Effects of Base Sequence on the Loop Folding in DNA Hairpins†

M. J. J. Blommers,‡ J. A. L. I. Walters,‡ C. A. G. Haasnoot,‡ J. M. A. Aelen,‡ G. A. van der Marel,§ J. H. van Boom,§ and C. W. Hilbers*,‡

Laboratory of Biophysical Chemistry, Faculty of Science, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands, and Gorlaeus Laboratories, State University Leiden, P.O. Box 9502, 2300 RA Leiden, The Netherlands

Received February 21, 1989; Revised Manuscript Received May 18, 1989

ABSTRACT: High-resolution NMR and UV–melting experiments have been used to study the hairpin formation of partly self-complementary DNA fragments in an attempt to derive rules that describe the folding in these molecules. Earlier experiments on the hexadecanucleotide d(ATCCTA-TTTT-TAGGAT) had indicated that within the loop of four thymidines a wobble T-T pair is formed (Blommers et al., 1987). In the present paper it is shown that if the first and the last thymines of the intervening sequence are replaced by complementary bases, sometimes base pairs can be formed. Thus for the intervening sequences -CTTG- and -TTTA- with the pyrimidine in the 5′-position and the purine in the 3′-position, a base pair is formed leading to a loop consisting of two residues. For the intervening sequences -GTTC- and -ATTT- with the purine in the 5′-position and the pyrimidine in the 3′-position, this turns out not to be the case. It was found that it made no difference when the four-membered sequence was closed by a G-C base pair or an A-T base pair. Replacement of the two central thymidine residues by the more bulky adenine residues limits the hairpin to a four-membered loop scheme. Very surprisingly, it was found from 2D NOE experiments that the T-A base pair, formed in the loop consisting of the -TTTA- sequence, is a Hoogsteen pair. It is argued that the pairing of the bases in this scheme may facilitate the formation of a loop of two residues, since the distance of the C1′ atoms in this base pair is 8.6 Å instead of 10.4 Å found in the canonical Watson–Crick base pair. Combination of the data obtained for the series of DNA fragments studied shows that the results can be explained by a simple, earlier proposed, loop folding principle which assumes that the folding of the four-membered loop is dictated by the stacking of the double-helical stem of the hairpin.

Extensive structural studies performed during the last decade (Sänger, 1984; Watson et al., 1987) have demonstrated that the conformations that can be adopted by nucleic acids are much more varied, complicated, and interesting than the regular double helix originally envisioned. It is often believed that the conformational space available to nucleic acids will somehow be important in cellular processes, and therefore, attempts have been made to explain the conformational behavior on the basis of nucleotide sequences. Among the polymorphic conformations, hairpin loop structures play a special role, particularly in RNA molecules although it has been found for DNA that sequences capable of hairpin formation are frequently occurring near regulation and promotion sites.

We have found that the homologous partly self-complementary DNA fragments d(ATCCTA-T_n-TAGGAT) (in which n varies from 0 through 7) form hairpins of which the stability is at a maximum when the loop of the hairpin comprises four or five nucleotides (Haasnoot et al., 1983a; Hilbers et al., 1985). This was in contrast to earlier studies of RNA hairpins for which optimal stability was found when there are six to seven nucleotides in the loop (Uhlenbeck et al., 1973; Gralla & Crothers, 1973a). This difference in RNA vs DNA behavior was rationalized (Haasnoot et al., 1985, 1986) in terms of structural models for the architecture of RNA and DNA hairpin loops. These models yield (Haasnoot et al., 1986) a simple framework in which the folding (i.e., structure) of the loop is dictated by the stacking of the bases in the double-helical stem of the hairpin; it is the base stacking pattern that allows for shorter "optimum" loops in B-DNA when compared to A-RNA.

For DNA hairpins our loop folding principle states (Haasnoot et al., 1985, 1986) a continued stacking of nucleotides in the loop section on top of the 3′-end of the B-type double-helical stem. The model predicts an optimum loop length of ca. four nucleotides. Notwithstanding its success in explaining the data availabe at that time, it was predicted (Haasnoot et al., 1986, 1987), however, that the *details* of the hairpin loop folding will depend on, e.g., base sequence in the loop and perhaps even on the base sequence in the stem. Since then, Orbons et al. (1986, 1987) discovered that the octanucleotide d(C*GC*GTGC*G), in which the starred cytidines are methylated at C5, is able to fold into a "mini-hairpin" structure in which the central -GT- part forms a loop consisting of only two nucleotides. At the same time it was assessed in our laboratory (Blommers et al., 1987; Hilbers et al., 1987) that the two terminal thymidines of the central four-membered T-stretch in the hairpin formed by d(ATCCTA-TTTT-TAGGAT) are involved in T-T wobble pairing. The base stacking pattern in the loops of these two hairpins is very similar and fulfills the basic criterion in the loop folding model (i.e., propagation of the stack from the 3′-end of the stem into the loop). Therefore, these findings can be interpreted (Haasnoot et al., 1987) in terms of a local optimization of the hairpin structure within the conformational space delineated by our loop folding principle.

