*THE EFFECT OF SUBSTITUENTS ON THE HYDROGEN BONDING OF ADENINE AND URACIL DERIVATIVES**

By Yoshimasa Kyogoku, R. C. Lord, and Alexander Rich

SPECTROSCOPY LABORATORY AND DEPARTMENTS OF CHEMISTRY AND BIOLOGY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated by Norman Davidson, December 27, 1966

It has proved possible to study the specificity of hydrogen bonding between the purine and pyrimidine constituents of nucleic acids by preparing derivatives which allow their interaction to be observed in a nonaqueous solvent. The nonaqueous solvent avoids the stacking interactions of the purines and pyrimidines and since the derivatives are uncharged, interionic forces are absent. In addition the interactions are not modified by the stereochemical constraints which are present when nucleotides are arranged in a polynucleotide chain. Thus these studies allow us to isolate the hydrogen bonding interactions from other factors which are important in polynucleotides. The initial infrared observations[1-3] showed that uracil and adenine derivatives form hydrogen bonds with each other preferentially as compared to the extent to which they hydrogen-bond with themselves. These observations have been extended to guanine and cytosine derivatives with both infrared and nuclear magnetic resonance techniques.[4-7] The latter studies show that the guanine and cytosine residues likewise form hydrogen-bonded dimers with each other much more readily than with themselves. Significantly, the affinity of adenine for uracil (or thymine) derivatives and of guanine for cytosine derivatives is highly selective; guanine or cytosine derivatives do not preferentially associate with adenine or uracil derivatives. Thus the hydrogen-bonding interactions of these molecules show a type of electronic complementarity which matches exactly the geometric complementarity that plays such an important role in our understanding of the structure of the double-stranded helical nucleic acid molecules.

In the present paper we report the results of a quantitative measurement of the association constants of seven uracil derivatives and six adenine derivatives. It is found that some substituents enhance the hydrogen-bonding affinity of the adenine-uracil association while others markedly decrease it. Some of the derivatives belong to the minor bases of transfer RNA; thus these studies allow us to measure their hydrogen-bonding affinities. Other derivatives allow us to choose between the two possible sites on adenine which can be involved in hydrogen bonding. In addition it is shown that those derivatives which enhance the association of adenine and uracil with each other nonetheless do not interact with guanine or cytosine derivatives.

Materials and Methods.—As in the previous study,[1, 2] derivatives of adenine with an ethyl group on N_9 and of uracil with cyclohexyl on N_1 were used. Seven types of uracil derivatives were studied: uracil, thymine, 3-methyluracil, 5,6-dihydrouracil, 5-bromouracil, 5-iodouracil, and 4-thiouracil. Six adenine analogues were studied: 6-aminopurine (adenine), 6-methylaminopurine, 6-dimethylaminopurine, 2-aminopurine; 2,6-diaminopurine, and 6-amino-8-bromopurine. These were purchased from the Cyclo Chemical Co., Los Angeles. Frequently they will be referred to without mentioning the 1-cyclohexyl or 9-ethyl substituents which are used to make them soluble in chloroform. All derivatives were soluble to more than 0.1 mole/liter except 2,6-diaminopurine and 8-bromoadenine, whose solubilities were found to be 0.005 and 0.01 M, respectively. Deutero-

chloroform (New England Nuclear Co., Boston) was used as solvent and was purified by passages through an alumina gel column 10-cm long. Infrared spectra were taken with a Perkin-Elmer model 521 double-beam spectrophotometer. Fused silica cells (American Instruments Co., and Beckman Instruments) ranging from 1 mm to 10 cm were used for the measurements above 3000 cm^{-1}. Six to ten spectra were recorded at different concentrations for the determination of each association constant. From the concentration dependence of the intensity of the monomer bands, association constants and absorption coefficients were calculated.

