Formaldehyde as a Probe of DNA Structure. II. Reaction with Endocyclic Imino Groups of DNA Bases†

James D. McGhee and Peter H. von Hippel*

ABSTRACT: We describe the equilibrium and kinetic aspects of the formaldehyde reaction with the endocyclic imino groups of derivatives of thymine, uracil, and a series of halogenated uracils, as well as poly(uridylic acid) and poly(inosinic acid). The main results are: (i) the equilibrium constants for forming a hydroxymethyl adduct remain quite constant at about 2–2.5 (M^{-1}) for all the compounds studied, independent of their pK_i; (ii) both forward and reverse rate constants with 5′-TMP are specific base catalyzed in the pH range of about 4–9; (iii) the response of the rate constants to temperature and to several solvent additives are measured; (iv) at neutral pH, for the series of pyrimidine compounds, a linear free energy relation is observed between the logarithm of both the forward and the reverse rate constant and the pK for deprotonation; (v) the unstructured polynucleotides, poly(U) and poly(I), react very similarly to their constituent monomers; (vi) a reaction mechanism is proposed; and (vii) some implications for polynucleotide studies are discussed. In an appendix, a method of spectral analysis is derived to obtain accurate estimates of the quite small equilibrium constants; this should be applicable to all similar two-component systems in which the final product is unobtainable, either by isolation or by saturation. Together with the results of the previous paper on the formaldehyde reaction with exocyclic amino groups (J. D. McGhee and P. H. von Hippel, preceding paper), these results form a reasonably comprehensive account of the basic chemical controls required to use formaldehyde as a quantitative probe of DNA structure.

In this paper we consider equilibrium and kinetic aspects of the formaldehyde reaction with nucleic acid components, such as thymidine, which possess an endocyclic imino group. The general aims of this study, and its relations to the overall study of DNA behavior and DNA–protein interactions, have been summarized in the preceding paper (McGhee and von Hippel, 1975).

Although Fraenkel-Conrat (1954) originally reported that the ultraviolet spectrum of uracil changed when mixed with formaldehyde, many workers subsequently have maintained that formaldehyde does not react with the endocyclic imino group of compounds such as thymidine, uridine, and inosine. This discrepancy presumably arose because (in contrast to the reaction with exocyclic amino groups of adenine, guanosine, and cytidine) the imino reactions are very fast, have smaller equilibrium constants, and are characterized by much smaller spectral changes. Nevertheless, Lewin (1962, 1964), by means of a "formol titration," demonstrated directly that these acid groups could indeed react with formaldehyde, and this finding was confirmed and extended by the spectrophotometric observations of Eyring and Ofengand (1967).

The marked difference between the reactions at the two classes of DNA sites (i.e., amino and imino) affords a unique opportunity to look at different aspects of the physical behavior of DNA, and in addition must obviously be considered in describing the equilibrium denaturation of DNA by formaldehyde. As in the preceding paper, we proceed by collecting equilibrium constants, rate constants, and spectral parameters for the various reactions, and investigate the dependence of these parameters on experimental variables. A number of nucleoside analogs are also studied, both because they provide insight into the reaction mechanism and because, in principle, one can take advantage of their distinctive optical and kinetic properties to incorporate them into polymers and observe their individual conformation- or environment-sensitive chemical reactions above the background of "regular" bases.

Materials and Methods

The various compounds used were obtained commercially as follows: 5′-TMP, thymidine, 5′-UMP, uridine, bromodeoxyuridine, fluorodeoxyuridine, and iododeoxyuridine from Calbiochem; inosine, 5′-IMP, and N^3-methyluridine from Sigma; N^6,N^6-dimethyladenine from Cyclo; and purine from Mann Research. These were checked for chromatographic purity in several solvents; bromodeoxyuridine and N^6,N^6-dimethyladenine showed trace contaminants, but were not purified further. Poly(uridylic acid) and poly(inosinic acid) were obtained from Miles, dissolved in 0.02 M phosphate–0.001 M EDTA (pH 6.95) by stirring in the cold, and dialyzed extensively against the 0.02 M phosphate buffer without EDTA. Neither polymer showed any temperature dependent absorbance change (indicative of structure) over the range of temperatures and salt concentrations of the present study.

Formaldehyde treatment and instrumentation are described in the preceding paper (McGhee and von Hippel, 1975). Unless stated otherwise, all experiments were carried out at a temperature of 24 ± 1° in buffer containing 0.01 M Na_2HPO_4 and 0.01 M NaH_2PO_4 at pH 6.95.

Results and Discussion

(*1*) *Structure of the Reaction Product.* The reaction of formaldehyde with nucleic acid components which lack an

† From the Institute of Molecular Biology and Departments of Biology and Chemistry, University of Oregon, Eugene, Oregon 97403. *Received August 8, 1974.* This investigation was supported in part by U.S. 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.

