Detection of Separated Amino Proton Resonance Signals of Adenine Derivatives at Low Temperature and Its Application to Estimation of Population of the Adenine–Uracil Dimers in Solution†

Hideo Iwahashi, Hiromu Sugeta, and Yoshimasa Kyogoku*

ABSTRACT: Splitting of the amino proton signals of 9-ethyladenine derivatives was found in proton nuclear magnetic resonance spectra at low temperature (ca. −30 °C). One of the separated signals corresponds to the syn amino proton relative to the N(1) nitrogen in the adenine ring and the other to the anti one. The phenomenon is ascribable to slowing down of the hindered rotation around the N(6)–C(6) bond, which has partial double bond character. On the addition of 1-cyclohexyluracil derivatives, one of the separated signals shifts downfield. From the analysis of the concentration dependence of the signals we could estimate the population of two kinds of adenine–uracil (AU) dimers that employ the syn and anti protons, respectively, i.e., the Watson–Crick-type and the Hoogsteen-type dimers. Independent of the substitution on the uracil ring, the Hoogsteen type is predominant at 70% and the Watson–Crick type at 30% (at −56 °C). On the other hand, with mixtures of several kinds of 9-ethyladenine derivatives with 1-cyclohexyluracil, the substituents on the adenine ring cause the population to deviate to extreme values; i.e., either the Watson–Crick-type or the Hoogsteen-type dimer predominates. 2-Chloro-9-ethyladenine and N^2-(dimethylamino)-9-ethyladenine take almost completely the Hoogsteen-type dimers, while 8-bromo-9-ethyladenine, N^2-(methylamino)-9-ethyladenine, and 2-amino-9-ethylpurine predominate in the Watson–Crick-type dimers.

In the double-stranded structure of deoxyribonucleic and ribonucleic acids, adenine forms specific hydrogen bonds with thymine (or uracil) and guanine with cytosine. Many works have been done to find the origin of the specificity in the hydrogen-bonded base pairs by the use of the synthetic poly- and oligonucleotides and purine and pyrimidine base derivatives (Brahms, 1965; Stevens & Felsenfeld, 1964; Blake & Massaulie, 1967; Gralla & Crothers, 1973; Borer et al., 1974). Some infrared (Hamlin et al., 1965; Küchler & Derkosch, 1966; Kyogoku et al., 1966, 1967a,b, 1969; Miller & Sobell, 1967) and proton nuclear magnetic resonance studies (Shoup et al., 1966; Katz & Penman, 1966; Katz, 1969; Engle & von Hippel, 1974; Morishima et al., 1977) clearly showed that even the base residues at the monomer level interact selectively through hydrogen bonding in solution. However, little is known about the structure of the dimer in solution. In polynucleotides it is believed that the Watson–Crick-type base pair is formed (Watson & Crick, 1953), but at the monomer level this is not always true. A ^{13}C NMR[1] study has showed that uracil and its derivatives interact with adenine in terms of hydrogen bonding with not only the C(4)-carbonyl group but also the C(2)-carbonyl group of the uracil derivatives as hydrogen-acceptor sites and the population of the two types of dimer depends on the substituents on the pyrimidine ring (Iwahashi & Kyogoku, 1977). According to X-ray analyses of the cocrystals of the adenine–uracil derivatives, the so-called Hoogsteen-type dimer often appears where N(7) is used as a proton-acceptor site instead of N(1) in the Watson–Crick-type dimer (Hoogsteen, 1963) (Figure 1). This means that the proton-acceptor site on the adenine ring is not fixed. In fact ^1H NMR investigation by observing the H(2) and H(8) proton signals of the adenine ring has showed that Hoogsteen-type and Watson–Crick-type dimers coexist in solution (Engel & von Hippel, 1974).

It is known that the hindered rotation about the N(4)–C(4) bond of cytosine allows us to observe the amino proton resonance as separated signals at low temperature (Becker et al., 1965; Shoup et al., 1966; Raszka & Kaplan, 1972). In the aqueous solution of a nucleoside monophosphate (GpC), the separated amino proton signals were observed (Young & Krugh, 1975), and one of the two signals showed concentration dependence but the other did not. In the present work similar separated signals of the adenine amino proton were observed at much lower temperature and could be respectively assigned to the syn and anti amino protons relative to the N(1) nitrogen. The population of the Watson–Crick-type dimer was calculated at monomer level by the use of the phenomenon. On the basis of this result, we will discuss the factors that govern the stability of each type of dimers. The existence of a stable structure of the base pair other than the normal Watson–Crick type pertains to the irregular base pair in the tertiary structure of tRNA and ribosomal RNA.

Experimental Procedures

Materials. 9-Ethyladenine (A), 8-bromo-9-ethyladenine (Br^8A), 2-chloro-9-ethyladenine (Cl^2A), N^2-(methylamino)-9-ethyladenine (Me^2A), N^2-(dimethylamino)-9-ethyladenine ($DiMe^2A$), 2-amino-9-ethylpurine (NH_2^2P), 1-cyclohexyluracil (U), 1-cyclohexylthymine (T), 1-cyclohexyl-5-bromouracil (Br^5U), 1-cyclohexyl-5,6-dihydrouracil (DhU), and 1-cyclohexyl-4-thiouracil (S^4U) were purchased from Cyclo Chemical Co., Los Angeles, CA. 1-Cyclohexyl-2-thiouracil (S^2U) was a gift from Dr. S. Higuchi, Mitsubishi Kasei Institute of Life Sciences. 9-Ethyladenine was recrystallized from a solvent mixture of carbon tetrachloride and chloroform, and the other compounds were used without further purification. ^1H NMR spectra were measured for their chloroform-d_1 ($CDCl_3$) solutions. Chloroform-d_1 was obtained from CEA, France.

Methods. ^1H NMR spectra were obtained at 100 MHz with a JEOL FX-100 pulse Fourier-transform NMR system

† From the Institute for Protein Research, Osaka University, Suita, Osaka, Japan. *Received May 29, 1981.*

[1] Abbreviations: ^{13}C NMR, carbon-13 nuclear magnetic resonance; ^1H NMR, proton nuclear magnetic resonance; ^{15}N NMR, nitrogen-15 nuclear magnetic resonance.

