Abstract: Cytosine and guanine have been shown previously to form Watson-Crick type base pairs in nonaqueous solvents, suggesting that the monomers can be used to understand and possibly to predict structures of the polymeric nucleic acids. Yet, the poor solubility properties of cytosine and guanine (and their corresponding nucleosides) have limited the utility of the monomeric model of polymeric nucleic acids. The 2'-deoxynucleosides, which are substituted at both ribose hydroxyls with triisopropylsilyl groups, have high solubilities (greater than 200 mM) in nonpolar solvents such as chloroform-d. These bases could be used as a monomeric model of polymeric nucleic acids. The 2'-deoxynucleosides can be used to understand and possibly to predict structures of the polymeric nucleic acids. A wide temperature range.

Dimers, Trimers, and Tetramers of Cytosine with Guanine

Nidhi Gupta Williams,† Loren Dean Williams,‡ and Barbara Ramsay Shaw*†

Contribution from Paul M. Gross Chemical Laboratory, Duke University, Durham, North Carolina 27706. Received December 27, 1988

Much of biological specificity is derived from hydrogen bonding. In nucleic acids, selective hydrogen bonds between complementary bases direct the fidelity of replication/transcription processes and also stabilize secondary and tertiary structures. Forces that stabilize double-stranded nucleic acids can be conceptually and experimentally decomposed into base-base hydrogen-bonding interactions (horizontal) and stacking interactions (vertical). The two types of interaction can be modeled by monomers in the appropriate environments. In relatively low dielectric solvents such as chloroform or even in dimethyl sulfoxide, the bases primarily form hydrogen bonds, while in aqueous solution the bases stabilize secondary and tertiary structures. Thus, the interactions of monomers are consistent with formation of two-, three-, and four-stranded nucleic acid polymers.
Another structure, the Watson–Crick base pair (Figure 1A), can form spontaneously between monomeric guanine and cytosine. This C:G dimer, stabilized by three hydrogen bonds, forms between monomers in nonaqueous solvents. However, the poor solubility properties of cytosine and guanine (and their corresponding nucleosides) limited the usefulness of this monomeric model of DNA.

We have recently described10,11 a well-behaved model system for studying the hydrogen-bonding interactions of these important biomolecules. Improved solubility of cytosine and guanine in chloroform-d and other low-dielectric solvents is achieved by addition of lipophilic trisopropylsilyl groups to both hydroxyls of the 2'-deoxynucleosides (Figure 2). Hence, hydrogen bonding of cytosine with guanine can be investigated over a wide temperature range up to relatively high concentrations (greater than 200 mM). In this DNA model system, 1H NMR shows cytosine and guanine form a stable Watson–Crick type dimer, as would be expected from previous studies of bases in higher dielectric solvents or at lower concentrations. We report here that the bases form such dimers and additional, more intricate hydrogen-bonded complexes shown in Figure 1B–D. Cytosine and guanine monomers form both trimers (cytosine:guanine) and tetramers (cytosine:guanine) in low-dielectric solution.

**Experimental Section**

The 2'-deoxy-3',5'-bis(trisopropylsilyl) derivatives of guanosine and cytidine were prepared and purified as described elsewhere.10,11 Chloroform-d (Stohler) was distilled over P2O5 under argon just prior to sample preparation. NMR samples, in 5-mm tubes, were prepared under anaerobic conditions under argon. The solutions were degassed by bubbling argon or by at least three freeze/pump/thaw cycles. Spectra were obtained on a Varian XL-300 spectrometer using 16K double-precision (32-bit) data points over a 5000-Hz spectral width. In the nuclear Overhauser experiments, for which an array of decoder offsets was used, spectra were obtained in the interleave mode. For all other spectra, the decoupler was used to suppress the large triisopropyl peak at approximately 1.0 ppm. Deuteriochloroform was used as a lock, and the trace nondeuterated chloroform contaminant was used as a reference. Ethylene glycol was used for high-temperature calibration of the probe, and methanol was used for low-temperature calibration. Probe temperatures were reproducible to ±1°C.