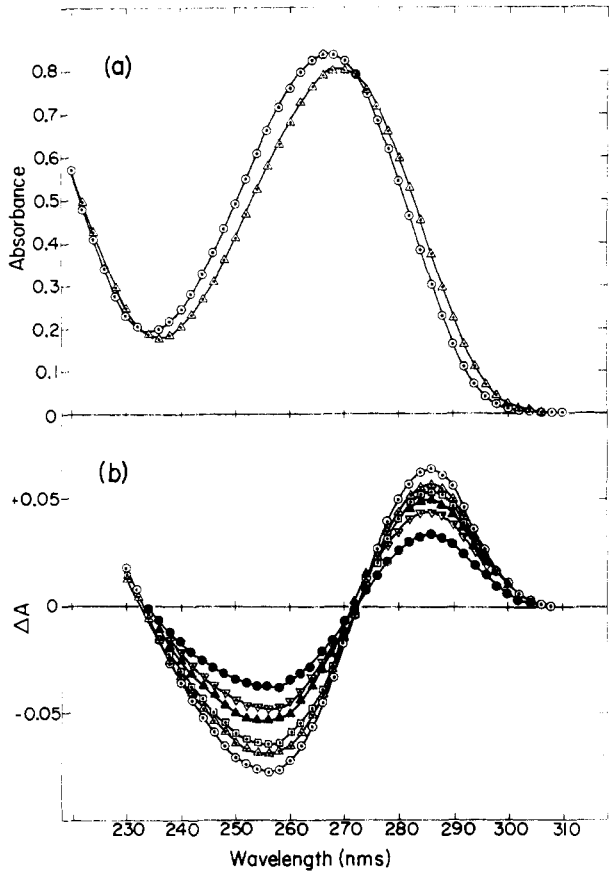
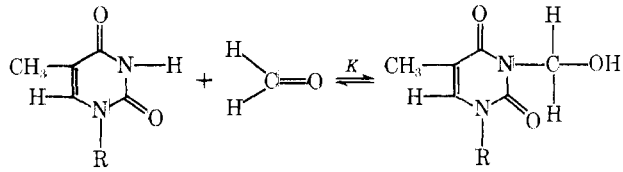


FIGURE 1: (a) Spectrum of thymidine: (○) control; (△) with 3.2 *M* formaldehyde. (b) Difference spectra of thymidine in the presence of various concentrations of added formaldehyde: (●) 0.3 *M*; (▽) 0.5 *M*; (▲) 0.7 *M*; (□) 1.1 *M*; (△) 1.3 *M*; (○) 2.1 *M* formaldehyde.

exocyclic amino group, but have an endocyclic "imino" group, can be written as follows (using thymidine as an example):



where the endocyclic N-3 proton is replaced by a hydroxymethyl or methylol group. The site of reaction is demonstrated by "formol titration" [in which formaldehyde is found to raise the apparent p*K* for the dissociation of this endocyclic proton (Lewin, 1964)], by analogy with the cyanoethylation reaction (Eyring and Ofengand, 1967), and by the observation (present study) that neither N^3-methyluridine (at neutral pH) nor 5′-TMP (at pH 12) shows any significant absorbance change on being incubated with formaldehyde. That the adduct is indeed a methylol or hydroxymethylol group is suggested from the arguments of the previous paper for the reaction with amino groups, by the similarity between the ultraviolet spectra of the reaction products and the corresponding methylated compounds, and by the chemical "unreasonableness" of the alternative cationic Schiff base.

(*2*) *Equilibrium Constants and Spectral Parameters.* Figure 1a shows the spectra for thymidine alone and in the presence of a high concentration (3.2 *M*) of formaldehyde; the spectral change on reaction with HCHO is quite small, but nevertheless significant. Difference spectra for thymidine at a number of intermediate formaldehyde concentrations (Figure 1b) show a peak at 286 nm, a valley at 256 nm, and a reasonably good isosbestic wavelength at 272–273 nm. As discussed in the preceding paper, the usual method of determining the equilibrium constant from such data is to measure, as a function of formaldehyde concentration, the absorbance changes at one particular wavelength (usually either at the peak or the valley of the difference spectra), and then to plot these changes according to standard mass action relations. However, for these small changes in absorbance such plots were found to be very sensitive to experimental error, and therefore not very effective in proving that there are indeed only two absorbing compounds in the system, as supposed in the above reaction scheme. To bypass these difficulties, a method of spectral analysis was devised in which a multi-wavelength spectrum **S**(F) of the nucleotide or nucleoside in the presence of a certain formaldehyde concentration, F, is expressed as a linear combination of two "base" spectra; **S**(O), the spectrum of unreacted starting compound, and **S**(H), the spectrum of the compound in the presence of a fixed high formaldehyde concentration, H. Algebraically

$$\mathbf{S}(\mathrm{F}) = \alpha\mathbf{S}(\mathrm{O}) + \beta\mathbf{S}(\mathrm{H})$$

where the coefficients α and β are functions of the formaldehyde concentration. It is shown in detail in the appendix that a plot of the ratio α/β vs. the reciprocal of the formaldehyde concentration should be linear, with equilibrium constants determined from the slope and intercept according to eq App-5. This method of analysis should be applicable to any two-component system in which the product spectrum is unobtainable, either by isolation or by saturation, and has the following advantages: the data from 25–30 wavelengths over the entire spectrum are used, with a considerable increase in the accuracy and reliability of the estimated equilibrium constants; the spectrum of the pure addition product can be obtained; and a consistency relation in the analysis indicates whether the experimental system does indeed conform to the two-component reaction scheme.