0006-2960/82/0421-0631$01.25/0

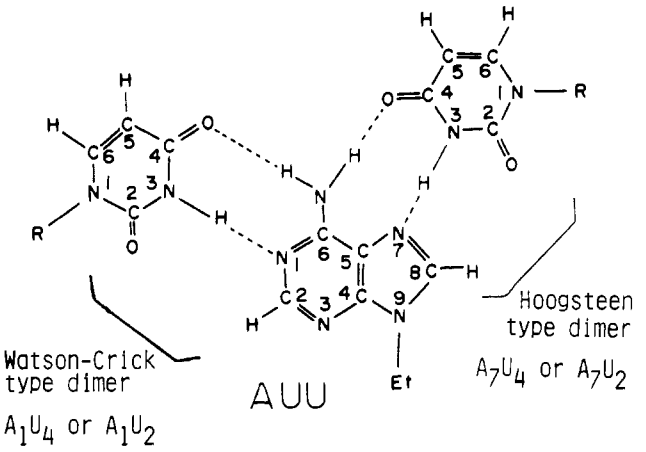

FIGURE 1: Schematic expressions for the AU dimers and AUU trimer. A_1U_4 is the Watson–Crick-type dimer. Its reversed type where the C(2)-carbonyl group of uracil is used instead of the C(4)-carbonyl group is possible to be formed. A_7U_4 is the Hoogsteen-type dimer. Its reversed type is also present.

locked on deuterium. Chemical shifts were measured relative to the ^1H resonance of internal tetramethylsilane. Spectra were run for solutions in a concentration range from 1 to 200 mM, at −56 °C and sometimes at −65 °C. The temperature of the sample tube was controlled by a temperature controller equipped with the NMR spectrometer, and the value of temperature was corrected by the chemical shift of methanol.

Procedures for Calculating the Population of Each Dimer Structure Using Chemical Shifts of the Separated Amino Proton Signals. We consider here the case of the adenine–uracil mixture. The procedure can also be applied to the systems of other derivatives. In the mixture solution of 9-ethyladenine (A) and 1-cyclohexyluracil (U), the following equilibria exist:

$$A + A \overset{K_{AA}}{\rightleftharpoons} AA \qquad K_{AA} = [AA]/[A]^2 \tag{1}$$

$$U + U \overset{K_{UU}}{\rightleftharpoons} UU \qquad K_{UU} = [UU]/[U]^2 \tag{2}$$

$$A + U \overset{K_{AU}}{\rightleftharpoons} AU \qquad K_{AU} = [AU]/([A]\,[U]) \tag{3}$$

$$AU + U \overset{K_{AUU}}{\rightleftharpoons} AUU \qquad K_{AUU} = [AUU]/([AU]\,[U]) \tag{4}$$

Equations 1 and 2 describe the processes of the self-association of A and U, respectively. Two types of association of A with U (1:1 and 1:2) are written in eq 3 and 4. We assume here that the formation of the other type of association, like AAA and AAU trimers, can be neglected. The observed chemical shifts of A and U are weighted averages of the intrinsic chemical shifts of the species (A, AA, AU, and AUU) for A and the species (U, UU, AU, and AUU) for U, since the molecular exchange among these chemical species is faster than the NMR time scale. The concentration dependence of the chemical shifts allows us to calculate the limiting shifts of the chemical species (δ_A, δ_U, δ_{AA}, δ_{UU}, δ_{AU}, and δ_{AUU}) and the association constants (K_{AA}, K_{UU}, K_{AU}, and K_{AUU}) (Sugeta, 1982). However, several types of structures of the dimers should be taken into account, because both A and U have two proton-acceptor sites. Taking all the possibilities into account, eq 3 can be rewritten as

$$A + U \rightleftharpoons P_{12}A_1U_2 + P_{14}A_1U_4 + P_{72}A_7U_2 + P_{74}A_7U_4 \tag{5}$$

where subscripts i and j in A_iU_j indicate the proton-acceptor sites in A and U, respectively, in the AU dimer (Figure 1); A_1U_2, reversed Watson–Crick type; A_7U_2, reversed Hoogsteen type; A_1U_4, Watson–Crick type; A_7U_4, Hoogsteen type; P_{ij} represents the mole fraction of the A_iU_j species ($\sum P_{ij} = 1$). The induced chemical shift Δ_{AU} of the AU dimer, which can be obtained from observed data, is expressed in terms of the induced shifts $\Delta_{A_iU_j}$ for the A_iU_j dimers:

$$\Delta_{AU} = \sum_{\substack{i=1,7 \\ j=2,4}} P_{ij}\Delta_{A_iU_j} \tag{6}$$

where the induced shift (Δ) is the displacement of the chemical shift due to the dimer formation from the limiting chemical shift of the monomer species, i.e., the difference of limiting shift of the dimer δ_D from the monomer status δ_F, $\Delta = \delta_D - \delta_F$. The macroassociation constant K_{AU}, which is obtainable from experimental data, is expressed in terms of the microassociation constants $K_{A_iU_j}$ for the A_iU_j formation:

$$K_{AU} = \sum_{\substack{i=1,7 \\ j=2,4}} K_{A_iU_j} \tag{7}$$

$$K_{A_iU_j} = [A_iU_j]/[A][U] = P_{ij}K_{AU} \tag{8}$$

As will be described in detail under Results and Discussion, observation of the separated amino proton signals of A gives information on the proton-acceptor sites in A although no information is available on the proton-acceptor sites in U in the present work. Accordingly, taking into account only the acceptor sites in A, eq 6–8 are rewritten as

$$\Delta_{AU} = P_W\Delta_{A_1U} + P_H\Delta_{A_7U} \tag{9}$$

$$K_{AU} = K_{AU}{}^W + K_{AU}{}^H \tag{10}$$

$$K_{AU}{}^W = P_WK_{AU} = K_{A_1U_2} + K_{A_1U_4} \tag{11a}$$

$$K_{AU}{}^H = P_HK_{AU} = K_{A_7U_2} + K_{A_7U_4} \tag{11b}$$

where

$$P_W = P_{12} + P_{14} \tag{12a}$$

$$P_H = P_{72} + P_{74} \tag{12b}$$

$$\Delta_{A_1U} = (P_{12}\Delta_{A_1U_2} + P_{14}\Delta_{A_1U_4})/(P_{12} + P_{14}) \tag{13a}$$

$$\Delta_{A_7U} = (P_{72}\Delta_{A_7U_2} + P_{74}\Delta_{A_7U_4})/(P_{72} + P_{74}) \tag{13b}$$

Here W and H refer to the Watson–Crick- and Hoogsteen-type associations; that is, the syn amino proton relative to N(1) and the N(1) atom of A participate in the Watson–Crick-type dimer, while the anti amino proton and N(7) participate in the Hoogsteen-type association. Δ_{A_1U} and Δ_{A_7U} are induced chemical shifts of the Watson–Crick- and Hoogsteen-type AU dimers, respectively.