**Results**

We have previously reported the preparation and utility of 3',5'-bis(trisopropylsilyl) derivatives of 2'-deoxycytidine and 2'-deoxyguanosine (Figure 2) for characterizing the thermodynamics of the C:G base pair with isoperibolic titration calorimetry.18 The same nucleoside derivatives were used in this spectroscopic study. The lipophilic trisopropylsilyl groups prevent interference of hydrogen bonding by ribose hydroxyls. These nucleoside derivatives provide a sensitive system for studying hydrogen bonding exclusively between the bases. For simplicity of description, the 3',5'-bis(trisopropylsilyl) derivatives of 2'-deoxyguanosine and 2'-deoxycytidine shall be referred to as G and C, respectively.

The nuclear Overhauser effect (NOE) was used here for structural elucidation and to aid in resonance assignment. Initially, we were somewhat surprised that in this model system the NOE's were exclusive. Negative NOE's were observed between hydrogens (such as between H1' and H8 of G, not shown) where exchange, scalar coupling, and the intermediacy of a third spin can be excluded as causes for a decrease in signal intensity. The negative sign of the NOE's indicates that these mononucleosides are not in the extreme narrowing limit (where T2D ≪ 1) but instead are behaving like much larger molecules. As described below, the formation of hydrogen-bonded and possibly stacked complexes explains the negative sign of the observed NOE's. In these experiments the decrease in N–H resonance intensities caused by selective saturation of other N-H resonances can result from two

---

Cytosine-Guanine Multimers

of Hb of C. The resonance marked with an asterisk at 2'-deoxy-3',5'-bis(triisopropylsilyl)guanosine

Figure 3. Spectra: (A) partial 300-MHz 1H NMR spectrum of 30 mM 2'-deoxy-3',5'-bis(triisopropylsilyl)guanosine plus 30 mM 2'-deoxy-

of all five N-H resonances of the base pair allows of both C and G are in slow rotation on the NMR time scale. As

expected from previous studies in dimethyl In

that hydrogen-bonded structures in addition to the C:G base pair form readily in mixtures of C

and G in chloroform-d. In 1:l mixtures of C and G (30 mM C plus 30 mM G), the

Hb proton of C appears to be involved in hydrogen bonding. As the temperature decreases from +16 to -69 °C, this resonance

moves downfield by over 350 Hz (over 1.2 ppm at 300 MHz; Figure 3A). Selective saturation of Hb (Figure 3D) caused NOEs to H2 and H1.

The difference in the resonance frequencies of H2 and H4 (over 1100 Hz at 300 MHz) is large considering that these two protons are covalently bound to the same nitrogen. The chemical environments of these two protons are radically different due to hydro- 

drogen bonding of H2 to O2 of C and steric protection of H4 by the substituted ribose moieties, which form an equivalent of the DNA minor groove.

The NOE data shown here confirm both the formation of the C:G dimer in chloroform-d and the resonance assignments of Figure 3A. Inter- and intramolecular NOE's were observed between protons expected to be located nearby in space in the base pair. However, as described in the following section, we have observed convincing evidence that hydrogen-bonded structures in addition to the C:G base pair form readily in mixtures of C and G in chloroform-d.

2. Tetramers. The complementary arrangement of hydrogen bond donors (Hb of C) and acceptors (N7 of G) suggests that pairs of Watson–Crick type dimers can associate to form tetraceramers (Figure 1B). Such hydrogen-bonded tetracerams have been observed previously in crystals10 obtained from 1:1 mixtures of G and C. Here, we report chemical shift evidence that C and G can also form tetracerams in chloroform-d solution.

In 1:1 mixtures of C and G (30 mM C plus 30 mM G), the Hb proton of C appears to be involved in hydrogen bonding. As the temperature decreases from +16 to -69 °C, this resonance moves downfield by over 350 Hz (over 1.2 ppm at 300 MHz; Figure 4). The temperature dependence of the Hb resonance frequency is consistent with increasing tetramer formation as temperature decreases. Similarly, the concentration dependence of Hb resonance frequency is consistent with increasing tetramer formation as the total concentration increases (data not shown). The involvement of the Hb proton in hydrogen bonding supports formation of (C:G)2 tetramer. This phenomenon cannot be explained by dimer formation alone.

