Biochemistry

Volume 16, Number 15 July 26, 1977

Formaldehyde as a Probe of DNA Structure. 3. Equilibrium Denaturation of DNA and Synthetic Polynucleotides†

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

ABSTRACT: The purpose of this and the following paper in this issue is to describe the use of formaldehyde as a probe of the conformational motility, or "breathing reactions", of native DNA. Formaldehyde, like other chemical probes, perturbs the DNA helix–coil transition, and the present paper attempts to establish the overall *equilibrium* thermodynamic description of the formaldehyde denaturation of DNA. Since the individual chemical reactions of formaldehyde with the DNA bases are reversible, formaldehyde should be a reversible denaturant of DNA. This is demonstrated using renaturable synthetic double-helical polynucleotides. It is also demonstrated that for single-stranded polynucleotides (with both stacked and unstacked bases), the equilibrium constants for the formaldehyde reaction with the various bases are the same as with the free monomers, and that the reactions are independent and noncooperative. Thus, it is possible to predict fairly closely the observed relation between DNA melting temperature and formaldehyde concentration using simple ligand-coupled helix–coil theory and numerical estimates of the various binding constants. These calculations define the conditions of temperature, salt, formaldehyde concentration, and pH under which DNA will be denatured *at equilibrium*. Such "phase diagrams" are needed to interpret kinetic experiments (McGhee, J. D., and von Hippel, P. H. (1977), *Biochemistry 16* (following paper in this issue)) and should also be helpful in practical use of the formaldehyde reaction, e.g., in denaturation mapping by electron microscopy. The present study also establishes formaldehyde as a very simple model for some equilibrium aspects of DNA–melting protein interactions.

The purpose of this series of papers is to provide a quantitative description of how a simple chemical reagent, formaldehyde, reacts with and denatures DNA. It is hoped that this will serve as a model system for some aspects of the behavior of more complicated ligands (e.g., drugs and proteins) which interact with nucleic acids during biological processes.

The first two papers of this series (McGhee and von Hippel, 1975a,b) described the reaction of formaldehyde with the monomeric components of nucleic acids. (Earlier related studies of others are summarized in McGhee and von Hippel (1975a,b) and in Feldman (1973).) It was shown that all amino and imino protons of the nucleic acid bases can be replaced with a hydroxymethyl (methylol) group ($-CH_2OH$). This means that there are sites available for reaction in denatured DNA which, in native DNA, are involved in hydrogen bonding. Addition of formaldehyde to native DNA should therefore pull the helix–coil equilibrium toward the coil form or, equivalently, lower the DNA-melting temperature, T_m. Furthermore, this lowering of the DNA-melting transition should, in principle, be reversible, since all the chemical reactions are themselves reversible (albeit, in some cases, slowly).

In this paper we show that polynucleotides can indeed be melted reversibly in the presence of formaldehyde. In addition, a quantitative relation between the equilibrium decrease in T_m and the formaldehyde concentration can be established through the use of simple ligand-coupled helix–coil theory and the known equilibrium constants for the formaldehyde reactions with the DNA monomers. Such a quantitative thermodynamic description of the overall denaturation reaction will be absolutely necessary to interpret the *kinetics* of DNA denaturation (McGhee and von Hippel, 1977).

It will also become apparent that formaldehyde does indeed serve as a simple, but quite appropriate, first model for the entire class of DNA–melting (or helix destabilizing) proteins (e.g., see Alberts et al., 1968; Kornberg, 1974). That is, formaldehyde has in common with melting proteins the capability of destabilizing and denaturing DNA double helices at equilibrium, but lacks some of the more complex characteristics of the real proteins (such as large binding-site size and protein-based cooperativity of binding). Indeed, the simple helix–coil theory outlined here for the formaldehyde–DNA

† From the Institute of Molecular Biology and Departments of Chemistry and Biology, University of Oregon, Eugene, Oregon 97403. *Received September 8, 1976; revised manuscript received April 12, 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.

system serves as a starting point for a theoretical description of a real melting protein–DNA system (Jensen et al., 1976; McGhee, 1976).

Materials and Methods

Formaldehyde solutions were prepared and standardized as described previously (McGhee and von Hippel, 1975a). Low concentrations of free formaldehyde (<50 μM) were estimated by a modification of the fluorescent assay of Belman (1963; see McGhee, 1975).

Synthetic polynucleotides, obtained from either Miles or P-L Biochemical Co., were dissolved in the cold in buffers containing 0.001 M EDTA,[1] extracted one to three times with redistilled buffer-saturated phenol, and then dialyzed extensively into the desired buffers (not containing EDTA). To prevent nuclease degradation of polyribonucleotides, buffers used with these materials were autoclaved, and glassware was oven-baked before use.

Ultraviolet absorbance readings were measured on a Cary 14 spectrophotometer interfaced to a Varian 620i computer. Temperature readings were taken with a thermocouple inserted into a dummy cuvette and are accurate to ±0.2 °C. Melting curves were measured on a Gilford 2000 recording spectrophotometer; solutions were placed in Teflon-stoppered cuvettes and briefly bubbled with water-saturated helium; all absorbance readings were corrected for the thermal expansion of water; weight losses during melting were 0.5% or less. Sedimentation coefficients, at both neutral and alkaline pH, were measured by band sedimentation (Studier, 1965) using a Model E analytical ultracentrifuge equipped with automatic scanner and UV optics.

Excess formaldehyde could be quickly removed from solutions of polynucleotides by a modification of the technique of Neal and Florini (1973). In outline, a short column of a gel, such as Bio-Rad P-2, is prepared in a centrifuge tube and the interbead (void volume) buffer removed by low-speed centrifugation. The sample is added directly to the gel bed and the tube recentrifuged. Small molecules in the sample can apparently exchange with the solvent inside the beads and are left behind, whereas large molecules are centrifuged through and recovered essentially undiluted. The free formaldehyde concentration is reduced by a factor of 10^4 to 10^5 in one pass (McGhee, 1975).

Results and Discussion

The equilibrium denaturation of DNA by a reagent such as formaldehyde can be formulated as:

$$\text{helix} \rightleftharpoons \text{coil} \overset{\text{HCHO}}{\rightleftharpoons} \text{reacted coil} \qquad (1)$$

This scheme divides the overall free-energy change accompanying denaturation into a component due to the unperturbed helix–coil transition and a component attributable to the chemical reaction of HCHO with the separated DNA strands.[2] Such a scheme is convenient and perfectly valid for equilibrium considerations. However, it says nothing about whether the denaturation kinetics do follow such a path.

The aim of the present paper is to describe the above reaction scheme quantitatively. In the first section below, we consider the equilibrium reaction of formaldehyde with single-stranded polynucleotides, as models for the chemical step in eq 1. The emphasis is on demonstrating that, in the absence of hydrogen bonding, single-stranded polynucleotides have the same reactive sites, with the same equilibrium affinity for formaldehyde, as do the constituent bases. The next step is to determine, as a function of HCHO concentration, the equilibrium depression of the melting temperature of several synthetic double-helical and renaturable polynucleotides; this provides a measure of the overall free-energy change involved in the denaturation. As a third step, the theoretical relation between the formaldehyde concentration and the depressed T_m is formulated, and is shown indeed to describe the experimental observations. Finally, a more involved theoretical relation is used to calculate the relation between HCHO concentration and the T_m of a natural heterogeneous DNA.