Now consider the induced chemical shifts, $\Delta_{AU}{}^{syn}$ and $\Delta_{AU}{}^{anti}$, for separated amino protons, syn and anti relative to N(1). In the Watson–Crick-type dimer the syn amino proton participates in hydrogen bonding while the anti amino proton is free from hydrogen bonding, vice versa in the Hoogsteen-type dimer. We assume that the shielding of the one amino proton participating in hydrogen bonding is different from that of the free state, while the shielding of the other amino proton free from hydrogen bonding does not change from that of the free-state situation. That is, $\Delta_{A_1U}{}^{syn} \gg \Delta_{A_7U}{}^{syn} \approx 0$ and $\Delta_{A_7U}{}^{anti} \gg \Delta_{A_1U}{}^{anti} \approx 0$. Then eq 9 for the separated amino proton signals becomes

$$\Delta_{AU}{}^{syn} \approx P_W\Delta_{A_1U}{}^{syn} \tag{14a}$$

$$\Delta_{AU}{}^{anti} \approx P_H\Delta_{A_7U}{}^{anti} \tag{14b}$$

In the AUU trimer both syn and anti amino protons participate in the hydrogen bond (Figure 1), and we assume that the

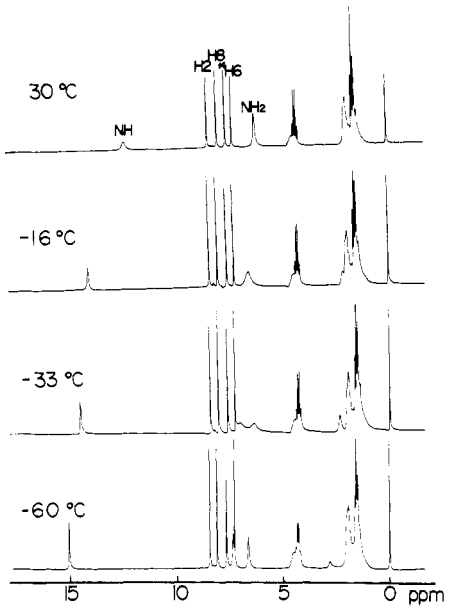

FIGURE 2: ^{1}H NMR spectra of the chloroform-d_1 solution of the 1:1 mixture of 9-ethyladenine (A) and 1-cyclohexyl-5-bromouracil (Br^5U) (50 + 50 mM) at different temperatures. The asterisk denotes the solvent signal.

induced chemical shifts, $\Delta_{AUU}{}^{syn}$ and $\Delta_{AUU}{}^{anti}$, of the separated amino protons are the same as those, $\Delta_{A_1U}{}^{syn}$ and $\Delta_{A_7U}{}^{anti}$, respectively, of the AU dimer formation:

$$\Delta_{AUU}{}^{syn} \approx \Delta_{A_1U}{}^{syn} \quad (15a)$$

$$\Delta_{AUU}{}^{anti} \approx \Delta_{A_7U}{}^{anti} \quad (15b)$$

By use of eq 14 and 15 the populations P_W and P_H of the Watson–Crick- and Hoogsteen-type association in the AU dimer can be obtained from the experimentally available chemical shifts Δ_{AU} and Δ_{AUU}:

$$P_W = \Delta_{AU}{}^{syn}/\Delta_{AUU}{}^{syn} \quad (16a)$$

$$P_H = \Delta_{AU}{}^{anti}/\Delta_{AUU}{}^{anti} \quad (16b)$$

For the systems where limiting shift in the trimer (Δ_{AUU}) cannot be obtained, the ratio ($\Delta_{A_1U}{}^{syn}/\Delta_{A_7U}{}^{anti}$) of the amino groups of all the adenine derivates is assumed to be constant. By use of the average ratio ($\Delta_{A_1U}{}^{syn}/\Delta_{A_7U}{}^{anti}$) of A and U derivatives, $\Delta_{A_1U}{}^{syn}$ and $\Delta_{A_7U}{}^{anti}$ of $DiMe^2A$ and Cl^2A are estimated, and the population of each dimer is obtained from eq 14. For NH_2^2P, which uses the 2-amino group for the dimer formation instead of the 6-amino group, this procedure was not applied.

Results

Mixture Systems of A with Several 1-Cyclohexyluracil Derivatives. Spectra of the chloroform-d_1 solution of 9-ethyladenine (A) and 1-cyclohexyl-5-bromouracil (Br^5U) at different temperatures are given in Figure 2. The amino proton signal of A splits into two peaks at lower temperature due to the hindered rotation around the C(6)–N(6) bond. The concentration dependence of the chemical shifts of the separated amino protons is presented for mixtures of A and U and of A and S^4U (Figure 3). For all other uracil derivatives, T, DhU, Br^5U, and S^2U, similar curves were observed in the plots of chemical shift vs. concentration. At the lower concentration of the uracil derivatives, one of the two peaks shifts downfield with the increase of concentration of the uracil derivatives, and the other peak shifts upfield to some extent rather than downfield. At these concentrations the lower field peak of the

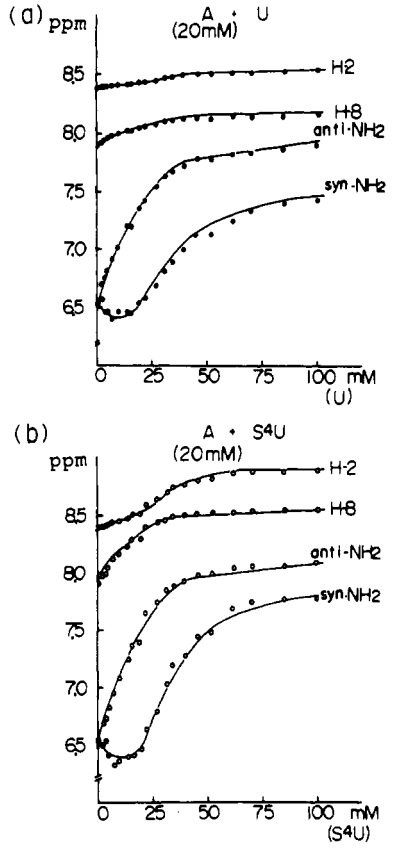

FIGURE 3: Chemical shifts of the separated amino proton signals of 9-ethyladenine (A) (20 mM) against the concentration of (a) 1-cyclohexyluracil (U) and (b) 1-cyclohexyl-4-thiouracil (S^4U) at −56 °C. The solid lines represent values calculated by the curve-fitting method.

separated amino proton signals is due to the proton that participates in the A–U hydrogen bonds, and the other higher field one arises from the amino proton free from hydrogen bonds. On addition of an excess amount of uracil derivatives, the higher field peak also begins to shift downfield, while the lower field one reaches the limit of the shift faster than the former. Therefore, the lower field proton is considered to be used in hydrogen bonds in the AU dimer, and the shift of the higher field peak on the addition of an excess amount of U is due to the formation of the AUU trimer where both anti and syn NH_2 protons are used in the hydrogen bonds.

By comparison of the concentration dependences of the chemical shifts of the H(2) and H(8) protons of A with those of the two amino protons, the lower field amino proton has a good correlation with that of the H(8) proton and the higher field amino proton with that of the H(2) proton. When the location of the H(2) and H(8) protons in the two types of the base pairing is taken into account, the lower field amino proton can be assigned to the anti NH_2 proton and the higher field one to the syn NH_2 proton. Since the syn NH_2 proton is used in the Watson–Crick-type dimer and the anti NH_2 proton in the Hoogsteen-type dimer, it seems that the Hoogsteen-type dimer is formed first, and the Watson–Crick-type bonding follows it.