The theory for the determination of association constants has been described in detail previously.[2, 8] For the self-association of a purine or pyrimidine derivative, the association constant K_{self} is defined by $K_{self} = C_d/C_m^2$, where C_m and C_d are the monomer and dimer concentrations in moles per liter. If the dimer is cyclic, absorbance A of the monomer band at a given frequency is related with K_{self} by the equation $A = (2K_{self})^{-1}[a_m^2\ell(C_0/A) - a_m\ell]$ where a_m is the absorption coefficient of the monomer band, ℓ is the path length in cm, and C_0 is the total concentration of the solute, which is equal to $C_m + 2C_d$ in dilute solutions. When A is plotted against C_0/A, a straight line should be obtained. K_{self} and a_m can be computed from the slope s and the intercept i of the line: $K_{self} = s/2i^2$; $a_m = -s/\ell i$. This procedure cannot be applied to the self-dimerization of 2,6-diaminopurine because a free amino group still remains in the cyclic dimer; in this case, when C_0 is very small compared with $1/K_{self}$, the expressions for K_{self} and a_m must be divided by 2. When equal amounts of two kinds of derivatives X and Y are mixed, the solution contains dimers XX, YY, and XY and monomers X and Y. If X and Y form a cyclic dimer and the monomer concentrations of X and Y are almost equal, the absorbance A of the monomer band of X is expressed by $A = (K_{XY} + 2K_{XX})^{-1}[a_m^2\ell(C_0/A)/2 - a_m\ell]$. The above assumptions are satisfied in the cases where K_{XY} is much larger than K_{XX} and K_{YY}, or K_{XX} is almost equal to K_{YY}. The K_{XY} and a_m are given by $K_{XY} = 2(s/i^2 - K_{XX})$ and $a_m = -2s/\ell i$. K_{XX} is the same as K_{self} of X, which can then be determined independently of K_{XY}. For the association of 2,6-diaminopurine with uracil derivatives, K_{XY} can also be derived as above but a_m in the last equation must be divided by 2. The methods described above have the advantage of not assuming any value of a_m. However, where K_{XY} is not much bigger than K_{XX} and K_{YY}, or where K_{XX} is much different from K_{YY}, the relation between A and K_{XY} is more complex. Even in this case, however, it is possible to get K_{XY} when a_m can be obtained from other experiments. It is estimated that the experimental errors of measurement yield relative association constants which are accurate to $\pm$ 10%.

Results.—Characteristic infrared absorption spectra are shown in Figure 1*A*, where dotted curve (*a*) shows the absorption spectrum of deuterochloroform in the region 3600–3200 cm^{-1}, and curve (*b*) shows the absorption spectrum of 9-ethyl-2,6-diaminopurine. Two strong bands are observed at 3526 and 3417 cm^{-1} which are associated with the antisymmetric and symmetric stretching vibrations of the nonbonded amino group. Although this molecule has two such groups, the frequencies of the two are so close that they cannot be observed separately. Figure 1*B* shows the strong band at 3385 cm^{-1} due to the free imino group of 1-cyclohexyl-5-bromouracil. Curves (*c*) show the absorption bands when they are replotted with the absorption spectrum of the solvent as the baseline. Figure 1*C* shows the absorption spectrum of an equimolar solution of 2,6-diaminopurine and 5-bromouracil. Strong association bands appear at 3492, 3350, and 3200 cm^{-1} (curve *c*). Curve (*d*) represents the calculated spectrum for a noninteracting mixture of the purine and pyrimidine derivatives. The deviations in the observed spectrum arise from hydrogen bonding between the bases. The association bands have a maximum intensity at a 1:1 mole ratio.

Curves similar to that seen in Figure 1*C*(*c*) are plotted in Figure 2 for 1:1 mixtures of several uracil derivatives with 9-ethyladenine and with 9-ethyl-2,6-diaminopurine. The uppermost curve in Figure 2*A* represents the spectrum of 9-ethyladenine plus 1-cyclohexyl-3-methyluracil. This curve matches the super-