In this paper we extend our investigations to obtain more insight regarding the loop architecture of DNA hairpins. In particular, we put to the test the prediction made (Haasnoot et al., 1987) on the basis of molecular mechanics calculations that -PyTTPu-[1] loop sequences can enhance the hairpin sta-

† Dedicated to Professor C. MacLean on the occasion of his retirement from the chair of Molecular Spectroscopy at the Free University in Amsterdam.

‡ University of Nijmegen.

§ State University Leiden.

0006-2960/89/0428-7491$01.50/0

bility by base pairing within the loop whereas -PuTTPy- loop sequences cannot.

Materials and Methods

The DNA fragments employed in this study were synthesized according to the phosphate triester method (van Boom et al., 1983). Samples for ultraviolet–melting studies and ^{1}H NMR spectroscopy were prepared by dissolving the oligonucleotides in 5 mM sodium phosphate buffer with 1 mM sodium cacodylate and 0.1 mM EDTA, pH 7.0; NaCl was added to give the desired sodium ion concentration. For samples used for NMR spectroscopy, the pH was adjusted to 6.0 in order to diminish proton exchange between the oligonucleotide and water.

UV–melting experiments were carried out at 260 nm on a Cary 118C spectrophotometer, in a constant-temperature cell (Hellma). The temperature was controlled by circulating water from a cryothermostat (Mettler, WKS) first through a thermostated cell holder and then through the cell. The heating rate was 1 °C/min. In our earlier studies (Tibanyenda et al., 1984; Haasnoot et al., 1985) transition enthalpies were calculated from the differential melting curves, according to Gralla and Crothers (1973b). In the present study we fitted the absorbance vs temperature curves with a nonlinear least-squares program, using ΔH, ΔS, $E_{H,0}$, $E_{C,0}$, ΔE_H, and ΔE_C as adjustable parameters. ΔH denotes the transition enthalpy, and ΔS is the transition entropy. $E_{H,0}$ stands for the absorption of the helix structure at 0 °C, $E_{C,0}$ is the absorption of the coil structure at 0 °C, ΔE_H and ΔE_C denote the linear temperature dependency of the absorption of the helix structure and the coil structure, respectively. In all cases excellent fits to the experimental data were obtained. All samples were melted four times, and the standard deviation of the T_m and ΔH values amounted to less than 1 °C and 1 kcal/mol, respectively. The thermodynamic data derived in this manner are more accurate and more reproducible than those derived by the earlier method. The concentration of the samples used in the NMR as well as in the UV–melting experiments is given in the legends to the figures.

The 500- and 600-MHz ^{1}H NMR 1D and 2D spectra were recorded on Bruker AM-500 and AM-600 spectrometers, respectively, which were interfaced to an ASPECT 3000 computer. Water suppression techniques, applied to enable observation of imino proton resonances, were as described earlier (Haasnoot & Hilbers, 1983). The relative intensities of the imino proton resonances were determined by integration; the estimated accuracy is 20% for the individual resonances. Only for the spectrum of d(ATCCTA-ATTT-TAGGAT) was the accuracy somewhat less, due to overlapping of some resonances.

Phase-sensitive 600-MHz NOESY spectra of the D_2O samples were recorded with a spectral width of 6000 Hz, time proportional phase incrementation being used (Marion & Wüthrich, 1983). Two different spectra were acquired, with 4K points in the t_2 direction and 512 points zero filled to 1K points in the t_1 direction, with mixing times of 0.075 and 0.3 s, respectively. Phase-sensitive 600-MHz NOESY spectra of the H_2O samples were recorded with a time-shared long pulse as an observation pulse (Haasnoot & Hilbers, 1983). The spectral width was 23 800 Hz; the carrier was placed at the low-field end of the spectrum. The spectrum was acquired, with 4K points in the t_2 direction and 512 points zero filled to 1K points in the t_1 direction, with a mixing time of 0.3 s. Before Fourier transformation each free induction decay was subjected to data shift accumulation.