Mixture Systems of Several 9-Ethyladenine Derivatives with U. Between U and several 9-ethyladenine derivatives, Br^8A, Cl^2A, $DiMe^2A$, Me^2A, and NH_2^2P, concentration dependences of the separated amino protons were also measured. The syn NH_2 and anti NH_2 protons of all the 9-ethyladenine derivatives were assigned in similar ways as those of A. The anti NH_2 protons of Cl^2A and $DiMe^2A$ shift downfield on addition of U (Figure 4). On the other hand, the syn NH_2

Table I: Limiting Shifts (δ_F) of Free States and Difference (Δ) in Those between Free and Bound States in Some Typical Mixture Systems of 9-Ethyladenine and1-Cyclohexyluracil Derivatives (ppm)

compound	signals	δ_F	Δ_{UU}	Δ_{AA}	Δ_{AU}	Δ_{AUU}
A + U[a]	imino H	8.062 (42)[c]	3.728 (33)		6.498 (74)	7.793 (272)
	syn NH_2	5.516 (30)		3.685 (85)	1.113 (68)	3.020 (131)
	anti NH_2	5.884 (26)		2.401 (61)	1.928 (46)	2.265 (66)
	H(2)	8.422 (16)		−0.025 (28)	0.033 (23)	0.271 (38)
	H(8)	7.933 (16)		0.017 (28)	0.197 (23)	0.335 (38)
A + Br^5U[a]	imino H	8.089 (28)	3.708 (31)		7.337 (81)	7.299 (66)
	syn NH_2	5.516 (30)		3.685 (85)	0.835 (67)	2.369 (40)
	anti NH_2	5.884 (26)		2.401 (61)	1.912 (66)	2.070 (30)
	H(2)	8.422 (16)		−0.025 (28)	0.000 (26)	0.229 (30)
	H(8)	7.933 (16)		0.017 (28)	0.244 (26)	0.309 (22)
A + S^4U[a]	imino H	9.418 (6)	3.042 (10)		5.968 (89)	6.084 (73)
	syn NH_2	5.516 (30)		3.685 (85)	0.856 (63)	2.618 (46)
	anti NH_2	5.884 (26)		2.401 (61)	2.228 (84)	2.225 (34)
	H(2)	8.422 (16)		−0.025 (28)	0.106 (32)	0.578 (24)
	H(8)	7.933 (16)		0.017 (28)	0.621 (32)	0.647 (24)
Me^2A + U[b]	imino H	8.598 (97)	3.299 (67)		5.639 (67)	5.786 (90)
	syn NH_2	5.269 (20)		3.725 (70)	3.236 (64)	2.709 (26)
	anti NH_2	5.751 (19)		2.788 (62)	0.058 (53)	1.996 (39)
	NHMe	4.881 (16)		1.836 (46)	2.976 (62)	2.569 (24)
	H(8)	7.463 (11)		−0.051 (21)	−0.020 (13)	0.247 (10)
$DiMe^2A$ + U[b]	imino H	8.598 (97)	3.299 (67)		6.468 (70)	
	syn NH_2	5.133 (9)		2.609 (48)	0.277 (12)	
	anti NH_2	5.614 (9)		2.516 (47)	2.219 (16)	
	H(8)	7.620 (12)		0.077 (30)	0.282 (10)	
NH_2^2P + U[b]	imino H	8.598 (97)	3.299 (67)		6.149 (302)	
	syn NH_2	5.062 (57)		2.010 (253)	2.578 (99)	
	anti NH_2	5.272 (52)		1.616 (212)	0.373 (54)	
	H(6)	8.776 (10)		−0.006 (38)	0.236 (11)	
	H(8)	7.915 (10)		−0.040 (38)	0.003 (11)	

[a] At −56 °C. [b] At −65 °C. [c] Figures in parentheses indicate standard deviation to the last digits in preceding figures.

Table II: Association Constants for Self-Association and Association between 9-Ethyladenine and 1-Cyclohexyluracil Derivatives (M^{-1})

compounds	K_{UU}	K_{AA}	K_{AU}	K_{AUU}	temp (°C)
A + U	91 ± 6	13 ± 1	1400 ± 370	47 ± 13	−56
A + T	38 ± 2		940 ± 230	64 ± 16	−56
A + DhU	34 ± 1		440 ± 70	30 ± 6	−56
A + Br^5U	40 ± 2		1800 ± 740	380 ± 140	−56
A + S^4U	19 ± 0.4		1000 ± 420	180 ± 61	−56
A + S^2U	380 ± 20		940 ± 230	130 ± 30	−56
Br^8A + U	91 ± 6	18 ± 4	480 ± 30	6.1 ± 3.1	−56
Cl^2A + U		22 ± 27	1600 ± 200	0	−56
Me^2A + U	160 ± 20	15 ± 1	4300 ± 1200	520 ± 180	−65
$DiMe^2A$ + U		4.3 ± 0.3	1800 ± 220	0	−65
NH_2^2P + U		4.3 ± 0.5	510 ± 130	0	−65

proton shows rather little change in its chemical shift even on addition of an excess amount of U. This indicates predominant formation of the Hoogsteen-type dimer and difficulty of AUU trimer formation. In the case of Br^8A, Me^2A, and NH_2^2P, the syn NH_2 proton signals shift downfield, and the anti NH_2 proton signal of Me^2A also shifts downfield on addition of an excess amount of U. The Watson–Crick-type base pair is stable for these derivatives, and the results agree quite well with the crystal structures of Br^8A and NH_2^2P with U complex obtained by X-ray analyses (Tavale et al., 1969; Mazza & Sobell, 1969).

Self-Association of 9-Ethyladenine Derivatives. The concentration dependences of the two amino protons in the self-association of A and its derivatives were measured. Although the H(2) and H(8) proton signals did not show any shift, syn and anti NH_2 protons could be assigned in comparison with the resonance positions in the A–U mixture system. In contrast to the AU association, the two amino protons show almost the same induced chemical shifts except for the case of Cl^2A (Figure 5). Thus, both of the two amino protons are almost equivalently used for hydrogen bonds in the self-associated dimers.

Quantitative Treatment of Association Constants and Limiting Chemical Shifts. The association constants and the limiting chemical shifts of each state of the self- and hetero-association of A and U derivatives were calculated from the data of concentration dependences of chemical shifts by applying curve-fitting methods (Sugeta, 1982). Although a slight amount of water (~10 mM) is present in the solutions, the association of water with adenine and uracil derivatives was not taken into account in the calculations. It is known that the association constant between water and the base derivatives is 10–100 times smaller than those between the derivatives (Kyogoku et al., 1969). The obtained limiting shifts of some typical adenine and uracil derivatives are given in Table I and association constants are in Table II.