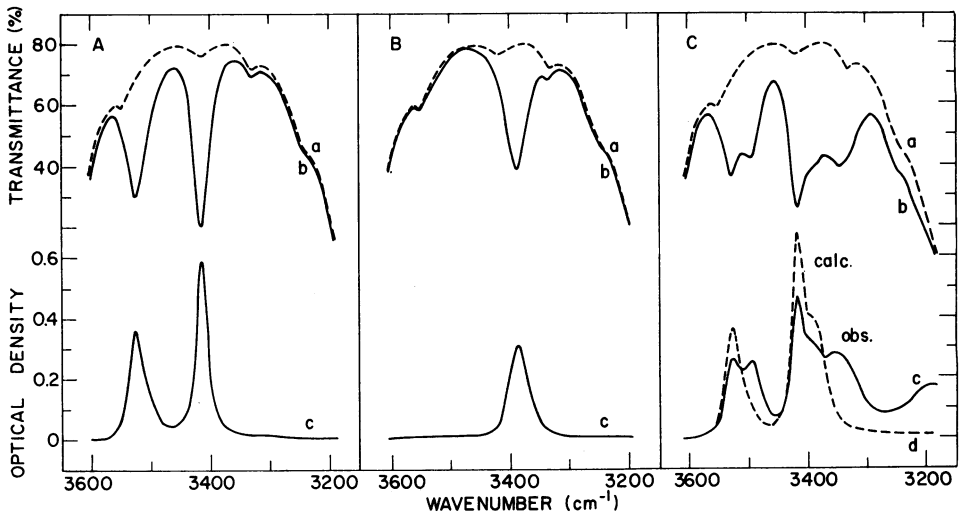

FIG. 1.—Infrared absorption spectra of (*A*) 2,6-diamino-9-ethylpurine (0.002 *M*), (*B*) 1-cyclohexyl-5-bromouracil (0.002 *M*), and (*C*) their 1:1 mixture (total concentration 0.004 *M*) in deuterochloroform solution in a 10-mm quartz cell.

(*a*) Absorption spectra of the pure solvent; (*b*) absorption spectra of the solution; (*c*) absorption spectra of the solution plotted as optical density with the solvent curve as a base line; (*d*) calculated sum of the optical density curves in *A*(*c*) and *B*(*c*).

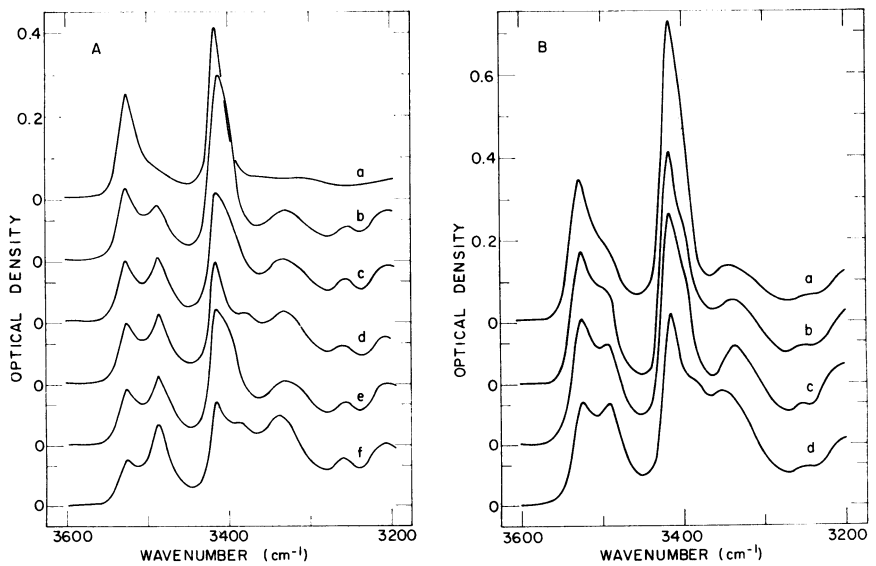

FIG. 2.—Infrared absorption spectra of the various 1:1 mixtures of some substituted 1-cyclohexyluracil derivatives and (*A*) 9-ethyladenine or (*B*) 9-ethyl-2,6 diaminopurine plotted as optical density with the solvent absorption as a base line. (*A*) Total concentration is 0.04 *M* and path length is 1.0 mm. All mixtures contain 9-ethyladenine plus (*a*) 3-methyluracil, (*b*) 5,6-dihydrouracil, (*c*) uracil, (*d*) 4-thiouracil, (*e*) thymine, (*f*) 5-bromouracil. (*B*) All mixtures contain 9-ethyl-2,6-diaminopurine plus (*a*) 5,6-dihydrouracil, (*b*) uracil, (*c*) thymine, (*d*) 5-bromouracil. Total concentration 0.004 *M* in 10.0-mm cell.

position of the two individual spectra and indicates that no association is occurring between the two molecules. This is not unexpected in view of the fact that the hydrogen of the imino nitrogen of uracil has been replaced by methyl and it cannot form a hydrogen-bonded cyclic dimer. The additional curves in Figure 2*A* show varying degrees of association as indicated by the increasing intensity of the bands at 3490 and 3350 in going from (*b*) to (*f*). Analogous spectra are shown in Figure 2*B* for mixtures of 9-ethyl-2,6-diaminopurine plus various uracil derivatives. Again varying degrees of association are shown. These results illustrate qualitatively the effect of various substituents on the degree of association. This phenomenon has been examined quantitatively and association constants determined for 9-ethyladenine with all of the uracil derivatives, as well as for 1-cyclohexyluracil with all of the adenine derivatives. The constants both for self-association and for the mixtures are shown in Figure 3 together with the structural formulae. The