Comparison of the Equilibrium Reaction of Formaldehyde with Single-Stranded Polynucleotides and with the Constituent Monomers

Previous work (McGhee and von Hippel, 1975a,b) has described in detail the equilibrium reaction of formaldehyde with both the nucleotide and nucleoside components of DNA and RNA. It was established that any proton attached to a nitrogen atom on a nucleic acid base can be replaced by a hydroxymethyl group (-CH_2OH). This includes the two protons of the exocyclic amino groups of adenine, guanine, and cytosine, and the single proton of the endocyclic imino group of thymine and guanine. All these reactions are reversible and are characterized by relatively small association constants, ranging from about 2 to 15 M^{-1}. The following experiments indicate that the *equilibrium* reaction of formaldehyde with bases in single-stranded nucleic acids is the same, within experimental error, as the equilibrium reaction of the bases in the free nucleosides and nucleotides.

Reactions with Unstacked Single-Stranded Polyribonucleotides. It was shown previously (McGhee and von Hippel, 1975b) that in polyribonucleotides such as poly(rU) and poly(rI), in which the bases are unstacked at room temperature, the endocyclic imino groups have an equilibrium constant for reaction with formaldehyde that is well within 5% of that for the constituent monomers. Similar observations have since been made with poly(dT) and 5′-dTMP (data not shown).

Reactions with Stacked Single-Stranded Polyribonucleotides. The bases in the single-stranded polynucleotides, poly(rA) and poly(rC), are partially stacked at room temperature (Leng and Felsenfeld, 1966; Leng and Michelson, 1968), and yet their exocyclic amino groups can be shown to react, at *equilibrium,* in a very similar fashion to those of the mononucleotides.

Poly(rA), poly(rC), 5′-rAMP, and 5′-rCMP, in a variety of different buffers, were incubated at 23 ± 1 °C with HCHO concentrations ranging from 0 to 5 M. After sufficient time to reach equilibrium (7–10 days), ultraviolet spectra were taken and corrected for the small HCHO only blank. Spectral changes seen on reaction of the polymers with HCHO are large and qualitatively similar to those observed with the relevant mononucleotides (McGhee and von Hippel, 1975a).

In order to extract equilibrium constants from such data, the absorbance changes are plotted according to a straightforward rearrangement of the mass action relation as previously described (McGhee and von Hippel, 1975a). Thus, at each HCHO concentration, the absorbance change at the peak of the difference spectrum, ΔA_λ, is determined and the ratio ΔA_λ/[HCHO] plotted vs. ΔA_λ. Such plots are illustrated in

[1] Abbreviations used are: EDTA, ethylenediaminetetraacetic acid; UV, ultraviolet.

[2] We use the designation HCHO to represent total solution formaldehyde, in both the hydrated and the unhydrated form (see McGhee and von Hippel, 1975a).

Figure 1a–d for poly(rA), 5′-rAMP, poly(rC), and 5′-rCMP, respectively, and are seen to be nonlinear, both for the mononucleotides and for the polynucleotides. In the case of the monomers, this nonlinear shape has been shown to arise from the formation of a dimethylol adduct at higher HCHO concentrations. Using a mass-action equation extended to allow for formation of these two adducts, a nonlinear least-squares curve-fitting procedure has been used to determine the two equilibrium constants quite accurately (McGhee and von Hippel, 1975a).

With the polymers, however, it can be shown that two factors contribute to this nonlinearity: (1) the formation of a dimethylol adduct, just as with the monomers; and (2) a hyperchromic effect due to base unstacking caused by a "solvent effect" of the high formaldehyde concentrations (up to 5 M) used in an attempt to reach reaction saturation.

Poly(rA) can be shown to form a dimethylol adduct simply by measuring the amount of formaldehyde bound to the polymer after equilibrating with a high concentration of HCHO. Thus, a solution of poly(rA) ($A_{257} \simeq 20$) was incubated to equilibrium at 23 °C with 5.3 M HCHO and excess unbound HCHO quickly removed by centrifuging twice through a short column of P-2 gel. This rapidly (<5 min) reduces excess unbound formaldehyde to the limit of detectability (<1–2 μM). Any formaldehyde still bound to the poly(rA) was released by boiling for 10 min, and then assayed by the fluorescence method of Belman (1963). Poly(rA) concentrations were measured by absorbance at 260 nm after alkaline hydrolysis (0.2 N NaOH, 37 °C, 20 h). The number of formaldehyde groups bound per adenine residue of the poly(rA) was thus estimated to be 1.9 ± 0.2. Using the monomer equilibrium constants for forming mono- and diadducts, the predicted number is 1.7. It is thus clear that dimethylol adducts can form with poly(rA), and that this must be responsible for at least part of the curvature of the equilibrium plots of Figure 1a,c.

It can also be shown that some portion of the absorbance changes observed with poly(rA) in high HCHO concentrations can be plausibly attributed to a "solvent effect" resulting in base unstacking. Reversal of adduct formation should be many orders of magnitude slower than reversal of a solvent effect and thus, with the same solution of poly(rA) as above (i.e., A_{257} of about 20 and equilibrated with 5.3 M HCHO), the absorbance at 266 nm was measured either: (1) undiluted in a 0.05-cm path-length cuvette, or (2) in a 1-cm path-length cell, after a rapid 20-fold dilution. The absorbance change in the latter procedure (linearly extrapolated back to zero time of mixing) measured about 18% lower than that observed by the former technique.[3]

Because of this "solvent" effect, the analysis of the formaldehyde binding plots of the polymers (Figure 1a,c) is not as straightforward as that for the monomers. Nevertheless, since this effect should become important only at high HCHO concentrations (>1 M), and most of the spectral changes are associated with monoadduct formation which occurs at low formaldehyde concentrations (<0.3 M), the binding plots of