For the assignment of the syn NH_2 and anti NH_2 protons of 9-ethyladenine and its derivatives, we employed the behavior of the induced chemical shifts of the H(2) and H(8) protons as mentioned previously. The obtained induced chemical shifts of the AU mixture system show a good correlation between the limiting shift of the the syn NH_2 proton (1.113 ppm) and that of H(2) (0.033 ppm) and between the anti NH_2 proton (1.928 ppm) and H(8) (0.197 ppm). Similar correlation in

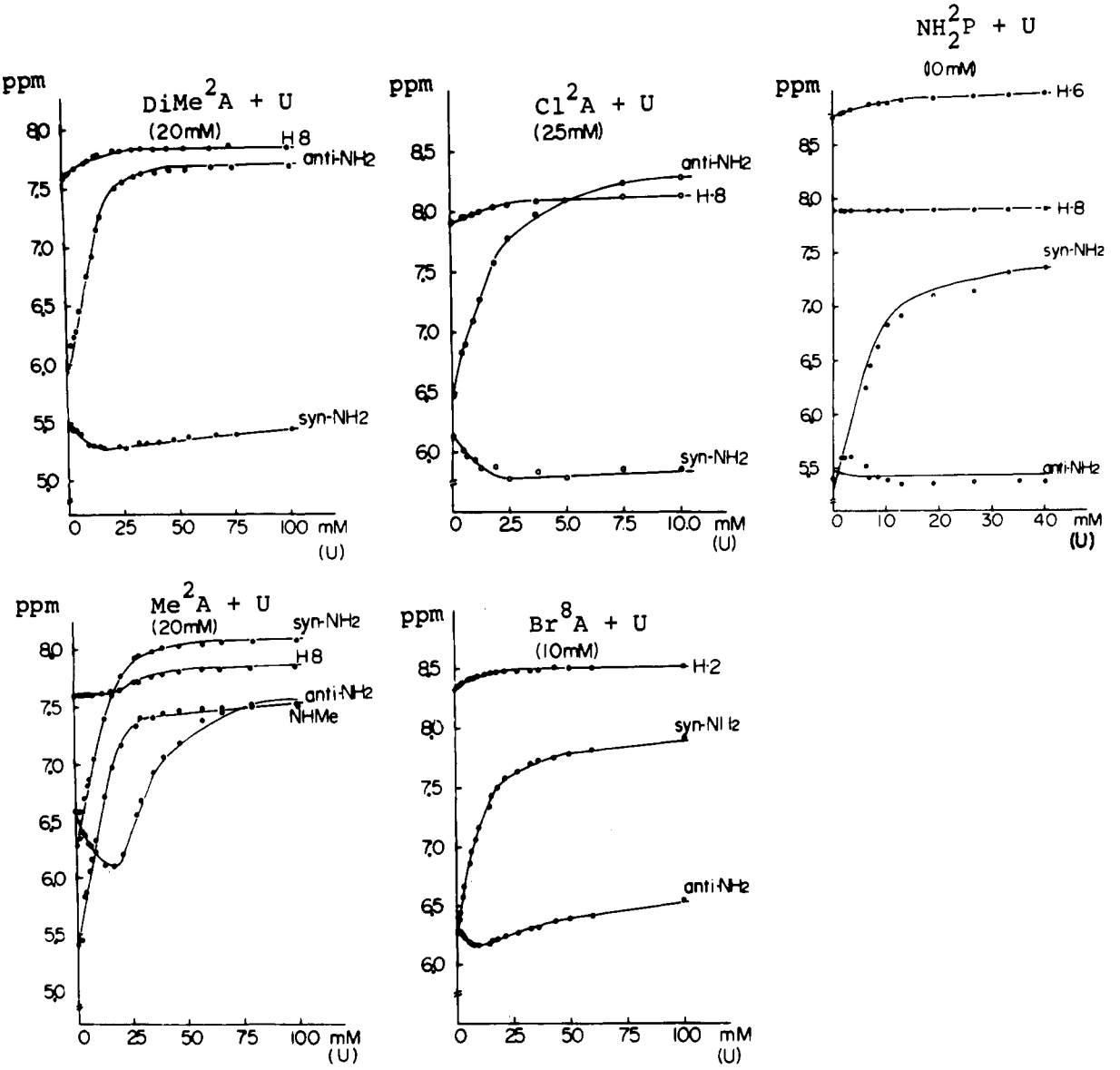

FIGURE 4: Chemical shifts of the separated amino proton signals of 9-ethyladenine derivatives (Cl^2A, Br^8A, $DiMe^2A$, Me^2A, and NH_2^2P) against the concentration of 1-cyclohexyluracil (U) at −56 °C (Cl^2A and Br^8A) and at −65 °C ($DiMe^2A$, Me^2A, and NH_2^2P). The solid lines represent values calculated by the curve-fitting method.

the induced chemical shifts between the amino protons and the H(2) or H(8) proton also holds for the other A derivatives.

The induced chemical shifts of the H(2) and H(8) protons of A in the AUU trimer where the proton-acceptor sites near the corresponding CH protons are completely occupied by hydrogen bonds depend on counterpart uracil derivatives. One of the two carbonyl groups of the uracil derivatives, which is not used in the hydrogen bond, locates more closely to the H(2) proton than to H(8) in the Watson–Crick-type dimer, while it locates more closely to the H(8) proton than to H(2) in the Hoogsteen-type dimer (Figure 1). The large downfield shifts of the H(2) and H(8) protons in the mixture with S^2U and S^4U (~0.6 ppm) may be due to more effective deshielding by the thiocarbonyl group than the carbonyl group (0.2–0.3 ppm). It is reasonable, therefore, that the H(2) and H(8) proton signals do not show any shift in the self-association of adenine derivatives (Figure 5) where no large chemical group that produces shielding or deshielding comes close to the position of H(2) or H(8).

By use of the association constants, the populations of the monomer (A), dimer (AA and AU), and trimer (AUU) were calculated. The concentration dependence of the fraction of each species is plotted in Figure 6. The AA dimer exists to some extent when U is not presented, and on addition of U to the solution, the fraction of the AA dimer decreases. The syn NH_2 proton is used in the hydrogen bond of the AA dimer in a considerable amount (Figure 5) but is not used very much in the AU dimer (Figure 3). The increase of the AU dimer in place of the AA dimer makes the syn NH_2 proton free from the hydrogen bond and results in a shift relatively upfield as shown in the low concentration range of U in Figures 3 and 4. Similar explanations can be done for all the other AU mixture systems. In the case of the NH_2^2P + U mixture (Figure 4), however, the deviations of the calculated shift values of the amino proton resonances from the observed ones are fairly large. They arise mainly from inaccurate reading of the resonance position due to the broadness of separated amino proton signals caused by residual rotation of the amino group of NH_2^2P at −65 °C.

