Shoup, R. R., Miles, H. T., and Becker, E. D. (1972), *J. Phys. Chem. 76*, 64.
Stevens, C. L. (1974), *Biopolymers 13*, 1517.
Stevens, C. L., and Rosenfeld, A. (1966), *Biochemistry 5*, 2714.
Studier, W. F. (1965), *J. Mol. Biol. 11*, 373.
Thrierr, J. C., Dourlent, M., and Leng, M. (1971), *J. Mol. Biol. 58*, 815.
Trifonov, E. N., Lazurkin, Yu. S., and Frank-Kamenetskii, M. D. (1967), *Mol. Biol. (Moscow) 1*, 164.
Van Holde, K. E., Brahms, J., and Michelson, A. M. (1965), *J. Mol. Biol. 12*, 726.

Formaldehyde as a Probe of DNA Structure. 4. Mechanism of the Initial Reaction of Formaldehyde with DNA†

James D. McGhee‡ and Peter H. von Hippel*

ABSTRACT: Formaldehyde is used as a probe of the dynamic behavior of native DNA. We first show experimentally how initial denaturation rates vary with temperature, formaldehyde concentration, DNA melting temperature, and DNA molecular weight. Electron microscopy of DNA from the initial phases of the reaction verifies that denaturation initiates at AT-rich regions in the interior of the DNA molecule. The overall denaturation rate is shown to increase with increasing pH. Since the only pH-dependent chemical reaction is at the imino group of thymine (and guanine) located directly in the middle of the Watson-Crick helix, it is concluded that interchain hydrogen bonds do indeed break prior to reaction. By studying denaturation rates as a function of temperature and pH, it is shown that under the usual reaction conditions denaturation involves adduct formation with the functional groups of both thymine and adenine. The thymine reaction (which is rapidly reversible) dominates the denaturation under conditions of high temperature and high pH; conversely, the adenine reaction can be considered to be effectively irreversible and analysis of reaction rates under adenine-reaction-dominated conditions is a vast simplification. By examining reaction rates with single-stranded polynucleotides as a function of temperature, we conclude that bases must unstack prior to reaction with formaldehyde; following unstacking, the reaction rates are essentially identical to those of mononucleotides. We also demonstrate that monohydroxymethylated adenine can form a base pair with thymine, the hydroxymethyl group lying coplanar with the base pair and protruding into the major groove of the double-helical DNA structure. However, such a substituted base pair is ~1.5 kcal/mol less stable than an unreacted AT pair. This stability difference can be quantitatively ascribed to simple stereochemistry and can be used to determine the number of neighboring base pairs which are denatured as a consequence of one chemical reaction. Thus, the initial reaction of formaldehyde with an adenine moiety in double-helical DNA proceeds as follows: (1) a small sequence of DNA base pairs denatures (i.e., interchain hydrogen bonds break and bases unstack) as a consequence of a local thermal fluctuation; (2) the exocyclic amino group of an adenine residue exposed in this spontaneous fluctuation reacts at the *same rate* as does the free mononucleotide under comparable reaction conditions; and (3) the reacted adenine either re-forms into a (less stable) hydroxymethylated AT base pair, or remains unstacked and unhydrogen bonded, depending on temperature and other environmental factors. Taking these three steps into account, reaction rates are predicted from simple helix-coil theory using the experimentally determined loop-weighting functions of Gralla and Crothers (Gralla, J., and Crothers, D. M. (1973), *J. Mol. Biol. 78*, 301). Agreement between calculations and observations is within a factor of two for denaturation rates and within ~1 kcal/mole for apparent activation energies. Conversely, the experimental data can be used to obtain independent estimates of loop-weighting functions describing the behavior of small open loops in DNA, as well as to calculate relative rates of reaction with HCHO at internal nicks and helix ends. The central conclusion from these calculations is that the most probable transiently denatured state of DNA at temperatures below T_m consists of "loops" containing only one open (unstacked and unhydrogen bonded) base pair. The definition, in this series of papers, of the initial reaction mechanism of DNA with formaldehyde should serve as a partial model for the more complex molecular pathways involved in processes of genome regulation, such as the interaction of melting proteins with initially native DNA sequences or RNA polymerase with initially "closed" promoter regions.

With this paper we complete our studies of the reaction of formaldehyde with DNA. In the first two papers of the series (McGhee and von Hippel, 1975a,b), we studied the basic chemistry of the formaldehyde reaction with the DNA bases. In the third paper (McGhee and von Hippel, 1977), we described, both experimentally and theoretically, the overall *equilibrium* denaturation of DNA by formaldehyde. Finally, in this paper we attempt to establish the complete sequence of chemical and conformational steps involved in the reaction of formaldehyde with an individual DNA base pair.

As described in preceding papers, two motivations underlie these studies. The first is to establish the thermodynamic behavior of a simple "melting-protein" model, as a prelude to

† From the Institute of Molecular Biology and Departments of Chemistry and Biology, University of Oregon, Eugene, Oregon 97403. *Received September 18, 1976; revised manuscript received April 4, 1977.* These investigations were supported in part by United States Public Health Service Research Grants GM-15792 and GM-15423. This work has been submitted (by J.D.M.) in partial fulfillment of the requirements for the Ph.D. degree at the University of Oregon.

‡ Present address: Laboratory of Molecular Biology, NIAMDD, National Institutes of Health, Bethesda, Md. 20014.

considering the thermodynamics of a "real" melting protein–DNA system (McGhee and von Hippel, 1977; Jensen et al., 1976). The second is to establish the complete kinetic pathway for the reaction of such a melting-protein-type probe with DNA, in order to define the nature of the (low temperature) conformational fluctuations or "breathing motions" of native DNA which could be utilized by, or at least are available to, melting proteins, polymerases, and nucleases in their interaction with DNA under physiological conditions.

The kinetics of DNA denaturation by formaldehyde have been studied many times in the past (e.g., see reviews by Feldman (1973) and by Frank-Kamenetskii and Lazurkin (1974)). However, all these previous studies have been indirect, in that a reaction mechanism has been surmised, given quantitative expression in a more-or-less complicated rate equation, and then numerical values for the reaction parameters associated with the postulated mechanism have been extracted from the overall kinetic curve. In no case has direct evidence been presented to describe the conformational and chemical nature of the reactive state.

In this paper we make use of various synthetic polynucleotides as simple and convenient models to define the major pathway for the initial events of DNA denaturation by formaldehyde. We provide direct evidence that, in order for DNA to react with formaldehyde, interchain hydrogen bonds must first break and bases must transiently unstack. Thus, the premelting fluctuations of the DNA helix which are monitored by the formaldehyde probe are indeed the low-temperature analogues of those conformations which, at high temperatures, dominate the overall helix–coil transition. Finally, by focusing on the low-pH, low-temperature (effectively irreversible) adenine reaction, we demonstrate that the observed denaturation rates can be calculated quantitatively through the use of simple helix–coil theory incorporating the experimentally determined loop-weighting functions of Gralla and Crothers (1973). These calculations should provide a precedent for quantitative analysis of the mechanisms whereby genome-regulating proteins interact with small loops in otherwise native DNA.

Materials and Methods

Formaldehyde solutions were prepared and standardized as described previously (McGhee and von Hippel, 1975a). Calf thymus, *Clostridium perfringens,* and *Micrococcus luteus* DNAs were obtained from Worthington Biochemical Corp., and synthetic polynucleotides from Miles or P-L Biochemical Co. Solutions of the nucleic acids were prepared and purified as described in the preceding paper of this issue (McGhee and von Hippel, 1977). Wild-type T7 bacteriophages were purified by polyethylene glycol precipitation and equilibrium banding in CsCl; DNA was isolated by four to six extractions with redistilled, buffer-saturated phenol. DNA from heat-induced $\lambda_{cI857susS7}$ was prepared in a similar manner. The instrumentation used has been described previously (McGhee and von Hippel, 1977).

To study the kinetics of DNA denaturation at different temperatures, 0.1 mL of preincubated formaldehyde solution was added with a prewarmed pipet to 1 mL of preincubated DNA solution. Temperature changes during the 10–15-s mixing period were shown to be, at the most, 0.1–0.2 °C (McGhee, 1975); this imposes a limit of approximately ±10% on the accuracy of the measured initial denaturation rates.

Results and Discussion

We first describe, very briefly, some general features of the denaturation kinetics, considering in particular how the initial

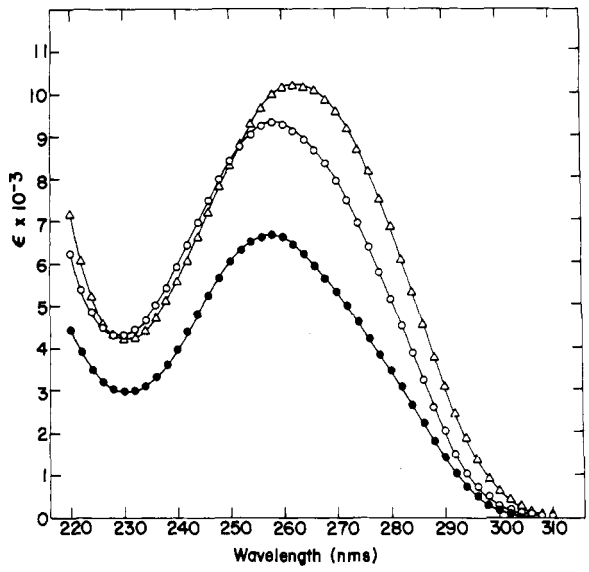

FIGURE 1: Spectra of T7 DNA, in 0.001 M phosphate, pH 7.0 (T_m = 54 °C); (●) 25 °C, no HCHO; (○) 75 °C, no HCHO; (△) 75 °C, 1.0 M HCHO.

rates depend on experimental variables such as temperature and formaldehyde concentration, as well as on DNA base composition, melting temperature, and molecular weight. It is only by understanding the interplay of all these connected variables that experimental conditions can be selected to isolate the individual steps of the reaction mechanism. In addition, a collection of experimental observations is provided which must be explained by any satisfactory reaction model.

Ultraviolet Spectra of DNA Denatured by Formaldehyde. By far the most convenient and easily interpretable method for observing the denaturation of DNA by formaldehyde is by following changes in UV absorbance. In this section, we investigate this procedure, consider its limitations, and indicate several possible errors in previous uses.

Figure 1 shows three ultraviolet spectra of T7 DNA (plotted as molar extinction coefficients). These spectra represent, respectively, native T7 DNA at 25 °C, denatured T7 DNA at a temperature above its melting temperature, and denatured T7 DNA at this same high temperature but reacted to equilibrium with 1 M HCHO. There are obviously two contributions to the increased absorbance (relative to native DNA) seen in the third spectrum: these are the absorbance increase due to DNA denaturation in the absence of any chemical reaction (i.e., the ultraviolet hyperchromicity due to the loss of the stacking interactions of the native double helix) and the increased absorbance due to the hydroxymethylation of the DNA bases themselves (McGhee and von Hippel, 1975a,b, and references therein). Similar spectra have been previously reported for a number of different DNAs and double-stranded polynucleotides (Haselkorn and Doty, 1961; Boedtker, 1967; Trifonov et al., 1967; Utiyama and Doty, 1971; von Hippel and Wong, 1971).

To separate the absorbance change due to denaturation from that due to the chemical reaction, the usual procedure has been to try to select a wavelength at which the absorbance increase reflects DNA denaturation independently of the chemical reaction of the bases. As can be seen in Figure 1, such a wavelength exists in the vicinity of 250 nm. However, a potentially serious objection to this approach is that the spectral properties of all regions of the DNA molecule are used to determine this wavelength, whereas, as will be shown below, the

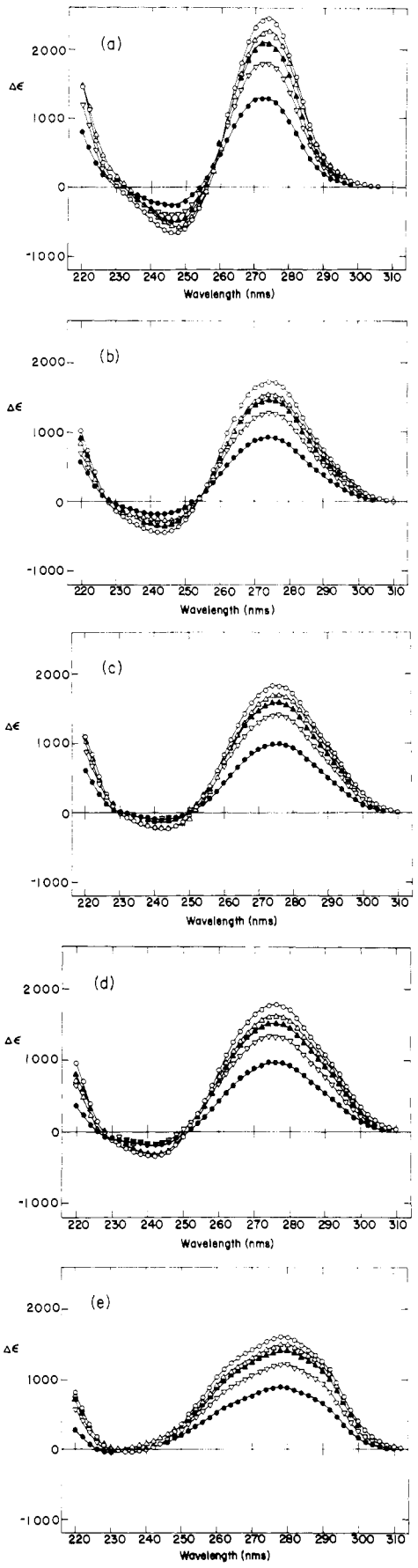

FIGURE 2: Difference spectra of DNAs of different base composition, equilibrated above the T_m with and without HCHO; (a) poly[d(A-T)]; (b) *Clostridium perfringens* (72% AT); (c) calf thymus (58% AT); (d) T7 (50% AT); and (e) *Micrococcus luteus* DNA (28% AT). In each graph, the symbols refer to different HCHO concentrations: (●) 0.20 M; (▽) 0.41 M; (▲) 0.62 M; (△) 0.82 M; (○) 1.02 M.

initial denaturation events occur in regions rich in AT base pairs.

To assess the magnitude of this potential error, a series of DNAs of different overall base composition (from 100 to 30% AT) was used as spectral models for the DNA regions that become progressively denatured during the time course of the reaction. The differences between spectra taken in the presence and absence of formaldehyde (and at temperatures above the T_m) are shown in Figure 2. For each type of DNA, the difference spectra do indeed show a fairly clearly defined wavelength at which absorbance increases upon DNA denaturation are largely independent of whether the bases are reacted; furthermore, this crossover wavelength varies only slightly with HCHO concentration (as well as only slightly with temperature). Most importantly, however, this crossover wavelength clearly depends upon DNA base composition, and moves to shorter wavelengths for increasingly GC-rich DNAs. Thus (at 75 °C), the wavelength at which DNA denaturation can be monitored independently of the chemical reaction is 254–256 nm for poly[d(A-T)] but falls somewhere between 230 and 240 nm for GC-rich *M. luteus* DNA.

