the experimental value. If the empirical procedure of Chou & Fasman (1974) is used to predict the structure from the amino acid sequence, one obtains 46% α, 30% β and 24% unordered, in good agreement with the estimates from circular dichroism. The prediction identifies the hydrophobic portion of the polypeptide chain as containing the β structure, and the polar ends of the chain (residues 5-20 and 39-45) as containing the α-helical portions. Pronase digestion of the detergent-solubilized protein confirms this assignment: the hydrophobic segment of the chain remains intact, but the chain ends are hydrolyzed. The detergent-solubilized residue has the circular dichroism spectrum of a β polypeptide and the ellipticity at 217 nm indicates at least 70% is in the β form (Chamberlain et al., 1978).

The protein was found to be dimeric in both detergents (Knippers & Hoffmann-Berling, 1966; Makino et al., 1975) and strenuous efforts to dissociate the dimer by using very high concentrations of sodium dodecyl sulfate have been unsuccessful. We assume that the inability to dissociate the dimer is related to the fact that the hydrophobic portion of the polypeptide, which must reside in the hydrocarbon interior of the detergent micelle, is in the β conformation. Hydrogen bonds needed for stabilization of the peptide groups in an apolar medium can then be formed only if several parallel or antiparallel segments are aligned in close proximity. A dimeric structure could form four adjacent strands if there is a bend in the middle (cf. Tanford & Reynolds, 1976; Figure 3D).

It should be noted that the conformation in sodium dodecyl sulfate is the same as in deoxycholate; i.e., the former does not appear to be a denaturing detergent for the coat protein. A plausible explanation for this will be provided in the Discussion. Of greater interest is the fact that essentially the same conformation is retained when deoxycholate is replaced by lysophospholipids or when the coat protein is inserted into diacylphospholipid vesicles, either directly by sonication of phage particles, or indirectly with intermediate solution in detergent (Nozaki et al., 1976; Chamberlain et al., 1978). We have used this as evidence to suggest that the 50% α conformation is the stable conformation of the protein when it is embedded in the membrane of its host bacterium, *Escherichia coil,* during in vivo infection. This suggestion has been questioned (Williams & Dunker, 1977) because there has heretofore been no proof that the 50% α conformation is in fact the equilibrium conformation under the conditions where it is formed. To provide such proof we need to show that the 50% α conformation can be formed from the randomly coiled polypeptide.

To carry out the experiment we need to add detergent to coat protein in concentrated GdmCl without reducing the GdmCl concentration, so as to minimize the risk of forming β polymer. This can be done if sodium dodecyl sulfate at an elevated temperature is used, as described under Materials and Methods. The circular dichroism of the product was found to be that of the 50% α form, as is illustrated by curve 3 of Figure 5.

β Oligomer and β Polymer. When the coat protein is dissolved in 6 M GdmCl, regardless of the form from which one starts, a conformation with the circular dichroism spectrum shown by curve 3 of Figure 3 is obtained. The spectrum arises predominantly from a β structure, with a pronounced trough at 217 nm. The particles with this structure are highly aggregated and in the ultracentrifuge they sediment to the bottom of the cell in minutes at rotor speeds of 40 000 rpm.

The formation of this β-polymer conformation appears to be irreversible. As shown by Figure 4 and the accompanying discussion, the polymer can be formed in the course of preparing coat protein solutions at concentrations of GdmCl greater than 6 M, and, if formed, shows no tendency to revert to the random coil form. On the other hand, the *intensity* of the spectrum appears to be variably responding slowly (over a period of days) to changes in GdmCl concentration. The minimal negative ellipticity at 217 nm (Figure 3, curve 2) is

−5200 deg·cm²/dmol. This value is obtained when β polymer is initially formed at GdmCl concentrations above 6 M (Figure 4) or when solutions originally at lower GdmCl concentrations are kept for a long period of time at ≤7.35 M GdmCl. The equilibrium value in 6 M GdmCl is −7000 deg·cm²/dmol, and at 4 M GdmCl it is −9000 deg·cm²/dmol. When succinylated protein is used, the minimal value of −5200 deg·cm²/dmol is retained to much lower GdmCl concentrations (see Figure 6 below). In terms of structure, the minimal ellipticity at 217 nm corresponds to 40–45% β structure, with the rest of the protein in a random coil conformation. Ellipticities of −7000 and −9000 deg·cm²/dmol correspond to 50 and 60% β structure, respectively.