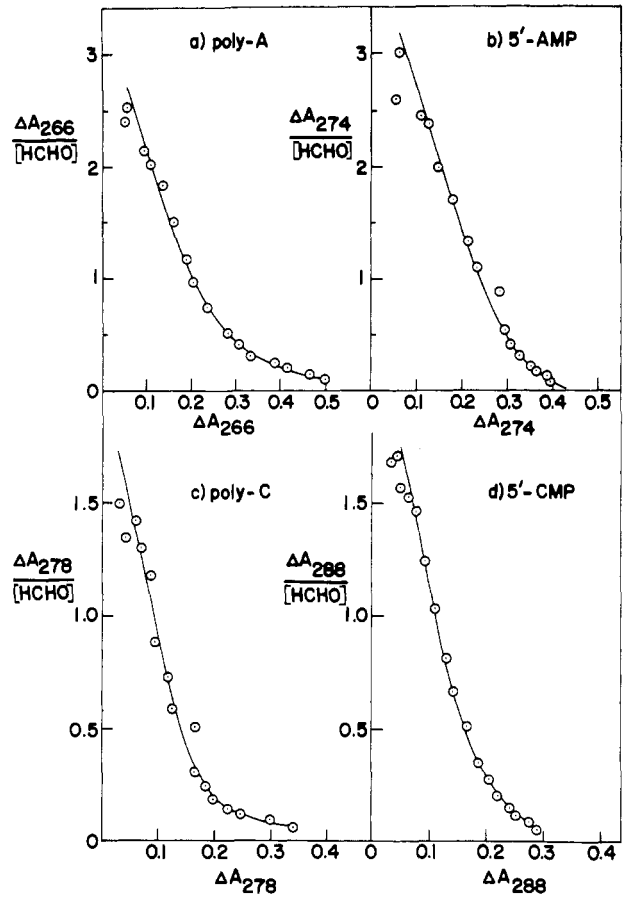

FIGURE 1: Plot of ΔA_λ/[HCHO] vs. ΔA_λ, where ΔA_λ is the equilibrium absorbance change measured at the peak of the difference spectrum in the presence of various concentrations of formaldehyde: (a) poly(rA); (b) 5′-rAMP; (c) poly(rC); (d) 5′-rCMP. As described in the text, analysis of such plots yields estimates of the equilibrium constants for formation of mono- and dimethylol adducts (Table I).

Figure 1 should still yield fairly accurate estimates of K_1, the equilibrium constant for monoadduct formation. Estimates of K_2, the equilibrium constant for diadduct formation, will obviously be less accurate. This presents no problem for the present studies, since all DNA melting experiments are carried out at very low HCHO concentrations (<0.1 M, where less than 3% of the residues are present as diadducts).

Estimated equilibrium constants, for several polynucleotides and in several solvents, are summarized in Table I. The values are, within an experimental error of about 15%, the same for all the amino groups, whether in a single-stranded polymer or in a mononucleotide (or indeed whether involving adenine or cytosine). The values of K_2 are more uncertain, but again are very similar for all compounds. From these results it is possible to conclude that, *at equilibrium,* base stacking neither inhibits the formaldehyde reaction, nor does the methylol group on the amino group inhibit base stacking.[4] It will, however, be shown

[3] We have also confirmed the finding of Stevens (1974) that poly(rA) in high HCHO shows a considerably reduced circular dichroism. For example, in 5 M HCHO, the ellipticity at 276 nm is only about 50% of that of the unreacted control (whereas at lower HCHO concentrations the ellipticity at this wavelength actually increases). By comparing the CD spectrum of a concentrated solution of poly(rA) in 5 M HCHO, taken either in a 0.05-cm path-length cell or in a 1-cm path-length cell after a 20-fold dilution, the ellipticity at 276 nm can be shown to revert, within 30 s after mixing, to about 90% of that of the unreacted control.

[4] These conclusions are in accord with earlier observations made with N^6-methylpoly(rA) and N^6,N^6-dimethylpoly(rA) (Griffin et al., 1964), N^6-hydroxyethylpoly(rA) (van Holde et al., 1965), and methylated poly(rC) (Brimacombe and Reese, 1966), all of which are partially stacked at room temperature and, on heating, show broad noncooperative increases in absorbance. We have also confirmed the earlier results of Stevens and Rosenfeld (1966) with poly(rA) and of Fasman et al. (1964) with poly(rC), that, after reaction with HCHO, both polymers still exhibit a gradual noncooperative rise in absorbance on heating; this would not be expected if the HCHO reaction abolished base stacking and the attendant hypochromicity.

TABLE I: Equilibrium Constants for Formaldehyde Reaction with Exocyclic Amino Groups in Single-Stranded Polynucleotides and the Constituent Mononucleotides.[a,b]

Compd		K_1 (M^{-1})	K_2 (M^{-1})
Poly(rA)		12.0[c]	0.57[c]
		9.4[d]	0.39[d]
		14.2[e]	0.61[e]
		13.7[f]	0.33[f]
	Mean:	12.3 ± 2.2[g]	0.48 ± 0.14[g]
5′-rAMP		12.3[c]	0.64[c]
		13.4[d]	0.62[d]
	Mean:	12.9 ± 0.8[g]	0.63 ± 0.01[g]
Poly(rC)		12.1[c]	0.35[c]
		13.9[d]	0.41[d]
	Mean:	13.0 ± 1.3[g]	0.38 ± 0.04[g]
5′-rCMP		16.4[c]	0.76[c]
		12.3[d]	0.46[d]
	Mean:	14.4 ± 2.9[g]	0.61 ± 0.21[g]

[a] Because of the nonspecific solvent-induced unstacking discussed in the text, the nonlinear least-squares curve-fitting procedure which worked well with monomers was found to converge poorly when presented with data from polymer titrations. The binding constants (as well as spectral parameters not listed above) were thus estimated from the limiting slopes and intercepts of binding plots such as Figure 1, as described earlier (McGhee and von Hippel, 1975a). [b] Temperature = 23 ± 1 °C. Under all buffer conditions used, no acid form of poly(rA) or poly(rC) would be expected (Adler et al., 1969; Sarocchi et al., 1970). This was verified by the absorption spectra and by the absence of cooperative melting curves. [c] 0.01 M Na_2HPO_4, 0.01 M NaH_2PO_4, pH 7.0. [d] 0.1 M Na_2HPO_4, 0.005 M NaH_2PO_4, pH 8.0. [e] 0.01 M methyl arsenate, pH 8.5. [f] 0.01 M methyl arsenate, pH 8.5, with added 0.5 M LiCl. [g] Unweighted mean ± standard deviation.

in the following paper of this issue (McGhee and von Hippel, 1977) that base stacking markedly slows the *kinetics* of adduct formation.

Reactions with Single-Stranded Polydeoxyribonucleotides. To study the HCHO reaction with the unstacked adenine and cytosine bases in single-stranded DNA, poly[d(A-T)] and poly[d(I-C)] in low salt ($1.5\text{–}3 \times 10^{-3}$ M Na^+; T_m less than 35 °C in all cases) were equilibrated with various concentrations of HCHO at temperatures (75 and 90 °C) well above T_m. Absorbance changes were measured at wavelengths which should be insensitive to either the thymine reaction (272 nm) or to the inosine reaction (266 nm) and then plotted as in Figure 1. The estimates of K_1 (equilibrium constant for monoadduct formation) were compared to values predicted at the same temperature, using monomer equilibrium constants and enthalpy changes (McGhee and von Hippel, 1975a). The average ratio of observed to predicted value of K_1 was 1.0 ± 0.4 (for details, see McGhee, 1975).