3. Trimers. The complementary arrangement of hydrogen bond donors (H1 and H2 of G) and acceptors (O4' and N7 of G) suggests that a second G can associate with a Watson–Crick C:G base pair to form a trimer of C:GtG (tG refers to the guanine in the extra-Watson–Crick orientation). The two possible conformations of C:GtG are shown in parts C and D in Figure 1. Formation of either C:GtG conformation would be favored in 1:2 mixtures of C and G. Trimer formation would exclude (C:G)2 tetramer formation (Figure 1B). Hence, certain conditions should promote tetramer formation (i.e., 1:1 mixtures, above), and certain con-
on the NMR time scale such that $H_1$ and $H_4$ are distinct resonances of $G$, whereas tetramer formation is temperature dependent (variable chemical shifts of $G$). However, below $-26 \, ^\circ C$, the amino bond of $G$ is in fast rotation on the NMR time scale and a single resonance (of integral 1) was not observed. Above $9 \, ^\circ C$, at least one of the amino bonds (of $G$ or $tG$) is in fast rotation on the NMR time scale and a single resonance (of integral 4, Figure 5) is observed for $H_2$ and $H_3$. The single amino peak in this temperature realm is the resonance of four rapidly exchanging protons: $H_2$ and $H_3$ of $G$; $H_2$ and $H_3$ of $tG$. However, below $-45 \, ^\circ C$, both amino groups of $G$ and $tG$ appear to be in slow rotation on the NMR time scale such that $H_1$ and $H_4$ are distinct resonances even though $G$ to $tG$ exchange is fast. In this slow-rotation realm, the observed resonance frequency of $H_4$ (Figure 5) should be the average of the frequencies of $H_1$ of $G$ and of $H_2$ of $tG$. The frequency of $H_4$ (Figure 5) is also an average.

The two hydrogen bond acceptors for $H_1$ in $C:G:tG$ are chemically similar. In 1:1 mixtures, essentially 100% of $H_1$ protons will bond with $O^\circ$ (of $C$) hydrogen bond acceptors. In 1:2 mixtures, only 50% of the $H_1$ protons will bond to $O^\circ$. The remaining 50% (i.e., the $H_2$ protons of $tG$) bond to either $O^\circ$ or $N^7$ of $G$. However, the three limiting frequencies (of the coalesced amino group, $H_2$, and $H_3$) show striking similarities in 1:1 compared to 1:2 mixtures (Figure 5). Although in C:G:tG the $tG$ amino resonances cannot be observed independently of those of $G$, we can infer that the resonance frequency of $H_2$ is nearly the same in $G$ as in $tG$ and the frequency of $H_3$ is also nearly the same in $G$ and as in $tG$. This analysis suggests that the chemical environment of the amino group of $tG$ is very similar to that of $G$. In Figure 1C, both $H_1$'s form hydrogen bonds to a carbonyl oxygen ($O^\circ$ of $C$ and $O^\circ$ of $G$). Further, both $N^7$'s form hydrogen bonds to basic nitrogen and both $H_3$'s are free from hydrogen-bonding interactions. The resonance frequencies of the amino protons of $G$ are most consistent with formation of $C:G:tG$ as shown in Figure 1C.

In addition to the amino exchange process, in 1:2 mixtures the $H_1$ proton of $G$ exchanges between two sites as $G$ exchanges with $tG$. This exchange is fast on the NMR time scale above $-22 \, ^\circ C$ and is slow on the NMR time scale below $-40 \, ^\circ C$ (Figure 6). Thus, in 1:2 mixtures above $-22 \, ^\circ C$, the observed frequency of the $H_1$ resonance is the average of the frequencies of $H_1$ of $G$ and $H_1$ of $tG$. However, below $-40 \, ^\circ C$, the frequency of $H_1$ at each of the two sites can be determined.

As shown in Figure 6, the downfield site resonates at a frequency near the frequency of $H_1$ in 1:1 mixtures. Thus, it is very likely that the downfield resonance in 1:2 mixtures is $H_1$ of $G$ while the upfield resonance is the $H_1$ of $tG$. The large difference in the frequency of $H_1$ of $G$ versus that of $tG$ (over 500 Hz) suggests some basic difference in chemical environment of the two $H_1$ protons of $G:G:tG$. As can be seen in parts C and D in Figure 1, one significant difference between $G$ and $tG$ (in either conformation of $C:G:tG$) is the absence of the hydrogen bond to the $O^\circ$ of $tG$.