Typical data from such analyses are shown for 5′-TMP, 5′-UMP, and for poly(U) in Figure 2a–c, respectively; all plots are seen to be linear as required by eq App-5. Equilibrium constants and spectral parameters are collected in Table I for a variety of compounds with reactive endocyclic nitrogens, and lead to the following general observations. (i) Except for the last two entries in Table I, the estimated equilibrium constants are quite low (2–3 M^{-1}), and adduct spectra are generally shifted 2–3 nm toward longer wavelengths, accompanied by slight intensity decreases. The equilibrium constants agree within experimental error with those reported for uridine (Eyring and Ofengand, 1967) and for 5′-UMP (Aylward, 1966). (ii) The last two entries in Table I, dimethyladenine and purine, where the reaction presumably takes place at the N-9 position, appear to fall into a separate class characterized by considerably higher equilibrium constants ($\geq$10 M^{-1}), in reasonable agreement with values determined (by titration) for adenine and purine (Lewin and Barnes, 1966). (iii) Based on comparisons of nucleosides and nucleotides, there appears to be no significant effect of the 5′-phosphate group on either the equilibrium constant or the spectral changes accompanying adduct formation. (iv) Uracil and thymine derivatives show virtually identical equilibrium and spectral parameters. (v) Increased ionic strength (up to 1 *M* NaCl) has no effect on

Table I: Equilibrium Constants and Spectral Parameters for the Formaldehyde Reaction.[a]

Compd	K^b (M^{-1})	Difference Spectrum			Extrapolated Product Spectrum	
		Isosbestic[c]	Peak	Valley	λ_{max} (nm)	$\epsilon \times 10^{-3}$
Thymidine	2.36	272.5	286	256	270	9.2
Thymidine + 1 *M* NaCl	2.47	272.6	286	256	270	9.2
5′-TMP	2.40	273.2	286	256	270	9.7
5′-TMP + 1 *M* NaCl	2.46	272.8	286	255	270	9.7
Uridine	2.35	267.1	281	251	264	9.7
5′-UMP	2.46	268.4	281	251	264	9.5
Poly(U)	2.30	266.4	280	250	262	8.4
Iododeoxyuridine	2.10	291.5	308	270	290	7.4
Bromodeoxyuridine	1.97	282.0	298	266	282	8.7
Fluorodeoxyuridine	1.93	269.8	288	252	272	9.1
Inosine	2.69	265.8	278	248	250	10.5
5′-IMP	2.70	265.7	278	248	250	10.4
Poly(I)	2.64	264.8	278	242	250	6.9
Purine	10.3	279.0	244	268	263	7.1
N^6,N^6-Dimethyladenine	11.4	247.0	268		278	17.7

[a] 0.02 *M* phosphate, pH 6.95; 24 ± 1°. [b] Estimated uncertainty about ±0.2. [c] Estimated by linear interpolation.

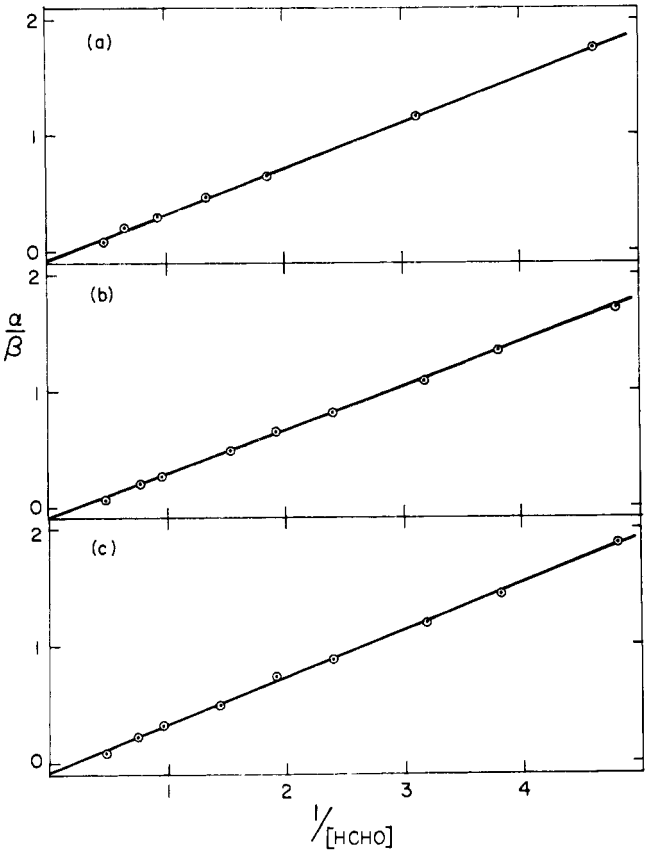


FIGURE 2: Plot of equilibrium spectral changes according to eq App-5 (see text): (a) 5′-TMP; (b) 5′-UMP; (c) poly(U).

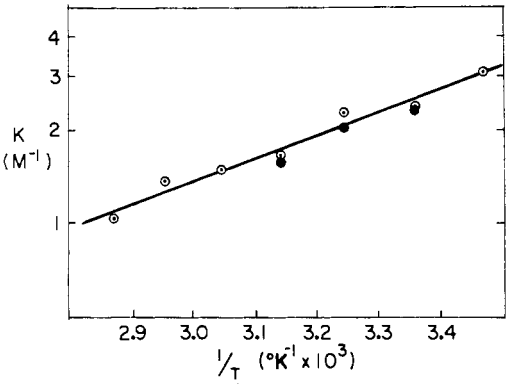


FIGURE 3: van't Hoff plot of the equilibrium constant for formaldehyde reaction as a function of temperature: (⊙) 5′-TMP; (●) thymidine.

the equilibrium of thymidine or 5′-TMP with formaldehyde. (vi) The equilibrium constants for the single-stranded polynucleotides poly(U) and poly(I) are unchanged from those of the constituent monomers. (vii) In the pyrimidine series, the equilibrium constant remains essentially unchanged as the pK for dissociation of the N-3 proton changes from about 8 to 10 (as will be shown below, the same change in acidity has profound effects on the reaction kinetics).