The amount of formaldehyde attached at equilibrium to denatured DNA can be estimated by the same type of experiment as described above for poly(rA). Calf thymus DNA (in 0.02 M phosphate, pH 6.95, T_m 75 °C) was mixed with various concentrations of HCHO (0.03, 0.1, and 0.3 M), sealed into an ampule, and heated in a boiling water bath for 15 min. After allowing 1 week at 23 ± 1 °C to ensure complete equilibration, excess unbound formaldehyde was quickly removed by the gel technique described under Materials and Methods, and any formaldehyde still bound to the DNA was released by placing the sample into a boiling water bath for an additional 15 min. The free formaldehyde was assayed by the fluorescence method of Belman (1963) and the amount of DNA recovered was measured by the diphenylamine technique (Giles and Myers, 1965). Thus, the measured amount of HCHO bound per base pair of the DNA could be compared with the amount calculated from the known monomer equilibrium constants and the base composition of calf thymus DNA (58% AT). (It was assumed that only the unstable thymine adduct would be lost in the gel procedure; in addition, a correction was made for the small amount of residual base pairing which can still take place in denatured DNA in these low HCHO concentrations.) The average ratio of the measured amount of HCHO bound per base pair to the amount calculated was 1.00 ± 0.05 (see McGhee, 1975).

The conclusion from the experiments described in this section is that, within experimental error (less than ±20%), the *equilibrium* reaction of HCHO with the bases in single-stranded nucleic acids can be quantitatively described by the equilibrium constants (and the associated enthalpy changes) determined for the nucleotides and nucleosides. Furthermore, all the data are consistent with the reactions being independent and noncooperative; i.e., reaction at one base in a single strand in no way influences the reaction at adjoining bases. This insensitivity of the HCHO equilibrium constants to the local electronic environment comprises a great experimental advantage in the application of this chemical probe.

The Reversible Denaturation of Double-Stranded Polynucleotides by HCHO

In this section, we show that HCHO can reversibly denature double-stranded polynucleotides (and by implication, natural DNA). We also attempt to measure the equilibrium melting temperature of such a polynucleotide as a function of HCHO concentration. The extent to which this melting temperature, T_m, is lowered from the unperturbed melting temperature, $T_m°$, is a measure of the overall free-energy changes, both conformational and chemical, associated with reaction 1 above.

It will become apparent that there are considerable experimental difficulties in obtaining an equilibrium melting curve of DNA in the presence of HCHO, both because the rates of the chemical reactions themselves are intrinsically slow, and because the DNA structure limits access to the potentially reactive amino and imino groups and thus further lowers rates of adduct formation. For example, even at relatively high temperatures (50 to 60 °C) in the presence of 0.1 M HCHO, the half-times for the forward and reverse reactions of HCHO with free adenine amino groups are in the range of 10 to 60 min. Thus, if HCHO is simply mixed with DNA and the solution heated at customary rates (0.2–0.5°/min), it is certain that the observed T_m will be very different from that which would be observed if the reactions were allowed to reach equilibrium at each temperature (see, e.g., Freifelder and Davison, 1963; Trifonov et al., 1967).

Nevertheless, it should be possible at least to obtain upper and lower bounds on the equilibrium T_m by using double-stranded, self-complementary polynucleotides, such as poly[d(A-T)] and poly[d(I-C)], which renature rapidly and completely after melting (Inman and Baldwin, 1962; Grant et al., 1972). Thus, if such a copolymer is heated in the presence of HCHO, the observed T_m should be lower than the unperturbed T_m because HCHO can react with groups that are exposed in the coil form, but buried in the helix; thus, the helix–coil equilibrium is pulled toward denaturation. However, the observed T_m should be an *overestimate* of the true equilibrium T_m; i.e., at each temperature less HCHO will have reacted with the polymer than would have if the reaction had

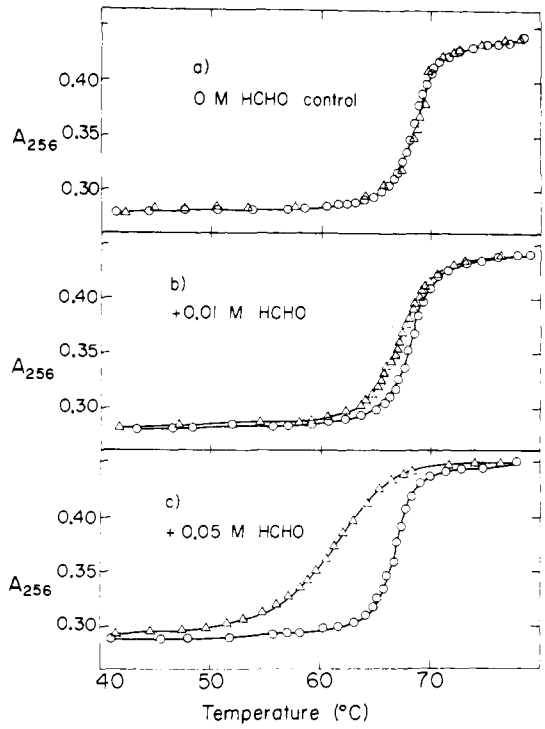

FIGURE 2: Poly[d(A-T)] (0.3 M NaCl, 0.02 M phosphate, pH 6.6) heated and cooled in the presence of various concentrations of HCHO. The absorbance change at 256 nm measures helix denaturation independent of the chemical reaction: (a) poly[d(A-T)]-only control; (b) 0.01 M HCHO; (c) 0.05 M HCHO. (⊙) Heating curve; (△) cooling curve.

been allowed to reach equilibrium. Conversely, if the poly[d(A-T)] is incubated above the melting temperature for sufficient time for the reaction to reach equilibrium and then recooled, the T_m observed on cooling should be an *underestimate* of the true equilibrium T_m. That is, the dissociation rate of the adduct is so slow that more adduct remains attached, and hence the polymer is more destabilized, than would be the case if the cooling rates were infinitely slow and equilibrium were attained at each temperature. (Fortunately, the equilibrium constants themselves change only slowly with temperature, about 2%/deg.)

Examples of such "hysteretic" melting curves are shown in Figure 2 for poly[d(A-T)] in 0.02 M phosphate, 0.3 M NaCl, pH 6.6. As will be described in the following paper of this issue (McGhee and von Hippel, 1977), absorbance at 254–256 nm should measure helix denaturation independently of the chemical reaction. Figure 2a shows that, for the poly[d(A-T)]-only control, heating and cooling curves are superimposable. If the cycle is repeated in the presence of 0.01 M HCHO (Figure 2b), the T_m is lowered slightly on heating and is further lowered on recooling. This difference between the heating and cooling curves (and between these and the unperturbed T_m) becomes greater at higher HCHO concentrations (Figure 2c). That the true equilibrium curve does indeed lie between such pairs of heating and recooling curves is shown by the experiment described in Figure 3, where, at temperatures close to the equilibrium T_m, the final absorbance is approached both from above and from below. This experiment also illustrates the very long time periods that would be required to attain equilibrium at each temperature.

