Formaldehyde as a Probe of DNA Structure. I. Reaction with Exocyclic Amino Groups of DNA Bases†

James D. McGhee and Peter H. von Hippel*

ABSTRACT: A comprehensive description is given of both the equilibrium and the kinetic aspects of the reaction of formaldehyde with the exocyclic amino groups of derivatives of adenine, cytosine, and guanine; the results extend previous data in the literature to the point where formaldehyde can now be used as a *quantitative* probe of DNA structure and dynamic behavior. The main results are: (i) the reaction product is proven (by isolation followed by nuclear magnetic resonance (NMR) spectroscopy) to be a hydroxymethyl group; (ii) a dihydroxymethyl adduct is shown to exist at high formaldehyde concentrations; (iii) equilibrium constants at 25° for forming the monoadduct with adenine and cytosine compounds are about 12 (M^{-1}), while those for forming the dihydroxymethyl adduct are about 0.4 (M^{-1}); (iv) the standard enthalpies for forming the monoadducts with adenine and cytosine compounds are about −4 to −6 kcal/mol; (v) indirect evidence is presented suggesting that a monohydroxymethyl group on adenine or cytosine derivatives exists preferentially as that rotational isomer which blocks Watson–Crick hydrogen bonding; (vi) in derivatives of guanine, it is shown that the N-1 endocyclic imino group can react with formaldehyde, as well as the amino group, the overall equilibrium constant being about 6 (M^{-1}); (vii) all rate constants are reported, as well as their response to temperature, pH, and various solvent additives known to perturb DNA structure; (viii) using a series of substituted anilines, a linear free energy relation is obtained between the logarithm of both the forward and the reverse rate constant for the formaldehyde reaction and the amine pK, over a range of 10^8 change in amine basicity; (ix) using this relation, the pK's for protonating the nucleoside amino groups are estimated to lie in the range of −2 to −4; (x) a reaction mechanism is proposed; and (xi) some implications of these results for polynucleotide studies are discussed.

The primary motivation of this and the following series of papers (McGhee and von Hippel, 1975; J. D. McGhee and P. H. von Hippel, manuscripts in preparation) is to develop an experimental system to investigate aspects of the physical behavior of DNA in solution which could be important in physiological processes. The two main sets of questions which we wish to approach are: (i) how does DNA interact with proteins or protein models; and (ii) what is the dynamic behavior of native DNA by itself? Answers to the first set of questions are obviously necessary to understand how genetic information is controlled, either by transcription or replication; answers to the second, i.e., information on the conformational motility or "breathing reactions" of native DNA, could be important in understanding recognition events at promoter sites, the kinetics and equilibria of the interaction of "open" DNA sequences with "melting proteins," or an initial DNA–DNA recognition event in recombination. A more detailed expression of this point of view is given in von Hippel and McGhee (1972). In addition many things learned about DNA–protein interactions may also apply to RNA–protein interactions, as, for example, in elucidating the structure and function of the ribosome.

Both of the above sets of questions can be approached using the simple reagent, formaldehyde. As we will demonstrate in later papers, since formaldehyde reacts *reversibly* with nucleic acid monomers, it is capable of bringing about the *reversible* melting of DNA. Thus, after the proton (and hydroxyl ion), HCHO is the simplest member of the class of DNA denaturants of which the T4 gene-32 protein is currently the most sophisticated (Alberts and Frey, 1970). However, since DNA–formaldehyde reactions are so very much simpler and more experimentally accessible than those of DNA with the biological proteins of ultimate interest (i.e., HCHO reacts noncooperatively, with one base pair at a time, and with only a modicum of base specificity), the reaction can be used both to extract quantitative information about DNA structure and stability, and to test statistical mechanical theories which can later be applied to real DNA–protein systems.

To study dynamic aspects of DNA structure, formaldehyde can be used kinetically as a typical "chemical probe" (e.g., Utiyama and Doty, 1971; Dean and Lebowitz, 1971; von Hippel and Wong, 1971). Here the overall reaction kinetics with the native structure are compared to rates with unstructured models to gain insight into both the nature and the frequency of appearance of reactive sites in the otherwise native DNA molecule. This approach provides information on the dynamic structure of the double helix which is complementary to that obtained by other techniques such as hydrogen exchange (e.g., McConnell and von Hippel, 1970; Teitlebaum and Englander, 1975).

In this and the following paper (McGhee and von Hippel, 1975), we characterize in detail the chemical reactions of formaldehyde with the monomeric components of nucleic acids, as the crucial chemical "control" on which studies of the nucleotide polymers must be based. To utilize a chemical probe effectively, one needs (ideally) to have the following information: (i) a complete description of the reaction site or sites, both in the polymers and in the monomers; (ii) equilibrium constants for reactions at each of the above sites; (iii) kinetic parameters for each site (i.e., rate con-

† 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.

stants and ideally complete reaction mechanisms, or at least steric requirements for reaction); (iv) complete descriptions of spectral or other changes characteristic of the monomer reactions which are needed to interpret polymer spectra; (v) information on the response of all the above to experimental variables, such as temperature, ionic strength, pH, and other solvent perturbants, which will allow us to distinguish between the conformational behavior of the polymer, and the different chemical reactions.

Formaldehyde has long been known to react with nucleic acids (Fraenkel-Conrat, 1954), and has been put to many uses in nucleic acid chemistry (see the recent review by Feldman, 1973), such as: (i) in viral inactivation and vaccine production; (ii) as an "irreversible" denaturant for studying single-stranded polynucleotides (Berns and Thomas, 1961; Inners and Inners, 1974) and to stabilize locally melted configurations for electron microscopy (Inman, 1966); (iii) as a means of quantitating the fraction of bases that are involved in hydrogen bonding in structures such as tRNA (Penniston and Doty, 1963) and viral RNA (Boedtker, 1967); (iv) to show that hydrogen bonds are not involved in the maintenance of the (presumably) single-stranded polynucleotide structures of neutral poly(rC) (Fasman et al., 1964) or poly(rA) (Stevens and Rosenfeld, 1966); (v) to quantitate the result of treatment of DNA by agents such as ultraviolet light, X-rays, nucleases, etc. (Lazurkin et al., 1970); (vi) as mentioned above, to study the dynamic behavior of native DNA (Freifelder and Davison, 1963; Lewin, 1964; Trifonov et al., 1967; Utiyama and Doty, 1971; Dean and Lebowitz, 1971; von Hippel and Wong, 1971); (vii) to study the detailed conformations of tRNA (Axelrod et al., 1969; Rosenfeld et al., 1970) and mRNA (Lodish, 1971); (viii) to study the features of more complicated structures such as viruses (Tikchonenko and Dobrov, 1969), ribosomes (Cox, 1969; Tal, 1970), and chromatin complexes (Brutlag et al., 1969; Clark and Felsenfeld, 1971; Li, 1972).

Although the basic chemistry of the reaction of formaldehyde with the nucleic acid monomers has been studied by several investigators (Haselkorn and Doty, 1961; Grossman et al., 1961; Feldman, 1964; Lewin, 1964; Lewin and Barnes, 1966; Boedtker, 1967), in most cases the literature data were found to be insufficient for present purposes, and, in several instances, to be based on erroneous methods of analysis; thus many of the above polymer studies have been inhibited or at least rendered much less incisive by this lack of fundamental quantitative information about the chemical reaction. We here attempt to provide such information, obtained under experimental conditions generally used for polymer studies, i.e., we consider primarily ranges of pH, temperature, etc., under which DNA is normally native. In this paper, we treat the reactions with nucleotides containing exocyclic amino groups (i.e., variously substituted adenines, cytosines, and guanines) and in the following paper (McGhee and von Hippel, 1975) the corresponding reactions with nucleotides containing endocyclic imino (or secondary amino) groups (i.e., variously substituted thymidines, inosines, and relevant analogs).

Materials and Methods

The various compounds used in this study were obtained commercially as follows: deoxyadenosine from Schwarz/Mann, 5′-dAMP and 5′-dCMP from Calbiochem; N^7-methyladenine and N^9-methyladenine from Cyclo; deoxycytidine, deoxyguanosine, 5′-dGMP, N^6-methyladenosine, 5-methyldeoxycytidine, N^1-methylguanosine, N^2-methylguanosine, N^1-methyladenosine, and N^3-methyldeoxycytidine from Sigma. These were checked for purity by chromatography on Eastman Chromagram cellulose sheets in three or four solvent systems. N^1,N^4-Dimethylcytosine was synthesized and kindly provided by Dr. J. Engel; 2-chloro-7-methyladenine and 2-chloro-9-methyladenine were gifts from Dr. R. Stewart.

Unless stated otherwise, the buffer used throughout consisted of 0.01 M Na_2HPO_4 and 0.01 M NaH_2PO_4 at pH 6.95; after Millipore filtration it was stored in the cold over a drop of chloroform. Doubly distilled water was used in all solutions. Room temperature was 24 ± 1°.

Mallinckrodt analytical reagent formalin, containing about 37% formaldehyde and 10–15% methanol, was treated as follows for each experiment: the clear solution was decanted from the small amount of solid polymeric precipitate which forms on standing (Walker, 1964), sealed in a glass ampoule, and placed in a boiling water bath for 15 min in an attempt to accelerate depolymerization (Freifelder and Davison, 1963; Walker, 1964). Any formic acid present was removed and the pH raised to about neutrality by passing the cooled solution through a small (~2 ml) column of Bio-Rad AG-2-X8 anion exchange resin prepared in the hydroxide form. The total formaldehyde content was then assayed, usually in triplicate, using a modified Na_2SO_3 method, back-titrating to pH 8.95 (Haslam and Squirrel, 1957). The average concentration of the stock solutions as assayed in these experiments was 13.3 ± 0.2 M. To bring this formaldehyde to the desired buffer concentration, four volumes of the stock were mixed with one volume of a fivefold concentrated buffer solution. Further dilutions were usually dispensed by weighing, and final concentrations are accurate to ~2%. We thus define our formaldehyde reagent as that treated as above and assayed by the sodium sulfite titration. Although this preparation is totally reproducible, it must be remembered that the original formalin solutions contain methanol and complex mixtures of polymers (Walker, 1964); for equilibrium experiments, at the concentrations present in final reaction mixtures, any effects of these constituents seem to cancel (Kallen and Jencks, 1966).

All spectra and most rate measurements were obtained on a Cary-14 spectrophotometer, interfaced to a Varian 620i computer. This apparatus permits the measurement of digitized spectra to high accuracy, as well as the automatic subtraction of base lines and calculation of difference spectra, etc.; a detailed description of this system is in preparation (W. B. Melchior *et al.*, to be submitted). Temperatures were measured with a thermocouple inserted in a dummy cuvet, and are accurate to ~±0.2°. Stopped-flow measurements were made on an instrument designed by E. Swanson and Dr. S. Bernhard and kindly made available by them; voltages were read directly from an oscilloscope, scope sweep time corrections were made when required and weighted first-order fits to the data were carried out on a PDP-10 computer. Some of the initial rate studies were performed on a Gilford Model 2000 recording spectrophotometer. The nuclear magnetic resonance (NMR) spectra were recorded at 25° on a Varian XL-100 spectrometer, using a deuterium lock arrangement.

Results and Discussion

(*1*) *The Structure of the Reaction Product.* As has been previously pointed out by Feldman (1973) and as will be-

come apparent in this and the subsequent paper (McGhee and von Hippel, 1975), formaldehyde can, in principle, replace any proton bonded to a nitrogen atom of a nucleic acid base; these include protons on the exocyclic amino groups, as well as protons attached to endocyclic nitrogens of the rings. Formaldehyde does not appear to react to a measurable extent under our conditions with either sugar hydroxyl or phosphate ester groups (Feldman, 1973).

Although the formaldehyde reaction with nucleic acid components has been in use for over 20 years, the exact structure of the reaction product still remains controversial (see Brown, 1974), the two main alternatives being a hydroxymethyl or methylol adduct ($RN(H)CH_2OH$) (Grossman et al., 1961; Utiyama and Doty, 1971) or a Schiff base ($RN{=}CH_2$) (Fraenkel-Conrat, 1954). The strongest evidence, favoring the hydroxymethyl alternative, is the similarity of the ultraviolet spectral properties of the adduct to those of the corresponding methylamino compounds. The principal obstacle to resolving this dilemma is that the formaldehyde adduct with the nucleic acid bases is quite unstable and is maintained in aqueous solution only in the presence of a substantial excess of free formaldehyde; thus it cannot be isolated in pure form for the required structural studies. As the adduct structure is obviously central to the whole problem, we have carried out a number of different experiments which overwhelmingly indicate that the major reaction product is indeed the hydroxymethyl derivative. These experiments also rule out the formation, even at very low levels, of Schiff bases or certain other types of derivatives.