To understand quantitatively how formaldehyde denatures DNA at different temperatures, the temperature dependence ($\Delta H°$) of these equilibrium constants must be determined. To this end solutions of 5′-TMP and thymidine were weighed into cuvets and preincubated, at various temperatures ranging from 15 to 75°, for at least 1 hr prior to taking spectra. Equilibrium constants obtained by the method of spectral analysis (summarized above and derived in the appendix) are plotted in the form of the van't Hoff relation in Figure 3. As found with the exocyclic amino compounds (see preceding paper), the reaction with endocyclic amines is exothermic, i.e., the equilibrium constants decrease with increasing temperature. From the slope of Figure 3, the enthalpy change for 5′-TMP has been estimated to be −3.4 ± 0.3 kcal/mol, very close to the value obtained with thymidine, and in quite good agreement with the value of −4.4 kcal/mol given by Aylward for 5′-UMP. Hence thermodynamic parameters estimated at 25° (i.e., $\Delta G° = -0.5$ kcal/mol; $\Delta S° = -10$ cal/deg) are probably similar for all the compounds of Table I (again with the exception of the last two entries).

(*3*) *Kinetic Constants.* Since formaldehyde is always in great excess, the kinetics of the reaction should be pseudo first order and the formalism used in the preceding paper should apply; i.e., a plot of ln $[(A_\infty - A_t)/(A_\infty - A_0)]$ vs. time should be linear, with slope equal to the pseudo-first-order rate constant, $k' = k_{12}[\text{HCHO}] + k_{21}$; where A_∞,

Table II: Forward and Reverse Rate Constants for Formaldehyde Reaction.[a]

Compd	pK_a	Forward, k_{12} ($sec^{-1} M^{-1}$)	Reverse, k_{21} (sec^{-1})
Thymidine	9.8[b]	0.077 ± 0.01	0.040 ± 0.004
5′-TMP	10.0[c]	0.028 ± 0.004	0.019 ± 0.001
Uridine	9.3[b]	0.22 ± 0.02	0.13 ± 0.01
5′-UMP	9.5[d]	0.073 ± 0.001	0.057 ± 0.003
Poly(U)		0.020 ± 0.002	0.011 ± 0.001
Iododeoxyuridine	8.2[e]	1.15 ± 0.07	0.70 ± 0.02
Bromodeoxyuridine	7.9[e]	2.0 ± 0.2	1.1 ± 0.05
Fluorodeoxyuridine	7.8[e]	2.7 ± 0.4	1.8 ± 0.1
Inosine		11.7 ± 0.4	4.0 ± 0.2
5′-IMP		5.8 ± 0.2	2.3 ± 0.1
Poly(I)		1.0 ± 0.08	0.44 ± 0.03

[a] 0.02 *M* phosphate, pH 6.95; 25.0 ± 0.2°. [b] Christensen et al. (1967). [c] Hurst et al. (1953). [d] Bock et al. (1956). [e] Berens and Shugar (1963).

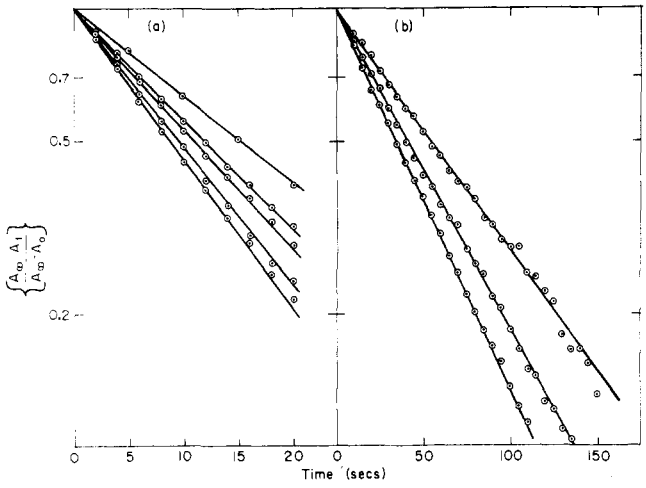


FIGURE 4: Pseudo-first-order plots of the reactions with various concentrations (0.1–0.5 *M*) of formaldehyde: (a) 5′-TMP; (b) poly(U).

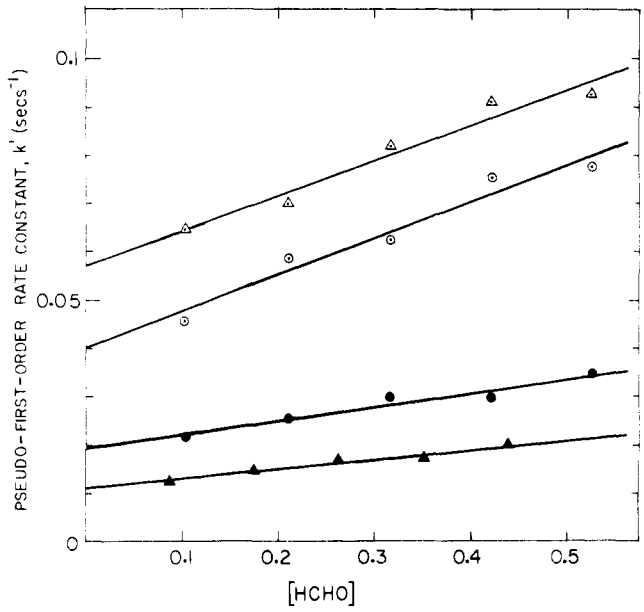


FIGURE 5: Plot of pseudo-first-order rate constant, *k′ vs.* formaldehyde concentration: (△) 5′-UMP; (⊙) thymidine; (●) 5′-TMP; (▲) poly(U).