The obtained association constants of the trimer (K_{AUU}) formation are almost zero for Cl^2A, $DiMe^2A$, and NH_2^2P as expected from the concentration dependences shown in Figure 4. For Br^8A, K_{AUU} is smaller than that of A but is not zero. The bromine atom at C(8) seems to partly disturb the trimer

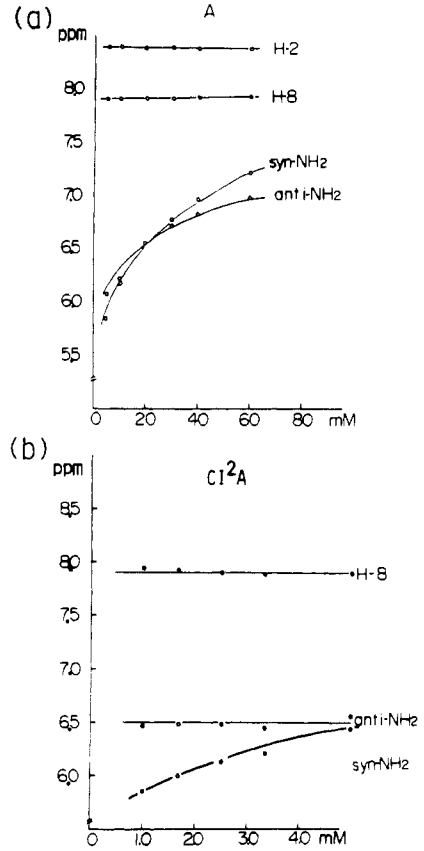

FIGURE 5: Concentration dependence of the chemical shifts of the separated amino proton signals of (a) 9-ethyladenine (A) and (b) 2-chloro-9-ethyladenine (Cl^2A).

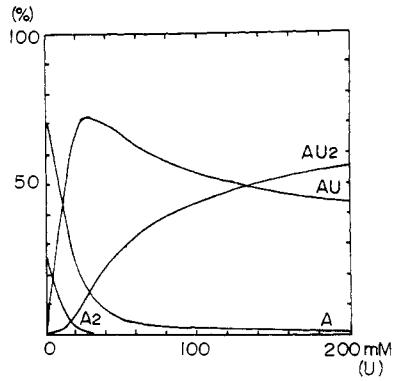

FIGURE 6: Population of each species (A, AA, AU, and AUU) plotted against the concentration of 1-cyclohexyluracil (U) (0–100 mM), keeping the concentration of 9-ethyladenine (A) at 20 mM.

formation. The association constant of the dimer formation between Me^2A and U is fairly large compared with those of other mixtures. The stabilization of the Watson–Crick-type dimer by three hydrogen bonds may be responsible for the strong association (Figure 7).

Discussion

When the method described under Experimental Procedures is followed, the populations of the Watson–Crick-type and the Hoogsteen-type dimers were calculated (Table III). If the assumptions in the procedure (eq 14–16) are held, the sum of P_W and P_H, both of which are independently calculated from the induced chemical shift of the syn NH_2 and anti NH_2 protons, should be 100. However, the sum ($P_W + P_H$) exceeds 100 for all derivatives. There are several reasons conceivable, for example, neglection of the $\Delta_{A_1U}{}^{anti}$ or $\Delta_{A_7U}{}^{syn}$ term in the

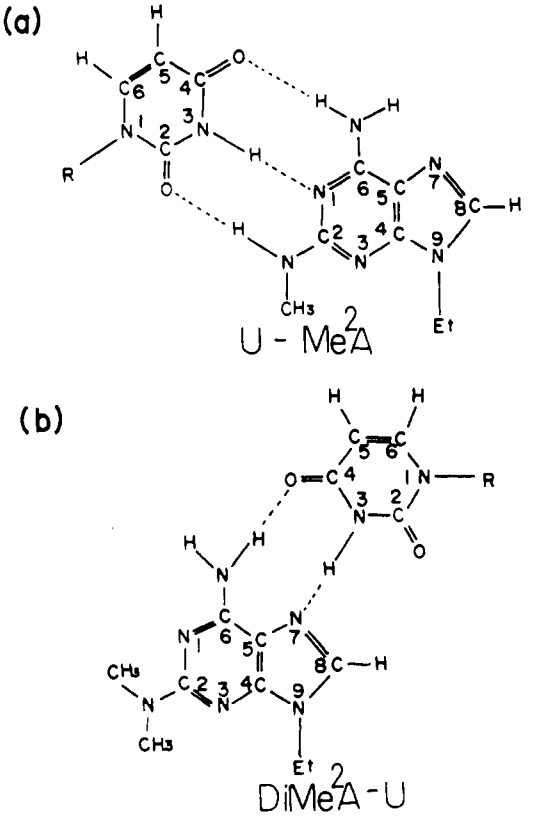

FIGURE 7: Structure of predominant dimers (a) between N^2-(methylamino)-9-ethyladenine (Me^2A) and 1-cyclohexyluracil (U) and (b) between N^2-(dimethylamino)-9-ethyladenine ($DiMe^2A$) and 1-cyclohexyluracil (U).

Table III: Estimation of Population of Watson–Crick-Type Base Pairs and Hoogsteen-Type Base Pairs in Mixtures of 9-Ethyladenine and 1-Cyclohexyluracil Derivatives (%)

compounds	P_W	P_H	$P_W{}^{N\,a}$	$P_H{}^{N\,b}$
A + U	37	85	30	70
A + DhU	43	93	32	68
A + T	36	97	27	73
A + Br^5U	32	92	26	74
A + S^4U	33	100	25	75
A + S^2U	31	93	25	75
Br^8A + U	129	9	93	7
Cl^2A + U[c]			7	93
Me^2A + U	119	3	98	2
$DiMe^2A$ + U[c]			9	91

[a] $P_W{}^N = P_W/(P_W + P_H)$. [b] $P_H{}^N = P_H/(P_W + P_H)$. [c] See Experimental Procedures.

derivation of eq 14, the assumptions $\Delta_{A_1U}{}^{syn} = \Delta_{AUU}{}^{syn}$ and $\Delta_{A_7U}{}^{anti} = \Delta_{AUU}{}^{anti}$, or an experimental effect such as contaminated water on the induced chemical shifts. If the formation of a hydrogen bond to one of the amino protons gives an effect to the chemical shift of the other amino proton which is free from hydrogen bonding, for example, by the change in charge distribution and/or lone pair electron shielding, the $\Delta_{A_1U}{}^{anti}$ and $\Delta_{A_7U}{}^{syn}$ terms cannot be neglected. However, we do not know such values, and, thus, we use the normalized $P_H{}^N$ and $P_W{}^N$ that were obtained by dividing the apparent P_H and P_W by the sum. The calculated populations of the Watson–Crick-type dimer are almost constant (25–30%) for the mixture systems of the several uracil derivatives with A. On the other hand, in the mixture systems of U with several A derivatives the population of the Watson–Crick-type or the Hoosteen-type dimer depends on the structure of A, predominant formation of the Watson–Crick type dimer for Br^8A,