In an effort to minimize errors due to this effect, we observe the initial denaturation of DNA by following the absorbance change at 254 nm (corresponding to the crossover wavelength for a DNA of about 75% AT) or at 255 nm (for poly[d(A-T)]. This should measure the *initial* extent of DNA denaturation independent of chemical reaction, to an uncertainty of about ±5%.

A further feature of the difference spectra of Figure 2 is that there is a wavelength (in the vicinity of 275 nm) at which the absorbance changes are maximally responsive to the hydroxymethylation reaction. The existence of these two wavelength regions, one apparently responsive primarily to conformational events and the other to the chemical reaction, has tempted many workers (e.g., Utiyama and Doty, 1971; von Hippel and Wong, 1971) to attempt to analyze the formaldehyde denaturation reaction in these terms. Thus, by monitoring absorbance changes at ~275 nm, as well as at the crossover wavelength of ~250 nm, Utiyama and Doty (1971) attempted to estimate the fraction of bases that had reacted, the fraction of base pairs that had been denatured but not reacted, and so forth, at various points during the overall formaldehyde-induced denaturation reaction. However, as can be seen from Figure 2, not only the exact wavelength maximally responsive to the chemical reaction, but also the apparent change in the DNA extinction coefficient as a function of extent of hydroxymethylation, depends upon DNA base composition. For example at 274 nm, the change in apparent extinction coefficient is about 60% larger for poly[d(A-T)] than for *M. luteus* DNA.

However, there is an additional, and more serious, objection to the type of two-wavelength analysis introduced by Utiyama and Doty (1971). This objection follows from the fact that the various reactive groups of DNA not only exhibit different spectral changes on reaction with HCHO, but also show markedly different chemical reaction *rates*. Thus, even in the homogeneous polynucleotide, poly[d(A-T)], both the thymine and the adenine bases can react with HCHO, and under the high pH conditions used by Utiyama and Doty, the chemical reaction with thymine can be 10^3 to 10^4 times faster than that with adenine (McGhee and von Hippel, 1975a,b). Furthermore, as will be shown below, under such conditions the thymine reaction is completely responsible for the denaturation process. Yet, at 274 nm, the absorbance change associated with the adenine reaction is about 50 times larger than that ac-

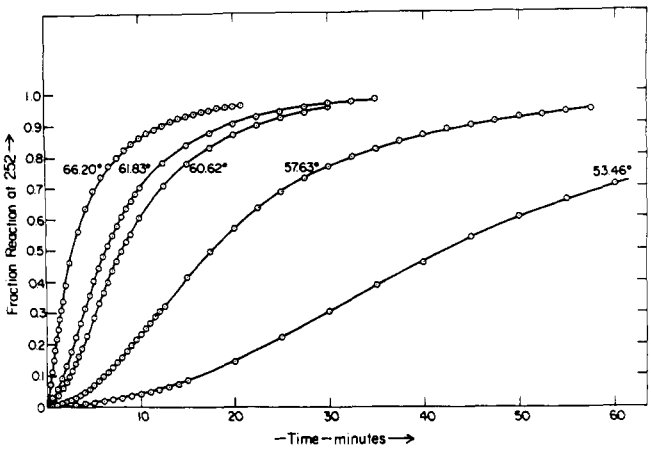

FIGURE 3: Kinetics of denaturation of calf thymus DNA by 1.06 M HCHO at various temperatures from 53 to 66 °C, as marked. Reaction observed by A_{252} and normalized to overall absorbance change; as described in the text, this wavelength measures (for calf thymus DNA) an average degree of helix denaturation independent of the chemical reaction. Buffer was 0.02 M phosphate, pH 6.95.

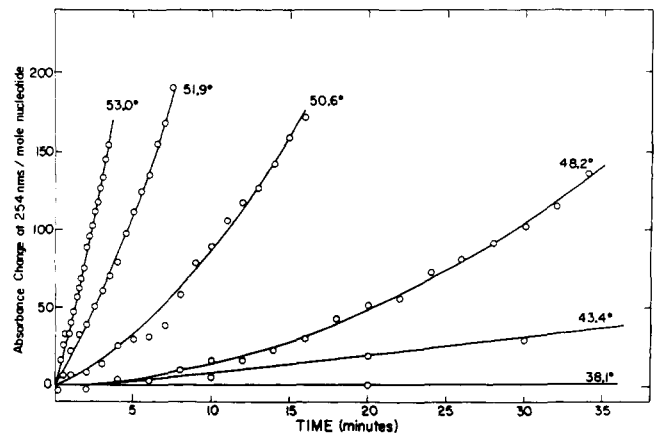

FIGURE 4: Initial phase of denaturation of T7 DNA by 1.0 M HCHO at temperatures ranging from 38 to 53 °C. Reactions observed at 254 nm and expressed as change in absorbance per mole of DNA nucleotide. Lines are least-squares best-fit quadratics, as described in the text (0.002 M phosphate, pH 7.1).

companying the thymine reaction (McGhee and von Hippel, 1975a,b). By following the chemical portion of the denaturation reaction at 274 nm, Utiyama and Doty assumed that the adenine reaction alone is responsible for helix denaturation, and thus were forced to conclude (erroneously) that each chemical reaction could induce the denaturation of large numbers (10–100) of neighboring base pairs.

Complete Time Course of Denaturation of DNA by Formaldehyde. In order to put subsequent measurements of initial denaturation rates into perspective, Figure 3 illustrates the complete time course of the denaturation of calf thymus DNA by 1 M HCHO, at a number of temperatures between 50 and 70 °C. As has been observed many times in the past (Berns and Thomas, 1961; Freifelder and Davison, 1963; Lewin, 1964; Trifonov et al., 1967; Lazurkin et al., 1970; Utiyama and Doty, 1971), the kinetics are sigmoidal or "autocatalytic" at low temperature. The usual interpretation of this curve shape (e.g., see Trifonov et al., 1967) is that there are two types of denaturation processes, an initial "nucleation" phase in which formaldehyde denatures isolated regions in the otherwise native DNA molecule, and a subsequent "growth" phase in which these denatured nuclei become larger.

However, as was first noted by Berns and Thomas (1961) and as can be seen in Figure 3, the curve shapes change with temperature, becoming more obviously sigmoidal at lower temperatures and more closely exponential as the reaction temperatures approach the DNA-melting temperature. In other words, the nucleation reaction has a greater temperature dependence than does the growth reaction. Indeed, the apparent activation energy of the initial rates is roughly twice that of the rates at 50% denaturation (McGhee, 1975).

Under the solvent conditions of Figure 3 (0.03 M Na^+), the unperturbed melting temperature of calf thymus DNA is 73 °C. Thus the reactions illustrated in this figure were run at temperatures fairly close to the unperturbed T_m of the DNA, but under conditions where, by optical criteria, the DNA is still completely native. Using the methods of the previous paper (McGhee and von Hippel, 1977) the *equilibrium* T_m of this DNA in the presence of 1 M HCHO should lie far below 0 °C, i.e., the DNA, at the end of all these reactions, should be completely denatured. This was indeed observed, the average relative change in A_{252} being 1.37 ± 0.006, essentially independent of the reaction temperature.

Experiments described in subsequent sections of this paper will focus on the initial "nucleation" phase of denaturation reactions such as those of Figure 3, and are designed to elucidate the mechanisms by which the *first* formaldehyde molecule reacts with otherwise native DNA. In order to estimate these initial rates objectively and with minimum error, the absorbance changes observed over the first 5–10% of the denaturation are fit by least-squares analysis to an equation of the form:

$$\Delta A_\lambda = a_0 + a_1 t + a_2 t^2$$

The "initial rate of denaturation", as used hereafter, is obtained by interpolation at 1% total denaturation. (This is a compromise between the more easily measured rates at 5 to 10% denaturation and the true "initial rate" at 0% denaturation, which is very sensitive to mixing artifacts; subsequent conclusions are all independent of the exact extent of denaturation at which these early rates are measured.) The data in Figure 4 cover the first 10% of the denaturation of T7 DNA (corresponding to a change of 0.02 in the absorbance at 254 nm); the solid lines are the least-squares best-fit quadratics, and are seen to be quite adequate representations of the data.

Dependence of Initial Denaturation Rate on Temperature. The initial denaturation rates measured in Figure 4 for T7 DNA, along with similar rates measured at lower HCHO concentrations, are plotted in Figure 5 as an Arrhenius plot. These rates cover the range that can be conveniently observed; at much higher temperatures the absorbance changes are faster than the mixing time, while at much lower temperatures the rates become so slow that measurements are inaccurate (the change in absorbance being on the order of 0.001/h, or even less). Over the region where the rates are most easily measured, the plots can be approximated as linear, with an apparent activation energy of about 80 kcal, as reported earlier by von Hippel and Wong (1971) and Fuchs and Daune (1974) for calf thymus DNA. However, over the more extended temperature range, the plots are discernibly nonlinear, the apparent activation energy varying up to about 120 kcal at temperatures close to the T_m and down to about 50 kcal at temperatures far below T_m. (The stoichiometric reaction unit to which this activation energy applies is, for the moment, left undefined.)

This large temperature dependence of the denaturation rates is also found with homogeneous polynucleotides; this will be demonstrated subsequently when we describe how these ap-

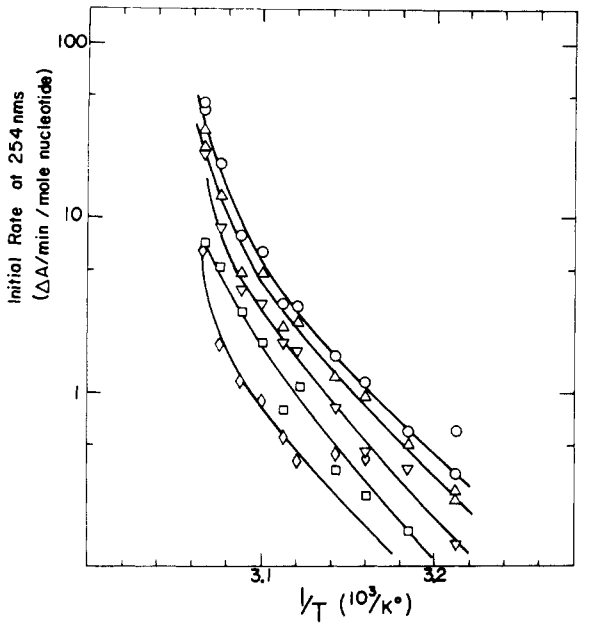

FIGURE 5: Arrhenius plot of initial denaturation rate of T7 DNA by various concentrations of HCHO: (○) 1 M; (△) 0.8 M; (▽) 0.6 M; (□) 0.4 M; and (◇) 0.2 M HCHO (0.002 M phosphate, pH 7.1).

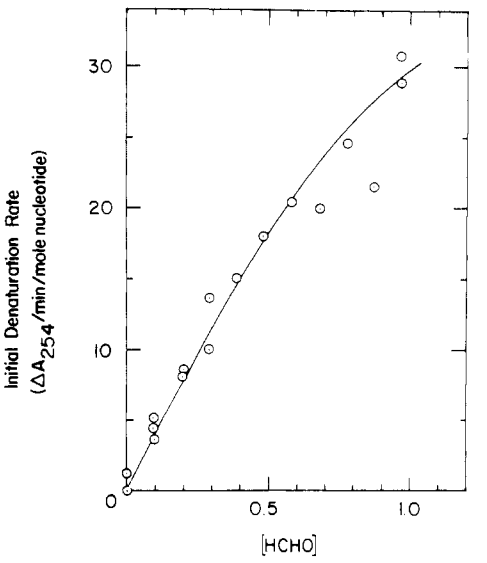

FIGURE 6: Dependence of initial denaturation rate of T7 DNA on formaldehyde concentration; temperature, 53.5 °C (0.002 M phosphate, pH 7.1).

parent activation energies themselves depend upon experimental conditions, particularly pH.

Dependence of Initial Denaturation Rate on Formaldehyde Concentration. Figure 6 plots the initial denaturation rate of T7 DNA as a function of formaldehyde concentration. At this temperature (54 °C, about 5 °C below the T_m) the denaturation rate monotonically increases with increasing formaldehyde. The relation is only approximately linear and appears to level off at high formaldehyde. At least under these conditions, the denaturation rate certainly does *not* increase as a function of some high power of the formaldehyde concentration, as might be expected from some types of cooperative models for the denaturation process.[1]

The Dependence of the Initial Denaturation Rate on DNA Melting Temperatures. At constant reaction temperature and formaldehyde concentration, the rate of DNA denaturation decreases as the DNA-melting temperature increases; i.e., as the DNA helix becomes more stable. This is illustrated in Figure 7a, which shows the first 10% of the denaturation of calf thymus DNA by 1 M formaldehyde at a constant reaction temperature (47.5 °C), as the T_m of the DNA is varied between 54 and 63 °C by means of added salt. Figure 7b plots the logarithm of these initial denaturation rates vs. the reciprocal of the melting temperature; the plots are linear with a slope that decreases with decreasing reaction temperature.

The shapes of the denaturation curves can also be seen to change with T_m, becoming more "sigmoidal" with T_m far above the reaction temperature, and more "exponential" as the two temperatures approach. This is the same pattern observed above (Figure 3) under conditions where reaction temperatures were varied and T_m values held constant. These two sets of results are both consistent with mechanisms in which nucleation of denatured regions has a stronger dependence on DNA-melting temperature than does the growth of such regions.

Initial Denaturation Occurs at AT-Rich Sites. Experiments were conducted to verify directly that, under the conditions of the kinetic experiments, denaturation begins at AT-rich sites in the DNA. To this end, phage λ DNA was mixed with 1 M HCHO at temperatures from 57 to 65 °C, and the absorbance followed as in Figures 3 and 4. At various times, samples were taken, chilled in ice, and immediately prepared for electron microscopy by a modified Kleinschmidt technique (Chattoraj and Inman, 1973). The degree of denaturation seen optically was always considerably greater than the degree of denaturation ultimately seen in the electron microscope, presumably because many small denatured loops go undetected in the EM procedure. Nevertheless, even for the earliest time points, the standard λ DNA denaturation pattern (Inman, 1967) was always seen, with only a very small fraction of the denaturation (<1%) occurring via the ends of the molecule. A histogram of such molecules measured at an average extent of denaturation of 10–15% is shown in Figure 8. Solvents which change the relative stabilities of AT and GC base pairs (Melchoir and von Hippel, 1973) completely alter this denaturation map (Chattoraj and McGhee, in preparation).