A_0, and A_t are the final, initial, and current absorbance readings at the wavelength of observation, and k_{12} and k_{21} are the forward and reverse rate constants, respectively. Such plots are shown in Figure 4a and b for 5′-TMP and for poly(U), and are seen to be linear as required. The forward and reverse rate constants are given by the slope and intercept (respectively) of plots of the pseudo-first-order rate constant (k') vs. formaldehyde concentration, as illustrated for several compounds in Figure 5. Furthermore it was shown that the estimated rate constants are not significantly dependent upon the concentration of TMP, over the range of 2–20 μM.

At the neutral pH initially chosen for this study, the reaction with most of the compounds of Table I was sufficiently rapid so that the rates had to be measured in a stopped-flow apparatus; this is in qualitative agreement with literature reports that the reaction is "instantaneous" and that no adduct of formaldehyde with poly(U) could be isolated by chromatography (von Hippel and Wong, 1971). The estimated rate constants, and their standard errors, are collected in Table II, together with literature values for the pK_a for deprotonation of the pyrimidine compounds (with no correction made for ionic strength, etc.). The following generalizations summarize the data. (i) Under the same experimental conditions of pH, etc., the rates of reaction of formaldehyde with the compounds of Table II are several orders of magnitude faster than the corresponding rates with the nucleotide exocyclic amino groups (McGhee and von Hippel, 1975). (ii) In the series of pyrimidine compounds, the reaction rates increase with decreasing pK_a. Indeed, when the logarithm of either rate constant is plotted against pK_a, as in Figure 6, there is seen to be a good linear correlation for both the forward rate (r = 0.988) and the reverse rate (r = 0.993). The equations for the best fit least-squares lines of Figure 6 are

$$\text{forward reaction: } \log k_{12}(M^{-1}\text{sec}^{-1}) = -(0.83 \pm 0.06)\text{p}K_a + (6.9 \pm 0.5)$$

$$\text{reverse reaction: } \log k_{21}(\text{sec}^{-1}) = -(0.83 \pm 0.04)\text{p}K_a + (6.6 \pm 0.4)$$

[As noted before, the equilibrium constant is essentially constant over a two orders of magnitude change in compound acidity.] The two inosine compounds would fall above the lines of Figure 6, and perhaps on a parallel line. (iii) The reaction rates with the two polynucleotides studied appear to be considerably slower than with their constituent monomers. However, taking the pK_a of poly(U) in 0.03 *M* Na^+ as 10.3 [Michelson and Monny, 1966), the poly(U) rates fall quite close to the lines of Figure 6. Similarly, within the large uncertainties of the rate–pK relation, the reaction rates of poly(I) are consistent with the monomer

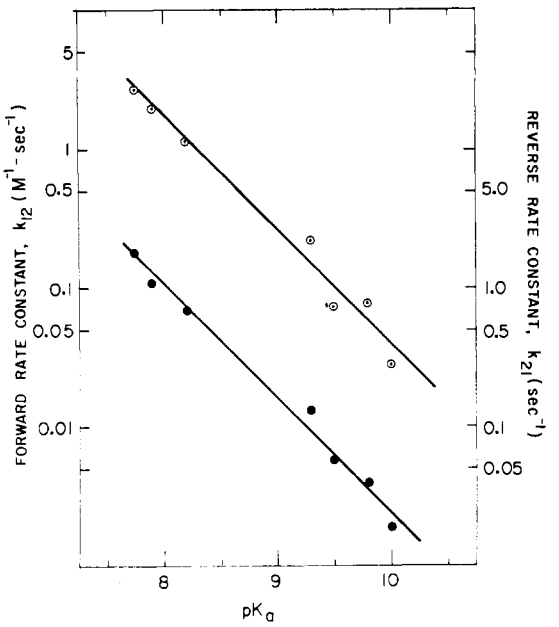


FIGURE 6: Plot of logarithm of rate constants vs. pK_a for the pyrimidine compounds of Table II: (⊙) forward rate constant, k_{12}; (●) reverse rate constant, k_{21}.

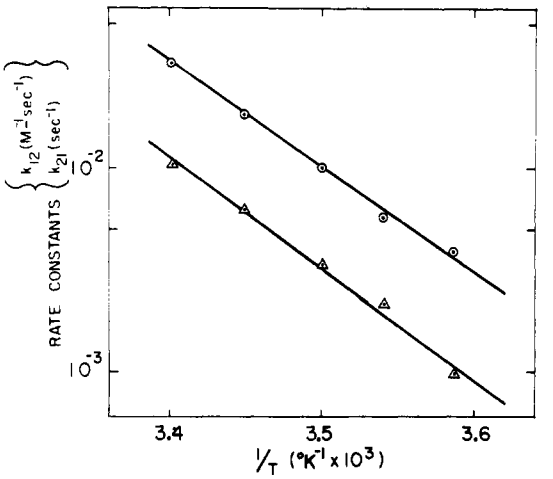


FIGURE 7: Arrhenius plots of rate constants for formaldehyde reaction with 5′-TMP: (⊙) forward rate constant, k_{12}; (△) reverse rate constant, k_{21}.