Although the isolated adduct is indeed unstable, a direct NMR spectrum of a formaldehyde–nucleotide adduct is obtainable since the decomposition rate of the adduct is quite slow (e.g., half-times $\simeq$8–10 hr for the dAMP adduct at 25°, as will be seen from the rate studies to follow). Thus samples of 5′-dAMP were preequilibrated with various amounts of formaldehyde, and quickly passed through a P-2 Bio-Gel column (1 × 40 cm) to separate the adduct (mol wt $\approx$350) from the excess unreacted formaldehyde. The emerging front was quick frozen, lyophilized, twice dissolved in D_2O and relyophilized, and finally dissolved in D_2O. The low-field portion of a typical NMR spectrum is shown in Figure 1a, for dAMP preincubated with 0.5 *M* formaldehyde. In this sample, based on the equilibrium constants reported below, about 5–10% of the nucleotide remains unreacted. To determine which peaks are due to the adduct, the sample was then placed in a boiling water bath for 5 min to decompose the adduct and a second spectrum recorded (Figure 1b); a control spectrum of unreacted dAMP is also shown (Figure 1c). The three spectra were aligned relative to a peak of added acetone (0.5%) and chemical shifts expressed as ppm downfield from acetone. We assign the following peaks in the adduct spectrum (Figure 1a):

(i) The peak at ~2.85 ppm is assigned to the two C–H protons of the hydroxymethyl group, since it is absent in the dAMP-only control (Figure 1c; the very small peak observed here is a spinning side band of the water peak centered at ~2.5 ppm). The hydroxymethyl C–H peak is almost abolished by boiling the solution (Figure 1b), and it falls close to the position of the corresponding peak seen in hydroxymethylimidazole (Dunlop et al., 1973). The resonances of the protons of the Schiff base alternative would be expected to fall much further downfield (Szymanski and Yelin, 1968; Utiyama and Doty, 1971).

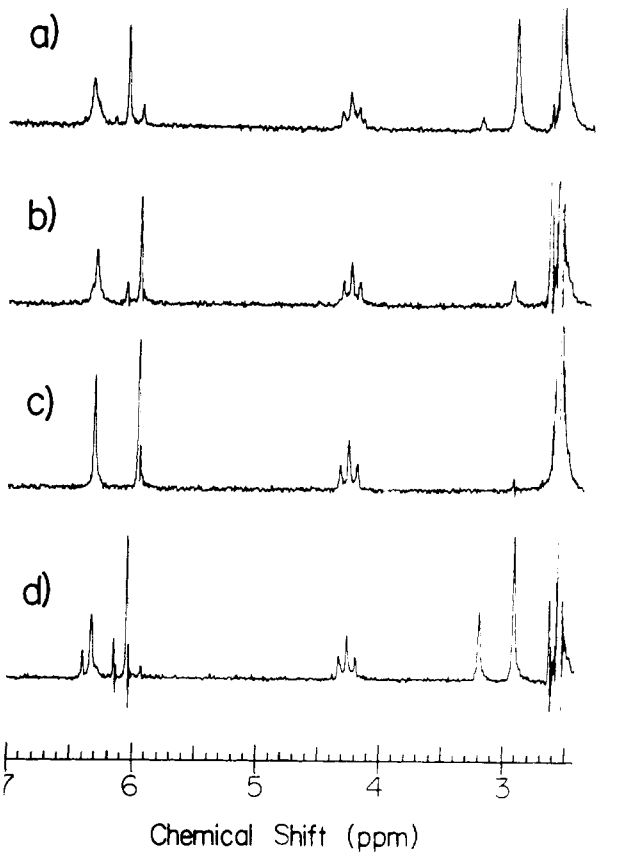

FIGURE 1: NMR spectra in D_2O at 25° and 100 MHz; expressed as ppm downfield from an internal standard of acetone: (a) 5′-dAMP, preincubated with 0.5 *M* HCHO and excess formaldehyde removed as described in text; (b) solution of spectrum (a) after boiling for 5 min and recooling to 25°; (c) 5′-dAMP only control; (d) 5′-dAMP, preincubated with 3 *M* HCHO and excess formaldehyde removed as described in text.

(ii) The peak at ~5.9 ppm is assigned to the C-2 proton of unreacted dAMP (T'so, 1974). The much larger peak at ~6.0 ppm is assigned to the C-2 proton of reacted dAMP, and is seen to revert to the dAMP control on boiling (Figure 1b). Within the accuracy of measurement and under a variety of conditions, the ratio of the adduct C-2 peak is, as required, about half that of the hydroxymethyl peak assigned above. Furthermore, if repeated spectra are taken at room temperature over a period of days, the relative proportions of the two peaks at 6.0 and 5.9 ppm change in a reciprocal fashion as the adduct slowly decomposes.

(iii) The peak at ~6.35 ppm is assigned to the C-8 proton of the reacted dAMP, with a shoulder at ~6.3 ppm arising from unreacted dAMP. This peak shifts slightly upfield on boiling and the same comments apply as for the C-2 peaks just discussed, except that there is much less splitting between the peaks of the adduct and unreacted material.

(iv) The peaks at 4.2–4.3 ppm are assigned to the C_1' protons of the adduct; these are shifted slightly downfield from the control and revert on boiling.

(v) The peak at ~2.6 ppm in Figure 1b is assigned to free formaldehyde (see also Dunlop et al., 1973). The area of this peak is seen to increase in direct proportion to the decrease of that of the hydroxymethyl group.

To correlate the above NMR spectra with solution studies, where formaldehyde is in great excess, we checked that the uv spectrum of the redissolved material used for Figure 1a had an absorption maximum at 265 nm, compared to

260 nm for dAMP and 265 nm for N^6-methyldeoxyadenosine (Sober, 1968).

If we accept that the primary reaction product is indeed a hydroxymethyl or methylol derivative, rather than a Schiff base, there now exists the possibility of forming a diadduct, i.e., a dimethylol derivative of formula $RN(CH_2OH)_2$, as has long been recognized to occur with aliphatic amines (French and Edsall, 1945; Walker, 1964; Kallen and Jencks, 1966). With nucleic acid components this possibility has either not been recognized or has been rejected (Scheit, 1965; Feldman, 1973). Nevertheless, with high formaldehyde concentrations, clear evidence can be obtained for the presence of a diadduct; for example, Figure 1d is a spectrum of dAMP preincubated with the relatively high concentration of 3 *M* formaldehyde, and then isolated as before. Three new distinct peaks arise, which we assign as follows. (i) The peak at ~3.15 ppm is assigned to the C-H protons of the dihydroxymethyl adduct; a small peak is also seen at the same position in Figure 1a. (ii) The peak at ~6.15 ppm is assigned to the C-2 proton of the dihydroxymethyl adduct; again a small counterpart peak is seen in Figure 1a. (iii) The peak at ~6.4 ppm is assigned to the C-8 proton of the dihydroxymethyl adduct.

Within the accuracies of measurement, the area of the peak at 3.15 ppm is, as required, about four times larger than the peaks at 6.15 and 6.4 ppm. In addition to this direct indication of a dihydroxymethyl adduct, there is a host of corroborative evidence, such as: (i) at an appropriately chosen wavelength of observation, a distinct slower second step can be seen in the reaction kinetics; (ii) as will be discussed in detail in the next section, uv spectra in the presence of varying amounts of formaldehyde do not exhibit isosbestic wavelengths, spectral rank analysis indicates the presence of three absorbing components, and the resulting equilibrium plots are nonlinear; and (iii) compounds such as N^6-methyladenosine react with formaldehyde, indicating that there is no absolute steric hindrance to formation of a diadduct.

While the above indicates that hydroxymethyl adducts are indeed the main reaction product, there remained the possibility that traces of Schiff base might be present, which should not only be taken into account for a complete analysis, but might be a potentially useful means for introducing methyl groups into nucleic acids, as has indeed been done with proteins (Means and Feeney, 1968; Paik and Kim, 1972). Nevertheless several attempts at borohydride reduction were unsuccessful (J. Engel, unpublished experiments) and, furthermore, uv spectra of concentrated nucleotide-formaldehyde solutions (OD_{max} = 25) in the region of 300–400 nm showed no significant absorbance which could possibly be attributed to the presence of a Schiff base.

A further possibility for the adduct structure, and one which might not be detected in either uv or NMR spectra above, is a paraformaldehyde polymer (Walker, 1964) of the form $RN(H)CH_2(—OCH_2)_nOCH_2OH$, of variable and unknown length. To determine the amount of formaldehyde that is actually attached to a nucleotide at equilibrium, dAMP, preequilibrated with about 1 *M* formaldehyde, was quickly passed through a Bio-Gel P-2 column, the various fractions were boiled for 15 min to reverse (>99%) the formation of adduct, and the free formaldehyde was assayed by a slight modification of the fluorometric method of Belman (1963). The ratio of formaldehyde bound per dAMP molecule was found to be within experimental error of the ratio expected if only mono- and dihydroxymethyl adducts were present. Thus we conclude that either the polymeric adducts decompose very rapidly on passing through the gel column (unlikely in view of the slow rate of free polymer decomposition; Walker, 1964), or, more likely, that such adducts are not formed to any appreciable extent with nucleic acid components.

A further type of adduct, which appears to be completely negligible under the conditions of the present study, is the formaldehyde cross-linked methylene dinucleotide (see Feldman, 1973 and earlier papers).

(*2*) *Equilibrium Constants and Spectral Parameters.* In light of the above, the overall reaction for formaldehyde with a nucleoside amino group can be written as follows:

$$R—\overset{H}{\underset{|}{N}}—H + CH_2O \overset{K_1}{\rightleftharpoons} R—\overset{H}{\underset{|}{N}}—CH_2OH + CH_2O \overset{K_2}{\rightleftharpoons} R—\overset{CH_2OH}{\underset{|}{N}}—CH_2OH$$

In this section, we measure the two equilibrium constants, K_1 and K_2 (M^{-1}), as well as the spectral parameters related to adduct formation. Previous estimates of these parameters appearing in the literature have not taken the second reaction above into account and could be substantially in error; in addition some literature estimates have been obtained from a mixture of rate and equilibrium data, which will be shown below to be unsatisfactory.

We first treat derivatives of adenine and cytosine, and then the more complex behavior of the guanine derivatives.

(a) ADENINE AND CYTOSINE DERIVATIVES: Typical experiments designed to determine equilibrium parameters consist of incubating constant amounts of the nucleoside or nucleotide with varying amounts of formaldehyde (concentration range: 0.01–5 *M*) for a sufficient time to reach equilibrium (3–5 days, at room temperature), followed by measurement of the uv spectra of the resulting mixtures. After subtraction of the small formaldehyde-only spectrum, representative spectra are shown in Figures 2a and 3a for 5′-dAMP and 5′-dCMP, respectively; the corresponding difference spectra, at several formaldehyde concentrations, are shown in Figures 2b and 3b. As noted many times previously, since it comprises one of the major advantages of formaldehyde as a chemical probe, the uv spectral changes on reaction are quite large, generally consisting of a shift of the absorption peak by about 5–10 nm to longer wavelengths, and an increase in peak intensity. As expected (if two products are formed), the spectra show no clear isosbestic wavelength; for example, with 5′-dAMP the crossover point of the difference spectra gradually shifts from about 255 to about 258 nm with increasing formaldehyde concentration.

The usual method of extracting binding constants from such data is to measure the absorbance change, ΔA, usually at the peak or valley of the difference spectra, as a function of free formaldehyde concentration, and then to plot these data according to an equation derived from mass action considerations. For example, when only *one* product is formed, the relevant equation is (see Lewin, 1964a)

$$\Delta A_\lambda/[\mathrm{HCHO}] = KC\Delta\epsilon_\lambda - K\Delta A_\lambda \qquad (1)$$

where ΔA_λ is the observed absorbance change at wavelength λ; [HCHO] is the free formaldehyde concentration ($\simeq$ to the total concentration because HCHO is present in large excess); C is the total nucleotide concentration; K

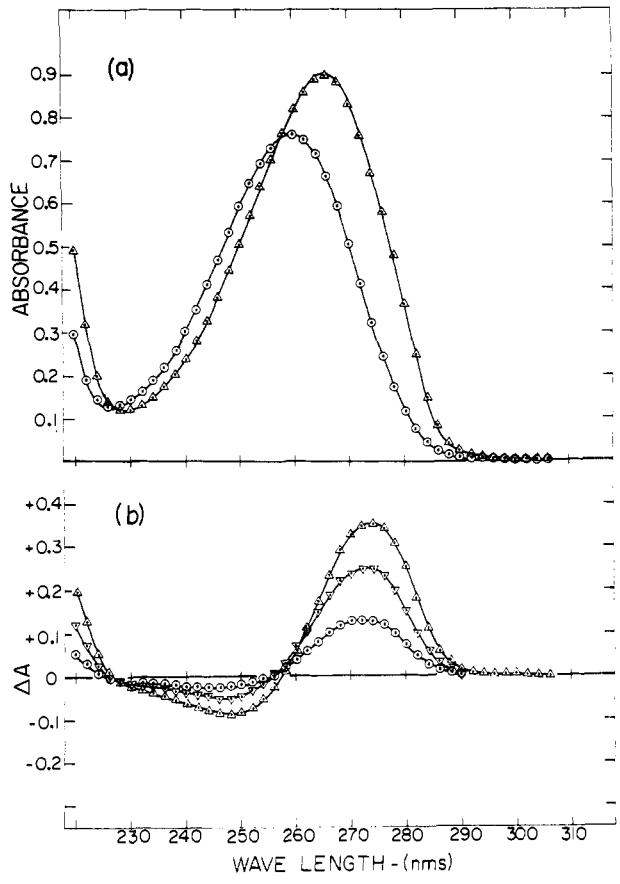

FIGURE 2: (a) Spectrum of 5′-dAMP: (⊙) control; (△) with 1.05 *M* formaldehyde. (b) Difference spectra of 5′-dAMP in the presence of added formaldehyde: (⊙) 0.05 *M*; (▽) 0.21 *M*; (△) 1.05 *M*.