Importance of Helix Ends. We compared the initial formaldehyde-induced denaturation rates of two DNA samples of relatively high molecular weight, one consisting of intact T7 DNA and the other of T7 half-molecules produced by mild shearing. (The T_m of each sample was 59 ± 0.1 °C; alkaline sedimentation revealed 1–2 nicks/strand.) Over a 10 °C range of reaction temperatures, the initial denaturation rates were always within 10% of one another. Thus, at least for high-molecular-weight DNA, only a negligible fraction of the denaturation reaction initiates via helix ends. Nucleation primarily from internal loci is not imposed solely by heterogeneity in DNA base composition, since, under solvent conditions

[1] With poly[d(A-T)], experimental conditions can be found under which the dependence of denaturation rates on HCHO concentration is more complex than in Figure 6, but also more revealing (McGhee, 1975). Especially at low temperatures, the relation seems to be biphasic; i.e., there is a region at low HCHO where the denaturation rate is essentially zero, followed by a region where the denaturation rate increases roughly linearly with increasing formaldehyde concentration. This behavior can be completely rationalized by the equilibrium considerations of the preceding paper in this issue (McGhee and von Hippel, 1977), where it was demonstrated that, for any particular temperature, there is some formaldehyde concentration below which the DNA will not be denatured *at equilibrium*.

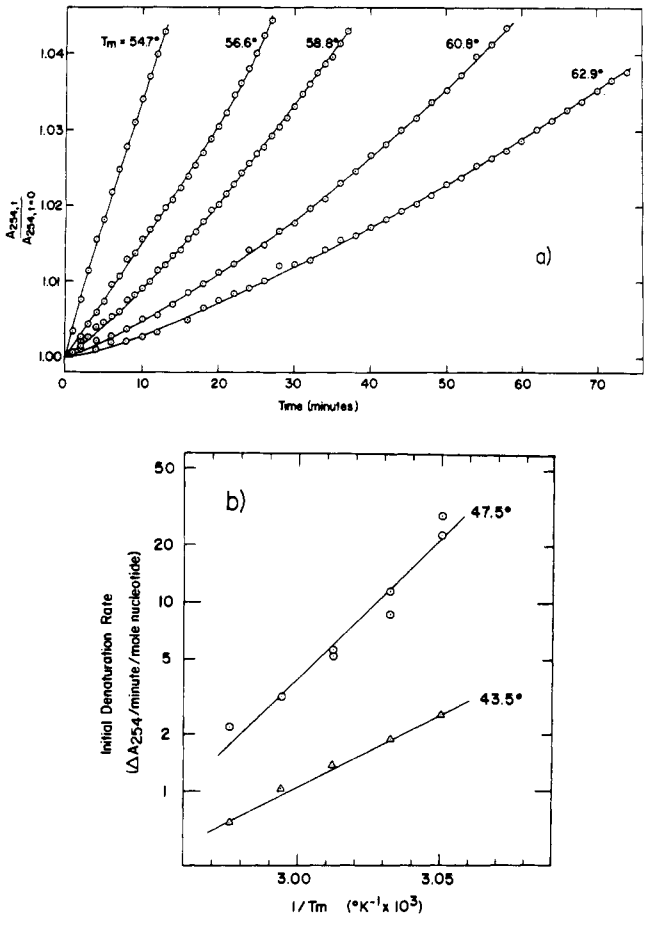

FIGURE 7: (a) Dependence of the initial denaturation rate of calf thymus DNA on melting temperature, T_m; 1.0 M HCHO; reaction temperature, 47.5 °C; buffer, 0.002 M cacodylate, 10^{-4} M EDTA, with various amounts of added NaCl to raise T_m from 54 to 63 °C, as marked on each curve; final pH, 6.2 ± 0.02. (b) Logarithm of the initial denaturation rate of calf thymus DNA (at a reaction temperature of either 43.5 or 47.5 °C) plotted vs. the reciprocal of the melting temperature (K^{-1}).

where the AT-GC stability differences are abolished, electron microscopy of partially denatured DNA indicates that the bulk of the denaturation is still internal and the ends are roughly as denatured as an average internal site (Chattoraj and McGhee, in preparation).

At a substantially lower molecular weight, the helix ends can and do serve as important sites for the initiation of denaturation (see Lazurkin et al., 1970). For example, the initial denaturation rates of a sample of sonicated T7 DNA (molecular weight of 800 000) was two- to fivefold greater, over a range of reaction temperatures, than the denaturation rate of a sample of molecular weight 16 000 000. However, considering that the decrease in molecular weight causes a 20-fold increase in the relative concentration of the helix ends, it can be calculated that at most 5 to 15% of the denaturation reaction of this higher-molecular-weight sample proceeds via helix ends.

It is much more difficult to assess the role of ends in the denaturation of poly[d(A-T)], since it is known (Inman and Baldwin, 1962) that this polymer can form internal self-complementary hairpin structures. Although Utiyama and Doty (1971) found little dependence of the denaturation rate on molecular weight for a series of nuclease-digested poly[d(A-T)] samples, it is not at all clear how the average length of a hairpin double helix is reflected in the molecular weight

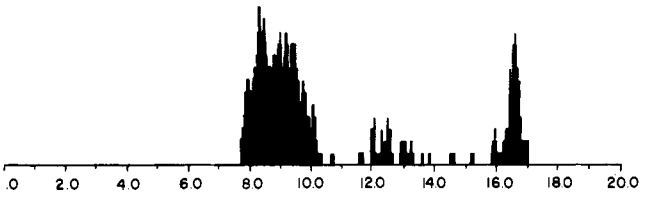

FIGURE 8: Histogram of 26 molecules of λ DNA partially denatured by 1 M HCHO at 58 °C, in 0.02 M phosphate, pH 6.95.

as determined by sedimentation analysis. In any event, at low temperatures the denaturation reaction of poly[d(A-T)] is clearly sigmoidal, suggesting the initial formation of internal nuclei. As argued above, due to its greater temperature dependence this nucleation process can be expected to become even more important at higher temperatures.

Possible Pathways of the First Reaction. The above experiments show that the initial reactions occur in AT-rich regions in the interior of the DNA molecule and now permit us to focus on the detailed chemical and conformational steps traversed by a single AT base pair in reacting with formaldehyde. In discussing possible reaction mechanisms, two things should be borne in mind. (1) All conformational changes of the DNA are extremely fast relative to the rates of the chemical reactions of HCHO with the DNA bases. Thus, the conformational state of the DNA can always be regarded as being at a temporary equilibrium.[2] (2) A hydroxymethyl group attached to the (planar) exocyclic amino group of adenine (and cytosine) can exist as two geometric isomers: either anti (trans) to N-7 and thus blocking normal Watson-Crick base pairing, or syn (cis) to N-7 and protruding into the large groove (see Engel and von Hippel, 1974). However, for steric reasons this group resides preferentially (by a factor of 10–20:1) as the anti isomer, preventing normal base pairing (Engel and von Hippel, 1974; Shoup et al., 1972; McGhee and von Hippel, 1975). Thus, a hydroxymethyladenine–thymine base pair should still be able to form, but should be less stable by at least the free-energy differences between the two isomeric positions. (This will be directly verified below.) As indicated for the other conformational processes, this isomerization can always be considered to be at equilibrium with respect to the chemical reactions.[3]

There are four possible pathways by which an initially double-helical (native) AT base pair can react with formaldehyde.

(1) The base pair "opens" and the thymine imino group reacts, completely preventing the re-formation of the base pair. (By "opening", we mean a process in which hydrogen bonds break and the bases move some distance apart; questions of how far apart the bases must be, whether they must be unstacked, how many neighboring base pairs are involved, etc., are considered in following sections.)

[2] For example, at 25 °C, neutral pH, and in 1 M HCHO, the fastest chemical reaction is that with the thymine imino group, which has a forward rate constant of about 0.1 s^{-1}. The slowest chemical reaction is that with the amino group of adenine, with a forward rate constant of about 10^{-4} s^{-1}, under the same conditions. To be compared with these rate constants is the rate constant for forming a base pair at the end of a helix, estimated to be in the range of 10^6 to 10^7 s^{-1} (Pörschke and Eigen, 1971; Craig et al., 1971; Pörschke, 1974). Simpler conformational steps, such as base unstacking (Pörschke, 1973), and rotation about backbone bonds (Akasaka, 1974), can be even faster.

[3] The lifetime of a methyl group in one isomeric position has been estimated to be approximately 1 to 10 ms at 50 °C (Engel and von Hippel, 1974).

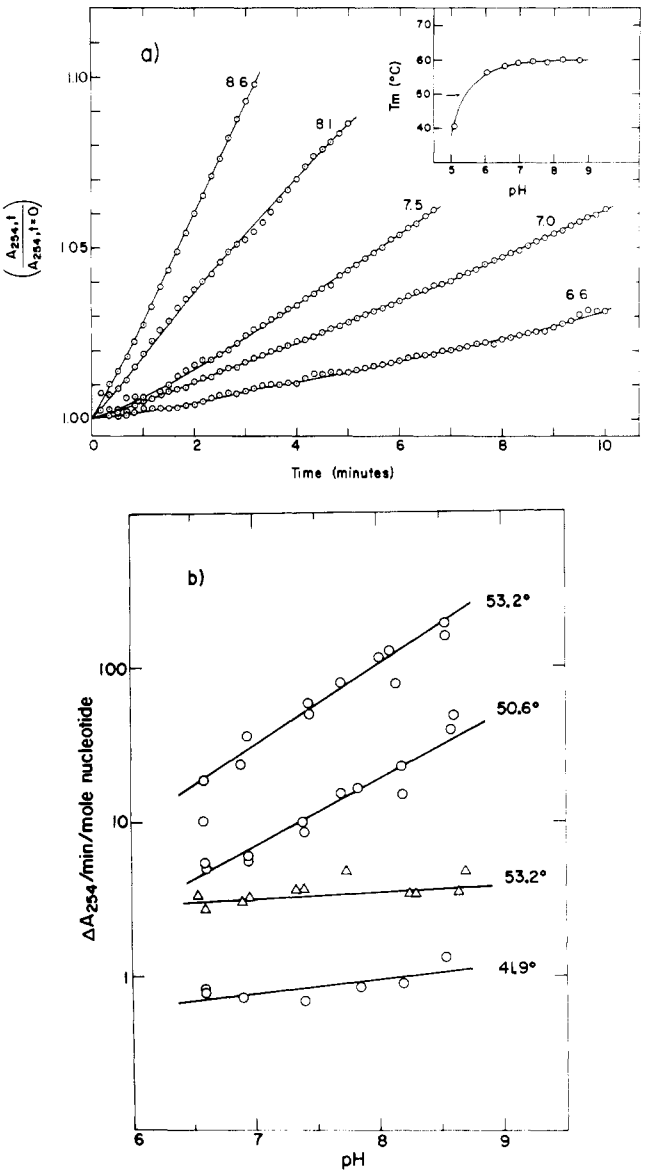

FIGURE 9: (a) pH dependence of the denaturation rate of T7 DNA by 1.06 M HCHO at 53.2 °C. [Na^+] = 0.003 M; pH as indicated on each curve. Inset shows relation of T_m to pH in the same buffers. (b) Logarithm of initial denaturation rate vs. pH for T7 DNA at temperatures and formaldehyde concentrations as marked: (△) 0.1 M HCHO; (○) 1.0 M HCHO; total [Na^+] = 0.003 M. Buffers used: acetate, cacodylate, phosphate, and borate.

(2) The base pair opens, the adenine amino group reacts, the hydroxymethyl group forms as the isomer anti to N-7, which again prevents base-pair formation by blocking Watson–Crick hydrogen bonds.

(3) The base pair opens, the adenine amino group reacts, but now the hydroxymethyl group forms as the rotational isomer syn to N-7, which thus allows the (now less stable) base pair to re-form.

(4) The base pair does not open, but rather formaldehyde simply replaces the outside nonbonded proton of the N^6 amino group of adenine via a direct "outside" reaction (e.g., see the proposal made by Lewin, 1966). As in alternative (3), the end result would be a hydroxymethyl group syn to N-7 and protruding into the large groove.

The number of possible steps subsequent to each of the above initial reactions increases rapidly. For example, after alternative 1, the thymine adduct could rapidly dissociate, allowing the base pair to re-form; after alternative 2, the hydroxymethyladenine group could isomerize into the syn position and allow the base pair to re-form; and after 3 and 4, the adducts either could accumulate (e.g., until some critical concentration has been reached), or the base pair could open to allow the hydroxymethyl group to isomerize into the favored anti position. The relative importance of these subsequent steps will be assessed in subsequent sections of this paper.

Reaction on the Outside of the Native Helix Is Unlikely. The basic chemistry of the reaction of formaldehyde with aromatic amines argues fairly strongly against alternative 4 above (i.e., that the helix does not have to open prior to the first reaction). The proposed mechanism for such reactions involves an initial attack of the carbonyl carbon of formaldehyde at the lone pair of electrons on the nitrogen atom (Jencks, 1964; McGhee and von Hippel, 1975a). The suggested transition state is a tetrahedral nitrogen atom, a geometry which would appear to require that the interbase hydrogen bonds be broken.[4]

In any case, "outside reactions" are not *necessary* for the denaturation process, since it has been demonstrated (McGhee, 1975) that HCHO can denature poly[d(N^6-methyl-A-T)], a double helix in which the aminomethyl group protrudes into the large groove and thus completely removes the possibility of either an outside reaction or the accumulation of outside adducts.

Hydrogen Bonds Break before Bases React. Over the pH range from 5 to 9, the rate of reaction of formaldehyde with the exocyclic amino groups of adenine, cytosine, and guanine is completely pH independent, whereas the reaction with the endocyclic imino groups of thymine (and guanine) is specific base catalyzed (McGhee and von Hippel, 1975a,b). Therefore, the only chemical reaction with a rate dependent on pH takes place at a group which is normally hydrogen bonded and buried in the center of the native Watson–Crick helix. Thus, if increasing the pH increases the rate of DNA denaturation by formaldehyde (under conditions where DNA stability as measured by T_m is independent of pH), this can be taken as direct evidence that the DNA helix is indeed opening as a first step and exposing the N3 group of thymine to the solvent.

Denaturation kinetics at different values of pH are shown for T7 DNA in Figure 9a; the insert shows that, in the same set of buffers, the T_m of the DNA is essentially pH independent. It is obvious that increasing the pH increases the rate of the helix denaturation, indicating that the helix does indeed open prior to reaction. Similar observations have been made with calf thymus DNA, poly[d(I-C)] and, as discussed in detail below, poly[d(A-T)].[5]

[4] This outside adduct could, in principle, form without breaking hydrogen bonds if the nonbonded amino proton simply dissociated ($pK_a \simeq 16$; Stewart and Harris, private communication), followed by a nucleophilic attack of this (planar) anion on the formaldehyde molecule (similar to the mechanism suggested for the reaction at the acid imino group (McGhee and von Hippel, 1975)). However, there is fairly strong evidence against such a mechanism, at least at the monomer level, viz., the lack of dependence of reaction rate on either pH (close to neutrality) or amino pK_a (McGhee and von Hippel, 1975a). Of course this outside pathway could conceivably become more important if either the amino group chemistry is substantially altered by interbase hydrogen bonding, stacking, etc., or a pathway that is minor in free mononucleotides becomes the major reaction pathway when other routes are (sterically) blocked by the DNA structure.