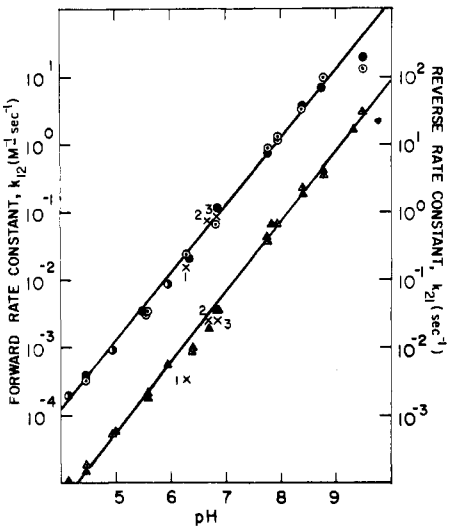


FIGURE 8: Plot of logarithm of rate constants of the formaldehyde reaction with 5′-TMP as a function of pH: (○,●) forward rate constant, k_{12}; (△,▲) reverse rate constant, k_{21}; open symbols at 0.1 *M* buffer; closed symbols at 0.02 *M* buffer; ionic strength, 1.0 *M*; symbols marked 1,2,3 represent reactions carried out in 5 *M* $NaClO_4$, 2.4 *M* tetraethylammonium chloride, and 3.0 *M* tetramethylammonium chloride, respectively.

rates. In both of these polymers, there are obviously no large steric effects on the rates of adduct formation. (iv) As noted with the exocyclic amino groups of the preceding paper, the ratio of forward to reverse rate constants is only approximately equal to the equilibrium constant. Thus for the compounds of Tables I and II these parameters are related by $k_{12}/k_{21} \simeq (0.8 \pm 0.2)K$. Possible reasons for this discrepancy are discussed in the preceding paper.

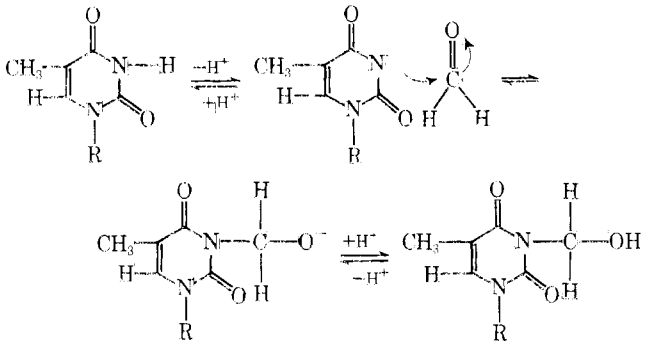
In the neutral pH buffer, the formaldehyde reaction with 5′-TMP is sufficiently slow so that, at temperatures below 25°, the kinetics can be followed in the Cary 14 spectrophotometer. Forward and reverse rate constants are plotted as an Arrhenius plot in Figure 7, and least-squares fits to these lines estimate the activation enthalpy for the forward rate constant as 23.4 ± 0.9 kcal/mol, and for the reverse rate constant as 25.1 ± 1.4 kcal/mol. (The difference of −1.7 ± 1.6 kcal/mol is in reasonable accord with the more accurate estimation of −3.4 kcal/mol determined above for the equilibrium enthalpy change.) If, as argued below, the reaction proceeds via a preequilibrium deprotonation step, then this observed $\Delta H^{\ddagger}$ will also include the equilibrium ΔH of ionization. Estimating the latter as 7.8 kcal/mol (Christensen et al., 1967) gives a $\Delta H^{\ddagger}$ ascribable to the reaction of HCHO with the anion of ~15.6 kcal/mol.

The kinetics of reaction with 5′-TMP were investigated at pH values ranging from 4 to about 9 (i.e., approaching the p*K* of deprotonation). (The buffers used, at concentrations of either 0.02 or 0.1 *M*, were acetate, cacodylate, phosphate, triethanolamine, methyl arsenate, and borate; ionic strength was maintained constant at 1.0 by addition of NaCl.] In Figure 8 the logarithms of both forward and reverse rate constants are plotted vs. pH. In contrast to the exocyclic amino compounds (McGhee and von Hippel, 1975), both the forward and reverse reactions are seen to be specific base catalyzed, the slopes of the least-squares line being 1.00 ± 0.02 and 1.03 ± 0.01, respectively. As required, the equilibrium constant is pH independent. There is little indication of general acid-base catalysis since the ratio of rate constants at the two buffer concentrations is 1.0 ± 0.1, for both the forward and the reverse reactions. A preliminary check of the inosine reaction at pH 5 indicates that it too is specific base catalyzed; in addition, a pH dependence has been reported for both 5′-UMP and poly(U) (Aylward, 1966).

After correcting for slight pH changes of the buffer, the reaction rates with TMP were found to increase only about twofold as the ionic strength was increased about 100-fold (from about 0.008 to 0.8 *M*); this is close to the change expected as a consequence of the equilibrium lowering of the acid p*K* by the added salt. Three points (marked 1, 2, and 3) are indicated on Figure 8 and represent the rate constants measured in the solvent systems of 5 *M* $NaClO_4$, 2.4 *M* tetraethylammonium chloride, and 3.0 *M* tetramethylammonium chloride, respectively. These are solvents that are found to have differential effects on the stability of DNA base pairs (Melchior and von Hippel, 1973), but are seen to have only minor effects on the TMP reaction kinetics.