There is evidence from work with synthetic polypeptides that the presence of unusually hydrophobic chain segments destabilizes the random coil conformation in concentrated GdmCl solutions and can lead to the formation of ordered structures which can include adjacent hydrophilic amino acids linked to the hydrophobic segments (Auer & Doty, 1966). We thus visualize the β polymer as having a core that arises from dense packing of the hydrophobic segments of many polypeptide chains, held together by lateral hydrogen bonds, as in the β-sheet structures originally proposed by Pauling & Corey (1951). If we assume that amino acid residues 21–42 participate in the formation of this core (as suggested in the caption to Figure 1), and that this is the only part of the polypeptide in β conformation, the calculated content of β structure would be 44%, in excellent agreement with the minimal content of β structure estimated from the circular dichroism data obtained under conditions that favor formation of a disordered state for the chain termini, i.e., maximization of the GdmCl concentration or the introduction of mutually repulsive charges by succinylation. The higher content of β structure observed for the unmodified protein at lower GdmCl concentrations suggests that parts of the polypeptide adjacent to the hydrophobic segment are readily drawn into the ordered structure as the GdmCl concentration is reduced.

It is of course physically impossible to form a highly polymerized state without the occurrence of lower aggregates as intermediates, but attempts to find intermediates, the formation of which might be reversible, have been unsuccessful. For example, when randomly coiled coat protein in ≥7.35 M GdmCl is diluted to 6 M GdmCl, the change in the circular dichroism spectrum to that of a β conformation is very fast, and virtually complete in a few minutes. If the reaction is allowed to proceed for only 30 s, it can be stopped by the addition of GdmCl crystals at a stage when only about 30% of the conversion to the β form has taken place. Even under these conditions, however, no reversal to the randomly coiled form could be observed.

By use of succinylated coat protein, which leads to a high net negative charge, we were able to circumvent the problem of rapid aggregation. Succinylation was carried out with the protein in the random coil form in concentrated GdmCl or with the protein in the 50% α form in deoxycholate, with no striking difference in the results. In reactions other than formation of the β polymer the succinylated protein behaved like the unmodified coat protein. Succinylation did not affect the circular dichroism spectrum in the various conformations: the spectrum of randomly coiled succinylated protein is indistinguishable from curve 1 of Figure 3; the spectrum of succinylated protein in the β conformation at 3.4 M GdmCl is the same as curve 2 of Figure 3; the spectrum in the 50% α conformation falls within the range of the spectra given in Figure 5. The only significant effect of succinylation proved to be in the transition of the randomly coiled protein to the β structure. Formation

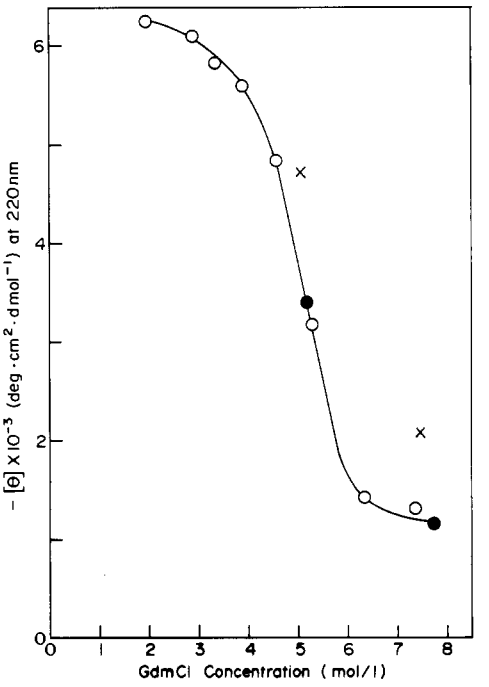

FIGURE 6: Interconversion of random coil and β form of succinylated coat protein in GdmCl solutions. Open circles show the change in ellipticity accompanying progressive dilution of the GdmCl. Filled circles are reversed points, obtained from a solution kept for 15 min at 2.7 M GdmCl; crosses were obtained using a solution maintained for several hours in the β conformation.

of the intractable β polymer was greatly slowed down and the formation of β structure per se proved to be reversible.