[5] If the complete course of denaturation is followed at different pHs, it is apparent (especially with a DNA like calf thymus) that both the rate and the curve shapes change with pH. At high pH, where nucleation by the imino reaction is fast, the reaction appears exponential; at low pH, where nucleation is slow, the reaction curves are more sigmoidal.

Initial denaturation rates, measured at a number of different pHs, temperatures, and formaldehyde concentrations, are plotted as a function of pH in Figure 9b. For T7 DNA, denatured by 1.0 M HCHO at high temperatures (the top two lines of Figure 9b), the rate increases about tenfold over 2 pH units, indicating that a substantial fraction (but not all) of the denaturation reaction proceeds by a specific base-catalyzed route.

However, both for high formaldehyde concentrations at low temperature (bottom line of Figure 9b), and for low formaldehyde concentration at high temperature (second line from bottom), the denaturation rates become almost pH independent. This has also been seen with poly[d(A-T)] (see McGhee, 1975; and below). One possible explanation is that, under these conditions, most of the denaturation reaction proceeds via the "outside" route (4 above) which does not involve helix opening. However, it was argued above, from a chemical standpoint, that such a reaction mechanism is unlikely, and, while it could conceivably explain the decrease in the dependence of the rate on pH at low temperature, it is not clear how it could explain the decreased dependence on pH at the same temperature but at lower formaldehyde concentrations. A more likely explanation for these effects is that under such conditions the helix must still open, but now the majority of the initial reaction is with the adenine amino group and is therefore pH independent.

The Relative Importance of the Adenine and Thymine Reactions Varies with Temperature and pH. Why the dependence of the denaturation rate on pH should itself depend on both the temperature and formaldehyde concentration emerges from the equilibrium considerations of the preceding paper in this issue (McGhee and von Hippel, 1977). Even though the thymine reaction can be 10^3- to 10^4-fold faster than the adenine reaction, the equilibrium constant is about fivefold smaller. Thus, the thymine reaction *by itself* is much less effective in lowering the DNA T_m *at equilibrium*. It can be calculated that under conditions for which we observe a strong dependence of denaturation rate on pH (e.g., Figure 9b), the formaldehyde concentration and temperature are such that the thymine reaction *by itself* is thermodynamically sufficient to denature the DNA. Conversely, where a very weak dependence of rate on pH is observed, we can calculate that the formaldehyde concentration and temperature are such that the DNA cannot be denatured *at equilibrium* by the thymine reaction alone.

To investigate more thoroughly this interplay between reaction temperature and pH, the initial denaturation rates of poly[d(A-T)] by 0.5 M HCHO have been measured at a number of different temperatures and in two different phosphate buffers, one set at pH 6.3, the other at pH 7.9. (The T_m of the DNA was 48.5 ± 0.5 °C in both buffers.) The results, expressed as d(rate)/d(pH), are plotted vs. the reaction temperature in Figure 10a; it is clear that at the same T_m and the same formaldehyde concentration the dependence of the denaturation rate on pH increases with increasing reaction temperature. The denaturation rate becomes specific base catalyzed (i.e., completely dominated by the thymine reaction and d(rate)/d(pH) = 10) at a temperature slightly below the T_m. The denaturation rates become pH independent (i.e., completely dominated by the adenine reaction and d(rate)/d(pH) = 1] at much lower temperatures, in this case about 25 °C.

If our interpretation is correct, the temperature at which the denaturation rate of poly[d(A-T)] becomes pH independent should correspond to the equilibrium melting temperature of this polynucleotide in 0.5 M HCHO *if only the thymine reaction were allowed.* Using the methods of the preceding paper in this issue (McGhee and von Hippel, 1977), the equilibrium T_m can be calculated to be about 31 °C, shown as the arrow in Figure 10a. The agreement between the two temperatures is probably acceptable, both because the calculations are necessarily somewhat imprecise and because, at these low temperatures, the denaturation rates become extremely slow and difficult to measure.[6]

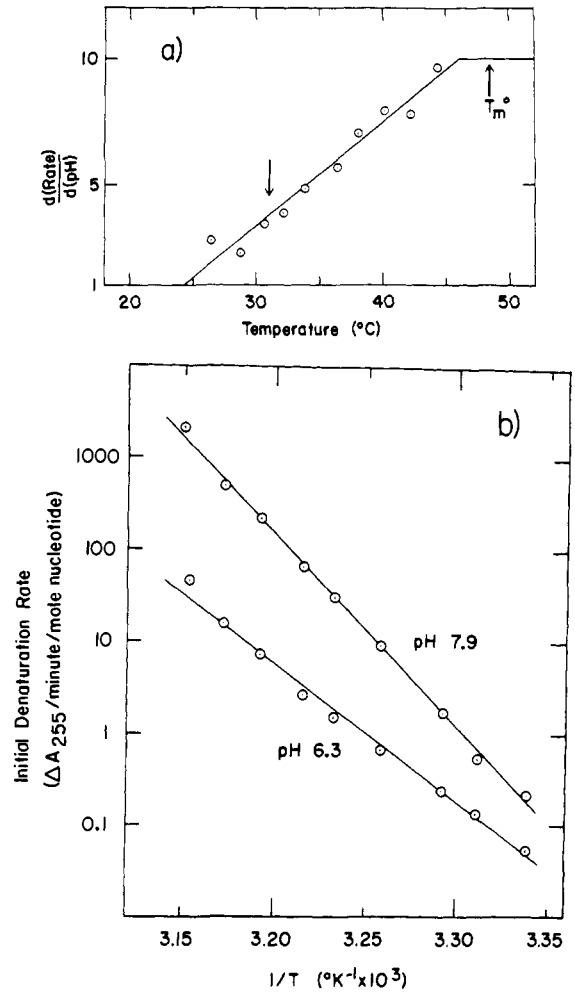

FIGURE 10: (a) pH dependence of the initial denaturation rates of poly[d(A-T)] as a function of temperature. Formaldehyde concentration = 0.48 M; [Na^+] = 0.02 M; phosphate buffers either at pH 6.3 or 7.9. The unperturbed $T_m°$ is indicated by the arrow at 48.5 °C. The arrow at 31 °C indicates the calculated equilibrium T_m of poly[d(A-T)] in the presence of 0.48 M HCHO, if only the thymine reaction could occur. (b) Arrhenius plots of the initial rate data used to construct Figure 10a.

Because the fraction of the denaturation reaction due to the thymine reaction increases at both high pH and high temper-

[6] It would not be surprising if this calculated equilibrium T_m did lie somewhat higher than the temperature at which the denaturation becomes pH independent; i.e., we might expect that the base-catalyzed thymine reaction is able to facilitate denaturation via the adenine reaction at temperatures where the thymine reaction cannot by itself cause DNA denaturation. In this regard, the thymine imino reaction could be viewed as analogous to the addition of alkali, which would lower the T_m without denaturing the DNA but would nonetheless increase the denaturation rates with formaldehyde, simply by decreasing the stability of the double helix. In other words, the adenine reaction would sense a lower DNA T_m, as a consequence of the *presence* of the thymine reaction. The agreement between calculated and observed temperatures in Figure 10a would indicate that this is probably not a major consideration.

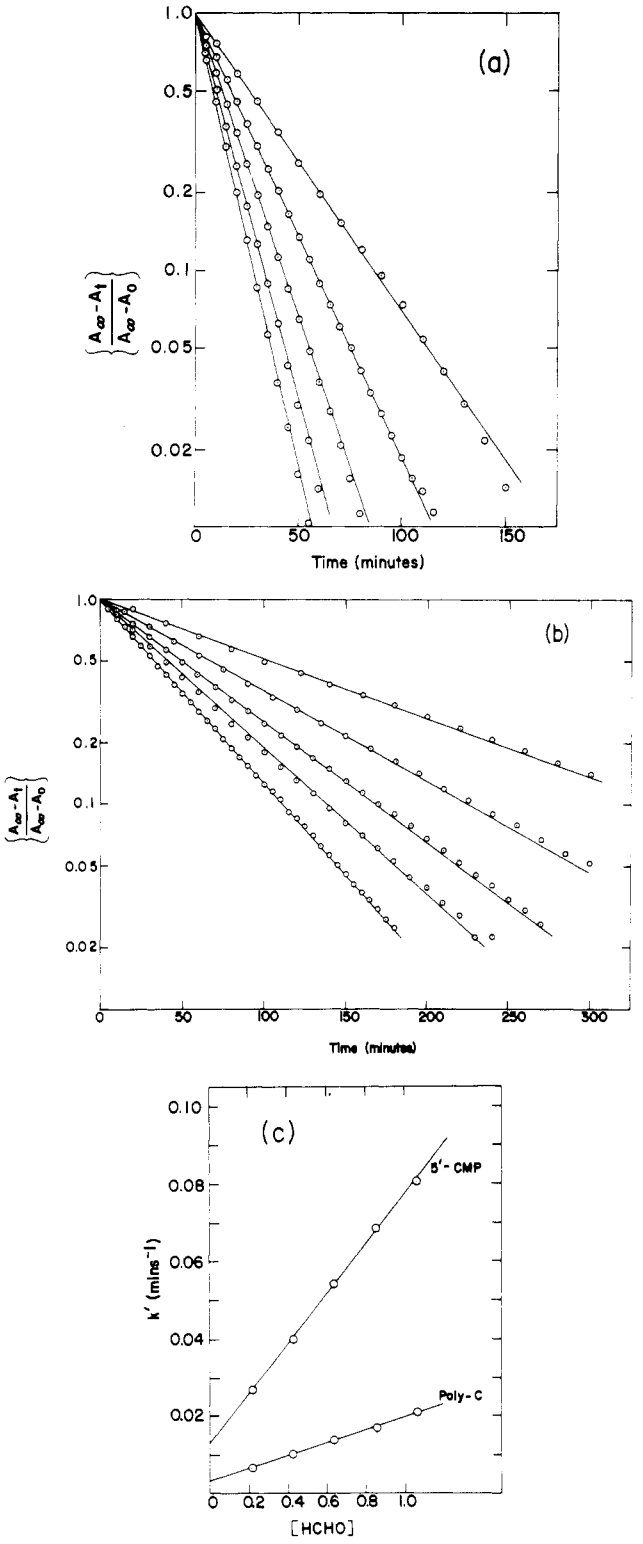

FIGURE 11: (a) First-order plot of reaction of 0.2–1.0 M HCHO with 5′-rCMP (288 nm). (b) First-order plot of reaction of 0.2–1.0 M HCH with poly(rC) (280 nm). (c) Pseudo-first-order rate constants (i.e., slope of lines in a and b) plotted vs. formaldehyde concentration for poly(rC) and 5′-rCMP (25 °C, 0.02 M phosphate, pH 7.0).

ature, the temperature dependence of the overall denaturation rate (i.e., the apparent activation energy) should also increase with increasing pH. This is clearly seen in Figure 10b, which shows an Arrhenius representation of the denaturation rates used to construct Figure 10a. The apparent activation energies are 96 and 69 kcal/mol at pH 7.9 and 6.3, respectively. In a subsequent section these results will be used to estimate the apparent activation energy which should apply under conditions where only the adenine or only the thymine reaction need be considered.

Thus, under the usual experimental conditions, the initial denaturation of DNA by formaldehyde is a combination of the denaturation caused by the thymine reaction and that caused by the adenine reaction. Furthermore, the relative importance of these two pathways varies with temperature, formaldehyde concentration, and pH; the higher the temperature or the pH, the greater the contribution of the thymine reaction.

The task of determining how the overall denaturation rates reflect the behavior of the unperturbed DNA molecule is greatly simplified if we consider (or extrapolate to) conditions where only one of these two chemical reactions need be considered. A serious objection to looking only at conditions where the thymine reaction causes denaturation is that the chemical reaction is quickly reversible; for example, at 40 °C and pH 8, the average lifetime of the thymine adduct is on the order of a tenth of a second (McGhee and von Hippel, 1975b). Thus, each adduct forms and dissociates many times during the overall denaturation process; this makes theoretical analysis overwhelmingly complex (e.g., see Craig et al., 1971).

In contrast to the thymine reaction, the dissociation of the adenine adduct is so slow that this reaction can (on the denaturation reaction time scale) be considered to be effectively irreversible; for example, at 40 °C the average lifetime of hydroxymethyladenine is about 100 min (McGhee and von Hippel, 1975a). As will be shown below, this irreversibility enormously simplifies theoretical analysis. In addition, as will be seen in the following two sections, model polynucleotides are available in which the stereochemistry of the adenine (and cytosine) reaction can be studied. Finally, the adenine reaction is dominant at temperatures far below the unperturbed DNA melting temperature, and thus closer to conditions of physiological interest. For all these reasons, we now concentrate on the denaturation caused by the formaldehyde reaction with adenine.

Bases Must Unstack Prior to Reaction with the Exocyclic Amino Groups of Adenine and Cytosine. After the interbase hydrogen bonds have been broken, a base can adopt a great number of positions relative to its former partner and to neighboring base pairs. In this section, we use single-stranded polynucleotide models to determine how base–base stacking interactions influence the rates of the chemical reaction.

Poly(dA), poly(rA), and poly(rC) are single-stranded polynucleotides in which the neighboring bases are extensively "stacked" on each other at low temperature; these stacking interactions break down in a gradual (noncooperative) fashion as the temperature is increased (see, e.g., Bloomfield et al., 1974). (The exact disposition of the bases is not known with any great degree of certainty, and thus the words "stacked" and "unstacked" are used rather loosely.) Haselkorn and Doty (1961) and Stevens and Rosenfeld (1966) have reported that the reaction of formaldehyde with poly(rA) and poly(rC) is slowed relative to the reaction with the component mononucleotides. We repeated these reactions at a number of temperatures in order to see if this slowing of the reaction rate could be ascribed to the temperature-dependent stacking interaction.

Reactions were followed at the peak of the absorbance difference spectra and typical pseudo-first-order plots (in this case at 25 °C and with formaldehyde concentrations ranging from 0.2 to 1 M) are shown in Figures 11a for 5′-rCMP and in Figure 11b for poly(rC). The formaldehyde reaction of both

TABLE I: Comparison of the Forward and Reverse Rate Constants for the Formaldehyde Reaction with Partially Stacked Single-Stranded Polynucleotides and with the Constituent Mononucleotides.[a]

Compd	Forward Rate Constant k_{12} (M^{-1} s^{-1} $\times 10^5$)	Forward Rate Constant $\frac{k_{12,polymer}}{k_{12,monomer}}$	Reverse Rate Constant k_{21} ($s^{-1} \times 10^4$)	Reverse Rate Constant $\frac{k_{21,polymer}}{k_{21,monomer}}$
Poly(rA)	4.4 ± 0.1	0.31	0.88 ± 0.03	0.29
5′-rAMP	14.2 ± 0.6		3.1 ± 0.3	
Poly(dA)	5.2 ± 0.1	0.36	1.5 ± 0.1	0.40
d(ApA)	9.0 ± 0.4	0.62	2.3 ± 0.2	0.64
5′-dAMP	14.5 ± 0.1		3.6 ± 0.1	
Poly(rC)	21.9 ± 2.8	0.22	3.7 ± 0.2	0.21
5′-rCMP	100.1 ± 4.2		17.7 ± 2.4	

[a] 25 ± 0.5 °C; 0.1 M Na_2HPO_4, 0.005 M NaH_2PO_4, pH 7.9; all polymers were verified to be single stranded, both from the wavelength of the absorbance maximum and from the absence of any cooperative change in absorbance with increasing temperature.

compounds accurately follows pseudo-first-order kinetics. The pseudo-first-order rate constants (i.e., the slopes of the lines of Figures 11a and 11b) are plotted vs. the corresponding formaldehyde concentrations in Figure 11c; the slope and intercept of such a plot correspond to the forward and reverse rate constant, respectively (McGhee and von Hippel, 1975a). Exactly similar data were obtained for poly(rA), 5′-rAMP, poly(dA), 5′-dAMP, and the dinucleotide, d(ApA). Forward and reverse rate constants (measured at 25 °C), as well as the ratios of the polymer to monomer rate constants, are collected in Table I. For all three polymers, the rate constants are slowed three- to fivefold relative to the constituent monomers, with the forward and reverse rates being slowed equally. The rate constants for the dimer, d(ApA), are intermediate between those for monomer and for polymer.