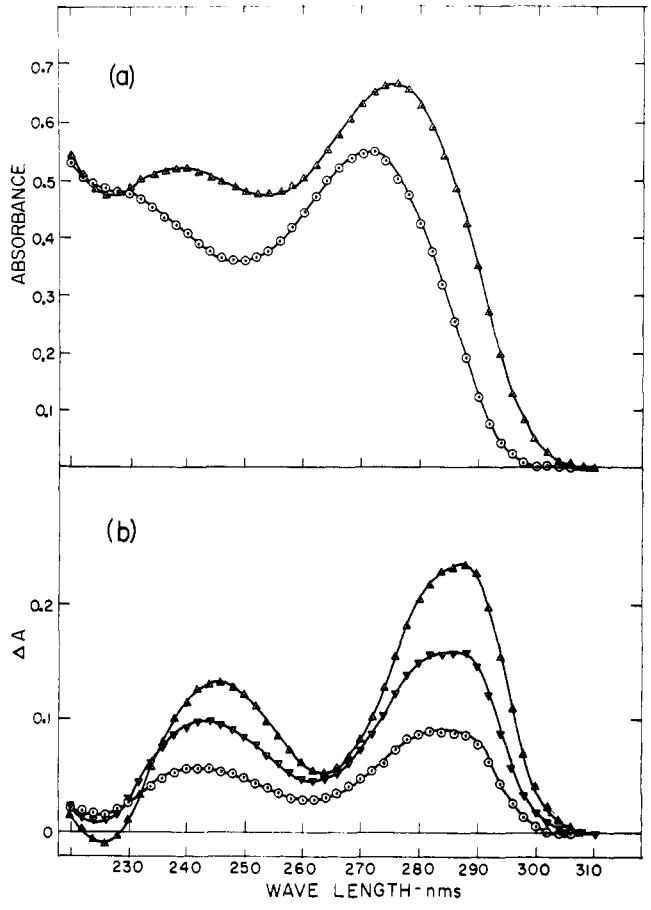

FIGURE 3: (a) Spectrum of 5′-dCMP: (⊙) control; (△) with 1.07 *M* formaldehyde. (b) Difference spectra of 5′-dCMP in the presence of added formaldehyde: (⊙) 0.05 *M*; (▽) 0.21 *M*; (△) 1.07 *M*.

(M^{-1}) is the equilibrium constant for the reaction; and $\Delta\epsilon_\lambda$ is the difference in the molar extinction coefficient between product and starting material at wavelength λ. Hence, *if* only one product is formed, a plot of $(\Delta A_\lambda/[\mathrm{HCHO}]$ vs. $\Delta A_\lambda)$ should be linear with slope = $-K$ and intercept on the Y axis = $KC\Delta\epsilon_\lambda$. Such plots are shown in Figure 4a and b and for dAMP and dCMP, respectively; both plots are seen to be definitely nonlinear, as would be expected if two products are formed. (These nonlinearities are not due to artifacts associated, for example, with high formaldehyde concentrations, since other compounds such as N^6-methyladenosine, where only the monoadduct can form, give linear plots under the same conditions.) Even though such plots show "linear regions," the slope of any one region does not give K_1 or K_2 independently of the other, and eq 1 simply does not apply to the present case (see Klotz and Hunston, 1971, for a discussion of the analogous case of macromolecule-ligand interactions).

Equation 1 above can be extended in a straightforward manner, to include the formation of a second adduct, and gives

$$\frac{\Delta A_\lambda}{[\mathrm{HCHO}]} = \frac{K_1C\Delta\epsilon_1 + K_1K_2C\Delta\epsilon_2[\mathrm{HCHO}]}{1 + K_1[\mathrm{HCHO}] + K_1K_2[\mathrm{HCHO}]^2} \quad (2)$$

where $\Delta\epsilon_1$ and $\Delta\epsilon_2$ represent the difference in extinction coefficient between the monoadduct and diadduct, respectively, and the starting material, at wavelength λ. In order to estimate the four parameters (K_1, K_2, $\Delta\epsilon_1$, and $\Delta\epsilon_2$) we considered the behavior of eq 2 in the limit as the formaldehyde concentration goes either to zero or to infinity; by repeated application of L'Hôpital's rule, the relevant expressions are

$$\lim_{\mathrm{HCHO}\to 0}\{\text{slope}\} = K_2(\Delta\epsilon_2/\Delta\epsilon_1) - K_1$$

$$\lim_{\mathrm{HCHO}\to 0}\begin{Bmatrix}\text{intercept}\\ \text{on } Y \text{ axis}\end{Bmatrix} = K_1C\Delta\epsilon_1$$

and in the other extreme

$$\lim_{\mathrm{HCHO}\to\infty}\{\text{slope}\} = K_2\Delta\epsilon_2/(\Delta\epsilon_1 - \Delta\epsilon_2)\cdot$$

$$\lim_{\mathrm{HCHO}\to\infty}\begin{Bmatrix}\text{intercept}\\ \text{on } X \text{ axis}\end{Bmatrix} = C\Delta\epsilon_2$$

In principle, the four required parameters can be estimated from the limiting behavior of the plots for each experiment; in practice, however, the data at either very low or very high formaldehyde are quite variable, and we use these limits only as initial estimates for a further nonlinear least-squares refinement utilizing all the data (Cleland, 1968; Draper and Smith, 1966). In Table I we collect estimates of these parameters as determined for both deoxynucleotides and deoxynucleosides of adenine and cytosine; the solid lines in Figure 4a and b are drawn using these values.[1]

[1] Due to the nature of the nonlinear curve fitting process we estimate ~±10% uncertainty in each of the parameters, although the actual error estimates emerging from the data analysis are usually substantially lower. As a partial check on the above analysis, $\Delta\epsilon_1$ can be estimated independently, by comparing initial rates with rate constants determined from the total kinetics; the average ratio of $\Delta\epsilon_1$ determined in this way to $\Delta\epsilon_1$ of Table I is 1.01 ± 0.06.

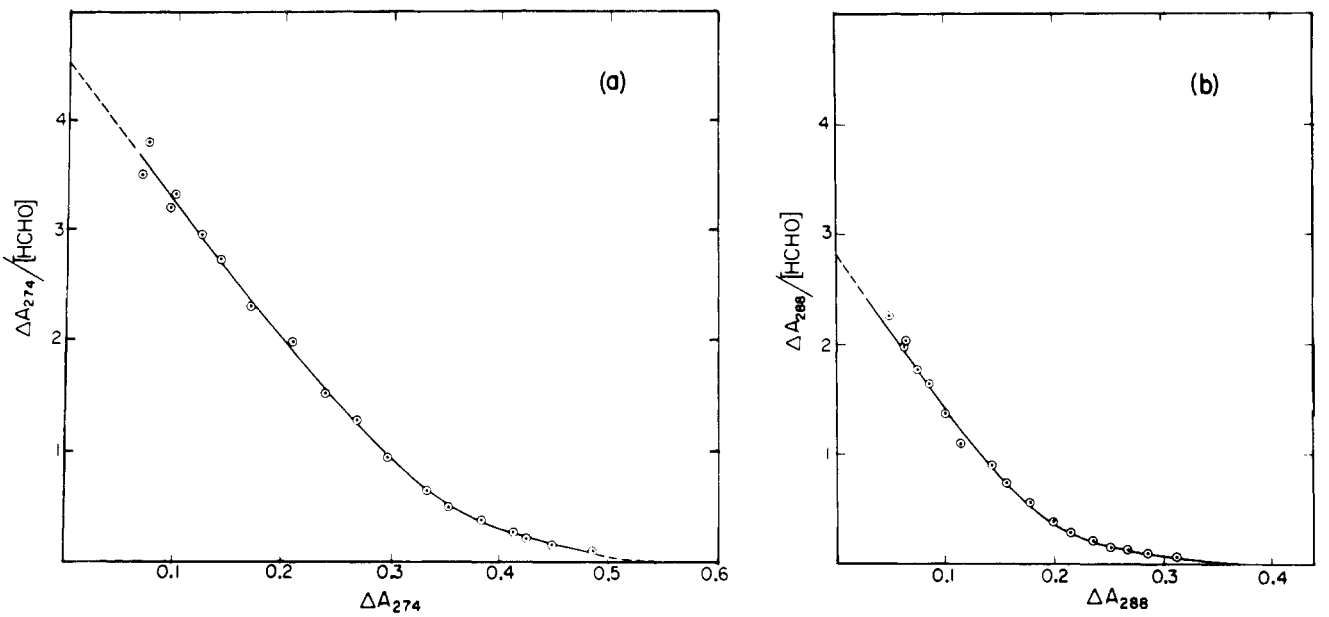

FIGURE 4: (a) Equilibrium absorbance changes (274 nm) of 5′-dAMP with added formaldehyde plotted according to eq 1 of text. (b) Equilibrium absorbance changes (288 nm) of 5′-dCMP with added formaldehyde, plotted according to eq 1 of text.

Table I: Equilibrium Constants and Difference Spectral Parameters of Formaldehyde Reaction with Adenine and Cytosine Compounds.[a,b]

Compd	K_1 (M^{-1})	K_2 (M^{-1})	λ_{max} (nm)	$\Delta\epsilon_1$ ($\times 10^3$)	$\Delta\epsilon_2$ ($\times 10^3$)
5′-dAMP	13.6	0.39	274	6.4	10.6
Deoxyadenosine	12.6	0.37	274	6.5	10.2
5′-dCMP	15.2	0.43	288	3.1	6.1
Deoxycytidine	14.1	0.37	288	3.0	6.1

[a] 0.02 M phosphate, pH 6.95, 23 ± 1°; for error estimates, see text. [b] Although unhydrated formaldehyde is shown as the reacting species in all reaction schemes, for operational convenience the rate and equilibrium constants are expressed with respect to total hydrated formaldehyde, i.e., as assayed by the bisulfite titration.

The data of Table I indicate that there is essentially no difference between deoxynucleotides and deoxynucleosides in their equilibrium behavior toward formaldehyde. Further, there is very little, if any, difference in either equilibrium constant between adenine and cytosine compounds. The equilibrium constant, K_2, for forming the second adduct, is about 35–40-fold lower than that for forming the monoadduct; this is of the same order as the ratio seen between the corresponding equilibrium constants with aliphatic amines (French and Edsall, 1945; Kallen and Jencks, 1966). Because the equilibrium constant, K_2, turns out to be small, our estimates of K_1 for dAMP and dCMP are within about 20% of values reported by Grossman et al. (1961) for the same compounds and by Feldman (1964) for adenosine and 5′-AMP, although both these previous studies did not consider a dimethylol adduct. However, the effect of K_2 is certainly not negligible; for example, in the presence of 1 M formaldehyde, the compounds of Table I would exist as about 25% dimethylol adduct, 70% monomethylol adduct, and 5% unreacted material. The changes in extinction coefficient at 274 nm on forming the monoadduct and diadduct from the adenine derivatives are about 15% lower than the corresponding values of about 7500 and 12,200 seen between the monomethyl and dimethyl derivatives and unsubstituted adenine compounds.

In order to use the binding constants of Table I to interpret quantitatively the depression of the melting temperature of DNA by formaldehyde, we must know how these binding constants depend upon temperature. To this end, reaction mixtures were incubated to equilibrium at various temperatures from 20 to 55°, quickly transferred to a prewarmed cuvet, and absorbance readings taken at the peak of the difference spectra. Because the nonlinear least-squares method used above to extract the four parameters from the binding plots is quite sensitive to the larger experimental errors incurred at the higher temperatures, an alternate analytic procedure was adopted. If $\Delta\epsilon_1$ and $\Delta\epsilon_2$, as given from Table I from the more accurate titrations at room temperature, are taken as temperature independent, eq 2 now has only the two unknowns, K_1 and K_2, and can be linearized as

$$\frac{\Delta A}{(C\Delta\epsilon_2 - \Delta A)}\frac{1}{[HCHO]^2} = K_1\frac{(C\Delta\epsilon_1 - \Delta A)}{(C\Delta\epsilon_2 - \Delta A)}\frac{1}{[HCHO]} + K_1K_2 \quad (3)$$

Thus a plot of $\Delta A/((C\Delta\epsilon_2) - \Delta A)[HCHO]^2)$ vs. $[(C\Delta\epsilon_1 - \Delta A)/(C\Delta\epsilon_2 - \Delta A)](1/[HCHO])$ should be linear, with slope K_1 and intercept K_1K_2. A typical plot is shown in Figure 5, for 5′-dAMP at 35°. K_1 and K_2 were determined by (unweighted) least-squares fits and log K_1 is plotted vs. $1/T$ for dAMP and deoxyadenosine in Figure 6a, and for dCMP and deoxycytidine in Figure 6b. In both cases the equilibrium constant for the major reaction decreases with increasing temperature [in agreement with qualitative observations of Boedtker (1967) on ribonucleotide adducts]. Both plots are linear over the restricted range of temperatures used (i.e., $\Delta H°_1$ is temperature independent). Again no significant difference in behavior is seen between nucleotides and nucleosides, and the data are combined to give the best least-squares estimate of the enthalpy change, $\Delta H°_1$, and estimates of the standard free energy and entropy changes at 25° (Table II). The estimates of the

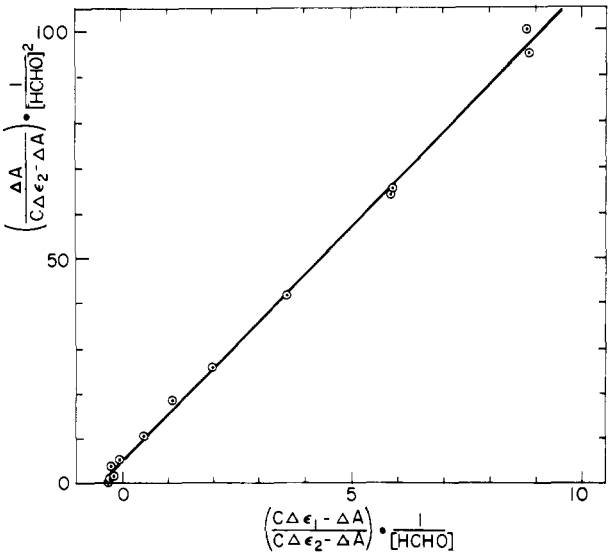

FIGURE 5: Equilibrium absorbance changes (274 nm) of 5′-dAMP with added formaldehyde, at 35°, plotted according to eq 3 of text.