General Discussion

(*1*) *Reaction Mechanism.* It has been demonstrated above, for the pyrimidine compounds studied, that there is a linear relation between the logarithm of either the forward or backward rate constant and the pK for deprotonation of the imino nitrogen, the rate increasing with increasing acidity with a slope of about 0.83. Given the uncertainty in some of the literature pK values (about ±0.2 unit), it is not clear that the slope actually differs from unity. If this slope *is* taken as unity, then the simplest reaction mechanism consistent with both this relation and the observed specific base catalysis is an equilibrium deprotonation step, followed by attack of the negatively charged nitrogen on the carbon atom of the unhydrated formaldehyde molecule, approaching perpendicular to the plane of the carbonyl group (Jencks, 1964). The final step would then be the fast reprotonation of the (now much more basic) ionized methylol group. This suggested mechanism can be diagrammed as follows, using thymidine as the example:



In contrast to the mechanism suggested in the preceding paper for the amino group reaction, where the initial nucleophilic attack appears to involve the neutral amino group which then passes through a tetrahedral transition state, the transition state of the current class of compounds must involve a planar nitrogen atom, since a tetrahedral nitrogen in the present case would abolish the resonance stabilization of the heterocyclic ring.

(*2*) *Implications for Polymer Studies.* The reaction with these imino compounds, even though their equilibrium constants are about fivefold lower than those of the amino-containing compounds, must obviously be considered in the quantitative description of DNA denaturation by formaldehyde. The equilibrium constants with the two polymers studied (poly(U) and poly(I)) are essentially identical with those of the constituent monomers. This observation, together with the general insensitivity of the equilibrium constants to large changes in compound acidity, gives us confidence in using these values (and their temperature dependencies) in a variety of different experimental situations, in particular in studying the properties of DNA molecules. In the class of compounds examined in this paper, the hydrogen atom replaced by the methylol group lies directly in the center of a normal Watson–Crick base pair. Thus if the mechanism suggested above is correct, and the reaction with formaldehyde only occurs after a prior deprotonation event, it would seem that reaction in a Watson–Crick polymer can occur only after this hydrogen bond has broken and the helix has "opened" or "breathed." Otherwise the steric requirements of the reaction seem modest (and, within the rather large uncertainties in correction for pK_a), the single-stranded polymers, poly(U) and poly(I), behave like their constituent monomers.

In conclusion, the position of the reactive (endocyclic) groups affords a unique opportunity to look into the center of a DNA helix, and this opportunity is exploited in subsequent papers (J. D. McGhee and P. H. von Hippel, in preparation).

Appendix

We present here a method of analysis by which equilibrium constants can be extracted from spectra of a nucleoside or nucleotide reacted with various concentrations of formaldehyde, assuming the simple reaction scheme given in the text. The need for such a multi-wavelength analysis arises because of the very small absorbance changes observed, and because the spectrum of the final reaction product cannot be obtained, either by isolation or (due to the very low equilibrium constants) by saturating with excess formaldehyde. Although the following discussion is explicitly in terms of a nucleotide–formaldehyde reaction, the same method of analysis can obviously be applied to other problems of the same type; for examples and a discussion of the general approach, see Lanczós (1956) and Magar (1972), among others.

The basic experiment consists of incubating a constant *total* concentration, C_0, of nucleoside or nucleotide, with varying *total* concentrations, [F], of formaldehyde; since the formaldehyde is always in great excess, [F] is also the free formaldehyde concentration. If $C_n(F)$ represents the concentration of unreacted nucleoside or nucleotide present at equilibrium in formaldehyde concentration [F], and $C_p(F)$ represents the corresponding concentration of product (i.e., hydroxymethyl adduct), then the initial nucleotide concentration is given by: $C_0 = C_n(F) + C_p(F)$, and the mass action relation for the reaction scheme gives

$$K = \frac{C_p(F)}{C_n(F)} \frac{1}{[F]} \qquad \text{(App-1)}$$

For any given formaldehyde concentration [F], and wavelength λ, the observed absorbance, $A_\lambda(F)$ can be expressed as

$$A_\lambda(F) = C_n(F)\epsilon_{n,\lambda} + C_p(F)\epsilon_{p,\lambda} \qquad \text{(App-2)}$$

where $\epsilon_{n,\lambda}$ and $\epsilon_{p,\lambda}$ are the molar extinction coefficients for nucleotide and product, respectively, both at wavelength λ. This reaction should hold at all wavelengths; thus by defining a spectrum, $\mathbf{S}(F)$, as an array of observed absorbance readings, and $\mathbf{E}_n$ and $\mathbf{E}_p$ as the corresponding array of molar extinction coefficients taken over the range of wavelengths of interest, eq App-2 can be extended as

$$\mathbf{S}(F) = C_n(F)\mathbf{E}_n + C_p(F)\mathbf{E}_p \qquad \text{(App-3)}$$

If $\mathbf{E}_n$ and $\mathbf{E}_p$ are known, this system of equations can then be solved for $C_n(F)$, $C_p(F)$ and hence the equilibrium constant; the difficulty in the present case is, as discussed above, that the vector $\mathbf{E}_p$ cannot be obtained. However, this difficulty can be bypassed by considering eq App-3 for one particular formaldehyde concentration, [F] = [H] (denoted by [H] since it is usually chosen as the highest concentration in the experiment). Thus

$$\mathbf{S}(H) = C_n(H)\mathbf{E}_n + C_p(H)\mathbf{E}_p$$

On rearranging, $\mathbf{E}_p$ can be expressed as a linear combination of $\mathbf{E}_n$ and $\mathbf{S}(H)$ both of which are known.