Figure 6 shows the change in circular dichroism when succinylated coat protein in 7.35 M GdmCl is progressively diluted. Measurements were made at 220 nm where the difference between the random coil and β conformations (Figure 3) is greatest. The transition from random coil to β structure occurs at a lower GdmCl concentration than for unmodified protein. The transition is reversible if the protein does not remain in the β form for too long, but only partial reversibility is obtained if the protein remains at low GdmCl concentrations for several hours.

Succinylated protein at 3.4 M GdmCl was subjected to ultracentrifugal analysis, and this provided some insight into the successive stages of aggregation, though exact quantitation was not possible because the protein consisted of a mixture of species, the composition of which was changing slowly with time. A short high-speed run of the freshly prepared solution showed that no extremely aggregated species was present initially.

Figure 7 shows the quasiequilibrium distribution of protein after about 16 h of centrifugation, which would have represented enough time to be within a few percent of true equilibrium in the absence of time-dependent changes in composition. A substantial fraction of the protein, in the top half of the cell, represents fairly homogeneous material with a molecular weight of about 30 000. Another scan, several hours earlier, showed the presence of a somewhat larger amount of this material, but no difference in the apparent molecular weight. No significant amount of material with molecular weight in the monomer–dimer range could have been present in either case. There is a fairly abrupt change in the slope of the plots near the middle of the solution column. The weight-average molecular weight in the lower half of the cell is well above 100 000, and extrapolation to the bottom of the cell accounts for all of the protein known to be present; i.e., there could not

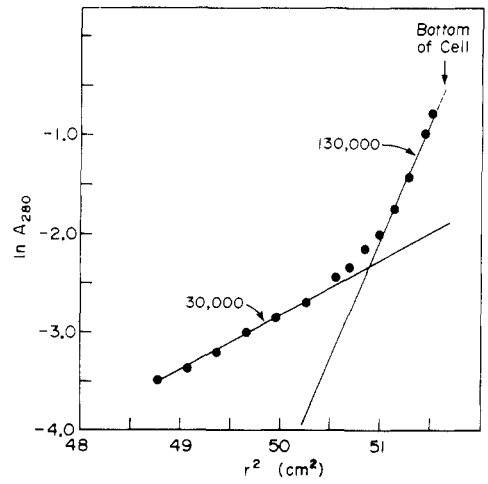

FIGURE 7: Quasiequilibrium distribution of succinylated coat protein in 3.4 M GdmCl after centrifugation for about 16 h at 20 000 rpm. The molecular weights given are the local weight-average molecular weights corresponding to the straight lines at the top and bottom of the plot. Extrapolation of the data to the bottom of the cell by extension of the lower linear portion (as shown by the dashed line) yielded an overall integrated optical density that accounted for all of the protein in the sample. Scans made after a longer period of centrifugation showed a loss of optical density (i.e., the presence of heavier material) if a similar linear extrapolation was made.

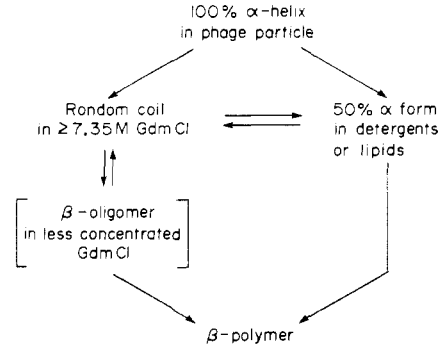

FIGURE 8: The known conformations of the coat protein, arrows representing the interconversions described in this paper. The β oligomer was observed only when succinylated protein was used, but is presumably a transient intermediate for unmodified protein. The 50% α form was obtained in all lipid environments investigated, and in ionic detergents. In the one nonionic detergent employed, the β-polymer form was obtained.

have been much material with extremely high molecular weight. In another scan, after 25 h, almost no material with molecular weight near 30 000 remained. The weight-average molecular weight over most of the cell was 170 000, but extrapolation to the bottom of the cell accounted for only about 70% of the known optical density, indicating that material with much higher molecular weight had been formed. This was confirmed by another high-speed run, which showed the presence of a rapidly sedimenting component that was not present in the freshly prepared solution.