The forward rate constants, measured at temperatures from 5 to 35 °C, are plotted as an Arrhenius plot in Figure 12a for poly(rA) and 5′-rAMP, and in Figure 12b for poly(rC) and 5′-rCMP. Throughout the temperature range, the polymer rates are always slower than the monomer rates, but *the degree of slowing decreases with increasing temperature,* as would be expected if stacking inhibited the chemical reaction. The activation energies estimated from the slopes of the lines of Figure 12 are 3–5 kcal/mol greater for the polymers than for the monomers.

The above results suggest that base stacking inhibits the formaldehyde reaction. In the temperature range from 5 to 35 °C, the fraction of bases that are unstacked at any one time ranges from about 0.1 to 0.5, and the polymer rates are roughly this same fraction of the monomer rates. Furthermore, the polymer rates approach the monomer rates at high temperatures, where all the bases come unstacked.[7]

In order to analyze the data quantitatively, we describe the simplest possible model for the reaction mechanism. The temperature-dependent structures of poly(rA) and poly(rC) can be treated to a fairly good approximation by the common two-state model, in which a base is considered either to be

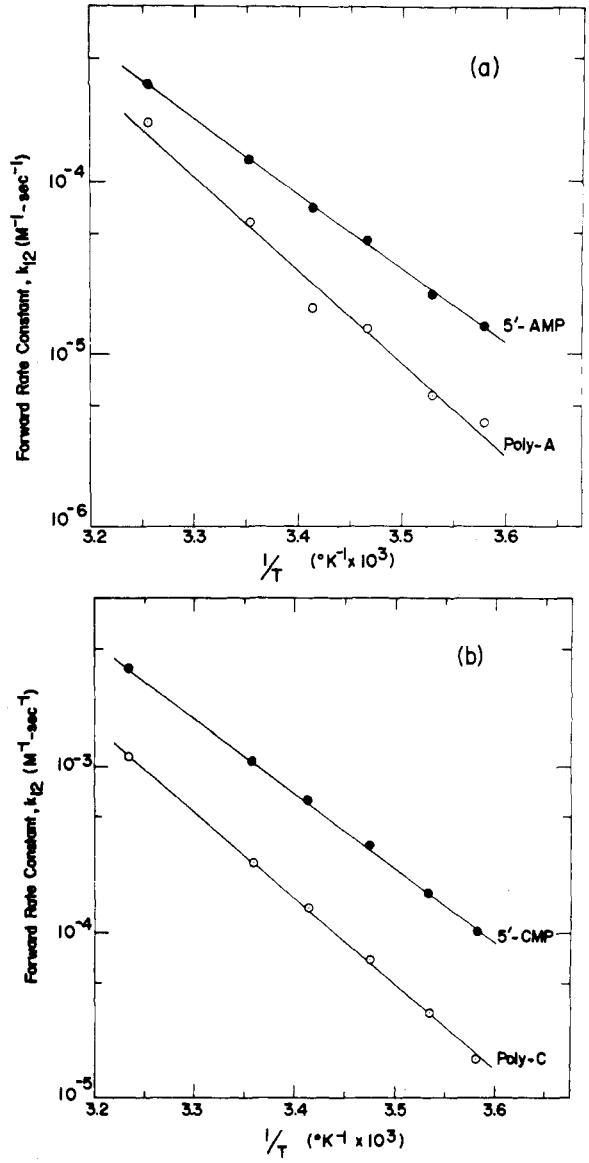

FIGURE 12: Arrhenius plot of k_{12}, the forward rate constant for formaldehyde reaction with (a) (○) poly(rA), (●) 5′-rAMP and (b) (○) poly(rC), (●) 5′-rCMP.

stacked or to be unstacked (Leng and Felsenfeld, 1966; Eisenberg and Felsenfeld, 1967; Lowe and Schellman, 1972). Further, the simplest assumption is that formaldehyde cannot react with a stacked base, but reacts with an unstacked base at the same rate as with the free mononucleotide.[8] Thus, the reaction scheme can be written as:

$$\underset{\text{stacked}}{\overset{s}{\rightleftharpoons}} \underset{\text{unstacked}}{\overset{k_{12}[HCHO]}{\underset{k_{21}}{\rightleftharpoons}}} \underset{\text{reacted}}{\overset{s}{\rightleftharpoons}} \underset{\text{reacted}}{}$$

This scheme implies that the equilibrium stacking tendency of a base is unchanged whether the base is reacted or not; this was shown to be true, within ± ~10%, in the previous paper of this issue (McGhee and von Hippel, 1977; see also Stevens, 1974). The scheme also implies that, as observed (see above),

[7] The reverse rate constants were also lower for the polymers than for the monomers throughout the temperature range investigated; however, the data were too scattered to make an analysis of their temperature dependence worthwhile. In addition, we found, with poly(rA) at low ionic strength, that the addition of LiCl (which is known to increase stacking; Leng and Michelson, 1968; Stannard and Felsenfeld, 1975) also decreases the reaction rate with formaldehyde.

[8] Since a stacking interaction occurs *between* two bases, this assumption means that one stacking interaction completely inhibits the reaction from the "top" of one base and from the "bottom" of the other base, or, on the average, inhibits the *overall reaction* of *one* base. However, this also implies that the rate of reaction of formaldehyde proceeding from the top of a base is the same whether or not the bottom of that base is involved in a stacking interaction, and this may or may not be so.

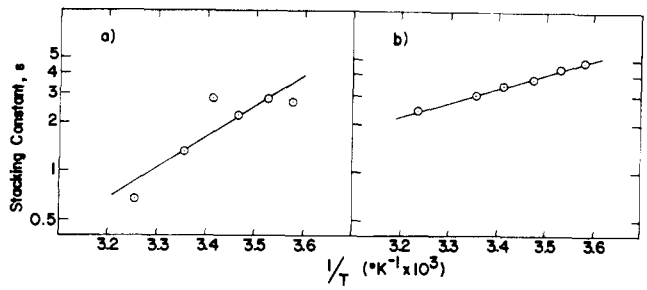

FIGURE 13: van't Hoff plot of the "stacking constants" for: (a) poly(rA) and (b) poly(rC), determined from the ratio of the rate constants as described in the text.

stacking should inhibit *both* the forward and the reverse rate constants.

Since the rates of interconversion of a base from a stacked to an unstacked state are many orders of magnitude faster than the chemical reaction rates (Pörschke, 1973), the forward rate constant for the polymer reaction, $k_{12,\text{polymer}}$, should then be given by:

$$k_{12,\text{polymer}} = k_{12,\text{monomer}} \begin{pmatrix}\text{equilibrium fraction of}\\ \text{unstacked bases}\end{pmatrix}$$

$$= k_{12,\text{monomer}} \left(\frac{1}{1+s}\right)$$

where s is the stacking equilibrium constant. Estimates of s at any temperature can thus be made by comparing the ratio of monomer and polymer rate constants; these are plotted as van't Hoff plots in Figure 13a for poly(rA) and in Figure 13b for poly(rC).

For poly(rA) the data are fairly scattered, primarily due to uncertainties in measuring the very slow reaction rates at low temperatures (which can correspond to an absorbance change of as little as 0.01–0.05 per day). Nevertheless, a least-squares fit to the data of Figure 13a yields an estimate of the stacking enthalpy of 8.6 ± 2.4 kcal/base, right in the middle of the range of values estimated by optical melting curves analyzed by a two-state model (Brahms et al., 1966; Leng and Felsenfeld, 1966; Davis and Tinoco, 1968). The van't Hoff plot for poly(rC) in Figure 13b yields an estimate for the unstacking enthalpy of 4.1 ± 0.1 kcal/base; this is comparable to, though lower than, the values of 9.6 kcal/base estimated for poly(rC) (Leng and Michelson, 1968) and 6.9 kcal/base estimated for r(CpC) (Davis and Tinoco, 1968). Taken together, these results indicate that a base must indeed be unstacked for the exocyclic amino group to react with formaldehyde, and that an unstacked base reacts at the same rate as the equivalent free nucleotide.[9]

The conclusion that an unstacked base reacts at the same rate as the free nucleotide can be verified by measuring the rate of reaction of DNA at temperatures above the melting temperature. To this end, poly[d(A-T)] and poly[d(I-C)] (both at 42 °C in 0.003 M Na^+ and thus above the respective melting temperatures) were reacted with formaldehyde; under these conditions, the reaction with the imino groups of thymine and inosine should be complete within 30 s, and any absorbance changes on a longer time scale should reflect reaction at the exocyclic amino groups. The observed kinetics were excellent pseudo-first-order (McGhee, 1975), and the average ratio of the forward rate constants observed to those calculated at the same temperature for the mononucleotides was 0.94 ± 0.02.

Stacking presumably inhibits the formaldehyde reaction by straightforward steric hindrance. As pointed out above, the initial step in the proposed reaction mechanism is an attack of the formaldehyde carbonyl carbon on the amino nitrogen lone pair from a direction perpendicular to the plane of the base (Jencks, 1964; McGhee and von Hippel, 1975a); detailed models of nucleoside dimers and single-stranded polynucleotides (see, e.g., Ts'o, 1974; Saenger et al., 1975; Arnott et al., 1976) indicate that such an attack would be inhibited by base stacking. The data on reaction rates of the dimer, d(dApA), listed in Table I also support the above interpretation. For example, if we assume that the fraction of stacked bases in d(ApA) is the same as in poly(dA) and is given by the ratio of the polymer to monomer rate constants (i.e., 0.38 ± 0.02, from Table I), then the reaction rate of the dimer can be calculated assuming that the HCHO attack on the amino group is unhindered from the direction of the two ends of the molecule, but is prevented from the direction of the middle if the bases are stacked. On this assumption, the reaction rate of d(ApA) relative to dAMP can be calculated as 0.5 + (0.38 × 0.5) = 0.69, reasonably close to the value of 0.63 ± 0.01 actually observed.

An alternative (and to some degree nonexclusive) explanation for the slowing of the polymer reaction rate is that base stacking lowers the pK_a of the amino groups, since the reaction rate is decreased about sevenfold by a decrease in amino pK of 1 unit (McGhee and von Hippel, 1975a). A more comprehensive study of dimers, trimers, and other polynucleotides could possibly resolve these two mechanisms, as well as yield fairly subtle information about polynucleotide geometries in solution.

As mentioned in the previous section, one of the reasons for focusing on the exocyclic amino reaction was that polynucleotide models were not available to study the effect of base stacking on the endocyclic imino reaction. Since the proposed mechanism for the latter reaction is an attack of the formaldehyde molecule *in the plane* of the ionized base (McGhee and von Hippel, 1975b), it is not clear whether base unstacking would indeed be required. What is clear, however, is that both forward and reverse rate constants for the formaldehyde reaction with the unstacked bases of poly(rU), poly(rI), and poly(dT) are the same as those for the constituent monomers, after allowance is made for the different pK_a values (McGhee, 1975).

Hydroxymethyladenine and Thymine Form a Weaker Base Pair. Continuing our consideration of conditions under which the adenine reaction dominates the denaturation, we shall now consider the steps immediately following the reaction of formaldehyde with a single open, unstacked adenine residue, and ask: does this reacted base pair remain denatured, does it re-form, or does it induce the denaturation of neighboring base pairs?

Since the lifetime of the hydroxymethyl adduct on the amino group of adenine is so long (about 10 min at 60 °C, 30 min at 50 °C, and about 12 h at 25 °C), it should be possible to investigate the behavior of such a derivative by equilibrating poly[d(A-T)] with formaldehyde at temperatures above the unperturbed T_m, followed by quick removal of excess formaldehyde (as well as of all the rapidly dissociating thymine adducts) by the high-temperature gel-filtration technique

[9] For poly(rC), where the data are fairly good, other models can be tested. For example, a model in which a base must be unstacked on both sides in order to react with the monomer rate implies that: $k_{12,\text{polymer}} = k_{12,\text{monomer}}(1/(1+s))^2$. Stacking constants estimated from this equation give a somewhat worse fit to a linear van't Hoff plot, and yield an even smaller estimate of 3.0 ± 0.1 kcal/base for the apparent enthalpy change. Furthermore, the model implies an unrealistically low value for the actual degree of base stacking as seen by hypochromicity.

(McGhee and von Hippel, 1977), and subsequent rapid cooling of the partially reacted polynucleotide. Any depression of the melting temperature from that of the unreacted control should reflect the denaturing ability of the remaining adenine hydroxymethyl adducts.

A series of such cooling curves is shown in Figure 14a (and described in more detail in the figure legend). It is seen that the greater the formaldehyde concentration during the preincubation period, the lower is the T_m and the broader the melting curve. Within experimental error, the overall hypochromicity is the same in all curves, and is essentially independent of formaldehyde concentration; thus, the poly[d(A-T)] helix must be completely re-formed.[10]

It can be shown that the hydroxymethyl groups are still attached to the native helix (and that they really are quite stable) both by direct chemical analysis of the bound formaldehyde (see below) and, perhaps more graphically, simply by remelting the reannealed samples. A series of three such successive reheating and recooling cycles is shown in Figure 14b (and described in more detail in the figure legend). In all the curves, the absorbance at 255 nm is expressed relative to the low-temperature absorbance observed at the start of the *first* cycle, and thus hypochromicity is completely regained between cycles. Also, the samples are completely melted far below the T_m of the original poly[d(A-T)]. The fact that each successive heating cycle of Figure 14b lies at a higher temperature undoubtedly reflects the dissociation of a certain fraction of the adducts during each cycle. After the third cycle of heating, if the heating is continued to 60 °C and the adducts are given sufficient time to dissociate, the T_m of the original poly[d(A-T)] is regained on cooling.[11]

It can also be demonstrated that the destabilization seen in Figure 14a,b is associated with a rapid process, such as isomerization about the exocyclic amino group. If experiments such as those of Figure 14a are repeated at a tenfold lower salt concentration, the reacted poly[d(A-T)] (after removal of excess formaldehyde) is still fully denatured at 25 °C. If NaCl is added to a final concentration of 1 M (thereby raising the T_m by about 50 °C), 98% of the overall helix hypochromicity is regained during a mixing time of about 30 s.