Additional support for formation of $C:G:tG$ in solution has been provided by our previous calorimetric studies, indicating that heats of dimerization when $[G] > [C]$ were not consistent with heats of dimerization when $[G] < [C]$, and by $^1H$ NMR continuous variation plots, suggesting formation of $C_{2n}G_{2n}$ complexes, although there the stoichiometry of interaction was am-
Cytosine-Guanine Multimers


Figure 7. Plot of frequency versus temperature of the H\textsuperscript{3} resonances of G in 1:1 mixtures (hollow circles) and 1:2 mixtures (full circles of C and G).

ambiguous. It is not possible to definitively determine whether one conformation (Figure 1C,D) is favored over the other. However, the data presented here support formation of C:G:tG as shown in Figure 1C.

4. Stacking. Vertical stacking interactions predominate in water,\textsuperscript{14} and to our knowledge, there have been no previous reports that nucleosides stack in nonaqueous solvents. However, in the experiments described here, some evidence for stacking at low temperatures in chloroform-d\textsuperscript{2} was observed. In 1:1 mixtures C and G, the H\textsuperscript{3} proton shows a biphasic change in frequency with temperature (Figure 7). In the lower temperature realm (below -46 °C), the H\textsuperscript{3} moves to lower frequency as temperature decreases. This observation is in analogy with the extensive shielding of H\textsuperscript{3} observed previously from stacking in aqueous solution.\textsuperscript{16} The shielding of H\textsuperscript{3} observed here may suggest a slight amount of stacking of the bases in chloroform-d.

Discussion

The Watson-Crick base pairs remain a most intriguing example of hydrogen bonding. Recognition of intact base pairs by external agents is thought to involve hydrogen bonding at non-Watson-Crick sites that are accessible from the major groove of double-stranded DNA.\textsuperscript{21} Thus, proteins can recognize DNA sequence by specific hydrogen bonding of amino acid side chains to sites in the major groove of DNA.\textsuperscript{22} Single- and possibly double-stranded nucleic acids also form hydrogen bonds to the same groove, resulting in sequence specific three- and four-stranded structures.\textsuperscript{23-27}

The interactions stabilizing such large and complex biological structures have an accessible physical and chemical basis. Indeed, we have observed three types of hydrogen-bonded complexes between monomers in chloroform-d solution: (a) the previously observed dimer, the C:G Watson-Crick base pair as shown in Figure 1,\textsuperscript{1-10} (b) a tetramer (formed by two base pairs) stabilized by two additional hydrogen bonds in a 2-fold symmetric pair of C:G base pairs (shown in Figure 1B), and (c) a trimer, (C:G:tG) stabilized by two hydrogen bonds between a G which is in a Watson-Crick base pair and a second G in a Hoogsteen-like orientation (shown in Figure 1C,D). Our results suggest that the conformation of Figure 1C may be favored over that of Figure 1D.

In the simple model system described here, hydrogen bonds are observed (to sites on Watson-Crick base pairs) that would stabilize specific three- and four-stranded nucleic acid complexes. The trimer (C:G:tG) observed in chloroform-d is analogous to that proposed for the triple-stranded rC:rG:rG\textsuperscript{25} and observed crystallographically in yeast phenylalanine tRNA.\textsuperscript{18} The tetramer has been proposed previously to explain DNA-DNA recognition processes.\textsuperscript{22,28} With the model system described here, the behavior of monomers can thus be used to understand, and hopefully to predict, that of biopolymers.

The patterns of hydrogen bonding in dimers, trimers, and tetramers formed between nucleic acid monomers in solution here appear to correspond with patterns of hydrogen bonding in nucleic acid polymers. This correspondence of monomeric interactions with those in polymers underscores the importance of hydrogen bonding in the specificity and stability of a variety of DNA secondary structures.

Acknowledgment. We thank Drs. Shelley Copley and Judith A. Swan for helpful discussions. This research was supported by American Cancer Society Grants PF-3062 to L.D.W. and NP-486 and BC-617 to B.R.S.