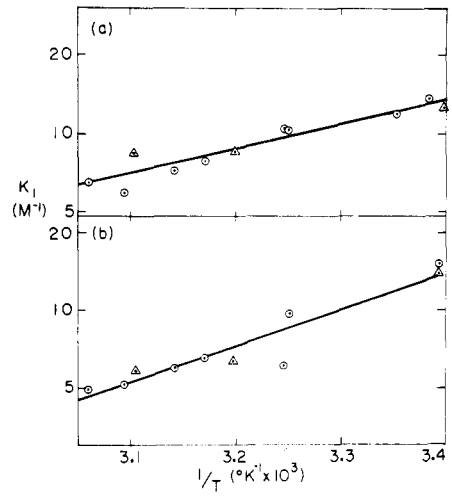

FIGURE 6: van't Hoff plot of the equilibrium constants, K_1, for the formaldehyde reaction: (a) ⊙, 5′-dAMP; △, deoxyadenosine; (b) ⊙, 5′-dCMP; △, deoxycytidine.

Table II: Estimates of Thermodynamic Parameters (25°) for Formaldehyde Reaction (Monoadduct Formation) with Adenine and Cytosine Nucleosides and Nucleotides.[a]

Compd	ΔH°_1 (kcal/mol)	K_1 (M^{-1})	ΔG°_1 (kcal/mol)	ΔS°_1 (cal/(mol deg))
dAMP, deoxyadenosine	−4.3 ± 0.5	12.2	−1.5	−10
dCMP, deoxycytidine	−6.5 ± 0.8	12.0	−1.5	−17

[a] Interpolated, at 25°, from the data of Figure 7.

second binding constants, K_2, are more error-ridden, but values of ΔH°_2 of −3 ± 2 kcal/mol for the adenine derivatives, and −1.5 ± 2 kcal/mol for the cytosine derivatives can be obtained.

(b) GUANINE DERIVATIVES. The reaction of formaldehyde with derivatives of guanine is more complex than with the derivatives of adenine and cytosine discussed above, since there are now three potential sites of reaction; i.e., the two protons of the amino group at C-2, and the one proton of the N_1 nitrogen of the heterocyclic ring. Furthermore, it is to be expected that the parameters characterizing these reactions will all be interdependent. In addition, considerable experimental complications resulted since the present study proved no exception to the common observation that guanine derivatives are experimentally intractable and, for unclear reasons, give variable results (Shapiro, 1968; Guschlbauer, 1972).

The fact that inosine (i.e., guanosine without the amino group) reacts with formaldehyde (Lewin, 1964b; Eyring and Ofengand, 1967; McGhee and von Hippel, 1974) strongly suggests that formaldehyde can also react at the N-1 position of the guanine ring. To distinguish this reaction from the reaction with the neighboring amino group, we can make use of the fact that the inosine reaction is quite fast, and is specific base catalyzed (McGhee and von Hippel, 1975), whereas, as will be shown in the next section, the formaldehyde reaction with the amino group of adenine and cytosine derivatives (and, by analogy, with the amino group of guanine derivatives) is much slower and moreover is independent of pH. From the data of Figure 14c below, the initial rate of absorbance change of 5′-dGMP in formaldehyde solutions is essentially pH independent below about pH 7 (and can be interpreted as the amino reaction dominating the absorbance change), but above pH 7, the rate markedly increases with increasing pH. Thus we conclude that the N-1 group of guanine derivatives can indeed react with formaldehyde and that at high pH this reaction dominates the initial absorbance change; however, this observation says nothing about the ultimate distribution of the reaction products.[2]

The overall absorbance changes were measured at equilibrium as a function of formaldehyde concentration, and representative spectra using 5′-dGMP are shown in Figure 7a, with difference spectra in Figure 7b. The observed absorbance changes at the peak of the difference spectra are plotted as before and are shown for 5′-dGMP in Figure 8. The experimental errors are seen to be far greater than with the adenine and cytosine derivatives; however, within these limits the plot appears to be linear and certainly gives no promise of yielding the three or more potential equilibrium constants. Thus the data are interpreted according to eq 1 above and "apparent equilibrium constants" are listed in Table III. From a comparison of the equilibrium constants with adenine and cytosine derivatives (above) to that of inosine (2.7 M^{-1}; McGhee and von Hippel, 1975), and as suggested by the results with methylated derivatives (below), the amino adduct is expected to be the major product under these conditions. There appears to be no significant difference in behavior between dGMP and deoxyguanosine, but the measured equilibrium constants are about half those for the adenine and cytosine derivatives (Table I). These values are close to the value of 5.3 (M^{-1}) reported by Grossman et al. (1961) for 5′-GMP but do not agree with their value of 45 (M^{-1}) for 5′-dGMP. N^1-Methylguanosine has a slightly lower value for the equilib-

[2] A seemingly more straightforward approach to the problem would be to conduct a "formol titration" as was originally done for inosine (Lewin, 1964b); however this was rendered ambiguous, not only because of compound insolubility, potential gel formation, and the nonspecific medium effects of the high formaldehyde concentrations on the ionization reaction (Kallen and Jencks, 1966), but also by lack of information on how reaction with the amino group would perturb the pK of the imino group being observed.

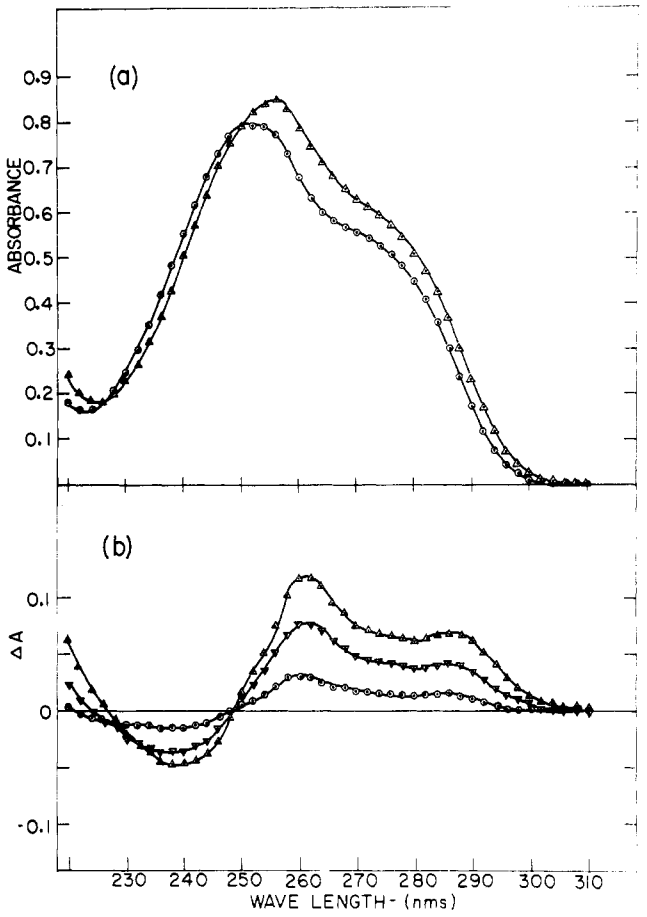

FIGURE 7: (a) Spectrum of 5′-dGMP: (⊙) control; (△) with 1.06 *M* formaldehyde. (b) Difference spectra of 5′-dGMP in the presence of added formaldehyde: (⊙) 0.05 *M*; (▽) 0.21 *M*; (△) 1.06 *M*.

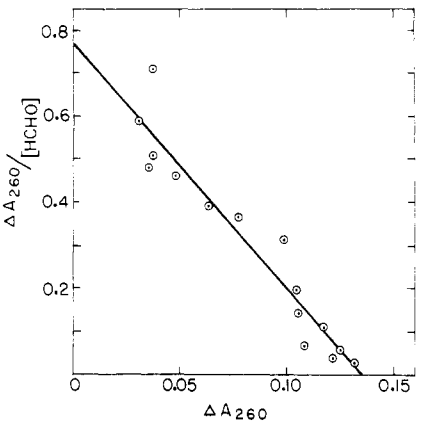

FIGURE 8: Equilibrium absorbance changes (260 nm) of 5′-dGMP with added formaldehyde, plotted according to eq 1 of text.

Table III: Equilibrium Constants and Difference Spectral Parameters for Formaldehyde Reaction with Guanine Derivative.[a]

Compd	K (M^{-1})	λ_{max} (nm)	$\Delta\epsilon$ ($\times 10^3$)
5′-dGMP	5.7 ± 0.5	260	2.3 ± 1
Deoxyguanosine	6.5 ± 0.4	260	2.3 ± 1
N^1-Methylguanosine	4.4 ± 0.3	260	1.1 ± 0.1
N^2-Methylguanosine	1.6 ± 0.2	264	4.2 ± 0.3

[a] 0.02 *M* phosphate, pH 6.95; 23 ± 1°.

rium constant than the unmethylated controls, but nevertheless indicates that the primary amino reaction product is the major adduct. The equilibrium constant with N^2-methylguanosine is considerably lower, and is approximately that expected for the N^1 reaction, in analogy with inosine. N^2,N^2-Dimethylguanosine shows no significant spectral change on incubation with formaldehyde; inspection of CPK models indicated severe steric hindrance to placing a hydroxymethyl group at N^1 in this compound, and thus a triply substituted guanine probably cannot exist.

c. STERIC EFFECTS IN THE MONOADDUCTS. Since our aim in these papers is to calibrate those aspects of the formaldehyde–monomer reactions which might be expected to be important at the polymer level, we now consider an additional, and highly relevant, feature of the above equilibrium studies. It is known that the exocyclic amino group of adenine and cytosine derivatives, like that of anilines and other aromatic amines, is coplanar with the base ring (see, e.g., Voet and Rich, 1970), and thus a single hydroxymethyl group can in principle exist in two isomeric forms; viz., for adenine, syn to N_1 (i.e., anti to N_7 and hence blocking normal Watson–Crick hydrogen bonding positions) or anti to N_1 (i.e., syn to N_7 and, in DNA terminology, sticking out into the large groove); for cytosine derivatives the corresponding isomers would be syn to N_3 (anti to C_5) and anti to N_3 (syn to C_5). In the corresponding methylamino compounds (e.g., N^6-methyladenine and N^4-methylcytosine) it has been shown by a combination of NMR and base-pairing studies in organic solution (Engel and von Hippel, 1974) that these two positions are indeed *not* equivalent but, apparently for steric reasons, the methyl group prefers to reside (by a factor of about 20:1) in the syn position blocking Watson–Crick hydrogen bonding. Since a hydroxymethyl group is expected to have even greater steric requirements than a methyl group, we thus ask the same question of the monoadduct of adenine and cytosine; viz., does a monohydroxymethyl group preferentially occur as the isomer that blocks normal Watson–Crick hydrogen bonding?

A direct test of this hypothesis by NMR was not possible, since the adduct spectra, in solvents such as chloroform, could not be obtained cleanly. Thus we resorted to the less incisive approach of measuring the equilibrium constant for formation of the monoadduct with a variety of adenine and cytosine derivatives, in which methyl groups are placed at different positions around the base ring. The expectation is that any increased steric hindrance introduced by the methyl group would decrease the observed binding constant, and should allow us to define the preferred rotamer position. Table IV collects equilibrium constants and spectral parameters for these various compounds.

The equilibrium constants for the two methylamino compounds, N^6-methyladenosine and N^1,N^4-dimethylcytosine, are reduced about tenfold from the unmethylated controls of Table I. This reduction is somewhat greater than is seen between methylated and unmethylated aliphatic amines (French and Edsall, 1945; Kallen and Jencks, 1966) and greater than the two- to threefold reduction seen between *N*-methylaniline and *N*-methyl-*p*-aminobenzoic acid and the unmethylated controls (McGhee and von Hippel, unpublished). This suggests that there is indeed some degree of steric crowding around the exocyclic amino group; thus to form an adduct in the adenine derivative, either the methyl or the hydroxymethyl group must be forced against the N^7 atom.

Table IV: Equilibrium Constants and Difference Spectral Parameters for Formaldehyde Reaction with Various Methylated Derivatives of Adenine and Cytosine.[a]

Compd	K (M^{-1})	λ_{max} (nm)	$\Delta\epsilon$ ($\times 10^3$)
N^6-Methyladenosine	1.4 ± 0.1	280	4.8 ± 0.1
N^7-Methyladenine	6.1 ± 0.2	277	4.3 ± 0.1
N^9-Methyladenine	14.5 ± 1	276	6.2 ± 0.4
N^1-Methyladenosine	0.27 ± 0.03	276	4.4 ± 0.2
N^1,N^4-Dimethylcytosine	1.6 ± 0.1	294	4.8 ± 0.1
5-Methyldeoxycytidine	7.7 ± 0.5	292	2.1 ± 0.1
N^3-Methylcytidine	0.24 ± 0.02	296	6.2 ± 0.2

[a] 0.02 M phosphate, pH 6.95; 23 ± 1°.