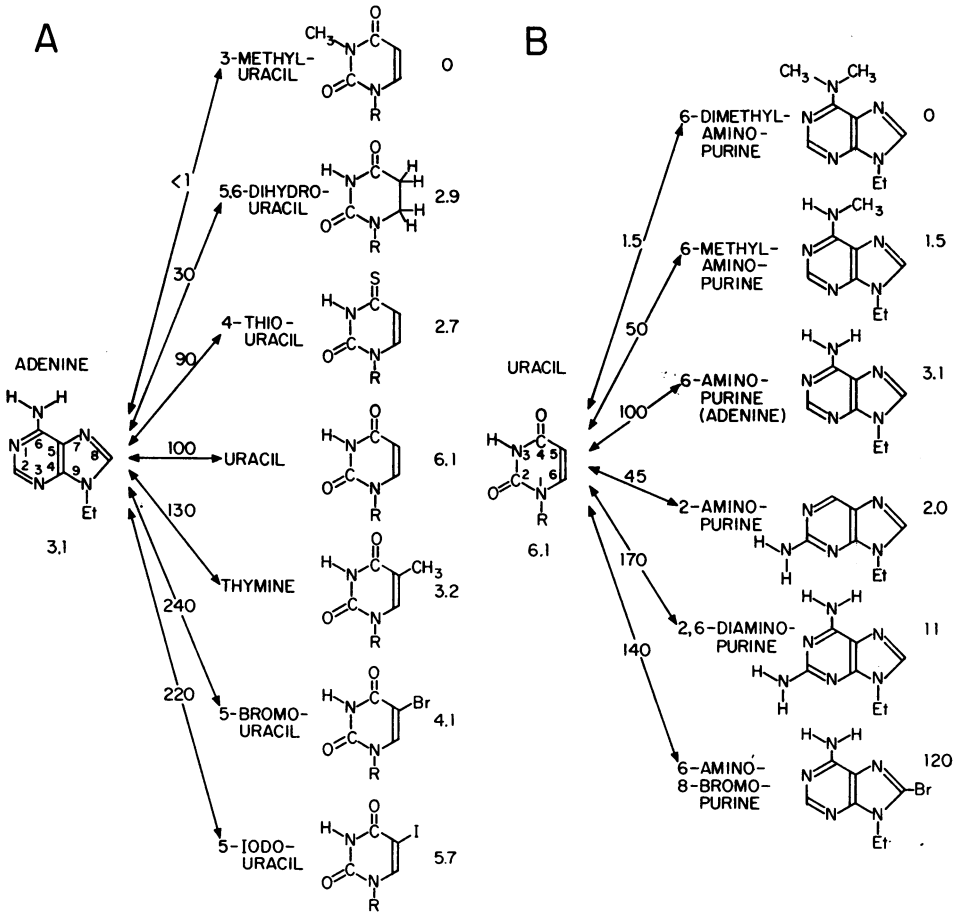

FIG. 3.—Association constants (liter/mole) between various substituted 1-cyclohexyluracil derivatives and 9-ethyladenine derivatives in deuterochloroform solution at 25°C. Figures near the double-headed arrows are the association constants between adenine and uracil derivatives, while the numbers shown to the right of the structural formulae are the self-association constants (*A*) Association of 9-ethyladenine with uracil derivatives. (*B*) Association of 1-cyclohexyluracil with adenine derivatives.

TABLE 1

CONSTANTS OF ASSOCIATION BETWEEN SOME 9-ETHYLADENINE AND 1-CYCLOHEXYLURACIL DERIVATIVES AT 25°C IN DEUTEROCHLOROFORM (IN LITERS/MOLE)

	5,6-Dihydrouracil	Uracil	Thymine	5-Bromouracil
2-Aminopurine	10	45	55	75
6-Methylaminopurine	15	50	70	100
6-Aminopurine (adenine)	30	100	130	240
2,6-Diaminopurine	100	170	210	550

strongest association is found with 5-bromo-1-cyclohexyluracil and 9-ethyladenine, and with 9-ethyl-2,6-diaminopurine and 1-cyclohexyluracil. It is of some interest to determine whether or not the association constants are related in an additive fashion to contributions from each of the partners. Accordingly, all 16 association constants were measured for a selected group of four uracil residues and four adenine residues as shown in Table 1. These constants vary in a systematic manner which can be understood on the basis of the individual contributions of the substituents in the purine and pyrimidine residues, as discussed below.