These experiments suggest that the β structure that is formed is an oligomer with a molecular weight corresponding approximately to an association of six polypeptide chains. We suggest that this species might be a soluble analogue of the "β barrel" often observed as a folded domain within the internal structure of globular water-soluble proteins (Richardson, 1977), with the difference that in this case the laterally hydrogen-bonded barrel arrangement is formed by association of chain segments from separate polypeptide chains, rather than by segments brought together by folding of a single chain. Aggregation beyond this first species to form β polymer is clearly a complex phenomenon, and the stage at which irreversibility sets in (as seen in Figure 6, for example) cannot be defined on the basis of these few experiments.

Formation of β Polymer in Detergent Solutions. In the procedure described previously (Nozaki et al., 1976) for transfer of coat protein from deoxycholate to other detergents with retention of the 50% α conformation, the deoxycholate is first removed by several successive dilution and ultrafiltration steps. This is a rapid procedure. The coat protein precipitates from solution in the course of the procedure and up to that point no change in circular dichroism spectrum is detected; i.e., the protein is presumably precipitated in the 50% α form. If the removal of deoxycholate is carried out slowly by dialysis, the 50% α conformation is not retained and substantial, irreversible conversion to β polymer is observed. This phenomenon probably also accounts for the small differences between the circular dichroism of individual preparations obtained by the rapid precipitation method, as seen in Figure 5 and in the previously published results (Nozaki et al., 1976); i.e., small amounts of β polymer are probably formed even when the rapid technique is used.

When the coat protein is dissolved in the nonionic detergent dodecyl octaethylene glycol monoether by any procedure, pure β polymer is obtained, with an ellipticity at 217 nm of −12 000 deg·cm²/dmol, corresponding to 70% β structure. Formation of the β structure is again irreversible: addition of sodium dodecyl sulfate to very high concentrations does not change the circular dichroism spectrum. Inability to retain the 50% α conformation in the nonionic detergent may be related to the amino acid sequence of the coat protein: the three lysine residues immediately adjacent to the hydrophobic segment (positions 40, 43, 44) may require a negative charge for stabilization in the 50% α conformation. When lysophosphatidylcholine or diacylphosphatidylcholine vesicles are used, the phosphate group can provide the requisite negative charge even though the overall net charge of the head group is zero.

Discussion

The fd coat protein has been shown to be capable of existing in five quite different conformational states, as summarized in Figure 8. No unusual solvents were employed, and the conformational versatility is remarkable, considering that the protein's single polypeptide chain has a length of only 50 amino acids. It should be noted that the hydrophobic segment of the protein probably has a β-type structure in the 50% α dimeric form in detergents or lipids, but that the necessity for hydrogen bonding between peptide groups suggests that there is at least one 180° turn within the segment so as to provide enough adjacent strands for a complete network of hydrogen bonds. Dimensional considerations lead to the same conclusions; i.e., an extended segment of 19 to 22 residues in the β conformation would be far too long to be accommodated by the hydrocarbon core of a phospholipid bilayer (Tanford & Reynolds, 1976; Figure 3). In the β oligomer or β polymer, on the other hand, there is no obvious need for a turn in the chain direction; i.e., the hydrophobic segment may be fully extended in these structures. The adoption of a β structure per se by the particular sequence of the hydrophobic segment is consistent with the empirical rules deduced from the known crystal structures of globular proteins (Chou & Fasman, 1974), and the apparent existence of the segment in an α-helical conformation in the viral particle constitutes an anomaly.

A noteworthy observation is the readiness with which the protein forms an apparently irreversibly aggregated state. The

formation of this intractable polymer has to be ascribed directly to the central hydrophobic segment of the polypeptide chain. Since many intrinsic membrane proteins may have polypeptide segments of this type, perhaps more than one such segment in a chain of much greater overall length (Henderson & Unwin, 1975; Rizzolo et al., 1976; Jenkins & Tanner, 1977), it would not be surprising to observe similar irreversible association processes with such proteins. We know of no experimental studies directly related to this problem.