At all the HCHO concentrations used (ranging up to 0.1 M), the poly[d(A-T)] helix is essentially completely re-formed on cooling, at least as monitored by ultraviolet hypochromicity (the average ratio of A_{256}, measured at 25 °C immediately

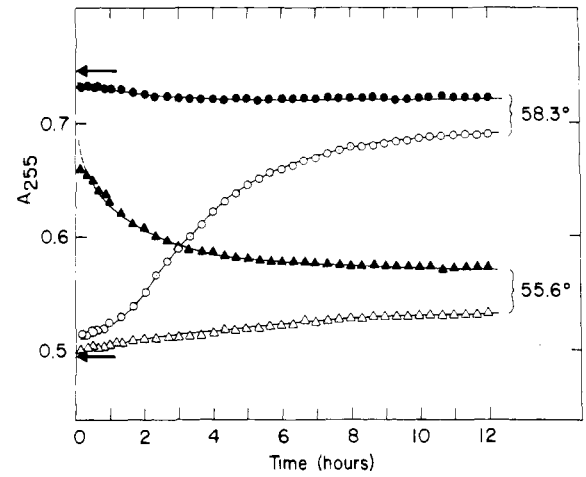

FIGURE 3: Denaturation and renaturation of poly[d(A-T)] at temperatures close to the equilibrium melting temperature. Poly[d(A-T)] was incubated with 0.10 M HCHO for 10 min either at 75 °C (filled symbols) or 25 °C (open symbols) and then transferred to 58.3 (○, ●) or 55.6 °C (△, ▲). Arrows indicate absorbance expected for fully native and fully denatured polynucleotide.

after and before the heating and cooling cycles, is 1.008 ± 0.007 for HCHO-containing samples, compared to 1.002 ± 0.005 for the poly[d(A-T)]-only controls). The T_m values observed on a number of such heating and cooling cycles are plotted vs. the corresponding HCHO concentration in Figure 4a. The lines are generated theoretically and will be discussed in the next section.[5]

Exactly analogous hysteretic melting curves were also measured with poly[d(I-C)] and with the (presumably double-stranded) high-salt form of poly(dI) (Inman, 1964). These over- and underestimates of the equilibrium T_m are plotted in Figure 4b,c for poly[d(I-C)] and poly(dI), respectively; again, the lines are theoretically generated (see next section). With both polymers, the double helix was essentially completely regained after melting, the residual hypochromicities being within 1–2% of the controls without formaldehyde.[6]

Thus synthetic polynucleotides seem indeed to be able to renature in the presence of HCHO, under conditions where a sizable fraction of coil bases would, if free, be reacted. Although (because of the slow reaction kinetics) the equilibrium melting temperature of these polynucleotides in the presence of HCHO could not be accurately measured, the observed T_m values do provide upper and lower bounds on the true equilibrium T_m.

Relation between Formaldehyde Concentration and Melting Temperature of a Homogeneous Double-Stranded Polynucleotide. Comparison of Theory with Experiment

In this section, we attempt to predict the quantitative relation between T_m and HCHO concentration, using numerical estimates of the parameters describing the HCHO reaction with the nucleic acid monomers. Agreement between theory

[5] The regain in hypochromicity is *not* due simply to the low HCHO concentration reacting with only a minute fraction of the bases. For example, at 40 °C, at the end of the cooling cycle of Figure 2c, about 30% of free adenine and about 10% of free thymine residues would be reacted.

[6] Formaldehyde was also observed to lower the melting temperature of the helical form of poly(U) from the control T_m of 10 °C (0.5 M CsCl in 0.01 M methyl arsenate buffer, pH 8.4), thus further indicating that the endocyclic imino groups are involved in maintaining this structure (Thrierr et al., 1971).

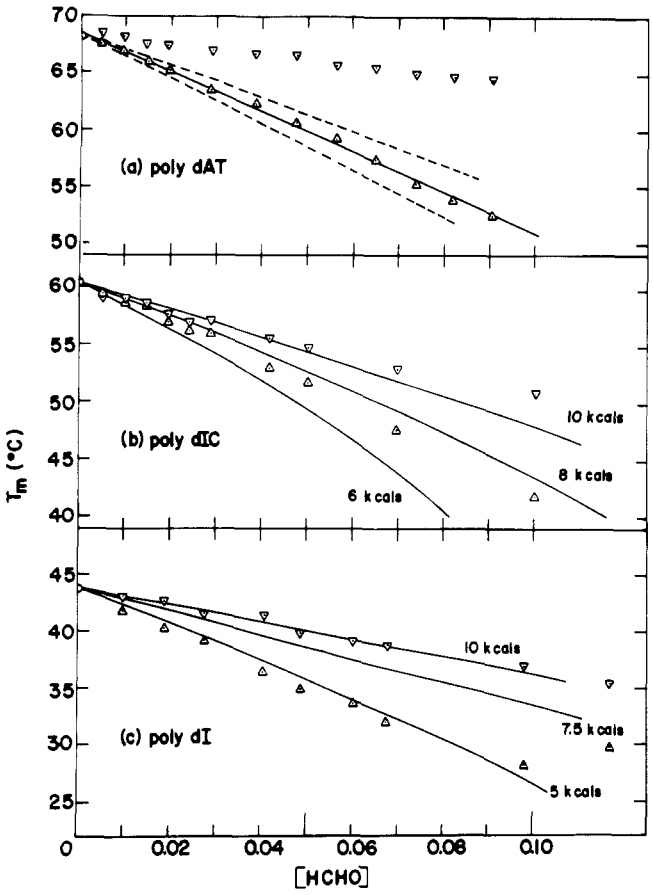

FIGURE 4: Plot of observed T_m vs. formaldehyde concentration for (a) poly[d(A-T)]; (b) poly[d(I-C)]; (c) poly(dI). (▽) T_m observed on heating cycle (should be an overestimate of the equilibrium T_m); (△) T_m observed on cooling cycle (should be an underestimate of the equilibrium T_m). Lines are generated theoretically as explained in the text. The buffers used were: (a) 0.01 M Na_2HPO_4, 0.01 M NaH_2PO_4, 0.3 M NaCl, pH 6.6; (b) 0.1 M Na_2HPO_4, 0.005 M NaH_2PO_4, 0.5 M NaCl, pH 7.7; (c) 0.02 M Na_2HPO_4, 0.001 M NaH_2PO_4, 1 M NaCl, pH 7.2.

and the results of the last section will allow us to use the theory to calculate the equilibrium T_m under other experimental conditions; this will turn out to be very useful in the following paper of this issue (McGhee and von Hippel, 1977).

We first consider the equilibrium denaturation of poly[d(A-T)] by formaldehyde. As derived in detail in the Appendix, the equation which should relate the poly[d(A-T)] melting temperature to the HCHO reaction with the various sites on the adenine and thymine bases is:

$$\frac{1}{T_m} - \frac{1}{T_m^\circ} = \frac{R}{\Delta H_{conf}} \ln \frac{(1 + K_{A,1}F + K_{A,1}K_{A,2}F^2)(1 + K_TF)}{(1 + K_{A,out}F)} \quad (2)$$

where the various symbols are: T_m, equilibrium melting temperature (kelvin) of the polynucleotide in the presence of formaldehyde concentration, F; T_m°, unperturbed melting temperature (kelvin) of the polynucleotide in the absence of HCHO; R, the gas constant; ΔH_{conf}, the enthalpy change for melting 1 mol of base pairs (for poly[d(A-T)] this has been measured by calorimetry as 7.9 ± 0.1 kcal/base pair (Scheffler and Sturtevant, 1969)); F, free formaldehyde concentration (since formaldehyde is always present in great excess over the polynucleotide, this is equivalent to the total HCHO concentration and is accurately known for each melting curve).