$$\mathbf{E}_p = \mathbf{S}(H)\frac{1}{C_p(H)} - \mathbf{E}_n\frac{C_n(H)}{C_p(H)} \qquad \text{(App-4)}$$

Now for any other formaldehyde concentration used in the experiment, $\mathbf{E}_p$ can be substituted from eq App-4 into

eq App-3 and thus the experimentally observed spectrum, $\mathbf{S}(F)$, can be expressed as a linear combination of $\mathbf{E}_n$, the original normalized nucleotide spectrum, and $\mathbf{S}(H)$, an observed spectrum at a particular formaldehyde concentration, [H]. When this substitution is performed [and $\mathbf{E}_n$ replaced by $\mathbf{S}(O)/C_0$, where $\mathbf{S}(O)$ is the experimentally measured spectrum of starting nucleotide in the absence of formaldehyde], the resulting equation is

$$\mathbf{S}(F) = \mathbf{S}(O)\left\{\frac{C_n(F)C_p(H) - C_n(H)C_p(F)}{C_O C_p(H)}\right\} + \mathbf{S}(H)\left\{\frac{C_p(F)}{C_p(H)}\right\}$$

For convenience we let the coefficient of $\mathbf{S}(O)$ be $\alpha(F)$, and the coefficient of $\mathbf{S}(H)$ be $\beta(F)$, where obviously both α and β are functions of formaldehyde concentration. (We note that $\alpha + \beta$ is identically equal to unity, and that this provides a measure of how closely the experimental system fits the two component model, as will be discussed below.)

By considering the ratio, α/β, of the two coefficients, and by using eq App-1 above for the equilibrium constant, the final relation to be used in the analysis is

$$\frac{\alpha(F)}{\beta(F)} = \frac{C_p(H)}{C_O \cdot K}\frac{1}{[F]} + \frac{C_p(H)}{C_O} - 1 \qquad \text{(App-5)}$$

Experimentally, spectra (usually 25–30 points at 2-nm intervals) are taken of the nucleotide or nucleoside after incubation to equilibrium in the presence of varying formaldehyde concentrations, ranging from 0 to about 5 *M* (more or less evenly spaced on a reciprocal scale). One such spectrum, usually at 3 *M* formaldehyde, is chosen as the basis vector, $\mathbf{S}(H)$, the other basis vector being the initial unreacted nucleotide spectrum, $\mathbf{S}(O)$. For each formaldehyde concentration, the system of 25–30 equations, $\mathbf{S}(F) = \alpha\mathbf{S}(O) + \beta\mathbf{S}(H)$, is solved for α and β by weighted least-squares methods. Finally the ratio α/β is plotted against $[F]^{-1}$ and the equilibrium constant determined from the slope and intercept according to eq App-5. In addition $C_p(H)$, and hence $C_n(H)$, can be determined and substituted into eq App-4 above, to yield E_p, the normalized spectrum of pure (and unobtainable) adduct.

As mentioned earlier, if the experimental system conforms to the two-absorbing-component model, not only should the plots of α/β vs. $[F]^{-1}$ be linear [i.e., eq App-5 and Figure 2] but the sum $(\alpha + \beta)$ should be unity. From 15 experiments using this type of spectral analysis on compounds expected to show only one product, the average value of the sum $(\alpha + \beta)$ was 0.9990 with a standard error of 0.0003. In contrast, analysis of the spectra of 5′-dAMP and formaldehyde, where two products can form, gave a nonlinear plot and a sum $(\alpha + \beta)$ ranging from 1.11 to 0.89 depending on the choice of basis spectra. N^6-Methyladenosine, on the other hand, where only one product is expected to be formed, gave a linear plot and a sum $(\alpha + \beta) = 1.004 \pm 0.002$.

References

Aylward, N. N. (1966), *J. Chem. Soc. B*, 627.

Berens, K., and Shugar, D. (1963), *Acta Biochim. Pol. 10*, 25.

Bock, R. M., Ling, N. S., Morell, S. A., and Lipton, S. H. (1956), *Arch. Biochem. Biophys. 62*, 253.

Christensen, J. J., Rytting, J. H., and Izatt, R. M. (1967), *J. Phys. Chem. 71*, 2700.

Eyring, E. J., and Ofengand, J. (1967), *Biochemistry 6*, 2500.

Fraenkel-Conrat, H. (1954), *Biochim. Biophys. Acta 15*, 307.

Hurst, R. O., Marko, A. M., and Butler, G. C. (1953), *J. Biol. Chem. 204*, 847.

Jencks, W. P. (1964), *Prog. Phys. Org. Chem. 2*, 63.

Lanczos, C. (1956), Applied Analysis, Prentice-Hall, Englewood Cliffs, N.J.

Lewin, S. (1962), *J. Chem. Soc. B*, 1462.

Lewin, S. (1964), *Experentia 20*, 666.

Lewin, S., and Barnes, M. A. (1966), *J. Chem. Soc. B*, 478.

Magar, M. E. (1972), Data Analysis in Biochemistry and Biophysics, New York, N.Y., Academic Press.

McGhee, J. D., and von Hippel, P. H. (1975), preceding paper.

Melchior, W. B., Jr., and von Hippel, P. H. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70*, 298.

Michelson, A. M., and Monny, C. (1966), *Proc. Natl. Acad. Sci. U.S.A. 56*, 1528.

von Hippel, P. H., and Wong, K. Y. (1971), *J. Mol. Biol. 61*, 587.