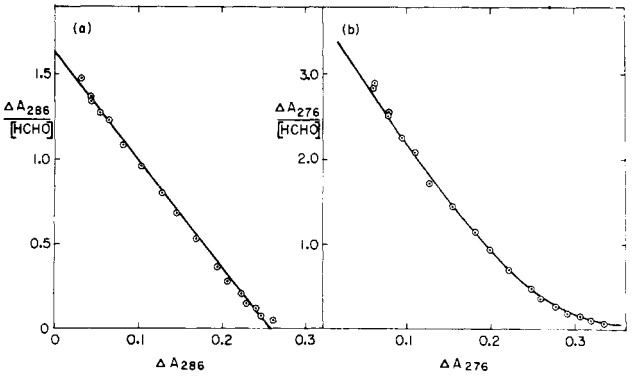

FIGURE 9: Equilibrium absorbance changes with added formaldehyde, plotted according to eq 1 of text: (a) 7-methyladenine (286 nm); (b) 9-methyladenine (276 nm).

The observed equilibrium constant for monoadduct formation in the presence of a methyl group on the N^7 of adenine or C^5 of cytosine is still appreciable. Moreover, the binding plot is now linear, as a second adduct cannot form, as shown in Figure 9a with N^7-methyladenine. (The control of N^9-methyladenine, where two adducts can form, is shown in Figure 9b.) On the other hand, placing the methyl group on the N^1 of adenine or the N^3 of cytosine (i.e., on the side postulated to be normally occupied by the adduct) is seen to lower the observed binding constant about 50-fold.

The assumption in this approach is obviously that changes in the measured binding constants reflect *only* changes in steric hindrance, and are not caused by differences in solvation, etc. This assumption is to some extent supported by observations with aliphatic amines (Kallen and Jencks, 1966) and with aromatic amines in which, as will be shown below, the formaldehyde equilibrium constants remain essentially unchanged as the amine basicity changes by a factor of 10^8. Nevertheless, the twofold reduction in equilibrium constant between N^7-methyladenine and 5-methyldeoxycytidine and the unmethylated controls is more than would be expected if the hydroxymethyl adduct existed *solely* as the isomer blocking Watson–Crick hydrogen bonding. Furthermore, it should be pointed out that both N^1-methyladenosine and N^3-methylcytidine are positively charged at the pH of the experiment. Because of these complications, we can only say that the results of Table IV seem to be compatible with the hydroxymethyl group residing preferentially on that side of the amino group that blocks Watson–Crick hydrogen bonding (and perhaps by the 20:1 factor expected from the methyl group measurements). Also the results in Table III on the methylated guanine derivatives are compatible with the hydroxymethyl group, like the methyl analog (Engel and von Hippel, 1974), preferentially existing in the position that does *not* obstruct the normal Watson–Crick hydrogen bonding.

(*3*) *Kinetic Constants.* In this section we measure the forward and reverse rate constants for forming the monohydroxymethyl adduct with the various nucleic acid components, and further estimate how this kinetic behavior is affected by known perturbants of DNA structure and stability.

At the formaldehyde concentrations at which the reaction kinetics are conveniently measurable (and at which most DNA experiments are done), i.e., from 0.1 to 0.5 M, it can be calculated from the equilibrium constants of Table I that, at equilibrium, 2–15% of adenine or cytosine nucleosides should exist as the diadducts, with 55–75% as the monoadduct; this puts limits on the accuracy to be expected from the following rate estimates. Nevertheless, since the reaction with the amino group to form the monoadduct is both stoichiometrically and spectrally the major reaction, the kinetics should approximate "pseudo-first-order" behavior (i.e., formaldehyde is in vast excess and present at essentially constant concentration throughout the reaction). When the reaction is followed by absorbance changes, the equation describing the kinetics may be written (Stevens and Rosenfeld, 1966)

$$\ln \left\{ \frac{A_\infty - A_t}{A_\infty - A_0} \right\} = -k't = -\{k_{12}[\mathrm{HCHO}] + k_{21}\}t \quad (4)$$

where A_0 and A_∞ are the initial and final absorbance and A_t is the absorbance at time t (all absorbances being measured at some particular wavelength, usually at the peak or valley of the difference spectrum); k_{12} is the (second order) forward rate constant for monoadduct formation; k_{21} is the (first order) reverse rate constant for adduct decomposition; and $k' = k_{12}[\mathrm{HCHO}] + k_{21}$, is the pseudo-first-order rate constant.

Typical kinetic data are plotted according to eq 4 in Figure 10. The data are seen to fit the relation quite well, with the dGMP kinetics being found to deviate slightly from the proposed behavior at later time, when supposedly the slower (at this pH) N-1 reaction is occurring. It was further checked that with dCMP, the rate constants were essentially constant (±5%) over a 100-fold range of nucleotide concentrations. The slopes of the plots of Figure 10 (as determined by weighted least-squares fits) give the pseudo-first-order rate constants, k', which in turn, when plotted as a function of formaldehyde concentration as in Figure 11, give straight lines with slope, k_{12}, and intercept, k_{21}. Table V collects rate constants determined by this method for the deoxynucleosides and deoxynucleotides, both in the standard buffer and in standard buffer with 1 M NaCl. Standard errors in the rate constants are estimated from the least-squares analysis of the plots of Figure 11 (Draper and Smith, 1966) but other effects (such as diadduct formation) may be expected to make larger contributions; a more realistic estimate of the uncertainty would be about ±10%, and in view of later discussion, perhaps even ±20%.

All the reactions are seen to be relatively slow; for example at a formaldehyde concentration of 1 M and at 25°, the half-times for reaction are about 10, 30, and 60 min for dCMP, dGMP, and dAMP, respectively. There appears to be no significant difference in rate constants between the deoxynucleotides and deoxynucleosides of adenine or cytos-

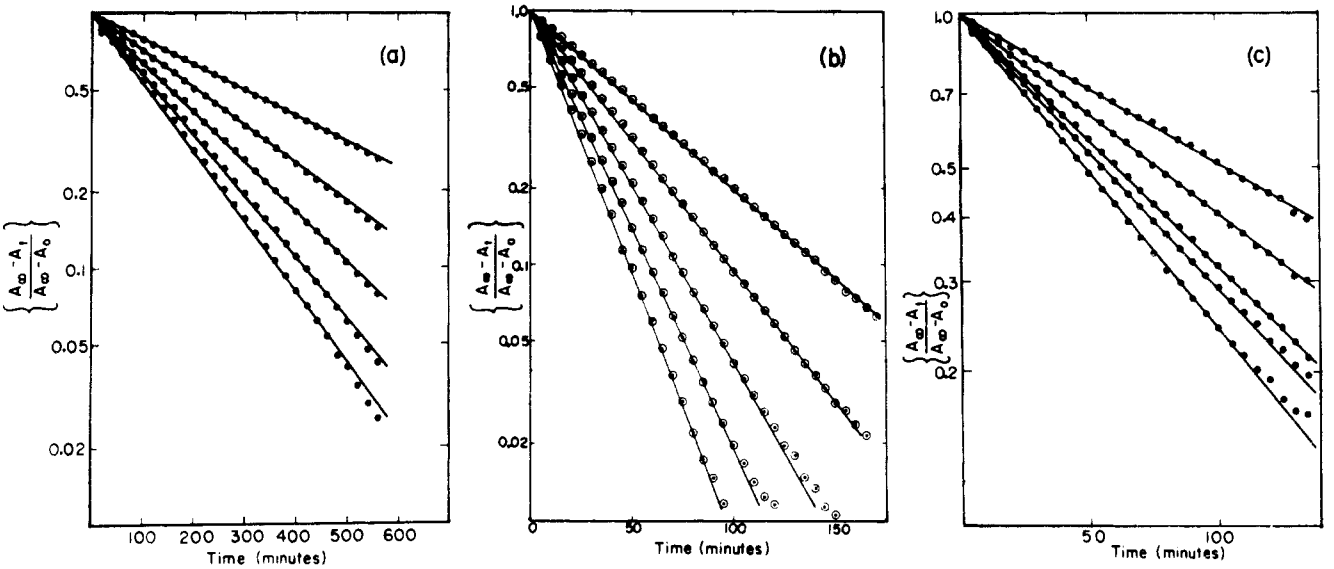

FIGURE 10: Pseudo-first-order plots of reaction of deoxynucleotides with 0.1, 0.2, 0.3, 0.4, and 0.5 *M* formaldehyde: (a) 5′-dAMP; (b) 5′-dCMP; (c) 5′-dGMP.

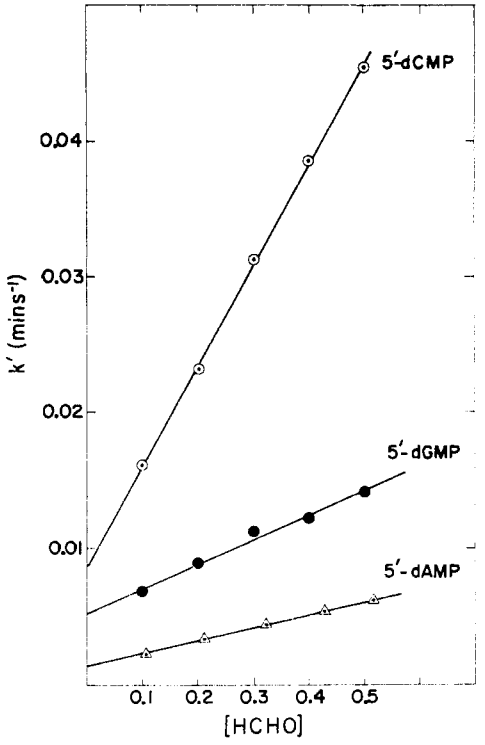

FIGURE 11: Plot of pseudo-first-order rate constant, *k′*, vs. formaldehyde concentration: (⊙) 5′-dCMP; (●) 5′-dGMP; (△) 5′-dAMP.

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

Compd	Forward, k_{12} (M^{-1} sec^{-1}) × 10^4	Reverse, k_{21} (sec^{-1}) × 10^4
Deoxyadenosine	1.52 ± 0.03	0.20 ± 0.01
Deoxyadenosine + 1 *M* NaCl	1.91 ± 0.05	0.27 ± 0.02
5′-dAMP	1.56 ± 0.03	0.23 ± 0.01
5′-dAMP + 1 *M* NaCl	1.85 ± 0.03	0.23 ± 0.03
Deoxycytidine	12.7 ± 0.2	1.6 ± 0.2
Deoxycytidine + 1 *M* NaCl	12.2 ± 0.2	1.5 ± 0.1
5′-dCMP	12.3 ± 0.2	1.4 ± 0.1
5′-dCMP + 1 *M* NaCl	12.6 ± 0.2	1.5 ± 0.1
Deoxyguanosine	1.15 ± 0.02	0.31 ± 0.01
Deoxyguanosine + 1 *M* NaCl	0.54 ± 0.04	0.19 ± 0.01
5′-dGMP	3.0 ± 0.2	0.9 ± 0.1
5′-dGMP + 1 *M* NaCl	2.9 ± 0.2	0.8 ± 0.1

[a] 0.02 *M* phosphate, pH 6.95; 25 ± 0.5°.

ine. The rate constants for the cytosine derivatives are essentially unchanged by 1 *M* salt, whereas those for adenine derivatives are marginally increased (about 20%). Deoxyguanosine appears to react about two- to threefold more slowly than 5′-dGMP, and its rate constants are roughly halved by the addition of salt; however, interpretation of these two phenomena is complicated by the fact that it is just at this pH that the N-1 imino reaction is beginning to become important. Thus, just as with the equilibrium constants of Table IV above, we are forced to interpret the rate data for guanine derivatives as "apparent rate constants."

The rate constants reported above can be used to calculate the pseudo-first-order rate constants at the one formaldehyde concentration used by Haselkorn and Doty (1961), and the results agree within about 30%. The rate constants reported by Grossman et al. (1961) are actually pseudo-first-order rate constants and thus apply only to the one formaldehyde concentration used; when their results are recalculated (Penniston and Doty, 1963), agreement with the values reported above is within a factor of about 2 and generally much closer. At 30°, our values for the forward rate constant with dAMP and deoxyadenosine are about 10% lower than that reported for adenosine (Stevens and Rosenfeld, 1966).

The second reaction step, for formation of the diadduct from the monoadduct, can be easily seen at high formaldehyde concentrations, especially at wavelengths close to the apparent isosbestic. By preequilibrating dAMP in relatively concentrated solutions of formaldehyde, and then diluting into lower formaldehyde concentrations, k_{23}, the appropriate rate constant, was estimated to be about 0.6 ± 0.3 × 10^{-4} (M^{-1} sec^{-1}), or about two- to threefold slower than the first reaction step.