Many membrane proteins have been solubilized in detergent solutions without loss of biological activity or significant structural change (Tanford & Reynolds, 1976), and the behavior of the coat protein is similar to the extent that the same conformation, the 50% α conformation, is observed in phospholipid vesicles and in several detergent solutions. This conformation is, however, not maintained in the nonionic detergent, dodecyl octaethylene glycol monoether, whereas detergents of this class have been among the most successful for maintaining membrane proteins in their native state. On the other hand, sodium dodecyl sulfate, which denatures virtually all proteins and dissociates them to their constituent polypeptide chains, does not disrupt the 50% α conformation of the coat protein, which is a dimer of the coat protein polypeptide.

Denaturation by nonionic detergents may be related to the specific amino acid sequence of the coat protein, as suggested in the Results section, but the stability of the 50% α conformation in sodium dodecyl sulfate may be a more general phenomenon, because there are theoretical reasons for expecting that an ordered structure normally stabilized by association with a phospholipid bilayer or by insertion into the micelles of a benign detergent may also be stabilized by sodium dodecyl sulfate if the length of polypeptide chain involved in the formation of the ordered structure is relatively small. This is because the addition of sodium dodecyl sulfate must shift the equilibrium between accessible states in the direction of greater binding of the detergent (Wyman, 1964). In the 50% α conformation, a dimer of the coat protein may be expected to associate with approximately one micelle of the detergent (Tanford & Reynolds, 1976), i.e., about 45 detergent molecules per polypeptide. An experimental measurement, not necessarily at saturation, gave a binding ratio of 30 dodecyl sulfate ions per chain (Makino et al., 1975). The binding of sodium dodecyl sulfate to denatured polypeptides is typically one detergent ion per two amino acid residues (Reynolds & Tanford, 1970), yielding an expected binding ratio of 25 dodecyl sulfate ions per coat protein chain. In this situation the detergent can be expected to stabilize rather than destabilize the 50% α conformation.

The 50% α conformation of the coat protein has been obtained directly from phage particles in sodium dodecyl sulfate or in deoxycholate (Makino et al., 1975) directly from phage particles and also indirectly via detergents in phospholipid vesicles (Chamberlain et al., 1978), and by transfer from deoxycholate solution to several lysophospholipids (Nozaki et al., 1976). We have shown in the present paper that the same conformation can be obtained starting with randomly coiled polypeptide, but the latter experiment was feasible only with sodium dodecyl sulfate as the recipient detergent. Because there is no reason to expect different conformations in sodium dodecyl sulfate and other detergent or lipid system, and because all of our experiments suggest that states other than the β polymer are readily interconvertible, the evidence is strong, but not conclusive, that the 50% α conformation is in fact the thermodynamically stable state of the protein in the host *E. coli* membrane. This does not necessarily mean that the 50% α conformation is actually formed in the in vivo situation; i.e., the newly synthesized protein might be inserted into the membrane in a nonequilibrium state. Reconstitution of the virus particle from separated DNA and coat protein is necessary to resolve this question.

Acknowledgments

We thank Dr. R. E. Webster and members of his laboratory for making preparations of fd phage available to us. We thank Drs. P. Y. Chou and G. D. Fasman for providing us with the predicted structure of the polypeptide chain.