The various equilibrium constants which appear in eq 2, and describe the HCHO reaction with poly[d(A-T)], are collected in Table II, together with similar data for other bases which are used in later calculations.[7]

The predicted relation between T_m and HCHO concentration for poly[d(A-T)] is shown as the solid line in Figure 4a, and is seen to pass close to the estimated lower bounds for the equilibrium melting temperature. The dashed lines of Figure 4a are calculated with all the parameters being changed by one standard error in a direction so as either to raise or to lower the estimated T_m. Considering all the experimental parameters (and their uncertainties) used in this calculation (and especially considering where the curve could have gone) the agreement between theory and experiment is adequate and probably as good as could be expected.

Equation 2 should also apply to the HCHO denaturation of poly[d(I-C)], with cytosine and inosine parameters replacing those of adenine and thymine, respectively. The various equilibrium constants and enthalpy changes are collected in Table II. Although the enthalpy change, ΔH_{conf}, for melting a base pair of poly[d(I-C)] has not been measured calorimetrically, it is unlikely to lie outside the range of 5 to 10 kcal/base pair (Ross and Scruggs, 1969; Hinz et al., 1970). As seen in Figure 4b, for a trial value of ΔH_{conf} = 8 kcal/base pair, the predicted relation agrees acceptably with the data; a value of ΔH_{conf} = 6 kcal/base pair seems too low, whereas 10 kcal/base pair seems slightly too high.

In the reaction of HCHO with the presumably double-stranded poly(dI), the only two reactive sites are the endocyclic imino groups, both of which are exposed in the coil form and buried in the helix form. Thus, eq 2 must be modified appropriately. Going through the same type of derivation as in the Appendix, it can be shown that the argument of the logarithm in eq 2 must simply be replaced by $(1 + K_IF)^2$, where K_I is the equilibrium constant for the reaction with the exposed inosine residues, as listed in Table II. Again the calorimetric ΔH_{conf} of melting is unknown, but, as seen in Figure 4c, reasonable values of ΔH_{conf} give reasonable agreement with the data.

We conclude from this section that, given all the uncer-

[7] One equilibrium constant requiring a more detailed description is that for the formaldehyde reaction with adenine when the bases still remain hydrogen bonded, designated $K_{A,out}$. In this reaction (the only one that can occur with the helical form of poly[d(A-T)]), the amino proton of adenine not involved in Watson–Crick hydrogen bonding is replaced by a methylol group, which then protrudes out into the large groove; it is shown in the following paper of this issue that such a group can indeed exist. Although difficult to measure, $K_{A,out}$ can be expected to be small, primarily due to steric hindrance at the monomer level, i.e., with the N-7 position of the adenine base (McGhee and von Hippel, 1975a; Engel and von Hippel, 1974). Such steric hindrance should also be expected to apply to the polymer. The numerical estimate given in Table II is the average value for several model compounds in which HCHO is forced to add in this unfavored isomeric position (McGhee, 1975). Exactly similar comments can be made about $K_{C,out}$, the equilibrium constant for reacting with a helical cytosine residue.

NMR studies (Engel and von Hippel, 1974) indicate that for *N*-2-methylguanosine, the methyl group prefers almost exclusively to residue as this outside isomer, and thus the best estimate for the outside equilibrium constant, $K_{G,out}$, is probably the same as for the reaction with the free amino group of *N*-1-methylguanosine, viz., 4.4 M^{-1}. Since it is unlikely, for steric reasons, that a dimethylol adduct can form with the guanosine amino group, the amino reaction is thus the same in both the double helix and the coil, and their effects cancel; they are included in the formulation only for completeness. As a slight qualification on the above assignments, there is some evidence, obtained with methylated guanosine polymers (Engel and von Hippel, 1974), suggesting that a methylol group attached to the amino group of guanosine may not fit easily into the small groove of the double-helical DNA structure. In this case $K_{G,out}$ would have a (small) net destabilizing effect.

TABLE II: Summary of Numerical Estimates of the Equilibrium Constants (25 °C) and Enthalpy Changes for the Reaction of Formaldehyde with Sites on Both Native and Denatured DNA.[a]

Reactive site	Equilibrium constant (M^{-1})	ΔH (kcal/mol)	Comments
Helix			
Adenine amino proton, syn to N7	$K_{A,out} = 0.7$	−4.3	See footnote 6.
Cytosine amino proton, syn to C5	$K_{C,out} = 0.8$	−6.5	See footnote 6.
Guanine amino proton, syn to N3	$K_{G,out} = 4.4$	−4.3	K taken to be same as for N-1-methylguanosine; enthalpy change taken to be same as for adenine amino group; see footnote 6.
Coil			
Adenine amino groups:			
Formation of monoadduct	$K_{A,1} = 12.2$	−4.3	Average of values determined for deoxyadenosine and 5′-dAMP; see also Table I.
Formation of diadduct	$K_{A,2} = 0.4$	−3.0	
Cytosine amino group			
Formation of monoadduct	$K_{C,1} = 12.0$	−6.5	Average of values determined for deoxycytidine and 5′-dCMP; see also Table I.
Formation of diadduct	$K_{C,2} = 0.4$	−1.5	
Guanine amino group			
Formation of monoadduct	$K_{G,2} = 4.4$	−4.3	K taken to be same as N-1-methylguanosine; enthalpy change taken to be same as adenine amino group.
Formation of diadduct	0		Probably no diadduct is formed because of steric hindrance at N-1 position.
Thymine imino group	$K_T = 2.4$	−3.4	Average of values determined for thymidine and 5′-TMP.
Inosine imino group	$K_I = 2.7$	−3.4	K is average of equilibrium constant determined for inosine, 5′-IMP and poly(I); enthalpy change taken to be same as thymine reaction.
Guanine imino group	$K_{G,1} = 2.7$	−3.4	Taken to be same as for inosine imino reaction.

[a] Estimates made mainly from McGhee and von Hippel (1975a,b); estimated uncertainty is ±10% in all values.

tainties, agreement between theory (eq 2) and experiment is as good as can be expected. With the more accurate equilibrium parameters determined from the nucleotide–HCHO reactions, eq 2 can thus be used to predict the T_m of a polynucleotide under a variety of conditions of HCHO, salt, etc.

Equilibrium Denaturation of Natural Heterogeneous DNA by Formaldehyde

When even a relatively simple heterogeneous DNA such as T7 DNA is mixed with modest amounts of formaldehyde (0.01–0.05 M) and subjected to the same cycle of slow heating and cooling used above with homogeneous polynucleotides, the denaturation and reannealing times are so long, and the difference between T_m values observed on heating and cooling is so great, as to make this approach essentially useless. Extensively cross-linking the DNA with mitomycin C did not help. Several other approaches to measuring the equilibrium T_m were also unsuccessful (McGhee, 1975). We must therefore calculate theoretically the equilibrium melting temperature of native DNA as a function of DNA base composition, formaldehyde, salt concentration, and other experimental variables.