Table IV: Estimation of Intrinsic Association Constants for Each Proton-Acceptor Site of Complexes between 1-Cyclohexyluracil and 9-Ethyladenine Derivatives (M^{-1})

compounds	$K_{AU}{}^{W}$	$K_{AU}{}^{H}$	temp (°C)
A + U	420	980	−56
Br^8A + U	450	30	−56
Cl^2A + U	110	1500	−56
Me^2A + U	4200	96	−65
$DiMe^2A$ + U	160	1600	−65

Me^2A, and NH_2^2P while predominant formation of the Hoogsteen-type dimer for Cl^2A and $DiMe^2A$. In addition to them, both types of dimer are almost equally produced in the self-association of all the A derivatives. We can consider two factors that induce this difference in the population of the two types of dimer, the intrinsic strength of a hydrogen bond and steric hindrance. In the cases of Cl^2A, Br^8A, and $DiMe^2A$, steric hindrance of substituted groups, chlorine and bromine atoms and the dimethylamino group, predominantly determines the structure of the AU dimer. The steric hindrance between the carbonyl group of the uracil derivatives and the atoms or groups at C(2) of A seems to disturb the Watson–Crick-type dimer formation. Since the van der Waals radius of the sulfur atom (1.85 Å) is larger than the oxygen atom (1.40 Å), the steric hindrance of thiouracil is more critical. Indeed, the population of the Hoogsteen-type dimer increases in the dimers of the S^2U and S^4U with A. The steric hindrance of Br^8A seems to be more critical than that of Cl^2A, since the bromine atom (1.95 Å) has a larger van der Waals radius than the chlorine atom (1.80 Å). However, Br^8A uses the Hoogsteen site in the hydrogen bond for the formation of AUU trimer while there is no AUU trimer in the mixture of Cl^2A and U. The steric hindrance between the free carbonyl group of U and the atom or the group at the C(2) position of A is critical for the stability of the AU dimer, but the hindrance between the carbonyl group of U and the group at the C(8) position of A seems not so serious.

The association constant of the AU dimer is the sum of those of the Hoogsteen-type and Watson–Crick-type dimers. The microassociation constants calculated by eq 10 are given in Table IV. The value of $K_{AU}{}^{W}$ is almost the same (ca. 500 M^{-1}) for A and Br^8A where the steric hindrance in the Watson–Crick site does not exist. The results seem to show that the Watson–Crick site has an intrinsic affinity for dimer formation, and it is hardly influenced by substituents that do not produce direct steric hindrance. A similar tendency is seen for the $K_{AU}{}^{H}$ of A and Cl^2A. It is found that the intrinsic association constant of the Hoogsteen-type base pair is about twice that of the Watson–Crick type at −56 °C.

Now we will estimate all the populations of the four possible structures of the AU dimers in solution. If each proton-acceptor site in A and U has affinity for hydrogen-bond formation independent of the donor group, each population of the four structures can be estimated by multiplying the intrinsic affinity of each proton-acceptor site. There is evidence that supports the above assumption in the following experimental results obtained previously. The C(4)-carbonyl group of U is used with almost the same frequency for both of the mixture systems with A (58%) (Iwahashi & Kyogoku, 1977) and NH_2^2P (present data) that form predominantly the Hoogsteen-type dimer and that Watson–Crick one, respectively. Also on the proton-acceptor sites of A, a similar tendency was seen, i.e., the Hoogsteen site of A is employed in the hydrogen-bond formation to almost the same extent (70%) with any uracil derivatives, U, T, Br^5U, DhU, S^4U, and S^2U. Although the experiments were done at a low temperature (−56 °C), a ^{15}N NMR experiment at room temperature agrees with the above result (Watanabe et al., 1981). Therefore, combination of the two results allows us to calculate the populations of the four AU dimer structures: Watson–Crick type, 17%; Hoogsteen type, 41%; reversed Watson–Crick type, 13%; reversed Hoogsteen type, 29%.

In spite of the predominant formation of the Hoogsteen-type dimer at monomer level, the Watson–Crick-type hydrogen bond is formed at polymer level, i.e., in double-stranded deoxyribonucleic acid and ribonucleic acid. Factors other than the intrinsic affinity between the bases may influence the structure of base pairs. The stable backbone conformation in polynucleotide chains may be an important factor. However, it is also known that the Hoogsteen-type base pair is realized in the tertiary structure of transfer RNA^{Phe} (Kim et al., 1974; Robertus et al., 1974). Also the Hoogsteen-type proton-acceptor sites may be important in other biological processes, for example, the recognition of operators by repressors, the binding of RNA polymerase at promoters, and the distinguishing of DNA base sequences by restriction endonucleases. The present results will give us some information on the factors for stabilizing the Hoogsteen-type base pair in real biological systems.

References

Becker, E. D., Miles, H. T., & Bradley, R. B. (1965) *J. Am. Chem. Soc. 87*, 5575.

Blake, R. D., & Massoulie, J. R. (1967) *J. Mol. Biol. 30*, 291.

Borer, P. N., Dengler, B., Tinoco, I., Jr., & Uhlenbeck, O. C. (1974) *J. Mol. Biol. 86*, 843.

Brahms, J. (1965) *J. Mol. Biol. 11*, 789.

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

Gralla, J., & Crothers, D. M. (1973) *J. Mol. Biol. 78*, 301.

Hamlin, R. M., Jr., Lord, R. C., & Rich, A. (1965) *Science (Washington, D.C.) 148*, 1734.

Hoogsteen, K. (1963) *Acta Crystallogr. 16*, 907.

Iwahashi, H., & Kyogoku, Y. (1977) *J. Am. Chem. Soc. 99*, 7761.

Katz, L. (1969) *J. Mol. Biol. 44*, 279.

Katz, L., & Penman, S. (1966) *J. Mol. Biol. 15*, 220.

Kim, S. H., Suddath, F. L., Quigley, G. J., MacPerson, A., Wang, J. L., Seeman, N. C., & Rich, A. (1974) *Science (Washington, D.C.) 185*, 435.

Küchler, E., & Derkosch, J. (1966) *Z. Naturforsch. B: Anorg. Chem., Org. Chem., Biochem., Biophys., Biol. 21B*, 209.

Kyogoku, Y., Lord, R. C., & Rich, A. (1966) *Science (Washington, D.C.) 154*, 518.

Kyogoku, Y., Lord, R. C., & Rich, A. (1967a) *J. Am. Chem. Soc. 89*, 496.

Kyogoku, Y., Lord, R. C., & Rich, A. (1967b) *Proc. Natl. Acad. Sci. U.S.A. 57*, 250.

Kyogoku, Y., Lord, R. C., & Rich, A. (1969) *Biochim. Biophys. Acta 179*, 10.

Mazza, F., & Sobell, H. M. (1969) *J. Mol. Biol. 43*, 407.

Miller, J. H., & Sobell, H. M. (1967) *J. Mol. Biol. 24*, 345.