References

Auer, H. E., & Doty, P. (1966) *Biochemistry 5,* 1716.
Chamberlain, B. K., Nozaki, Y., Tanford. C., & Webster, R. E. (1978) (submitted for publication).
Chen, Y. H., Yang, J. T., & Chau, K. H. (1974) *Biochemistry 13,* 3350.
Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry 13,* 211, 222.
Cohn, E. J., & Edsall, J. T. (1943) *Proteins, Amino Acids and Peptides,* p 370, Reinhold, New York, N.Y.
Corcoran, D., & Strittmatter, P. (1977) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 36,* 897.
Fish, W. W., Mann, K. G., & Tanford, C. (1969) *J. Biol. Chem. 244,* 4989.
Greenfield, N., & Fasman, G. D. (1969) *Biochemistry 8,* 4108.
Hade, E. P. K., & Tanford, C. (1967) *J. Am. Chem. Soc. 89,* 5034.
Henderson, R., & Unwin, P. T. N. (1975) *Nature (London) 257,* 28.
Ikehara, K., Utiyama, H., & Kurata, M. (1975) *Virology 66,* 306.
Jenkins, R. E., & Tanner, M. J. A. (1977) *Biochem. J. 161,* 139.
Klapper, M. H., & Klotz, I. M. (1972) *Methods Enzymol. 25,* 536.
Knippers, R., & Hoffmann-Berling, H. (1966) *J. Mol. Biol. 21,* 281.
Lee, J. C., & Timasheff, S. N. (1974) *Biochemistry 13,* 257.
Makino, S., Reynolds, J. A., & Tanford, C. (1973) *J. Biol. Chem. 248,* 4926.
Makino, S., Woolford, J. L., Jr., Tanford, C., & Webster, R. E. (1975) *J. Biol. Chem. 250,* 4327.
Marvin, D. A., & Wachtel, E. J. (1975) *Nature (London) 253,* 19.
Marvin, D. A., Pigram, W. J., Wiseman, R. L., Wachtel, E. J., & Marvin, F. J. (1974) *J. Mol. Biol. 88,* 581.
Nakashima, Y., & Konigsberg, W. (1974) *J. Mol. Biol. 88,* 598.
Nozaki, Y. (1972) *Methods Enzymol. 26,* 43.
Nozaki, Y., & Tanford, C. (1971) *J. Biol. Chem. 246,* 2211.
Nozaki, Y., Chamberlain, B. K., Webster, R. E., & Tanford, C. (1976) *Nature (London) 259,* 335.
Pauling, L., & Corey, R. B. (1951) *Proc. Natl. Acad. Sci. U.S.A. 37,* 251, 729.
Reynolds, J. A., & Tanford, C. (1970) *Proc. Natl. Acad. Sci. U.S.A. 66,* 1002.
Richardson, J. S. (1977) *Nature (London) 268,* 495.
Rizzolo, L. J., le Maire, M., Reynolds, J. A., & Tanford, C. (1976) *Biochemistry 15,* 3433.
Smilowitz, H., Carson, J., & Robbins, P. W. (1972) *J. Su-*

pramol. Struct. 1, 8.
Tanford, C., & Reynolds, J. A. (1976) *Biochim. Biophys. Acta 457,* 133.
Tomita, M., & Marchesi, V. T. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72,* 2964.
Webster, R. E., & Cashman, J. S. (1973) *Virology 55,* 20.
Williams, R., & Dunker, A. K. (1977) *Biophys. J. 17,* 71a.
Wong, K. P., Roxby, R., & Tanford, C. (1971) *Anal. Biochem. 40,* 459.
Wyman, J. (1964) *Adv. Protein Chem. 19,* 224.

O^2- and O^4-Alkyl Pyrimidine Nucleosides: Stability of the Glycosyl Bond and of the Alkyl Group as a Function of pH†

B. Singer,* M. Kröger,‡ and M. Carrano

ABSTRACT: Alkylation of the O^2 of deoxypyrimidine nucleosides greatly destabilizes the glycosyl linkage. Within the pH range studied, cleavage of the glycosyl bond of O^2-alkyl thymidine and O^2-alkyl deoxyuridine is fastest in acid pH but also occurs with little or no pH dependence over the range pH 4.0–13.0. The glycosyl bond of O^2-alkyl deoxycytidine is the least stable of any O^2-alkyl derivative at neutrality and is cleaved without significant pH dependence. The glycosyl bond of O^2-alkylated ribopyrimidine nucleosides is cleaved at a much slower rate than that of the deoxy ribo derivatives. The relative stability, at pH 7.0, of the glycosyl bond of the ethyl derivatives studied is O^2-Et-dCyd $\ll$ O^2-Et-dUrd < O^2-Et-Thd < O^2-Et-Cyd $\ll$ O^2-Et-Urd. Those methylated derivatives investigated were found to behave similarly. The stability of the glycosyl bond of O^2-Et-dCyd and O^2-Et-Thd was also compared with that of 7-Et-dGuo at pH 7.0, 37 °C. In contrast, all of the O^4-alkylated deoxyribo- and ribopyrimidine nucleosides resemble the unmodified nucleosides in the stability of the glycosyl bond. All *O*-alkyl groups are labile, particularly in strong acid and base, and the rate of their hydrolysis was studied at various pHs. pH dependence was similar to that for glycosyl bond cleavage; dealkylation being much faster in acid than in alkali for all except O^2-alkyl deoxycytidine. The relative stability, at pH 1.5, of the ethyl groups is O^4-Et-Urd < O^4-Et-Thd < O^4-Et-dUrd < O^4-Et-Ura < O^2-Et-Urd < (O^2-Et-dUrd) O^2-Et-Ura < (O^2-Et-Thd) O^2-Et-Thy $\ll$ O^2-Et-Cyd < O^2-Et-Cyt. pK_a values were determined for the new derivative, O^2-Et-Thy, and for O^2-Et-Thd, O^2-Et-dUrd, O^4-Et-Thd, O^4-Et-dUrd, and O^2-Et-Ura.