In order to determine the effect of a variety of experimental variables on the reaction kinetics, the forward rate constant, k_{12}, was determined from the initial rates of

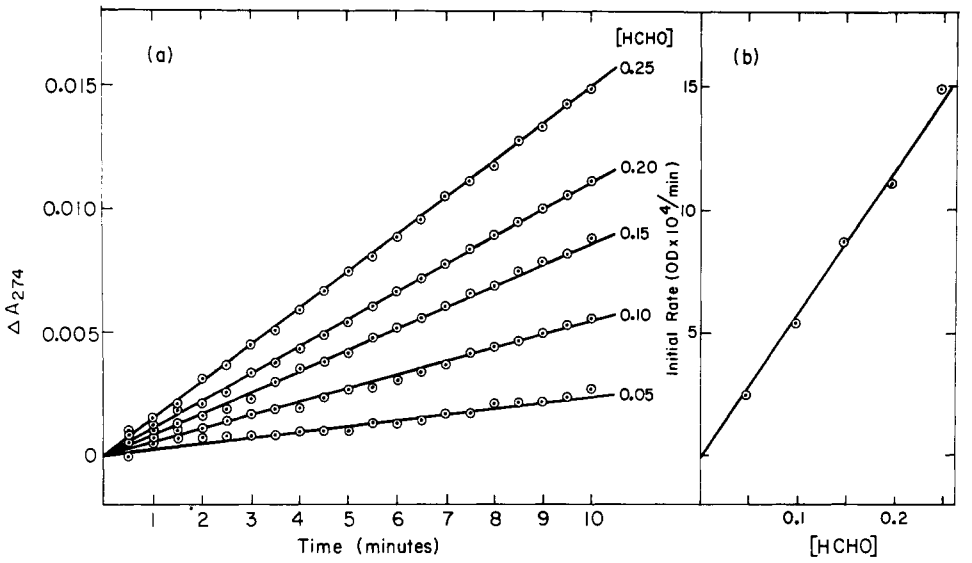

FIGURE 12: (a) Initial change in absorbance (274 nm) in the reaction of 5′-dAMP with various formaldehyde concentrations; (b) plot of initial rates in (a) vs. formaldehyde concentration.

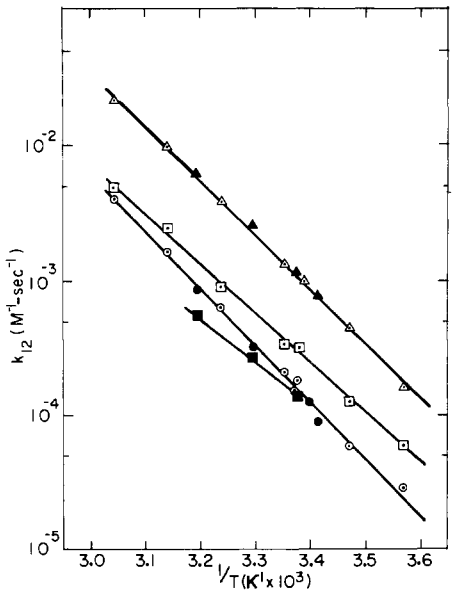

FIGURE 13: Arrhenius plots of forward rate constant (k_{12}) for formaldehyde reaction with deoxynucleosides and deoxynucleotides: (△) 5′-dCMP; (▲) deoxycytidine; (□) 5′-dGMP; (■) deoxyguanosine; (○) 5′-dAMP; (●) deoxyadenosine.

change of absorbance, rather than from the complete (and lengthy) time course. The expression for the absorbance change, ΔA, at wavelength λ as a function of time after formaldehyde addition, is

$$\Delta A_\lambda = k_{12}[\text{HCHO}]C\Delta\epsilon t \qquad (5)$$

where C is the total nucleotide concentration, and $\Delta\epsilon$ is the change in extinction coefficient on forming the monoadduct. Figure 12, for 5-dAMP with 0.05–0.25 M formaldehyde, shows that, at early times, the absorbance change certainly can be considered linear with time. The slopes of these lines ($k_{12}C\Delta\epsilon$[HCHO]) are then plotted vs. [HCHO] in Figure 12b; knowing C and $\Delta\epsilon$ from Table I, k_{12} can be determined. (Alternatively, as implied above, $k_{12}\Delta\epsilon$ can be determined from the slope of Figure 12b, k_{12} from the complete kinetics under the same conditions, as from Table IV, and thus $\Delta\epsilon$ determined independently and as a check on the nonlinear least-squares procedure used above to fit the equilibrium curves.)

Table VI: Activation Enthalpies for Forward Reaction with Formaldehyde.

Compd	$\Delta H^\ddagger$ (kcal/mol)	Compd	$\Delta H^\ddagger$ (kcal/mol)
5′-dAMP	19.3 ± 0.4	Deoxycytidine	18.8 ± 0.5
Deoxyadenosine	19.8 ± 0.9	5′-dGMP	16.9 ± 0.4
5′-dCMP	18.6 ± 0.3	Deoxyguanosine	14.7 ± 0.6

To compare with the extreme temperature dependence observed for the formaldehyde reaction with DNA (von Hippel and Wong, 1971), forward rate constants were measured at temperatures from 5 to 65°, and the resulting Arrhenius plots shown in Figure 13. Over the range of temperatures studied, the plots are linear and thus $\Delta H^\ddagger$ can be taken as constant; Table VI collects these activation enthalpies and estimated standard errors. There is little significant difference between the estimates for the cytosine and adenine derivatives; estimates for dGMP and deoxyguanosine are somewhat lower (again the apparent rate constant with dGMP is approximately double that for deoxyguanosine, throughout the temperature range studied). The activation enthalpies of Table VI are about 10% higher than those reported by Haselkorn and Doty, who erroneously used pseudo-first-order rate constants. Nevertheless, this agreement provides independent corroboration of the small values for the equilibrium enthalpy change estimated above since a plot of log (k') *vs.* $1/T$ can *only* be linear if the equilibrium enthalpy change is zero. Using Tables VI and II, the activation enthalpy for the reverse reaction can be estimated as about 15 kcal/mol for adenine derivatives, and about 12 kcal/mol for cytosine derivatives.

To compare with the specific base catalysis in the formaldehyde reaction with 5′-TMP (McGhee and von Hippel, 1975), the forward rate constants for dAMP, dCMP, and dGMP were measured at pH values ranging from about 4 to 9, i.e., keeping within the pH range of maximal DNA stability. (The buffers used, at either 0.02 or 0.1 M total concentration, were acetate, cacodylate, phosphate, and

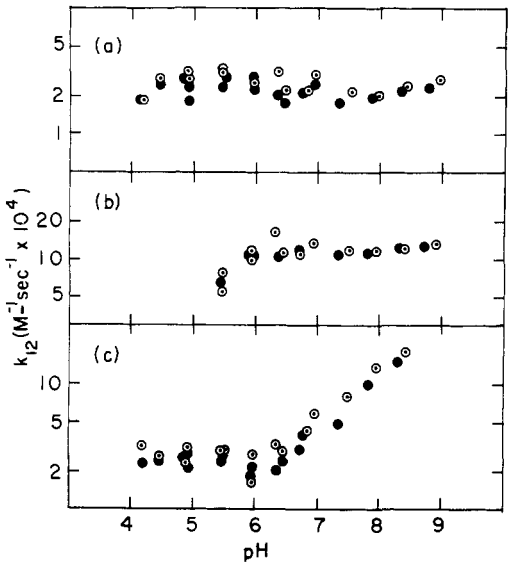

FIGURE 14: Plot of forward rate constant (k_{12}) vs. pH: (a) 5′-dAMP; (b) 5′-dCMP; (c) 5′-dGMP; (⊙) 0.1 *M* buffer; (●) 0.02 *M* buffer; ionic strength constant at 1.0 *M*.

methyl arsenate; ionic strength was kept constant at 1.0 by addition of NaCl; pH was measured directly in the reaction solutions.) The results are plotted in Figure 14 as log k_{12} vs. pH. Within the rather large errors arising from the initial rate measurements, it is seen that: (i) for dAMP, k_{12} is essentially constant from pH 4 to 9; (ii) for dCMP, k_{12} is constant from about pH 6 up to 9 (below about pH 5.5 base protonation occurs and the initial rates can no longer be used); (iii) with dGMP, k_{12} remains constant from pH 4 up to about 6.5–7, thereafter increasingly sharply, presumably due to the base-catalyzed N-1 reaction now becoming faster than the uncatalyzed amino reaction. (Rough checks indicate that, for both dAMP and dCMP, the forward rate is accelerated about tenfold at about pH 11.) The average ratio of the rate constants determined in 0.02 *M* total buffer to that determined at 0.1 *M* total buffer was calculated as: 0.87 ± 0.09 for dAMP (all pH's); 0.98 ± 0.15 for dCMP (pH's above 6); and 0.85 ± 0.19 for dGMP (pH's below 7). Thus there is some indication of general acid-base catalysis. This was investigated in more detail only with a series of cacodylate buffers, ranging in total concentrations from 0.1 to 0.5 *M* and with 5′-dAMP; the results indicate that there is indeed general acid-base catalysis, that the rate constant for buffer catalysis is approximately equal to that for solvent catalysis, and that the rate constant for catalysis by undissociated cacodylic acid is approximately fourfold greater than that for the conjugate base. Although these observations open up a new dimension for probing DNA behavior, we have not, at the moment, extended them; for the usual buffer concentrations used in DNA experiments, this effect can be neglected.

In order to be able to differentiate between the effect of DNA base pair heterogeneity on thermal stability and the effect on the chemical reaction rates, we have looked at several solvent systems that perturb the relative stabilities of A·T and G·C base pairs (Melchior and von Hippel, 1973).

Table VII collects the forward rate constants measured with the deoxynucleotides in 2.4 *M* tetraethylammonium chloride, 3.0 *M* tetramethylammonium chloride, and 4 *M* sodium perchlorate. In spite of the rather gross changes in solvent represented by these additives, the observed rate constants change remarkably little, decreasing about 10–30% in the first two solvents and increasing about two- to threefold in 4 *M* sodium perchlorate. Several other solvent additives were studied, the only (minor) feature of interest being that 20% dioxane, which has little or no effect on the reaction rate with guanine or cytosine derivatives, roughly halves the reaction rate with adenine derivatives (dAMP, deoxyadenosine, and 9-methyladenine).[3]

(*4*) *Aromatic Amines as Model Compounds.* To gain insight into the mechanism and, in particular, the stereochemistry of the reaction, and to attempt to delineate those properties of polynucleotides that influence the reaction

[3] Before leaving this section, we should mention a discrepancy in the present results and one which has also gone unnoticed in the literature. In principle, the ratio of the forward to the reverse rate constant should equal the equilibrium constant, K_1, for monoadduct formation, obtained by measuring the overall absorbance changes at equilibrium. However, the measured ratio k_{12}/k_{21} is always considerably lower; *i.e.*, comparison of this ratio from Table V with the appropriate entries from Tables I to IV indicates that $k_{12}/k_{21} = (0.57 \pm 0.04)K_1$. To discover the source of this discrepancy we considered the following. (i) Perhaps as much as 5–10% (i.e., up to 25% of the overall discrepancy) could be attributed to technical errors, such as differences in reaction temperature, etc. (ii) Perhaps a further 10–20% of the discrepancy could be assigned to errors in the rate constant analyses, such as neglect of the second adduct (which as seen above can be appreciable); nevertheless this cannot be the total explanation since compounds which only form monoadducts, such as N^6-methyladenosine, show the same discrepancy. On the other hand, with these singly reacting compounds, the discrepancy is less and with the eight compounds treated in the following paper, $k_{12}/k_{21} = (0.80 \pm 0.16)K$. (iii) The nonlinear least-squares analysis used to determine the equilibrium constants is unlikely to be at fault since the same discrepancy is seen with monoreacting compounds analyzed by the much simpler eq 1 above. (iv) Kinetic constants were usually determined with formaldehyde concentrations of 0.1–0.5 *M*, whereas the equilibrium determinations used concentrations ranging up to 5 *M*, where various nonidealities such as polymerization reactions, lowered water activity, etc., might be expected to become more severe. However, not only have Kallen and Jencks (1966) shown that most of these effects appear to cancel, but using smaller estimates of formaldehyde concentrations in the calculations (i.e., concentrations more closely approximating actual activities) would lead to an increase in the estimate of the equilibrium constant, where a decrease is needed to explain the discrepancy. (v) The rate of methylene glycol dehydration to form the reactive monomer is expected to become rate limiting only when reactions are several orders of magnitude faster than the present ones and moreover only when the nucleotide concentrations far exceed the optical concentrations used here (Walker, 1964). (vi) Methanol is present in the commercial solutions used, at a molar ratio of about one methanol to three formaldehyde monomers, and thus could potentially cause the apparently overslow forward reaction rate. However, with 5′-dCMP as a test case, doubling this ratio up to about 2 mol of methanol per formaldehyde had no significant effect (±2%) on the forward rate (although, not surprisingly, massively increasing the methanol up to about 20% dropped the reaction rate about tenfold). (vii) Formaldehyde depolymerization reactions, following dilution into the reaction mixtures, seemed a likely candidate to explain the discrepancy, as these reactions are known to be quite slow (Walker, 1964). Nevertheless, and again using dCMP as a test case, the estimated forward rate constants were independent of whether concentrated dCMP was added to dilute formaldehyde, or concentrated formaldehyde added to dilute dCMP, and moreover were not significantly different (±2%) if the formaldehyde was used immediately after dilution from a concentrated stock or if it was diluted and then allowed to sit at room temperature for about 2 weeks prior to use.

We are thus left pleading an accumulation of errors as the rather unsatisfactory explanation of the discrepancy; at the moment, we feel this is somewhat to be preferred to postulating a second step in the forward reaction. We have not pursued this question further, both because our aim is mainly to "calibrate" the chemical reaction (and, as seen, even the discrepancy is quite reproducible); and, moreover, considering the very complicated nature of formaldehyde-methanol-formaldehyde polymers solutions, we are pleased to find that the discrepancy is no larger.