Morishima, T., Inubushi, T., Yonezawa, T., & Kyogoku, Y. (1977) *J. Am. Chem. Soc. 99*, 4299.

Raszka, M., & Kaplan, N. O. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 2025.

Robertus, J. D., Lander, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C., & Klug, A. (1974) *Nature (London) 250*, 546.

Shoup, R. R., Miles, H. T., Becker, E. D. (1966) *Biochem. Biophys. Res. Commun. 23*, 194.

Stevens, C. L., & Felsenfeld, G. (1964) *Biopolymers 2*, 293.

Sugeta, H. (1982) *Bull. Chem. Soc. Jpn.* (in press).
Tavale, S. S., Sakore, T. D., & Sobell, H. H. (1969) *J. Mol. Biol. 43*, 375.
Watanabe, M., Sugeta, H., Iwahashi, H., Kyogoku, Y., & Kainosho, M. (1981) *Eur. J. Biochem. 117*, 553.
Watson, J. D., & Crick, F. H. C. (1953) *Nature (London) 171*, 737.
Young, M. A., & Krugh, T. R. (1975) *Biochemistry 14*, 4841.

Primary Structure of Monkey Osteocalcin†

Peter V. Hauschka,* Steven A. Carr,‡ and K. Biemann

ABSTRACT: The complete 49-residue amino acid sequence of osteocalcin from the old world monkey *Macaca fascicularis* has been determined by efficient combination of gas chromatography–mass spectrometry and Edman techniques. This vitamin K dependent protein of bone matrix contains three γ-carboxyglutamic acid residues at positions 17, 21, and 24, as well as a disulfide-bonded loop (23–29). Features of the sequence which apparently are required for the binding of Ca^{2+} have been strongly conserved throughout evolution.

Osteocalcin[1] is an abundant Ca^{2+}-binding protein indigenous to the organic matrix of bone, dentin, and possibly other mineralized tissues (Hauschka et al., 1975; Hauschka & Gallop, 1977; Gallop et al., 1980). This protein contains 47–51 amino acid residues (M_r 5200–5900) depending on the species (Price et al., 1979; Poser et al., 1980; Linde et al., 1980; Carr et al., 1981b). Osteocalcin is distinguished by its content of three γ-carboxyglutamic acid (Gla)[2] residues. Gla appears in the protein as a result of posttranslational, vitamin K dependent carboxylation (Lian & Friedman, 1978) of specific glutamic acid residues at sequence positions 17, 21, and 24 (Price et al., 1979; Poser et al., 1980; Carr et al., 1981b). Biosynthesis and vitamin K dependent carboxylation of osteocalcin occur in bone tissue in organ culture (Lian & Heroux, 1979) and in isolated bone cells in tissue culture (Nishimoto & Price, 1980). Osteocalcin appears coincident with the onset of calcium phosphate mineral deposition in developing embryonic bone and is probably a specific product of cells differentiated with respect to bone formation (Hauschka & Reid, 1978; Hauschka, 1979b). While the function of osteocalcin is not known, this protein exhibits several interesting properties: (1) specific Gla-dependent binding of Ca^{2+} ions (Hauschka & Gallop, 1977; Poser & Price, 1979); (2) Ca^{2+}-induced transition to the α-helical conformation (Hauschka, 1981); (3) tight association with hydroxylapatite and inhibition of brushite → hydroxylapatite metamorphosis in vitro (Hauschka et al., 1975; Hauschka & Gallop, 1977); (4) low-level circulation in blood plasma (Price & Nishimoto, 1980); (5) 1,25-$(OH)_2$-vitamin D_3 stimulation of its biosynthesis (Price & Baukol, 1981); and (6) derivation from higher molecular weight precursor proteins (Lian & Heroux, 1979; Hauschka, 1979b; Nishimoto & Price, 1980). Among osteocalcins of different species, the homology of the Gla positions and the disulfide-bonded loop argues for the functional importance of these conserved regions of the protein. However, numerous structural questions remain, as well as delineation of the calcium binding sites, the role of hydroxyproline in position 9, and the apparent lack of requirement for Gla_{17} in human osteocalcin (Poser et al., 1980).

The protein sequencing technique of gas chromatography–mass spectrometry (GC–MS), recently reviewed by Biemann (1980), has proved advantageous for Gla-containing proteins. Decarboxylation of the intact protein in DCl–D_2O readily converts Gla to γ,γ-dideuterioglutamic acid (Carr, 1980; Hauschka, 1979a; Carr & Biemann, 1980; Carr et al., 1981b), a chemically stable and mass spectrometrically unique derivative of Gla. Following partial enzymatic or acidic hydrolysis and derivatization, the complex mixtures are analyzed (without isolation of individual peptides) by GC–MS (Biemann, 1980; Carr et al., 1981a). Peptide fragments in which Gla residues were present show sequence ions in their mass spectra corresponding to those of glutamic acid, but shifted upward by 2 mass units per Gla (Carr, 1980; Carr & Biemann, 1980; Carr et al., 1981b). Incomplete posttranslational carboxylation at Gla positions may also be determined precisely from the relative abundance of characteristic sequence ions. Chicken bone osteocalcin was the first Gla-containing protein to be sequenced by this technique (Carr, 1980; Carr et al., 1981b), while cow, swordfish, and human osteocalcins have been elucidated by standard Edman methods (Price et al., 1979; Poser et al., 1980). Sequential Edman degradation is complementary to GC–MS because the former provides information on long-range order, amide assignments, His positions, and differentiation of Ile from Leu. This paper presents the complete structure of monkey osteocalcin as determined by GC–MS and Edman methodology.

Materials and Methods

Sacrifice of young adult male cynomolgus monkeys (*Macaca fascicularis*) by sodium pentothal injection was followed by

† From the Department of Orthopaedic Surgery, Children's Hospital Medical Center, and the Department of Oral Biology, Harvard School of Dental Medicine, Boston, Massachusetts 02115 (P.V.H.), and the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 (S.A.C. and K.B.). *Received August 20, 1981.* Supported by the March of Dimes Birth Defects Foundation (1-533) and National Institutes of Health Grants AM 15671 and AM 16754 to the Children's Hospital Medical Center and GM 05472 and RR 00317 to the Massachusetts Institute of Technology.

‡ Present address: Department of Biological Chemistry, Harvard Medical School, Boston, MA 02115.

[1] Osteocalcin has also been referred to as "bone Gla protein" or "BGP" in the literature (Nishimoto & Price, 1980).

[2] Abbreviations: Gla, γ-carboxyglutamic acid; CmCys, *S*-carboxymethylcysteine; Hyp, 4-hydroxyproline; PTH, phenylthiohydantoin; GC–MS, gas chromatography–mass spectrometry; HPLC, high-performance liquid chromatography; xLeu, Leu or Ile from GC–MS data.

0006-2960/82/0421-0638$01.25/0