Earlier studies showed that 9-ethyladenine and 1-cyclohexyluracil associate preferentially with each other, as do cytosine and guanine derivatives. Nonetheless the adenine and uracil derivatives were found not to associate with the guanine and cytosine derivatives. This striking complementarity is of great interest in that it parallels the geometrical complementarity seen in double-stranded nucleic acids. It is therefore of interest to ascertain whether or not the 5-bromouracil and the 2,6-diaminopurine derivatives with the highest association constants are able to interact with the guanine and cytosine derivatives. A careful study was carried out with 2′,3′-benzylidine-5′-trityl-guanosine and -cytidine;[6] the results clearly indicate that there is no association of these residues with the 5-bromouracil and 2,6-diaminopurine derivatives.

Discussion.—In the detailed analysis[2] of the interaction of 9-ethyladenine with 1-cyclohexyluracil, several arguments were put forth for interpreting the 1:1 association of these molecules as due to the formation of a cyclic dimer in solution. Most of these points are relevant to the present study since the association bands seen here also have a maximum at a 1:1 mole ratio. In this regard, it should be mentioned that the association band at 3490 cm^{-1} of the amino group arises from the NH bond which remains free when the other NH is used for hydrogen bonding. If an adenine derivative formed a trimer with two molecules of uracil bonding onto the amino group, the free NH band would not be observed. Such a trimer may predominate at much higher concentrations, as has been suggested by nuclear magnetic resonance studies.[4] This is especially of interest in the case of the interactions with the 9-ethyl-2,6-diaminopurine since a 1:2 crystal has been obtained with this molecule and 1-methyl-5-fluorouracil (Chandross and Rich, unpublished observations) as well as 1-methyl-5-iodouracil (H. Sobell, personal communication). Dimers can exist in both open and cyclic forms. However, thermodynamic considerations indicate that formation of the cyclic dimer is much more likely than formation of an open dimer. This is confirmed by the negligible association of 9-ethyladenine and 1-cyclohexyl-3-methyluracil, where the presence of a methyl group on the imino nitrogen of the latter prevents the formation of a cyclic dimer but not of a linear dimer. In addition, the infrared band due to the free NH_2 group in an adenine derivative as well as that due to the free NH group in a uracil

derivative can be measured in 1:1 mixtures of various total concentrations. These bands decrease at a *constant intensity ratio* with increasing concentration, showing that both the NH_2 group of the adenine derivatives and the NH group of the uracil derivatives are involved in dimer formation at a constant ratio. Because self-association is negligible except at the higher concentrations, adenine NH_2 bonds only to uracil and uracil NH bonds only to adenine so that the dimer must be cyclic.

The data in Figure 3*A* show that the self-association constants of the uracil derivatives are very small and vary little with substitution. The only exception is 3-methyluracil which cannot self-associate by hydrogen bonding and for which no association was detected. On the other hand, the association constants of these derivatives with 9-ethyladenine vary considerably from that of unsubstituted uracil. Hydrogenation and its attendant loss of aromaticity in 5,6-dihydrouracil result in a substantial decrease. Hydrogenation doubtless causes puckering of the planar uracil molecule, which could weaken the hydrogen bonds of the cyclic dimer. More significant, probably, is the alteration in the acidity of the imino group, whose pK changes from 9.4 in uracil to 11 or larger in 5,6-dihydrouracil.[9-11] Since stronger hydrogen bonds are formed to a given acceptor by more acidic protons, increase in pK for the imino group should reduce its hydrogen bonding power. In Figure 3*A*, 4-thiouracil has $K = 90\ M^{-1}$, which is the same as the value for uracil within experimental uncertainty. Sulfur is much less electronegative than oxygen, and is thus a much weaker acceptor atom. Hence it would be expected that the adenine NH_2 group will bond to the oxygen of the 2-carbonyl rather than to the 4-thio group. If the 2- and 4-carbonyl sites in uracil itself are roughly equivalent, elimination of one of them should reduce K to about one half its value. However, we expect the pK of the proton to be more acidic. In 4-thiouridine, pK = 8.2,[12] compared with 9.4 for uridine.[11] Apparently this significant decrease in pK is sufficient to offset the loss of the 4-carbonyl oxygen.