Table VII: Forward Rate Constant of Formaldehyde Reaction with Nucleotides in the Presence of Various Solvents which Perturb DNA Stability.[a]

Solvent	Forward Rate Constant k_{12} (M^{-1} sec^{-1} × 10^4) 5′-dAMP	5′-dCMP	5′-dGMP
2.4 *M* Tetraethylammonium chloride	1.1 ± 0.2	8.0 ± 0.3	
3.0 *M* Tetramethylammonium chloride	1.4 ± 0.1	9.3 ± 0.3	
4 *M* Sodium perchlorate	5.0 ± 0.1	22.7 ± 0.9	7.5 ± 0.7

[a] Also with 0.02 *M* phosphate, 25 ± 1°.

rate, we investigated a series of aromatic amines with pK's for protonation ranging from pH +4.6 to (H_0 =) −4.2, as determined by either aqueous or nonaqueous titrations.[4]

The compounds used were chiefly substituted anilines, but included some diphenylamines and α-naphthylamine. Preliminary equilibrium experiments indicated that the spectral changes associated with the formaldehyde reaction were usually quite large, and wavelengths for observation were chosen as the peak or valley of the difference spectrum; formaldehyde concentrations were kept as low as possible to minimize diadduct formation, and all experiments were done in the standard 0.02 *M* phosphate buffer (pH 6.95). Reactions were followed with either the Cary 14 or in a stopped-flow apparatus. Typical data from the two time domains are shown in Figure 15a and 15b for 2,4-dinitroaniline (pK = −4.2) and 2-bromoaniline (pK = +2.5), respectively; both reactions are seen to be quite good pseudo-first-order. Pseudo-first-order rate constants are plotted as a function of formaldehyde concentration in the inserts to Figure 15 and rate constants determined as above. Table VIII collects the compounds studied, their pK values from the literature, the wavelength of observation, and the estimated forward and reverse rate constants. The logarithm of the forward rate constant is plotted against pK in Figure 16a; the plot is linear over a million-fold change in rate constant, with correlation coefficient, r = 0.997. The least-squares equation describing the relation is

$$\log_{10} k_{12}(M^{-1}\,\text{sec}^{-1}) = \text{p}K(0.73 \pm 0.01) - (1.64 \pm 0.04)$$

Apparently the formaldehyde dehydration reaction begins to become rate limiting only for the fast reactions with aniline, and here the observed rates were extrapolated to zero amine concentrations (i.e., so as still to maintain pseudo-first-order conditions). The linear relation appears to hold for all ortho-substituted compounds, for diphenylamines, and for α-naphthylamine, although the three isomeric aminobenzoic acids, in which the carboxyl group is ionized at the experimental pH, fall quite far above the main line and were not used to fit the equation. Furthermore, there is no gross irregularity in the kinetic behavior of either anthranilic acid or the ester, methyl anthranilate, where there is the potential of forming an intramolecular hydrogen bond.

As shown in Figure 16b, there is also a reasonably good, but less striking correlation (r = 0.97) between the logarithm of the reverse rate constant and pK. Hence, the equilibrium constant, as estimated by the ratio of the two rate constants, is essentially unchanged over about eight orders

[4] A similar study has been made by Drs. W. R. Abrams (Abrams, 1972) and R. Kallen (personal communication).

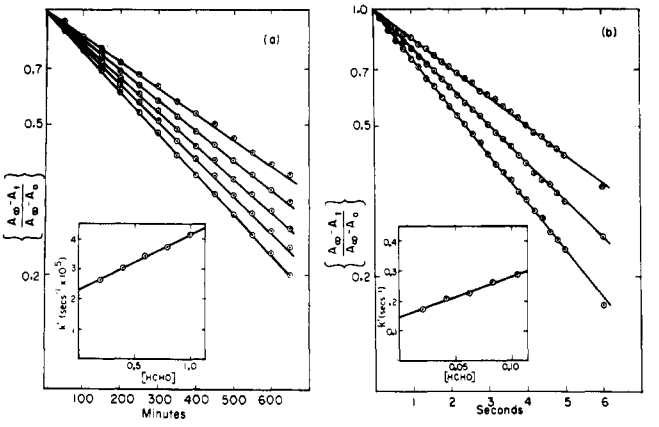

FIGURE 15: Pseudo-first-order plots for reaction with various amounts of formaldehyde: (a) 2,4-dinitroaniline; (b) 2-bromoaniline. Inserts are plots of pseudo-first-order rate constants, k', vs. formaldehyde concentration.

of magnitude change in amine basicity, ranging between about 0.5 and 30 (M^{-1}), with no obvious trend.

We also looked for a correlation between rate constants and the pK for deprotonating the neutral amines. In this case, several nucleoside analogs were available whose pK's had been determined by nonaqueous titration (Harns and Stewart, 1975); for completeness, the rate constants for the various compounds are collected in Table IX, although little correlation was apparent, either for the forward (r = 0.42) or the reverse rate constant (r = 0.31).

General Discussion

(1) Mechanism of Reaction. An understanding of the reaction mechanism and its steric requirements is necessary to interpret the observed effect of double-helical polynucleotide structure on the formaldehyde kinetics. The mechanism of formaldehyde addition to the (planar) aromatic amines of the present study can be described by analogy with the accepted mechanism of simple carbonyl group additions to aliphatic amines (Jencks, 1964). The initial step is the nucleophilic attack of the nitrogen lone pair of electrons (present in an orbital perpendicular to the base ring) on the partially positive carbon atom of (unhydrated) formaldehyde, with the nitrogen approaching perpendicular to the plane of the carbonyl group; the final step is then a proton transfer (Jencks, 1964). The very good correlation observed in Figure 16 between the logarithm of the rate constants and amine pK_a (as a measure of amine nucleophilicity) is consistent with either the initial formation of the C–N bond, or the subsequent proton transfer step, being rate limiting. This linear free energy relation suggests a close similarity between the structure of the protonated amine (i.e., having a tetrahedral nitrogen bearing a positive charge) and either the transition state of the reaction (if the C–N bond formation is rate limiting) or the zwitterionic intermediate (if the proton transfer is rate limiting) [Leffler and Grunwald, 1963; Jencks, 1969; see Sayer et al., 1974, for a discussion of rate-limiting proton transfer steps]. The overall reaction can be diagrammed as:

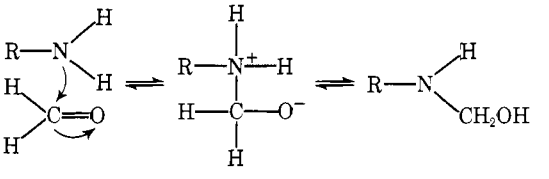

Table VIII: Collected Rate Constants for Formaldehyde Reaction with Aromatic Amines.[a]

No.	Compd	pK_a	λ (nm)	Forward, k_{12} (sec^{-1} M^{-1})	Reverse, k_{21} (sec^{-1})
1	2,4-Dinitroaniline	−4.3[b]	210	$1.86 \pm 0.06 \times 10^{-5}$	$2.26 \pm 0.04 \times 10^{-5}$
2	4-Nitrodiphenylamine	−3.1[b]	430	$1.3 \pm 0.5 \times 10^{-4}$	$2.3 \pm 0.3 \times 10^{-4}$
3	2,5-Dichloro-4-nitroaniline	−1.8[b]	410	$9.9 \pm 0.9 \times 10^{-4}$	$3.4 \pm 0.3 \times 10^{-4}$
4	4-Chloro-2-nitroaniline	−1.1[b]	240	$3.7 \pm 0.2 \times 10^{-3}$	$1.56 \pm 0.07 \times 10^{-3}$
5	3-Chlorodiphenylamine	−0.5[b]	240	$1.04 \pm 0.04 \times 10^{-2}$	$9.9 \pm 0.1 \times 10^{-3}$
6	2-Nitroaniline	−0.3[b]	415	$9.3 \pm 0.3 \times 10^{-3}$	$3.0 \pm 0.1 \times 10^{-3}$
7	Diphenylamine	0.8[c]	285	$9.3 \pm 0.4 \times 10^{-2}$	$3.5 \pm 0.2 \times 10^{-2}$
8	4-Nitroaniline	1.0[c]	360	$1.6 \pm 0.2 \times 10^{-1}$	$7.8 \pm 1.0 \times 10^{-3}$
9	2,5-Dichloroaniline	1.5[b]	248	$3.3 \pm 0.1 \times 10^{-1}$	$2.6 \pm 0.1 \times 10^{-3}$
10	2,4-Dichloroaniline	2.05[c]	310	0.73 ± 0.08	$6.8 \pm 0.5 \times 10^{-2}$
11	Anthranilic acid	2.09[d]	330	20 ± 2	0.8 ± 0.1
12	Methyl anthranilate	2.36[d]	340	0.6 ± 0.07	0.13 ± 0.01
13	*p*-Aminobenzoic acid	2.41[d]	286	12 ± 1	0.48 ± 0.05
14	Methyl *p*-aminobenzoate	2.45[d]	298	2.5 ± 0.2	$8.2 \pm 0.01 \times 10^{-2}$
15	2-Bromoaniline	2.53[c]	298	1.4 ± 0.1	0.145 ± 0.005
16	2-Chloroaniline	2.65[c]	298	1.6 ± 0.1	0.19 ± 0.005
17	*m*-Aminobenzoic acid	3.07[d]	318	18 ± 1	0.3 ± 0.1
18	3-Chloroaniline	3.46[c]	298	9.1 ± 0.8	0.3 ± 0.1
19	Methyl *m*-aminobenzoate	3.58[d]	334	17.2 ± 0.3	0.49 ± 0.02
20	α-Naphthylamine	3.92[c]	334	8.6 ± 0.7	1.05 ± 0.05
21	4-Chloroaniline	4.15[c]	304	25.3 ± 1.9	0.4 ± 0.1
22	Aniline	4.63[c]	292	46	0.59

[a] 0.02 *M* phosphate, pH 6.95; 25.0 ± 0.5°. [b] Stewart and Dolman, 1967. [c] Handbook of Chemistry and Physics, 1970. [d] Christensen *et al.*, 1970.

Table IX: Rate Constants for Formaldehyde Reaction with Substituted Adenines.

Compd	pK_a[a]	Rate Constant: Forward k_{12} (M^{-1} sec^{-1})	Rate Constant: Reverse k_{21} (sec^{-1})
2-Chloro-7-methyladenine	14.0	3.2×10^{-5}	6.1×10^{-6}
7-Methyladenine	15.5	3.9×10^{-4}	8.7×10^{-5}
2-Chloro-9-methyladenine	15.8	7.0×10^{-6}	3.0×10^{-6}
9-Methyladenine	17.2	2.1×10^{-4}	3.0×10^{-5}

[a] From Harns and Stewart (1975).

As remarked earlier, the model compound data also indicate the remarkable insensitivity of the equilibrium constants to amine basicity, a result also observed in other formaldehyde reactions (Kallen and Jencks, 1966).

(*2*) *Implications for Polymer Studies.* Three practical implications of the above results apply to the way in which formaldehyde is generally used in nucleic acid chemistry. (i) From the rate constants of Table V, it can be seen that the rates for decomposition of the methylol adduct are very slow; for example, at 0° the half-time of decomposition of the adenine adduct is about 60 hr. Thus the usual procedure of removing reversibly bound formaldehyde from nucleic acids by dialyzing 1–2 days in the cold is seen to be quite inadequate to remove all the bound methylol groups, let alone any methylene cross-links which might have formed. (ii) The procedure usually used to bring the (quite slow) nucleic acid–formaldehyde reactions to completion is to heat at 100° for several minutes, followed by cooling back to the initial temperature. While this does indeed increase the reaction rates, it also changes the *extent* of reaction at the high temperature, since the equilibrium constants are temperature dependent. For example, at a formaldehyde concentration where adenine compounds are 50% reacted at room temperature, they are only about 20% reacted at 100°; this may or may not have practical consequences, depending on the experimental context. (iii) Caution should be used in applying the quantitative results of this paper to systems other than pure nucleic acids. In more complicated systems such as ribosomes and chromatin, completely new types of reactions can occur and, indeed, might even be expected to dominate. For example, it has long been known that formaldehyde leads to histone–DNA cross-links (Brutlag et al., 1969), and more recently, Siomin et al. (1973) have reported a considerable acceleration in reaction rate of formaldehyde with nucleotides in the presence of amino acids and histones. We have observed such rate accelerations in the presence of the simpler amine, morpholine. This apparent rate increase is probably not due to general acid-base catalysis (see above), nor to nucleophilic catalysis (Jencks, 1969), but rather to the formation of a completely new species. This compound presumably is the nucleotide, cross-linked to the amine through a methylene bridge; i.e., a new spot appears on chromatograms if dAMP or dCMP is incubated in the presence of *both* formaldehyde and morpholine (McGhee and von Hippel, unpublished).

Further implications of the present results for polymer studies will be dealt with more fully in later papers; however, some of the implications can be summarized now so that the validity of any extrapolations can be assessed.