To analyze the T_m depression quantitatively, the amount of formaldehyde actually bound to the poly[d(A-T)] at each T_m must be estimated. At the optical concentrations of polynucleotide usually used, direct measurement of the bound formaldehyde by chemical assay or spectral analysis turned out to be inaccurate, and we thus calculated the amount of bound formaldehyde, using the known equilibrium constants and dissociation rates for the monomer reactions and the demonstration (above) that the reaction of the bases in single-stranded DNA is quite accurately described by the reaction of the free monomers.[12]

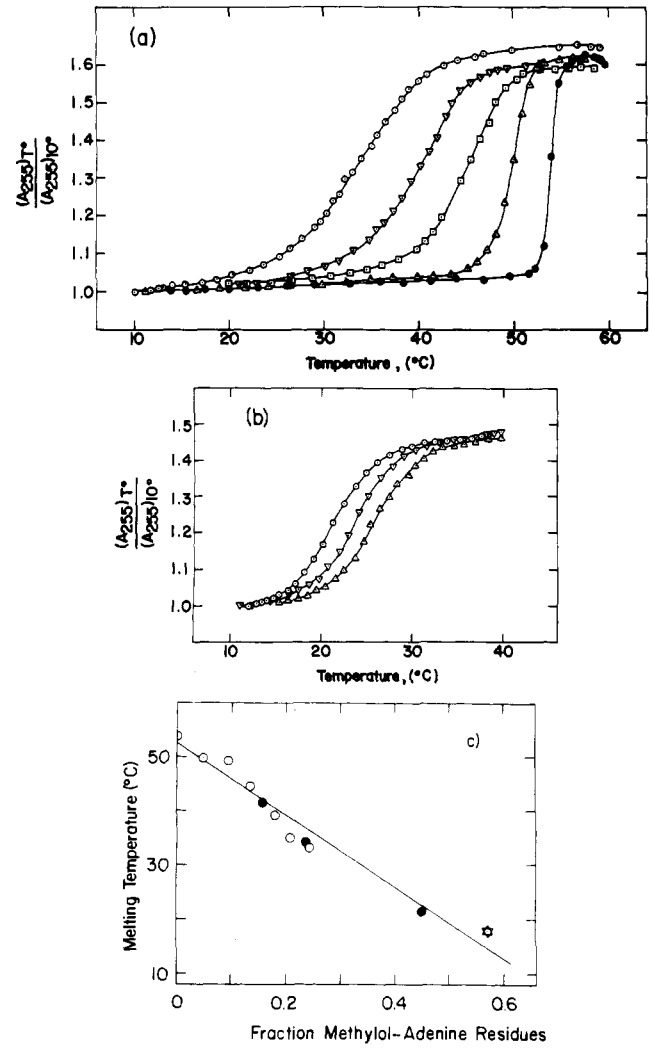

FIGURE 14: (a) Cooling curves of hydroxymethylated poly[d(A-T)]. Poly[d(A-T)] in 0.02 M phosphate, pH 7, incubated at 60 ± 0.10 °C for 1 h, with either (●) 0 M, (△) 0.03 M, (□) 0.14 M, (▽) 0.28 M, or (○) 1.07 M HCHO. Excess unbound formaldehyde was then quickly removed by centrifuging the sample twice through a short column of P-2 gel as previously described (McGhee and von Hippel, 1977). During the 2-min centrifuge runs, the gel tubes were kept in a 60 °C water jacket and reequilibrated for 1 min at 60 °C both between and after the two spins. The poly[d(A-T)] samples, in which the free formaldehyde concentration had been reduced to several micromolar or less, were transferred to prewarmed cuvettes (60 ± 0.5 °C) in the Gilford spectrophotometer and rapidly cooled (at a rate of 2.0 ± 0.1 °C/min), generating the indicated curves. (b) Three successive heating curves of hydroxymethylated poly[d(A-T)] originally incubated at 60 °C for 1 h with 1.07 M HCHO and excess formaldehyde removed: (○) sample cooled to 10 °C and reheated to 40 °C, (▽) sample recooled to 10 °C and reheated, (△) sample cooled to 10 °C and heated to 40 °C a third time. (c) Plot of observed T_m vs. fraction of hydroxymethyladenines present at the T_m. T_m values estimated from curves such as in Figure 14a; hydroxymethyladenine concentration estimated as described in the text: (○) cooling curves, (●) heating curves. For the point marked ☆, the experiment was done with tenfold concentrated poly[d(A-T)] and the amount of bound formaldehyde determined by chemical assay.

Figure 14c is a plot of the observed T_m vs. the fraction of adenine residues that are estimated to be still hydroxymethy-

[10] Any differences in hypochromicity actually seen between the runs in Figure 14 can be completely ascribed to a small (but variable) number of gel particles coming through the retaining mesh during the high-temperature gel filtration. This slight variability was also seen in the formaldehyde-only controls.

[11] These repeated heating and cooling cycles rule out the possibility that the low-melting temperatures seen in Figure 14a are simply due to the long times required for the adducts to dissociate, followed by re-formation of the unreacted double helix. Another possible alternative is that the low and broad melting curves are due to the short average helix length resulting from ejection of reacted residues into short hairpin loops which could then become kinetically trapped. However, the data of Elson et al. (1970) indicate that such hairpins would have to be less than 10 base pairs long to account for the observed low T_m, and this would certainly be detected as a reduced hypochromicity. Such convoluted metastable conformations are also not obtained with the poly[d(N^6-methyl-A-T)] analogue (Engel, 1975; Engel and von Hippel, in preparation).

[12] Full details of the calculations are given in McGhee (1975). The dissociation rates are calculated at each temperature using an activation enthalpy of 23.9 kcal/mol for the adenine reaction (and 25.2 kcal/mol for cytosine). These are calculated as the differences between the activation

lated at the T_m. Also shown are points from samples that were passed through the gel at high temperature, quickly cooled, and then reheated. One point in Figure 14c is from an experiment in which the poly[d(A-T)] was incubated at a concentration tenfold higher than usual, in order to permit the bound formaldehyde to be measured directly by chemical analysis (McGhee and von Hippel, 1977). All these data can be reasonably represented by a straight line; a least-squares fit gives a slope of −67 °C with a standard error of ±4 °C. However, a more reasonable error limit, in view of the large uncertainties in estimating the fraction of reacted residues, would probably be ±15 °C.

The interpretation of the experiments of Figure 14 can be summarized as follows. After the high-temperature gel treatment to remove excess formaldehyde, some of the adenines of the poly[d(A-T)] remain reacted as hydroxymethyladenines, the fraction increasing with increasing formaldehyde concentration. Hydroxymethylated base pairs are apparently less stable than their unreacted counterparts, but, at sufficiently low temperatures, can be incorporated into the Watson-Crick helix by isomerizing the hydroxymethyl group syn to N-7 (i.e., into the large groove). Since the partially reacted poly[d(A-T)] now represents a fixed sequence heteropolymer, the transition is broadened and the melting temperatures depend linearly upon the mole fraction of hydroxymethyladenine-thymine base pairs. Figure 14c would seem to be completely analogous to the classical Marmur-Doty plot in which melting temperatures of natural DNAs are linearly related to the mole fraction of AT base pairs (Marmur and Doty, 1962). Thus, hydroxymethyladenine and thymine simply form a weaker base pair than do adenine and thymine, and a pure hydroxymethyladenine-thymine polynucleotide would have a melting temperature 67 °C below that of unreacted poly[d(A-T)]. This quantitative measure of the destabilizing effect of a reacted adenine residue will be utilized in the subsequent theoretical calculations.[13]

General Discussion

In the first part of this paper, we considered general features of the denaturation of DNA by formaldehyde, and how the initial denaturation rates depend on experimental variables such as reaction temperature, formaldehyde concentration, and DNA melting temperatures. Denaturation was shown to initiate in AT-rich regions in the interior of the DNA molecule. It was also shown that, under the usual reaction conditions of neutral pH and temperatures close to the unperturbed T_m, the overall denaturation is due in part to the adenine reaction, and in part to the thymine reaction. The higher the pH and the closer the reaction temperature to the unperturbed T_m, the more important is the thymine reaction.

In order to obtain unambiguous information about the dynamic behavior of the unperturbed DNA, we decided to focus on experimental conditions of low pH and low temperature where only the adenine reaction need be considered. The reasons for concentrating on the adenine-induced denaturation were: (1) in contrast to the rapidly reversible thymine reaction, the adenine reaction can be considered to be irreversible, enormously simplifying theoretical analysis; (2) appropriate model polynucleotides are available to investigate both the effect of base stacking on the reaction rate and the behavior of the reacted base pair; and (3) the adenine reaction dominates at temperatures far below T_m, corresponding most closely to the physiological conditions of ultimate interest here.

The reaction mechanism by which an initially helical AT base pair reacts with formaldehyde can be described as follows. In order for formaldehyde to react with adenine, interchain hydrogen bonds must first break, and bases must unstack. Thus, the "reactive state" of the DNA which is probed by formaldehyde corresponds to a denatured site as described by conventional helix-coil theory. Furthermore, since the conformational steps leading to this reactive state are many orders of magnitude faster than the subsequent chemical reaction, the population of these reactive states should be calculable by *equilibrium* helix-coil theory as applied to the unreacted DNA molecule.

It was also shown that the unstacked adenine now reacts at rates essentially identical to the rates of free mononucleotides, but, once formed, the hydroxymethyl adduct dissociates so slowly that (on the time scale of the denaturation reaction) it can be considered as a permanent fixture, exactly like a methyl group. Since the exocyclic amino group can isomerize so that the hydroxymethyl group protrudes into the major groove, a hydroxymethyladenine-thymine base pair can re-form, especially at low temperatures. However, such a substituted base pair is considerably weaker than a normal AT base pair, and at high temperature will tend to be denatured, as well as to induce the denaturation of neighboring base pairs. Very importantly, these postreaction conformational events should also be described by *equilibrium* helix-coil theory, with the

enthalpies measured for the forward reaction, and the equilibrium enthalpy change associated with forming the monomethylol adduct, and agree within 1 kcal/mol with values directly measured for ribonucleotides (as in Figure 13 above). Values reported previously were too low because of a calculation error (McGhee and von Hippel, 1975a).

The overall correction for the dissociating adducts is *large* and only 30-50% of the original hydroxymethyl groups survive to the T_m. By far the greatest error in the estimated correction arises from temperature uncertainties during the gel procedure. Assuming that the sample spends 5 min at 60 °C and 5 min at 55 °C (rather than 10 min at 60 °C) undoubtedly underestimates the dissociating fraction and increases the final estimated adduct concentration by about 20%.

[13] This difference of about 1.5 kcal in the free energy of stabilization of the reacted and unreacted base pairs can be predicted from the behavior of the monomeric components. The simplest assumption is that a hydroxymethyladenine-thymine base pair is less stable than an adenine-thymine base pair solely because of the steric preference of the hydroxymethyl group to reside anti to N-7 and thus block hydrogen bonding. As mentioned above, estimates of this preference ratio are about 10-20 to 1, as directly measured for a methyl group (Engel and von Hippel, 1974), and are probably of similar magnitude for the hydroxymethyl group (McGhee and von Hippel, 1975a). Taking the enthalpy change for melting a poly[d(A-T)] base pair as 7.9 kcal (Scheffler and Sturtevant, 1969), a steric preference ratio of 10:1 or 20:1 would predict a decrease in T_m of 54 or 65 °C, respectively, certainly of the same magnitude as the destabilization actually observed.

There is a possibility however, that this steric-preference effect might be even larger in the polymer than in the monomer. Inspection of a CPK model of hydroxymethyladenine indicates that, in the geometric isomer syn to N-7 (i.e., the isomer which can be incorporated into the native helix), there are three possible rotational isomers of the hydroxymethyl group on the exocyclic nitrogen atom. Two of these isomers have the hydroxyl group protruding far above the plane of the adenine ring. Inspection of an alternating TAT sequence in the structure of B DNA (either on a CPK model or in a computer graphics display generated using published atomic coordinates (H. Schellman, unpublished results)) suggests that this hydroxyl group would interfere with the stacking of the neighboring thymine on one side of the reacted adenine residue, but probably not on the other side. Thus, assuming that only two of the three rotational isomers can be incorporated into the native helix, the equilibrium preference ratio favoring the isomer anti to N-7 would effectively be increased to 15:1 or 30:1. These ratios would correspond to a T_m depression of 60-71 °C, quite close to the value of 67 °C measured in Figure 14c. Thus, it appears that nothing more than steric hindrance need be invoked to explain the destabilizing effect of a hydroxymethyl group.

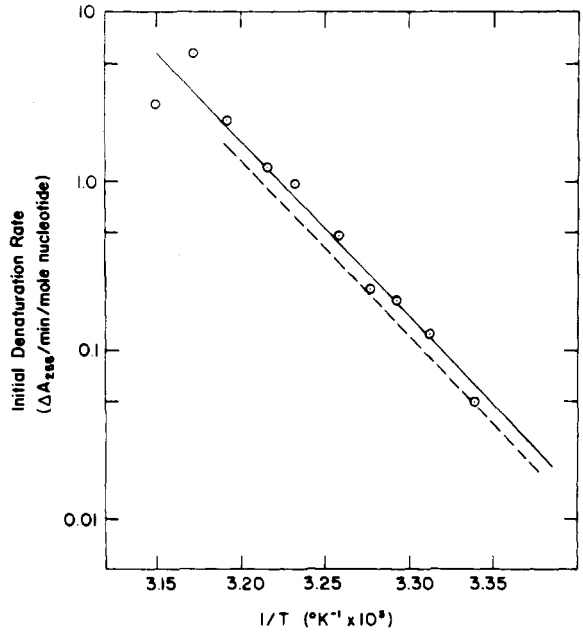

FIGURE 15: Arrhenius plot of the pH-independent initial denaturation rate of poly[d(A-T)] by 0.5 M HCHO, determined from the data of Figure 10 above. The dashed line is calculated theoretically and is described in the text.