Thymine associates slightly more strongly than does uracil ($K = 130$ vs. 100), which is understandable in view of the slight difference in electronegativities of —H and —CH_3. The methyl group is a weak electron donor and may improve the basicity of the 4-carbonyl oxygen a little, while raising the pK of the imino proton (pK of thymidine = 9.9[13] vs. 9.4 of uridine). These two weak effects will work in opposite directions, and it is not surprising that the K's for thymine and uracil are so close. A more substantial change is produced by 5-substitution of bromine or iodine, which more than doubles the association constant. The electronegativity of the halogens raises the acidity of the imino proton markedly (pK = 7.8 for 1-methyl-5-bromouracil[14] vs. 9.7 for 1-methyl-uracil[9, 11]) and on this score we would expect an increased association. However, we have not allowed for the effect of the halogens on the basicity of the carbonyl oxygens, which should be poorer acceptors here than in uracil. It appears that the latter effect is insufficient by a considerable amount to offset the large increase in the acidity of the NH group.

In Figure 3*B* are listed the results of measurements of association of several purine derivatives with themselves and with uracil. The self-association constants are all small except for 6-amino-8-bromopurine. There are two sites on adenine for self-association, involving the amino group and N_7 or N_1. The former site is occupied in the solid state in adenine hydrochloride,[15] while both sites are used in crystals of 9-methyladenine.[16] It is possible that bromine substitution in the 8-

position changes the site occupancy. The resonance effect would enhance the negative charge on N^7, while the inductive effect would have the opposite result. The latter effect, however, would increase the acidity of protons attached to the amino group and thereby increase the association. However, the increase in the self-association constant is unexpectedly large.

The two sites on adenine mentioned above are also available for cyclic dimer formation with uracil. Many crystal structures are found which use the adenine N_7 as an acceptor for the uracil imino group hydrogen and either of the uracil carbonyl groups as acceptor for an adenine NH_2 proton.[17–20] However, the purine N_1 has also been seen as a hydrogen-bond acceptor in one crystal analysis.[21] Site multiplicity is important in understanding the association of uracil with the adenine derivatives (Figure 3*B*). The association is almost completely eliminated by the loss of both protons on the amino group of adenine in 6-dimethylaminopurine. However, replacing one proton by a methyl group results in an approximate halving of the association constant. One or the other of the two binding sites is blocked by the methyl group. Thus it seems likely that both sites have half occupancy, and a halving of the association constant is not unexpected. Another interesting effect is seen when the amino group is placed in the 2-position of the purine ring. Here the only cyclic dimer which can form involves N_1, as is found in the crystal structure of the 9-ethyl-2-aminopurine:1-methyl-5-fluorouracil complex.[21] For this derivative also, the association constant is one half of that for adenine. These observations suggest the possibility that both N_1 and N_7 are equally involved in hydrogen bonding in adenine.

A substantial increase in the association constant is seen in the interaction of the uracil derivative with the 9-ethyl-2,6-diaminopurine. Both sites, N_1 and N_7, are readily available for hydrogen bonding with the imino NH of uracil. However, bonding to N_1 gives a structure capable of forming three hydrogen bonds and thus one would anticipate a substantial increase in the stability of the dimer compared to that seen for adenine itself. In the analysis of the guanosine-cytidine dimer with three hydrogen bonds, an incremental enthalpy of 2–3 kcal/mole is seen compared to the adenine-uracil interaction which is stabilized by two hydrogen bonds. Such an increase in enthalpy should cause the association constant to be 30–150 times larger than that found with the dimer containing only two hydrogen bonds, if we assume that the entropy change is the same in both cases. However, the association constant for 2,6-diaminopurine is only 1.7 times that of the adenine derivative, which suggests an enthalpy difference of 0.3 kcal/mole.