(i) Within the range of error in determination of the equilibrium constants and enthalpy changes (and coupled with the results of the following paper on the imino reac-

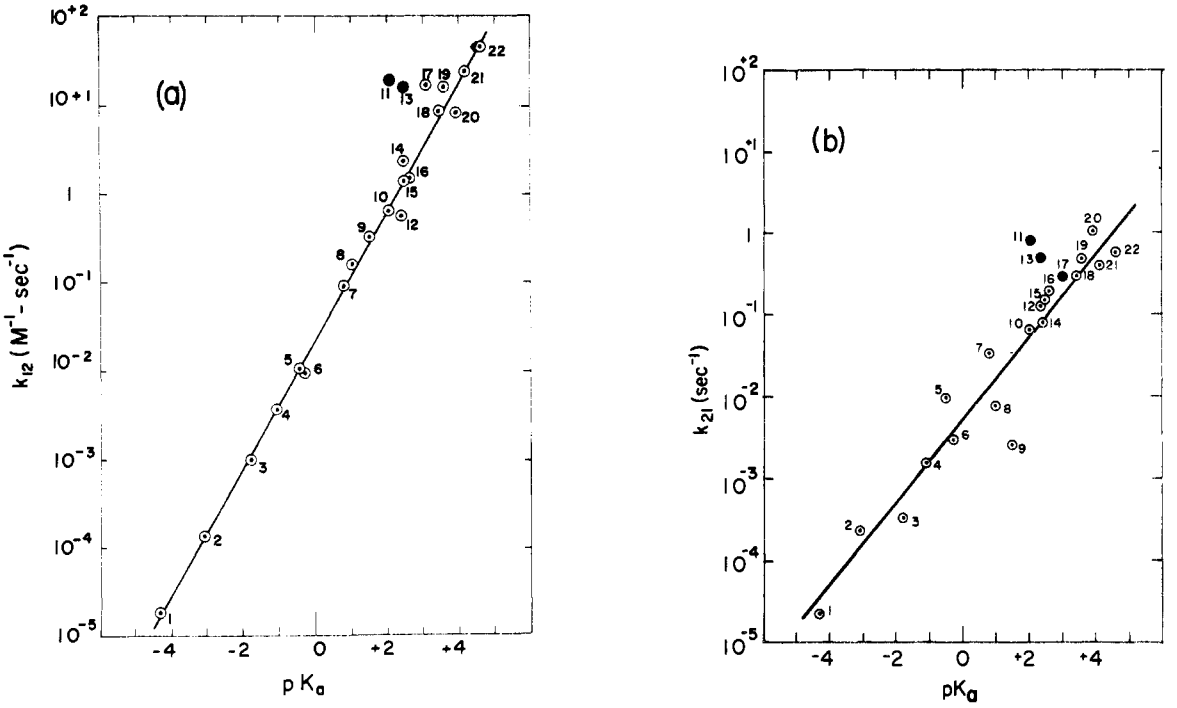

FIGURE 16: Plots of the logarithms of the (a) forward, and (b) reverse rate constants for formaldehyde reaction *vs.* pK_a of the aromatic amines of Table VIII (filled symbols are carboxylic acids, see text).

tion), the equilibrium denaturation of DNA by formaldehyde can now be described quantitatively for the first time, and (in principle) can now be used as a tool to elucidate unknown structures in polynucleotides.

(ii) The various rate constants and their temperature dependencies form crucial controls for the use of formaldehyde as a chemical probe (Lazurkin et al., 1970; Utiyama and Doty, 1971; Dean and Lebowitz, 1971; von Hippel and Wong, 1971). Furthermore, the proposed reaction mechanism (whereby the formaldehyde molecule must approach the amino group from either above or below the plane of the base) is central to our understanding of the character of "reactive states" of a native DNA molecule. For example, these results suggest that bases would have to unstack transiently to allow reaction (McGhee and von Hippel, in preparation).

(iii) The demonstration that the reaction adduct is indeed a methylol group, as had long been assumed, and that furthermore a dimethylol adduct can form, raises the distinct possibility that formaldehyde can react with a native DNA double helix without necessarily disrupting a hydrogen bond, i.e., by extending into the large groove (if a monomethylol adduct of adenine or cytosine) or into the small groove (if a monomethylol adduct of guanine). Nevertheless, by analogy with the corresponding methylamino compounds (Engel and von Hippel, 1974) and in keeping with the above results on model compounds, steric hindrance within the base structure is expected to exist such that the methylol group of adenine or cytosine "prefers" to exist in that isomeric position which necessarily disrupts the normal Watson-Crick hydrogen bonding. For guanine adducts, on the other hand, the same argument suggests that the methylol group would preferentially extend into the small groove, although whether such a group can indeed exist in a native DNA molecule will be determined by the dictates of the polymer structure, e.g., "groove geometry" (see Engel and von Hippel, 1974).

(iv) As noted earlier, the apparently "normal" kinetic behavior of the methyl anthranilate compound in Figure 16 suggests that the presumed (see Eigen, 1964) involvement of the amino group in an (intramolecular) hydrogen bond (as opposed to its "unavailability" within a structure not accessible to water) does not by itself drastically slow the reaction rate. Thus hydrogen bonding of the bases in a polymer structure cannot be the sole reason for the great reduction of the rate of reaction with formaldehyde. This lack of detectable effect in the model compounds is probably a reflection of the fact that the amino groups of the control compounds are always hydrogen bonded to water molecules anyway.

(v) If the linear free energy relations observed in Figures 16a and b between the reaction rates and basicity of the aromatic amines can be taken to apply even approximately to the heterocyclic nucleic acid bases, then the observed kinetics of the latter can be used to estimate their respective pK's. These pK values are critically needed to interpret polynucleotide hydrogen exchange studies (McConnell and von Hippel, 1970; Teitlebaum and Englander, 1974) and are unobtainable by titration, due to prior protonation of the heterocyclic ring. Thus using the forward rate constants listed in Table V, and using the least-squares relation illustrated in Figure 16a, the pK's for protonating the amino groups of dAMP, dCMP, and dGMP can be estimated as −3.0, −1.7, and −2.6, respectively. From the somewhat less satisfactory relation observed with the reverse rate constants, the corresponding estimates are −4.6, −3.1, and −3.5. It would require a much more thorough analysis to decide how reliable these pK values actually are; nevertheless they do lie in the range predicted by McConnell and Seawell (1972) from the pH dependence of NMR exchange rates of the amino protons. As a further extension, these rate-pK relations indicate that if the monomer amino pK is decreased by one unit by incorporation into a polymer structure, the forward reaction rate with formaldehyde can

be expected to decrease at least fivefold.

Acknowledgments

The authors should like to thank Dr. J. Engel for help with the nmr experiments, and for many helpful discussions. We should also like to thank Ms. Pamela O'Conner for help with the calculations and some of the initial rate studies.

References

Abrams, W. R. (1972), *Chem. Abstr. 76,* 139826m.
Alberts, B. M., and Frey, L. (1970), *Nature (London) 227,* 1313.
Axelrod, V. D., Feldman, M. Ya., Chuguev, I. I., and Bayev, A. A. (1969), *Biochim. Biophys. Acta 186,* 33.
Belman, S. (1963), *Anal. Chim. Acta 29,* 120.
Berns, K. I., and Thomas, C. A., Jr. (1961), *J. Mol. Biol. 3,* 289.
Boedtker, H. (1967), *Biochemistry 6,* 2718.
Brown, D. M. (1974), in Basic Principles of Nucleic Acid Chemistry, Vol. II, Ts'o, P. O. P., Ed., New York, N.Y., Academic Press.
Brutlag, D., Schlehuber, C., and Bonner, J. (1969), *Biochemistry 8,* 3214.
Christensen, J. J., Rytting, J. H., and Izatt, R. M. (1970), *Biochemistry 9,* 4907.
Clark, R. J., and Felsenfeld, G. (1971), *Nature (London), New Biol. 229,* 101.
Cleland, W. W. (1968), *Adv. Enzymol. Relat. Areas Mol. Biol. 29* .
Cox, R. A. (1969), *Biochem. J. 114,* 743.
Dean, W. W., and Lebowitz, J. (1971), *Nature (London), New Biol. 231,* 5.
Dunlop, P., Marini, M. A., Fales, H. M., Sokoloski, E., and Martin, C. J. (1973), *Bioorg. Chem. 2,* 235.
Draper, N. R., and Smith, H. (1966), Applied Regression Analysis, New York, N.Y., Wiley.
Eigen, M. (1964), *Angew. Chem., Int. Ed. Engl. 3,* 1.
Engel, J. D., and von Hippel, P. H. (1974), *Biochemistry 13,* 4143.
Eyring, E. J., and Ofengand, J. (1967), *Biochemistry 6,* 2500.
Fasman, G. D., Lindblow, C., and Grossman, L. (1964), *Biochemistry 3,* 1015.
Feldman, M. Ya. (1964), *Biochemistry (USSR) 29,* 616.
Feldman, M. Ya. (1973), *Prog. Nucleic Acid Res. Mol. Biol. 13,* 1.
Fraenkel-Conrat, H. (1954), *Biochim. Biophys. Acta 15,* 307.
Freifelder, D., and Davison, P. F. (1963), *Biophys. J. 3,* 49.
French, D., and Edsall, J. T. (1945), *Adv. Protein Chem. 2,* 277.
Grossman, L., Levine, S. S., and Allison, W. S. (1961), *J. Mol. Biol. 3,* 47.
Guschlbauer, W. (1972), in The Purines, Bergmann and Pullman, Eds., New York, N.Y., Academic Press.
Handbook of Chemistry and Physics (1970), 50th ed, Cleveland, Ohio, Chemical Rubber Publishing Co.
Harns, M., and Stewart, R. (1975), *Chem. Commun.* (in press).
Haselkorn, R., and Doty, P. (1961), *J. Biol. Chem. 236,* 2738.
Haslam, J., and Squirrel, D. C. M. (1957), *Analyst 82,* 511.
Inman, R. B. (1966), *J. Mol. Biol. 18,* 464.
Inners, D., and Inners, R. R. (1974), *Biophys. Soc. Annu. Meet. Abstr. No. 1131.*
Jencks, W. P. (1964), *Prog. Phys. Org. Chem. 2,* 63.
Jencks, W. P. (1969), Catalysis in Chemistry and Enzymology, New York, N.Y., McGraw-Hill.
Kallen, R. G., and Jencks, W. P. (1966), *J. Biol. Chem. 241,* 5864.
Klotz, J. M., and Hunston, D. L. (1971), *Biochemistry 10,* 3065.
Lazurkin, Y. S., Frank-Kamenetskii, M. D., and Trifonov, E. N. (1970), *Biopolymers 9,* 1253.
Leffler, J. E., and Grunwald, E. (1963), Rates and Equilibria of Organic Reactions, New York, N.Y., Wiley.
Lewin, S. (1964a), *J. Chem. Soc. B,* 792.
Lewin, S. (1964b), *Experentia 20,* 666.
Lewin, S. (1966), *Arch. Biochem. Biophys. 113,* 584.
Lewin, S., and Barnes, M. A. (1966), *J. Chem. Soc. B,* 478.
Li, H. J. (1972), *Biopolymers 11,* 835.
Lodish, H. F. (1971), *J. Mol. Biol. 56,* 627.
McConnell, B., and von Hippel, P. H. (1970), *J. Mol. Biol. 50,* 297, 317.
McGhee, J. D., and von Hippel, P. H. (1975), following paper.
Means, G. E., and Feeney, R. E. (1968), *Biochemistry 7,* 2192.
Melchior, W. B., Jr., and von Hippel, P. H. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 298.
Paik, W. K., and Kim, S. (1972), *Biochemistry 11,* 2589.
Penniston, J. T., and Doty, P. (1963), *Biopolymers 1,* 145, 209.
Rosenfeld, A., Stevens, C. L., and Printz, M. P. (1970), *Biochemistry 9,* 4971.
Sayer, J. M., Pinsky, B., Schonbrunn, A., and Washtien, W. (1974), *J. Am. Chem. Soc. 96,* 7998.
Scheit, K. H. (1965), *Tetrahedron Lett. 15,* 1031.
Shapiro, R. (1968), *Prog. Nucleic Acid Res. Mol. Biol. 8,* 73.
Siomin, Yu. A., Siminov, V. V., and Poverenny, A. M. (1973), *Biochim. Biophys. Acta 331,* 27.
Sober, H. A., Ed. (1968), Handbook of Biochemistry, Cleveland, Ohio, Chemical Rubber Co.
Stevens, C. L., and Rosenfeld, A. (1966), *Biochemistry 5,* 2714.
Stewart, R., and Dolman, D. (1967), *Can. J. Chem. 45,* 925.
Szymanski, H. A., and Yelin, R. E. (1968), NMR Band Handbook, New York, N.Y., Plenum Publishing Co.
Tal, M. (1970), *Biochim. Biophys. Acta 224,* 470.
Teitlebaum, H., and Englander, S. W. (1974), *J. Mol. Biol.* (in press).
Tikchonenko, T. I., and Dobrov, E. N. (1969), *J. Mol. Biol. 42,* 119.
Trifonov, E. N., Lazurkin, Yu. S., and Frank-Kamenetskii, M. D. (1967), *Mol. Biol. (USSR) 1,* 142.
Ts'o, P. O. P. (1974), in Basic Principles in Nucleic Acid Chemistry, Vol. I, Ts'o, P. O. P., Ed., New York, N.Y., Academic Press.
Utiyama, H., and Doty, P. (1971), *Biochemistry 10,* 1254.
Voet, D., and Rich, A. (1970), *Prog. Nucleic Acid Res. Mol. Biol. 10,* 183.
von Hippel, P. H., and McGhee, J. D. (1972), *Annu. Rev. Biochem. 41,* 231.
von Hippel, P. H., and Wong, K. Y. (1971), *J. Mol. Biol. 61,* 587.
Walker, J. F. (1964), Formaldehyde, 3rd ed, New York, N.Y., Reinhold.