The various formaldehyde reactions that can occur with DNA, both helix and coil, are listed in Table II. Statistical weights of the different bases in the presence of formaldehyde are assigned in a manner exactly analogous to that described in the Appendix for adenine and thymine. The DNA model used for the calculations is the nearest-neighbor Ising model, with heterogeneity being treated in the arithmetic mean or "pseudo-chemical" approximation. This approach has been described many times previously (see, e.g., Poland and Scheraga, 1970), and a full presentation of the present calculations is given in McGhee (1975).

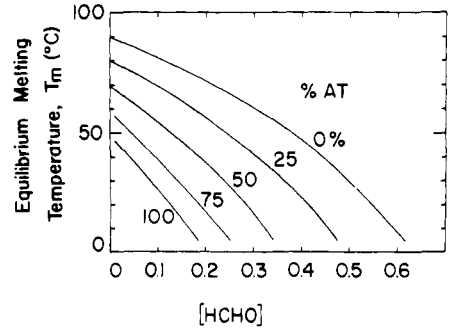

FIGURE 5: Calculated relation between DNA T_m and formaldehyde concentration, for DNAs of base composition ranging from 0–100% AT as marked on each curve. Calculations are described in the text and correspond to a solvent of approximately 0.01 M Na^+.

Figure 5 is the required "phase diagram" relating DNA melting temperature to the formaldehyde concentration, for DNAs with base compositions ranging from 0 to 100% AT. The parameters chosen to describe the DNA are: T_m of pure AT DNA = 50 °C; T_m of pure GC DNA = 90 °C; these parameters are close to those which apply to 0.01 M Na^+ (Schildkraut and Lifson, 1965). Based on the great number of uncertainties involved in the calculations, the curves of Figure 5 should probably be taken as accurate to only about ±10 °C at the higher HCHO concentrations.

It can be calculated that, at 25 °C, the destabilizing effect of formaldehyde is roughly the same for a GC base pair as for an AT base pair. At 0.1 M HCHO, this amounts to a free-energy contribution of about 0.57 kcal for an AT pair and about 0.59 kcal for a GC pair. This equivalence has long been implicitly assumed in the use of formaldehyde as a denaturant

to map DNA regions of differing base composition. However, at higher temperatures, since the equilibrium enthalpy change for the cytosine reaction is estimated to be larger than that for the adenine reaction, formaldehyde preferentially destabilizes AT-rich regions. For example, at 75 °C and 0.1 M HCHO, it can be calculated that an AT base pair would be destabilized by about 0.30 kcal, whereas a GC pair would be destabilized by about 0.23 kcal. However, this is still only a minor effect, and formaldehyde really shows remarkably little base specificity. The curves of Figure 5 for GC-rich DNA are flatter than those for AT-rich DNA simply because *all* the formaldehyde binding constants decrease with increasing temperature.

Since the equilibrium constants for the formaldehyde reaction depend hardly at all on ionic strength, and change only slowly with temperature, the curves of Figure 5 can also be applied to other salt conditions by simply shifting the T_m axis of the graph.

General Discussion

Formaldehyde, like other chemical probes of DNA structure, perturbs DNA conformational equilibria even as it monitors them. The present paper serves to establish an overall thermodynamic framework for the DNA–HCHO interaction; within this framework the kinetics of denaturation are considered in the following paper of this issue (McGhee and von Hippel, 1977).

We have shown, within the limits imposed by the very slow chemical reaction rates, that formaldehyde is indeed a *reversible* denaturant of DNA. We have also shown that the equilibrium reaction of formaldehyde with the bases in single-stranded DNA is the same (within experimental error) as in free mononucleotides. These two facts have allowed us to establish a simple theoretical relation (eq 2) describing quantitatively the dependence of the melting temperature of a double-stranded homopolynucleotide on formaldehyde concentration.

This relation, in turn, permits other approaches. For example, it allows us to calculate equilibrium melting curves for double-helical polynucleotides under experimentally inaccessible conditions, to establish experimental conditions under which denaturation will actually occur, and (if the reaction does not proceed) to determine whether this is due to a kinetic block or to an unfavorable final equilibrium.

One point emerging from eq 2, which will be important in subsequent discussions of the denaturation kinetics, is that not all the denatured base pairs need have reacted with HCHO, even in the middle of the formaldehyde-perturbed DNA transition. This means that one reacted base pair has the potential to "induce" the denaturation of a number of neighboring base pairs, the ratio of denatured to reacted base pairs increasing with increasing temperature.

Equation 2 can also be used to dissect the overall equilibrium T_m depression into contributions from the reaction of HCHO with the individual groups of the DNA bases. For example, we can calculate that if sufficient formaldehyde is present to depress the overall T_m of poly[d(A-T)] by 10 °C, the reaction with thymine will contribute only 20% of this destabilization, the remainder coming from the adenine reactions. Since (as will be shown in the following paper of this issue) the adenine and thymine reactions are kinetically distinguishable, such calculations will be used to sort out denaturation pathways.

We have argued that equilibrium melting curves should also exist for natural heterogeneous DNA reacting with formaldehyde. Although such curves could not be demonstrated experimentally for several reasons, they can be easily calculated from simple helix–coil theory. Such calculations have a number of practical uses. For example, in electron microscopy and hydrodynamic studies, formaldehyde has long been used to prevent DNA strand renaturation, but with no certain knowledge of how much formaldehyde is actually required or whether the concentrations used might be inadequate under conditions of low temperature and high salt. Appropriate reaction conditions for such experiments can now be calculated unambiguously.

Finally, the "phase diagrams" (Figures 4 and 5) relating DNA melting temperatures to formaldehyde concentrations should serve as prototypes of similar phase diagrams describing the interactions of DNA with melting proteins. While the hysteretic effects observed for DNA melting in the presence of formaldehyde are due to the slow rate of chemical reaction of HCHO with the bases, and as such do not have an exact analogue in DNA–melting protein interactions, hysteretic effects are also seen with phage T4 coded gene 32 protein. In this case, it is the cooperative-binding properties of the protein which kinetically block the melting of DNA even under conditions where the overall free energy of the system strongly favors the melted form (Jensen et al., 1976).

Appendix: Relation between Equilibrium Melting Temperature of a Homopolynucleotide and Formaldehyde Concentration

We wish to derive the theoretical relation between the melting point depression of a helical homopolynucleotide and formaldehyde concentration, based on the known association constants of HCHO with the various reactive sites of the bases. Similar equations have previously been derived by many investigators to treat the general case of any helix (polypeptide or polynucleotide) being denatured by reversible ligand binding (Schellman, 1958; Peller, 1959; Bixon and Lifson, 1966; Lazurkin et al., 1970; Crothers, 1971; Schellman, 1975). The present equation is derived specifically for the case of formaldehyde denaturing a homogeneous DNA such as poly[d(A-T)]. With minor changes, the same equation also applies to formaldehyde denaturation of other homopolynucleotides.