The addition of a bromine atom to the adenine C_8 results in an enhancement of the association constant. Furthermore, we propose that the real enhancement is greater than would appear by a simple comparison of the experimental association constants of uracil with adenine (100) and 8-bromoadenine (140). This is due to the fact that bonding by uracil to N_7 and the amino group is sterically inhibited by the large bromine Van der Waals radius of 1.95 Å. This prevents the formation of a hydrogen bond between the uracil N_3 and adenine N_7, due to the steric constraint arising from the carbonyl oxygen of uracil which is not involved in hydrogen bonding. The distance of closest approach between uracil N_3 and adenine N_7 is over 3.5 Å, which is far from a hydrogen bond distance. Accordingly, only the site involving the adenine N_1 and the amino group can be occupied by uracil. As seen in the

examples above, 2-aminopurine and 6-methylaminopurine halve the association constants relative to adenine itself. The observed association constant of 140 thus implies an increase by a factor of 2 to 3 in the extent of hydrogen bond formation for the N_1-amino group site. This might be anticipated by considering the inductive effect due to electron-attracting power of the bromine atom which would make the amino group nitrogen atom slightly positive.

The incentive for continuing this study in a nonaqueous solvent is the belief that some aspect of the interactions studied here may be related to the interactions which occur in double-stranded nucleic acid structures in aqueous media. In the center of the helical polynucleotides where the bases are stacked, there is in fact a region which is reasonably well shielded from the intrusion of water molecules. The present studies of hydrogen bonding allow us to make quantitative measurements of association phenomena, and thereby obtain information which may increase our understanding of polynucleotide interactions. Several of the derivatives studied here, especially the methylated bases, are found in transfer RNA, and the results may be relevant to its helical structure. In the discussion we have made qualitative statements about the effect of various substituents on the hydrogen-bonding association, but a more formal and precise treatment is required. The theoretical basis of the phenomenon of electronic complementarity needs to be well understood, since it may be of importance in completing our understanding of the molecular basis of information transfer in biological systems.

Note added in proof: Qualitative infrared measurements of some of the associations described above have also been made by J. Miller and M. M. Sobell (*J. Mol. Biol.*, to be published). Their results are similar to ours.

* This research was supported by grants from the National Institutes of Health, the Office of Naval Research, and the National Aeronautics and Space Adminstration.

[1] Hamlin, R. M., Jr., R. C. Lord, and A. Rich, *Science*, **148**, 1734 (1965).
[2] Kyogoku, Y., R. C. Lord, and A. Rich, *J. Am. Chem. Soc.*, in press.
[3] Kuechler, E., and J. Derkosch, *Z. Naturforsch.*, **21B**, 209 (1966).
[4] Katz, L., and S. Penman, *J. Mol. Biol.*, **15**, 220 (1966).
[5] Shoup, R. R., H. T. Miles, and E. D. Becker, *Biochim. Biophys. Res. Commun.*, **23**, 194 (1966).
[6] Kyogoku, Y., R. C. Lord, and A. Rich, *Science*, **154**, 518 (1966).
[7] Pitha, J., R. N. Jones, and J. Pithova, *Can. J. Chem.*, **44**, 1044 (1966).
[8] Lord, R. C., and T. J. Porro, *Z. Elektrochem.*, **64**, 672 (1960).
[9] Levene, P. A., L. W. Bass, and H. S. Simms, *J. Biol. Chem.*, **70**, 229 (1926).
[10] Shugar, D., and J. J. Fox, *Biochim. Biophys. Acta*, **9**, 199 (1952).
[11] Jonáš, J., and J. Gut, *Collection Czech. Chem. Commun.*, **27**, 546 (1962).
[12] Lipsett, N. M., *J. Biol. Chem.*, **240**, 3975 (1965).
[13] Fox, J. J., and D. Shugar, *Biochim. Biophys. Acta*, **9**, 369 (1952).
[14] Katritzky, A. R., and A. J. Warning, *J. Chem. Soc.*, 1540 (1962).
[15] Cochran, W., *Acta Cryst.*, **4**, 81 (1951).
[16] Stewart, R. F., and L. H. Jensen, *J. Chem. Phys.*, **40**, 2071 (1964).
[17] Hoogsteen, K., *Acta Cryst.*, **12**, 822 (1959); *ibid.*, **16**, 907 (1963).
[18] Haschemeyer, A. E. V., and H. M. Sobell, these PROCEEDINGS, **50**, 872 (1963).
[19] Mathews, F. S., and A. Rich, *J. Mol. Biol.*, **8**, 89 (1964).
[20] Katz, L., K. Tomita, and A. Rich, *J. Mol. Biol.*, **13**, 340 (1965).
[21] Sobell, H. M., *J. Mol. Biol.*, **18**, 1 (1966).