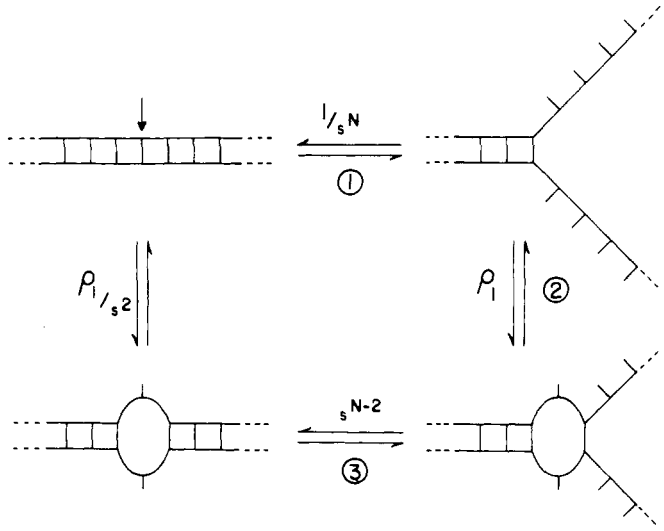

FIGURE 16: Definition of the equilibrium constant for denaturing a single base pair. A particular helical base pair (indicated by the arrow at upper left) is denatured in a loop of one base pair with an equilibrium constant, ρ_1/s^2, as indicated on the left. This process can be decomposed into the following steps. Step 1: N base pairs are opened starting from the (remote) helix end and including the selected base pair. This process has the equilibrium constant, $1/s^N$. Step 2: A loop of one denatured base pair is formed. This step has the equilibrium constant, ρ_1, and defines the loop-weighting function. Step 3: All the denatured base pairs outside of the loop zip up to re-form the intact helix. The equilibrium constant for this step is s^{N-2}. Thus, the overall equilibrium constant is: $(1/s^N)(\rho_1)(s^{N-2}) = \rho_1/s^2$, as indicated at the left.

hydroxymethyl-AT base pair being treated simply as a base pair with a lowered (and known) melting temperature.

In order to extrapolate to experimental conditions where only the adenine reaction need be considered (and where the above reaction mechanism will apply), the overall denaturation rates of poly[d(A-T)] from Figure 10 above have been expressed, at each temperature, as the sum of a pH-independent and a pH-dependent reaction. The calculated pH-independent (i.e., adenine dominated) reaction rates are plotted as an Arrhenius plot in Figure 15; the apparent activation enthalpy is 47.4 ± 3.6 kcal.

As expected, the pH-dependent (i.e., thymine-initiated) portion of the denaturation reaction has a much higher apparent activation enthalpy of 98.7 ± 2.0 kcal. The thymine reaction *must* be considered to be at least partially reversible, not only because it is fast but because the lowest reaction temperatures studied were close to the temperature at which the thymine reaction by itself cannot denature the poly[d(A-T)]. It has been previously demonstrated how such reversible denaturations can give rise to large apparent activation enthalpies (Saunders and Ross, 1960).

Prediction of Denaturation Rates for Poly[d(A-T)] from Helix-Coil Theory. We now attempt to put the above qualitative model into quantitative terms, and to predict the rates of denaturation of poly[d(A-T)] by formaldehyde under conditions where only the adenine reaction need be considered. That is, we try to calculate Figure 15.

Since all the measured initial rates are expressed as ΔA_{255} min^{-1} (mol of nucleotide)$^{-1}$, we essentially need to calculate the denaturation rate initiating at one particular base or base pair. There are *three* temperature-dependent steps that must be considered in the analysis. These steps are summarized in the following equation:

initial rate of change in A_{255} per mole of nucleotide
= (probability that a particular base pair is in a reactive, denatured state)
× (rate constant for the forward reaction with denatured adenine, at the particular formaldehyde concentration and reaction temperature)
× (change in extinction coefficient at 255 nm per reacted adenine base) (2)

In more detail, the three temperature dependent factors are the following.

(1) The probability that any particular base pair is in a reactive, i.e., denatured, state at any particular instant is the *equilibrium* fraction of coil, θ, calculable by a simple helix-coil theory. Any particular base pair can be denatured by itself, i.e., in a loop of one base pair, or along with one or more of its neighboring base pairs. Although larger loops are expected to be less likely on energetic grounds, their importance will be increased by a statistical factor (i.e., there are three ways in which a particular base pair can be in a three base pair loop, etc.), and we must a priori include them in the calculations. The exact contribution of these larger loops will be assessed later on.

The equilibrium constant for forming a loop of one base pair can be written as ρ_1/s^2, where s is the usual stability constant (i.e., the equilibrium constant for forming a base pair at the end of a helix, equal to unity at the T_m and increasing at lower temperatures according to the van't Hoff relation), and ρ_1 is the loop-weighting function for forming a loop containing one denatured base pair. The definition of this equilibrium constant is diagrammed in Figure 16 (see also Gralla and Crothers, 1973). The equilibrium constant for everting the particular base pair in a loop of two denatured base pairs is $2\rho_2/s^3$, and in a loop of three base pairs is $3\rho_3/s^4$. Experimental values for higher loop-weighting functions are not available and, in any case, will be shown below to be unimportant. Thus, we can

write the probability that a particular base pair is open and can react as:

$$\theta = \frac{\frac{\rho_1}{s^2} + \frac{2\rho_2}{s^3} + \frac{3\rho_3}{s^4}}{1 + \frac{\rho_1}{s^2} + \frac{2\rho_2}{s^3} + \frac{3\rho_3}{s^4}} \quad (3a)$$

Numerical values will be assigned below.[14]

(2) Since it has been shown above that the rate of reaction of an unstacked adenine base is essentially the same as that of a free mononucleotide, the chemical reaction rate is thus simply k_{12}[HCHO], where k_{12} is the forward rate constant for the reaction of 5′-dAMP or deoxyadenosine.

(3) The average change in extinction coefficient per reaction, $\Delta\epsilon$, depends strongly on temperature, since it arises from the fact that the number of base pairs held open by one reacted adenine varies with temperature. Since we have shown above (see Figure 14c) that hydroxymethyladenine can be treated simply as if it formed a weaker base pair (and since the half-life for adduct dissociation is about 10 h at 25 °C where the rates will be calculated), the number of base pairs denatured in the vicinity of the reacted adenine can be calculated by the same equilibrium methods used to calculate the probability of the initial opening event. Thus, the equilibrium constant for denaturing the particular reacted base pair can be written as ρ_1/ss^*, where ρ_1 and s are the same as in Figure 16, but s^* refers to the equilibrium constant for forming a hydroxymethyladenine–thymine base pair. As deduced from Figure 14c, s^* will be equal to unity at 67 °C below the T_m of the unreacted poly[d(A-T)] and will vary with temperature according to the van't Hoff relation. In exactly similar fashion, the equilibrium constant for everting the reacted base pair in a loop of two base pairs can be written as $2\rho_2/s^*s^2$, and for everting the reacted base pair in a loop of three base pairs is $3\Delta_3/s^*s^3$. A statistical factor appears in these equilibrium constants; e.g., the reacted base pair can be arranged three ways in a three base pair loop, etc.

It is important to realize again that, although the large loops are energetically less likely, they are favored not only by the statistical factor, but also because larger loops have a greater associated absorbance change. Since the best current approximation is that the hyperchromicity due to helix denaturation is proportional to the number of stacking interactions that are lost (see, e.g., Bloomfield et al., 1974), we can thus write the average change in extinction coefficient per adenine reaction as:

$$\overline{\Delta\epsilon} = \frac{\frac{\rho_1}{ss^*}(2\Delta\epsilon_0) + \frac{2\rho_2}{s^2s^*}(3\Delta\epsilon_0) + \frac{3\rho_3}{s^3s^*}(4\Delta\epsilon_0)}{1 + \frac{\rho_1}{ss^*} + \frac{2\rho_2}{s^2s^*} + \frac{3\rho_3}{s^3s^*}} \quad (3b)$$

where $\Delta\epsilon_0$ is the change in extinction coefficient (at 255 nm) per lost stacking interaction, as determined from the hyperchromicity of high-molecular-weight poly[d(A-T)].

Inserting expressions 3a and 3b into eq 2 above, the equation to be used to calculate the initial denaturation rate is:

$$\text{initial denaturation rate} = \left(\frac{\frac{\rho_1}{s^2} + \frac{2\rho_2}{s^3} + \frac{3\rho_3}{s^4}}{1 + \frac{\rho_1}{s^2} + \frac{2\rho_2}{s^3} + \frac{3\rho_3}{s^4}}\right) \times (k_{12}[\text{HCHO}]) \times \left(\frac{\frac{\rho_1}{ss^*}(2\Delta\epsilon_0) + \frac{2\rho_2}{s^2s^*}(3\Delta\epsilon_0) + \frac{3\rho_3}{s^3s^*}(4\Delta\epsilon_0)}{1 + \frac{\rho_1}{ss^*} + \frac{2\rho_2}{s^2s^*} + \frac{3\rho_3}{s^3s^*}}\right) \quad (4)$$

We now insert numerical values for the parameters in eq 4 and calculate the expected denaturation rate of poly[d(A-T)], under the conditions of the adenine-dominated reaction as plotted in Figure 15. We first try to predict the observed denaturation rate at 25 °C, since the experimental loop-weighting functions were estimated at this temperature (Gralla and Crothers, 1973). Numerical values for each term of eq 4 are assigned as follows.

(1) To calculate θ from eq 3a, we take the loop-weighting functions for small loops formed in poly[d(A-T)] from the study of Gralla and Crothers (1973) on oligoribonucleotides. Differences between ribophosphate and deoxyribophosphate chains are unlikely to introduce appreciable errors (Olson and Flory, 1972). Gralla and Crothers reported the loop-weighting functions for loops containing the equivalent of one, two, and three base pairs but bounded by GC base pairs; the only value for loops bounded by AU base pairs was for a loop of size one base pair, obtained by analyzing the data of Martin et al. (1971). To obtain the higher loop-weighting functions for AU-bounded loops, we have taken the difference between the free energies for forming a GC-bounded and an AU-bounded loop of one base pair, and then subtracted this difference from the free energies for forming a GC-bounded loop of two and three base pairs. (In any case, it will be shown below that loops larger than one base pair are not too important at low temperature.) Thus, the numerical values to be used for the loop-weighting functions are $\rho_1 = 4.8 \times 10^{-2}$, $\rho_2 = 3.8 \times 10^{-3}$, $\rho_3 = 8.3 \times 10^{-1}$.

These values are taken to be temperature independent (Gralla and Crothers, 1973) and moreover to be independent of salt concentration, clearly only a first approximation.

The only other parameter needed to calculate θ is the stability constant, s. This is taken to be unity at the T_m of the poly[d(A-T)] (48.5 °C for the data of Figure 15) and to increase with decreasing temperature according to the van't Hoff law using an enthalpy change of 8.0 kcal/base pair (Scheffler and Sturtevant, 1969; after correcting for their slightly lower T_m of 40 °C by assuming constant entropy of melting); thus, at 25 °C, $s = 2.68$. Hence,

$$\theta = \frac{6.7 \times 10^{-3} + 4.0 \times 10^{-4} + 4.8 \times 10^{-5}}{1 + 6.7 \times 10^{-3} + 4.0 \times 10^{-4} + 4.8 \times 10^{-5}} = 7.1 \times 10^{-3}$$

(2) At 25 °C, the forward rate constant for reaction with 5′-dAMP and deoxyadenosine is $9.2 \pm 0.2 \times 10^{-3}$ (M^{-1} min^{-1}) (McGhee and von Hippel, 1975a). For the data of Figure 15, the formaldehyde concentration was 0.48 M.

(3) To calculate the average change in extinction coefficient per reaction, the values of the loop-weighting functions and of the stability constant are kept the same as those used to cal-

[14] We must use this approach, rather than an expression for the fraction coil of the overall DNA molecule, in order to be able to insert experimental values for the loop-weighting functions and to describe the behavior of the reacted base pairs. As the temperature approaches the main helix–coil transition and the loops get larger, the present approach will lead to overcounting the number of denatured base pairs. However, the two treatments can be shown to approach each other as the temperature falls below T_m.

culate θ. From Figure 14c, we take the T_m of pure poly(hydroxymethyladenine–thymine) to be 67 °C below that of poly[d(A-T)]. The enthalpy change for melting this altered base pair is taken to be 8.0 kcal/base pair, the same as that of the poly[d(A-T)]. (We feel at least partially justified in using the same enthalpy change, since the equilibrium constant for isomerization of the exocyclic amino groups, which was argued above to be the sole cause of the base pair destabilization, was found to be temperature independent (Engel and von Hippel, 1974).) Thus, at 25 °C, $s^* = 0.10$.

The only other number required is $\Delta\epsilon_0$, the change in extinction coefficient at 255 nm for denaturing a base pair in poly[d(A-T)]. From the data of Figure 2a, this can be estimated to be $3.0 \pm 0.1 \times 10^3$ (cm^{-1} (mol of nucleotide)$^{-1}$) and is taken to be temperature independent. Thus at 25 °C,

$$\overline{\Delta\epsilon} = \frac{1076 + 96 + 16}{1 + 0.18 + 1.1 \times 10^{-2} + 1.3 \times 10^{-3}} = 997$$

At this low temperature, the reacted base pair clearly is predicted to be re-formed most of the time.

Combining these three factors, our predicted rate of denaturation of poly[d(A-T)] by 0.48 M formaldehyde, at 25 °C, and caused solely by the adenine reaction is:

initial denaturation rate

$$= (7.1 \times 10^{-3})(9.2 \times 10^{-3})(0.48)(997)$$
$$= 3.1 \times 10^{-2}(A_{255}\ \text{min}^{-1}\ \text{(mol of nucleotide)}^{-1})$$

This is to be compared with the experimentally observed value of 4.3×10^{-2} obtained from Figure 15. This agreement is far better than we could have expected considering the uncertainties in the various parameters used in the calculations. Probably the least certain number is s^*, the stability constant for a hydroxymethyladenine–thymine base pair; changing the estimated T_m by ±10 °C changes the calculated rates by about 50%.

Perhaps a less capricious test of the model is to calculate the apparent activation energy, since this should depend on relative rather than absolute values for the parameters. Thus, the above calculations were repeated at other temperatures; the calculated Arrhenius plot is indicated as the dashed line in Figure 15, and lies extremely close to the experimental plot. The predicted activation enthalpy is 48.3 kcal/reaction, compared to the observed value of 47.4 ± 3.6 kcal/reaction. Thus, both the reaction rate and its temperature dependence are predicted extremely closely by our simple model. We emphasize that all parameters in the calculations were determined completely independently of the kinetic experiments.

Implications of the Reaction Mechanism. If this model is correct, the data of Figure 15 can actually be used to give numerical estimates of the loop-weighting functions. Thus eq 4 can be written with only two unknowns, ρ_1 and ρ_2 (ρ_3 being taken as zero). If this is done at two temperatures (25 and 35 °C), the two equations can be solved to yield: $\rho_1 = 6.0 \times 10^{-2}$; $\rho_2 = 2.1 \times 10^{-3}$. These loop-weighting functions are quite close to those estimated from the data of Gralla and Crothers (1973) and listed above.

There are several additional implications of the model and of the above calculations. At the temperatures used for the calculations, loops of only one base pair make by far the largest contribution, both to the initial opening event and to the postreaction extinction-coefficient change. For example, at 25 °C (i.e., 23 °C below the T_m) the term for a one base-pair loop contributes 94 and 91% to the calculated value of θ and $\Delta\epsilon$, respectively. The relative contribution of one base-pair loop decreases with increasing temperature as expected, but, nevertheless, even at 40 °C (8.5 °C below the T_m and certainly at the upper limit of the simple model), one base-pair loops still contribute 88 and 82% of the calculated values of θ and $\overline{\Delta\epsilon}$, respectively.