The process of forming an AT base pair at the end of a long helical stack can be written as

$$A + T \overset{s}{\rightleftharpoons} A{:}T$$

where A:T signifies the helical base pair, A + T signifies the two separated bases in the coil regions, and s is the usual "stability constant" (Poland and Scheraga, 1970).

In the helical A:T base pair, there is only one group which can still react with formaldehyde, namely, that proton on the adenine amino group which is not involved in hydrogen bonding and which projects out into the large groove. (Evidence is presented in the following paper of this issue that this group can indeed be reacted in the intact helix.) If the equilibrium constant for this reaction is $K_{A,out}$, the reaction can be denoted:

$$A{:}T + HCHO \overset{K_{A,out}}{\rightleftharpoons} {}^{f}\backslash A{:}T$$

where f is used to designate a methylol group.

There are three reactive sites for formaldehyde in the separated base pairs. The adenine amino group can react with formaldehyde (with an equilibrium constant $K_{A,1}$) to form a monomethylol adduct, and this monoadduct can react further

(with an equilibrium constant $K_{A,2}$) to form a dimethylol adduct. That is:

$$A + HCHO \overset{K_{A,1}}{\rightleftharpoons} {}^{f}\backslash A + HCHO \overset{K_{A,2}}{\rightleftharpoons} {}^{f}\backslash A {}^{/f}$$

Finally, the endocyclic imino group of thymine residues can react, with an equilibrium constant, K_T:

$$T + HCHO \overset{K_T}{\rightleftharpoons} T\text{-}f$$

Using the mass-action expressions for the various equilibrium constants, it is straightforward to show that:

total concentration of coil base pairs =
(concentration of unreacted A + T coil base pairs) ×
$(1 + K_{A,1}F + K_{A,1}K_{A,2}F^2)(1 + K_TF)$

where F is the free formaldehyde concentration.

Similarly:

total concentration of helix base pairs =
(concentration of unreacted A:T helical base pairs) ×
$(1 + K_{A,out}F)$

At T_m, the concentration of all species of helix base pairs is equal to the concentration of all species of coil base pairs. That is,

$$[A:T](1 + K_{A,out}F) = [A + T](1 + K_{A,1}F + K_{A,1}K_{A,2}F^2)(1 + K_TF)$$

Also the relation between the concentration of *unreacted* helical base pairs and the concentration of *unreacted* coil base pairs is given by the stability constant

$$s = [A:T]/[A + T]$$

where $s = 1$ at the unperturbed melting temperature of the DNA ($T_m°$) and, at any other temperature, will be given by the van't Hoff relation.

$$\ln\left(\frac{s_T}{s_{T_m°}}\right) = \int_{T_m°}^{T_m} \frac{\Delta H_{conf}}{R}\, d\left(\frac{1}{T}\right) = \frac{\Delta H_{conf}}{R}\left(\frac{1}{T_m} - \frac{1}{T_m°}\right)$$

where ΔH_{conf} is the enthalpy change for melting 1 mol of base pairs. Combining the last three equations yields the final desired relation:

$$\frac{1}{T_m} - \frac{1}{T_m°} = \frac{R}{\Delta H_{conf}} \ln \frac{(1 + K_{A,1}F + K_{A,1}K_{A,2}F^2)(1 + K_TF)}{(1 + K_{A,out}F)}$$

This equation was derived for the formation of a base pair at the end of a long helical stack, and the cooperative nature of the DNA transition has not been considered. However, it has been shown (Bixon and Lifson, 1966; Lazurkin et al., 1970; Crothers, 1971) that, since the HCHO reactions are assumed independent, the usual stability constant, s, can be replaced by

$$s^* = \frac{s(1 + K_{A,out}F)}{(1 + K_{A,1}F + K_{A,1}K_{A,2}F^2)(1 + K_TF)}$$

and s^* can then be used to replace s in any particular model for helix–coil theory desired. This equation will still be valid as long as cooperativity parameters, etc. are taken to be temperature independent. This is probably a fairly good approximation, especially at the high temperatures and for the limited range of T_m depressions dealt with in the present paper.

References

Adler, A. J., Grossman, L., and Fasman, G. D. (1969), *Biochemistry 8,* 3846.

Alberts, B. M., Amodio, F. J., Jenkins, M., Gutman, E. D., and Ferris, F. L. (1968), *Cold Spring Harbor Symp. Quant. Biol. 33,* 289.

Belman, S. (1963), *Anal. Chim. Acta 29,* 120.

Bixon, M., and Lifson, S. (1966), *Biopolymers 4,* 815.

Brimacombe, R. L. C., and Reese, C. B. (1966), *J. Mol. Biol. 18,* 529.

Crothers, D. M. (1971), *Biopolymers 10,* 2147.

Engel, J. D., and von Hippel, P. H. (1974), *Biochemistry 13,* 4143.

Fasman, G. D., Lindblow, C., and Grossman, L. (1964), *Biochemistry 3,* 1015.

Feldman, M. Ya. (1973), *Prog. Nucleic Acid Res. Mol. Biol. 13,* 1.

Freifelder, D., and Davison, P. F. (1963), *Biophys. J. 3,* 49.

Giles, K. W., and Myers, A. (1965), *Nature (London) 206,* 93.

Grant, R. C., Kodama, M., and Wells, R. D. (1972), *Biochemistry 11,* 805.

Griffin, B. E., Haslam, W. J., and Reese, C. B. (1964), *J. Mol. Biol. 10,* 353.

Hinz, H. J., Haar, W., and Ackermann, T. (1970), *Biopolymers 9,* 1923.

Inman, R. B. (1964), *J. Mol. Biol. 9,* 624.

Inman, R. B., and Baldwin, R. L. (1962), *J. Mol. Biol. 5,* 172.

Jensen, D. E., Kelly, R. C., and von Hippel, P. H. (1976), *J. Biol. Chem. 251,* 7215.

Kornberg, A. (1974), DNA Synthesis, San Francisco, Calif., W. H. Freeman.

Lazurkin, Yu. A., 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.

McGhee, J. D. (1975), Ph.D. Thesis, University of Oregon.

McGhee, J. D. (1976), *Biopolymers 15,* 1345.

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* (following paper in this issue).

Neal, M. W., and Florini, J. R. (1973), *Anal. Biochem. 55,* 328.

Peller, L. (1959), *J. Phys. Chem. 63,* 1199.

Poland, D., and Scheraga, H. A. (1970), Theory of Helix–Coil Transitions in Biopolymers, New York, N.Y., Academic Press.

Ross, P. O., and Scruggs, R. L. (1969), *J. Mol. Biol. 45,* 567.

Sarocchi, M.-T., Coutois, Y., and Guschlbauer, W. (1970), *Eur. J. Biochem. 14,* 411.

Scheffler, I. E., and Sturtevant, J. M. (1969), *J. Mol. Biol. 42,* 577.

Schellman, J. A. (1958), *J. Phys. Chem. 62,* 1485.

Schellman, J. A. (1975), *Biopolymers 14,* 999.

Schildkraut, C., and Lifson, S. (1965), *Biopolymers 3,* 195.

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.