Depyriminidation of deoxyribonucleotides has been found to occur in neutral solution, although very much slower than depurination (Shapiro & Kang, 1969; Lindahl & Karlstrom, 1973). Both purine and pyrimidine base release proceeds by the same mechanism, namely, hydrolytic cleavage of the glycosyl bond without opening of the sugar ring. Although the rate of depurination of dAdo[1] and dGuo is a linear function of the hydrogen ion concentration, the hydrolysis rate of dCyd and Thd is virtually pH independent over the range pH 3.0–7.0 (Shapiro & Kang, 1969; Shapiro & Danzig, 1972).

The glycosyl bond of the pyrimidine ribonucleosides is more stable than that of the corresponding deoxyribonucleosides (reviewed by Kochetkov & Budowsky, 1972), and at pH 7.0, 37 °C, the rate constants for hydrolysis of Cyd and Urd are two orders of magnitude less than for dCyd and dUrd. Various substituents on the C-5 of dUrd (but not dCyd) increase the rate of glycosyl cleavage to a slight but measurable extent (Kochetkov & Budowsky, 1972). No comparable data appear to exist for other positions of ring substitution or for any ribopyrimidine nucleoside modified on the ring.

Recent studies on the reaction of nucleic acids with carcinogens of the *N*-nitroso type showed that *O*-alkyl pyrimidines were major products (Singer, 1976a). There were clear indications that O^2-alkyl thymidine and O^2-alkyl deoxycytidine were converted to the corresponding alkyl base under mild conditions (Singer, 1976a). It thus appeared important for the study of the biological role of O-alkylation to systematically study the effect of the position of substitution on the stability of the glycosyl linkage and of the alkyl group for both ribo- and deoxyribopyrimidine nucleosides.

We now report that alkylation of the O^2 of all pyrimidines greatly destabilizes the glycosyl bond at all pHs investigated, while alkylation of the O^4 of Urd, dUrd, or Thd has no measurable effect on glycosyl bond cleavage.

Experimental Section

Preparation of Alkylated Nucleosides and Bases. Nucleosides and bases were alkylated using several methods. Thd, dUrd, Urd, dGuo, and Cyt were reacted in anhydrous solution with both radioactive and unlabeled diazoalkanes according to Kuśmierek & Singer (1976). The following derivatives used in this work were isolated from reaction mixtures using paper chromatography (Whatman 3MM) or thin-layer chromatography on silica gel plates (Eastman No. 6060): O^2-alkyl-Thd, O^4-alkyl-Thd, O^2-Et-dUrd, O^4-Et-dUrd, O^2-alkyl-Urd, O^4-alkyl-Urd, 7-Et-dGuo, 1,7-diEt-Guo, and O^2-alkyl-Cyt. The paper chromatographic systems were 80 1-butanol, 10 ethanol, 25 water (solvent A); or 75 ethanol, 30 1 M pH 7.5 ammonium acetate (solvent B). Silica gel plates were devel-

† From the Department of Molecular Biology and Virus Laboratory, and Space Sciences Laboratory, University of California, Berkeley, California 94720. *Received September 8, 1977.* This investigation was supported by National Institutes of Health research Grant CA 12316.

‡ M.K. is a Fellow of the Deutsche Forschungsgemeinschaft.

[1] Abbreviations used: Me, methyl; Et, ethyl; Ado, Guo, Urd, and Cyd refer to ribonucleosides; dAdo, dGuo, dUrd, dCyd, and Thd refer to deoxyribonucleosides; Ade, Gua, Ura, Cyt, and Thy refer to bases. A single letter abbreviation refers to the base moiety in a polynucleotide.

0006-2960/78/0417-1246$01.00/0