We can now estimate, at least approximately, the relative magnitudes of the rates of the nucleation and growth reactions. If nucleation is taken to occur only via a one base-pair loop, and growth occurs by opening a single base pair either on one side or the other (and taking the extinction coefficient change to be the same for both reactions), then the relative rates of growth vs. nucleation can be written as:

$$\frac{\text{rate of growth reaction}}{\text{rate of nucleation reaction}} = \frac{\text{probability of forming a two base-pair loop, given a one base pair loop}}{\text{probability of forming a one base-pair loop}} \cong \frac{\frac{2\rho_2}{s^3} \Big/ \frac{\rho_1}{s^2}}{\frac{\rho_1}{s^2}}$$

At 25 °C, this ratio is calculated as 3 or 8, depending on whether we use the loop-weighting functions of Gralla and Crothers (1973) or those determined above.

Moreover, as can be seen from the above equation, the importance of nucleation relative to growth increases both with increasing reaction temperature and with decreasing DNA melting temperature, as actually observed in Figures 3 and 7.[15]

Using the same type of calculation employed to estimate the relative rates of growth and nucleation, we can also calculate the expected rate of reaction at a base pair bordering a nick,

[15] Thus the growth reaction is indeed predicted to be more rapid than the nucleation reaction. However, the expected ratio is several orders of magnitude less than the factor of 10^4 to 10^5 emerging from the analysis of Lazurkin et al. (1970), which has been made the foundation of a proposed method for detecting low concentrations of DNA defects (Lazurkin et al., 1970; Frank-Kamenetskii and Lazurkin, 1974; Vologodskii and Frank-Kamenetskii, 1975). The emphasis of these workers on the growth reaction relative to the nucleation reaction derives from their use of a nearest-neighbor Ising model to analyze the initial opening event. This model places *all* the free energy associated with DNA cooperativity on the formation of a one base-pair loop and neglects the necessity of assigning individual loop-weighting functions to each size loop. Such a model is unrealistic not only for treating large DNA (see, e.g., Crothers and Zimm, 1964) but can be much more misleading in treating small loops (Baldwin, 1971; Gralla and Crothers, 1973). Moreover, these earlier treatments by the Russian group (see footnote 16) take no account of the detailed chemistry of the reaction, such as the interplay between pH-dependent and pH-independent reactions, the existence of equilibrium melting temperatures, the ability of a reacted base pair to re-form, and so on. Even on a purely phenomenological level, the curve shape of the overall denaturation reaction was seen to change both with pH and with proximity of the reaction temperature to the melting temperature, and not solely due to the presence of "defects".

Indeed, we would suggest that the sigmoidal denaturation kinetics, especially of natural DNA at low temperatures, may not primarily reflect the difference in reaction rates of a one vs. a two base-pair loop, but rather may reflect much more long-range events, for example, the merging of large loops as takes place during the conventional helix–coil transition. As can be seen in Figure 3, the maximum rate of increase of the denaturation rate occurs around 10–20% of the overall reaction; at this extent of denaturation, electron microscopy reveals a typical "denaturation map" consisting of very long loops, a far cry from the one or two base-pair loops used in the analysis of the *initial* denaturation rates described above.

as compared to a base pair in an intact helix. Thus, at 25 °C, and assuming the base pairs stack across the nick:

$$\frac{\text{reaction at a nick}}{\text{reaction at an intact base pair}} \cong \frac{1/s^2}{\rho_1/s^2} \cong 20$$

Similarly,

$$\frac{\text{reaction at a helix end}}{\text{reaction at an intact base pair}} \cong \frac{1/s}{\rho_1/s^2} \cong 50$$

Thus, although reactions at nicks and ends are indeed favored (and more so at lower temperatures), the ratio of the rates is predicted to be really rather small, and large-molecular-weight DNAs (where there are many thousand internal base pairs for each helix end) are expected to react primarily from internal sites. This has been observed directly by electron microscopy, even under solvent conditions which abolish the different stability of the different base pairs (Chattoraj and McGhee, in preparation).

Prediction of Kinetic Parameters for Natural DNA. The same qualitative model for the initial steps of the denaturation should also apply to the formaldehyde reaction with natural DNA at low temperatures and at low pH. Although rigorous calculations become much more complicated due to base-pair heterogeneity, very simple calculations can be made with the principal assumptions that: (1) only loops of size one base pair need be considered; (2) the DNA base sequence is random; (3) the known sequence dependence of the stacking interactions (Gralla and Crothers, 1973) can be approximated by assuming that if the stability constant for forming an AT base pair on a preexisting AT base pair is s_{AT}, and that for forming a GC base pair on a preexisting GC base pair is s_{GC}, then that for forming at AT base pair on a preexisting GC base pair is $\sqrt{s_{AT}s_{GC}}$; and (4) a hydroxymethylcytosine–guanine base pair is destabilized as much as a hydroxymethyladenine–thymine base pair.

With these assumptions and by methods completely analogous to those just used for a homopolynucleotide, denaturation rates have been calculated for both T7 DNA and calf thymus DNA at 20 to 25 °C below the T_m (where the denaturation rates are pH independent and should correspond to reaction at amino groups). Calculated denaturation rates agreed with the observed denaturation rates within a factor of about 2 to 7. Although, given all the various uncertainties, this agreement is probably acceptable, much more work will be required to obtain really quantitative agreement between theory and experiment.[16]

Summary

In this series of papers, we have described the use of the simple chemical reagent, formaldehyde, as a probe of the conformational behavior of DNA. Since the use of such a probe requires a complete understanding of the *chemistry* of the underlying interactions, the first two papers of the series were devoted to developing a fairly complete picture of the equilibria, kinetics, and spectroscopic consequences of the interaction of formaldehyde with the various bases found in nucleic acids. In the third paper, we explored the equilibrium effects of formaldehyde adduct formation on the conformational stability of DNA. All this together, finally, made it possible for us, in this paper, to interpret the kinetics in terms of a detailed reaction pathway, to tie this reaction pathway firmly to the predictions of helix–coil theory, and thus to obtain a picture of the conformational equilibria of DNA at temperatures far below the unperturbed melting transition. Since the formaldehyde probe interacts with an "open" state (bases unstacked and unhydrogen bonded) which seems to have a great deal in common with the open states involved in DNA interactions with melting proteins (Jensen et al., 1976), we hope that the elucidation of the reaction pathway for this simple probe will help to shed some light on the much more complicated interactions involved in replication, recombination, and transcription.

[16] A recent paper by Lukashin et al. (1976) uses Monte-Carlo methods to calculate possible sequences of events during the denaturation of a natural heterogeneous DNA by HCHO. These calculations graphically and elegantly illustrate not only the pivotal role of the thymine reaction demonstrated above, but moreover the immense complexities which arise when all bases and all reactions are considered.

Acknowledgments

The authors are very grateful to Dr. D. K. Chattoraj for carrying out the electron microscopic examinations described in this paper.

References

Akasaka, K. (1974), *Biopolymers 13,* 2273.
Arnott, S., Chandrasekaran, R., and leslie, A. G. W. (1976), *J. Mol. Biol. 106,* 735.
Baldwin, R. L. (1971), *Acc. Chem. Res. 4,* 265.
Berns, K. I., and Thomas, C. A., Jr. (1961) *J. Mol. Biol. 3,* 289.
Bloomfield, V. A., Crothers, D. M., and Tinoco, I., Jr. (1974), Physical Chemistry of Nucleic Acids, New York, N.Y., Harper and Row.
Boedtker, H. (1967), *Biochemistry 6,* 2718.
Brahms, J., Michelson, A. M., and von Holde, K. E. (1966), *J. Mol. Biol. 15,* 467.
Chattoraj, D. K., and Inman, R. B. (1973), in DNA Synthesis *In Vitro,* Wells, R. D., and Inman, R. B., Ed., Baltimore, Md., University Park Press, p 33.
Craig, M. E., Crothers, D. M., and Doty, P. (1971), *J. Mol. Biol. 62,* 383.
Crothers, D. M., and Zimm, B. H. (1964), *J. Mol. Biol. 9,* 1.
Davis, R. C., and Tinoco, I. (1968), *Biopolymers 6,* 223.
Eisenberg, H., and Felsenfeld, G. (1967), *J. Mol. Biol. 30,* 17.
Elson, E. L., Scheffler, I. E., and Baldwin, R. L. (1970), *J. Mol. Biol. 54,* 401.
Engel, J. D. (1975), Ph.D. Thesis, University of Orgeon.
Engel, J. D., and von Hippel, P. H. (1974), *Biochemistry 13,* 4143.
Feldman, M. Ya. (1973), *Prog. Nucleic Acid Res. Mol. Biol. 13,* 1.
Frank-Kamenetskii, M. D., and Lazurkin, Yu. S. (1974), *Annu. Rev. Biophys. Bioeng. 3,* 127.
Freifelder, D., and Davison, P. F. (1963), *Biophys. J. 3,* 49.
Fuchs, R. P. P., and Daune, M. P. (1974), *Biochemistry 13,* 4435.
Gralla, J., and Crothers, D. M. (1973), *J. Mol. Biol. 78,* 301.
Haselkorn, R., and Doty, P. (1961), *J. Biol. Chem. 236,* 2738.
Inman, R. B. (1967), *J. Mol. Biol. 28,* 103.
Inman, R. B., and Baldwin, R. L. (1962), *J. Mol. Biol. 5,* 172.
Jencks, W. P. (1964), *Prog. Phys. Org. Chem. 2,* 63.
Jensen, D. E., Kelly, R. C., and von Hippel, P. H. (1976), *J. Biol. Chem. 251,* 7215.
Lazurkin, Yu. S., Frank-Kamenetskii, M. D., and Trifonov,

E. N. (1970), *Biopolymers 9,* 1253.
Leng, M., and Felsenfeld, G. (1966), *J. Mol. Biol. 15,* 455.
Leng, M., and Michelson, A. M. (1968), *Biochim. Biophys. Acta 155,* 91.
Lewin, S. (1964), *Experientia 20,* 666.
Lewin, S. (1966), *Arch. Biochem. Biophys. 113,* 584.
Lowe, M. J., and Schellman, J. A. (1972), *J. Mol. Biol., 65,* 91.
Lukashin, A. V., Vologodskii, A. V., Frank-Kamenetskii, M. D., and Lyubchenko, Y. L. (1976), *J. Mol. Biol. 108,* 665.
Marmur, J., and Doty, P. (1962), *J. Mol. Biol. 5,* 109.
Martin, F. H., Uhlenbeck, O. C., and Doty, P. (1971), *J. Mol. Biol. 57,* 201.
McGhee, J. D. (1975), Ph.D. Thesis, University of Oregon.
McGhee, J. D., and von Hippel, P. H. (1975a), *Biochemistry 14,* 1281.
McGhee, J. D., and von Hippel, P. H. (1975b), *Biochemistry 14,* 1297.
McGhee, J. D., and von Hippel, P. H. (1977), *Biochemistry 16* (preceding paper in this issue).
Melchior, W. B., Jr., and von Hippel, P. H. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 298.
Olson, W. K., and Flory, P. J. (1972), *Biopolymers 11,* 57.
Pörschke, D. (1973), *Eur. J. Biochem. 39,* 117.
Pörschke, D. (1974), *Biophys. Chem. 2,* 97.
Pörschke, D., and Eigen, M. (1971), *J. Mol. Biol. 62,* 361.
Saenger, W., Riecke, J., and Suck, D. (1975), *J. Mol. Biol. 93,* 529.
Saunders, M., and Ross, P. D. (1960), *Biochem. Biophys. Res. Commun. 3,* 314.
Scheffler, I. E., Elson, E. L., and Baldwin, R. L. (1970), *J. Mol. Biol. 48,* 145.
Scheffler, I. E., and Sturtevant, J. M. (1969), *J. Mol. Biol. 42,* 577.
Shoup, R. R., Miles, H. T., and Becker, E. D. (1972), *J. Phys. Chem. 76,* 64.
Stannard, B. S., and Felsenfeld, G. (1975), *Biopolymers 14,* 299.
Stevens, C. L. (1974), *Biopolymers 13,* 1517.
Stevens, C. L., and Rosenfeld, A. (1966), *Biochemistry 5,* 2714.
Trifonov, E. N., Lazurkin, Yu. S., and Frank-Kamenetskii, M. D. (1967), *Mol. Biol. (Moscow) 1,* 164.
Ts'o, P. O. P. (1974), *Basic Princ. Nucleic Acid Chem. 2.*
Utiyama, H., and Doty, P. (1971), *Biochemistry 10,* 1254.
Vologodskii, A. V., and Frank-Kamenetskii, M. D. (1975), *J. Theor. Biol. 55,* 153.
von Hippel, P. H., and Wong, K. Y. (1971), *J. Mol. Biol. 61,* 587.

Structure–Function Relationships in TPN-Dependent Isocitrate Dehydrogenase. I. Electron Paramagnetic Resonance Studies of the Interaction of Enzyme-Bound Mn(II) with Substrates, Cofactors, and Substrate Analogues†

Ronald S. Levy and Joseph J. Villafranca*

ABSTRACT: Electron paramagnetic resonance (EPR) spectra were obtained for various isocitrate dehydrogenase–Mn(II) complexes. The qualitative effects of the binding of substrates, nucleotides, and substrate analogues on the isotropic character of the electronic environment of enzyme-bound Mn(II) were subsequently investigated. The addition of isocitrate produces a markedly anisotropic spectrum whereas α-ketoglutarate does not alter the spectrum of enzyme–Mn(II) substantially. This suggests direct coordination of isocitrate to the Mn(II) but perhaps a different mode of binding for α-ketoglutarate. Other studies demonstrated mutually exclusive binding relationships between TPN and TPNH, between Mn–isocitrate and TPNH, and between HCO_3^- (CO_2) and formate or thiocyanate. Indirect evidence supporting CO_2 rather than HCO_3^- as the actual reactive species which binds to the enzyme in the reductive carboxylation reaction is presented, on the basis of the results of the formate and thiocyanate studies. From the EPR results recorded for ternary, quaternary, and quinary enzyme–substrate complexes, correlations between the appearance of fine structure signals and the binding of individual substrates and/or nucleotides are found, and tentative assignments of such signals are made on this basis. Additional studies were conducted to determine binding constants for Mg(II), Co(II), and Co–isocitrate, and a comparison was made with kinetically determined binding constants.

TPN-dependent isocitrate dehydrogenase (*threo*-D_S-isocitrate:TPN oxidoreductase (decarboxylating), EC 1.1.1.42) is found in both the mitochondria and the cytoplasm of mammalian tissues. It is composed of only one subunit, and its molecular weight is approximately 60 000 (Plaut et al., 1957).

The enzyme catalyzes the oxidation of the D_S isomer of isocitrate (I) to produce oxalosuccinate (II) which is then decarboxylated to α-ketoglutarate (III). The reaction proceeds

† From the Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802. *Received September 17, 1976; revised manuscript received February 3, 1977.* This work was supported by National Science Foundation Grant No. PCM7609688.