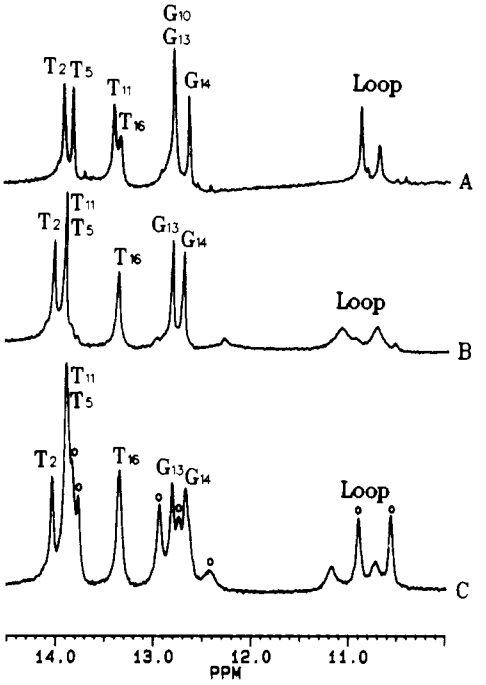

FIGURE 1: 500-MHz imino proton spectra of the hairpins d(ATCCTA-CTTG-TAGGAT) (A) and d(ATCCTA-GTTC-TAGGAT) (B) at 5 mM NaCl. The imino proton resonances between 15 and 12 ppm belong to base pairs and those near 11 ppm to unpaired residues in the loop region. The spectra were recorded at 274 K; the total strand concentration amounted to 1 mM DNA. The numbering of the resonances corresponds with the base numbering given in Figure 5. After the salt concentration was adjusted to 200 mM NaCl, spectrum B changes to a mixed hairpin/dimer spectrum (C). The resonances originating from dimer species are indicated with an open dot.

Results

The choice of the oligonucleotides, studied in the UV–melting and ^{1}H NMR experiments presented below, was strongly guided by the outcome of earlier molecular mechanics calculations (Haasnoot et al., 1987). These modeling studies suggested that in the hairpin with -PyTTPu- loop sequences the terminal Py/Pu bases might be involved in base pairing. In contrast, in a hairpin loop with a -PuTTPy- sequence no low-energy structure displaying the same feature could be found. Such a proposition can be checked by studying the imino proton spectra of the hairpins considered (Hilbers et al., 1987). The imino proton NMR spectra of d(ATCCTA-CTTG-TAGGAT) and d(ATCCTA-GTTC-TAGGAT) recorded in the presence of 5 mM NaCl are shown in spectra A and B of Figure 1, respectively. The assignments of the resonances given in this figure are based on a comparison of the spectra with that of d(ATCCTA-TTTT-TAGGAT), (Haasnoot et al., 1983a) (cf. Figure 3A) supplemented by one-dimensional Overhauser experiments and NMR–melting studies. Apart from the resonances near 11.0 ppm, which originate from the unpaired residues in the loop, the integrated peak intensities in spectrum A account for seven resonances, whereas in spectrum B six resonances are found (Figure 1). In agreement with UV–melting experiments (at NMR concentrations and at 0.2 M NaCl; to be discussed below) we observe for d(ATCCTA-GTTC-TAGGAT), the less stable hairpin, a mixture of hairpin and dimeric molecules when the ionic strength is raised (Figure 1C). The extra resonances in

[1] Abbreviations: NOE, nuclear Overhauser enhancement; Pu, purine; Py, pyrimidine; NOESY, nuclear Overhauser enhancement spectroscopy.

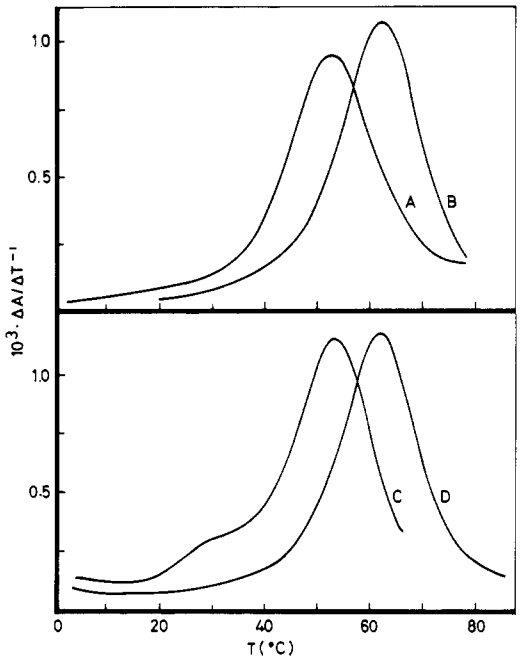

FIGURE 2: Differential UV-melting profiles measured for the hexadecanucleotides d(ATCCTA-CTTG-TAGGAT) and d(ATCCTA-GTTC-TAGGAT) (200 mM NaCl). In the upper drawing the melting of d(ATCCTA-CTTG-TAGGAT) (curve B) is compared with the melting of d(ATCCTA-GTTC-TAGGAT) (curve A) at a total strand concentration of 1 μM. In the lower drawing the same comparison is made for 1 mM total strand concentration (NMR concentration). At both conditions the hairpins formed by d(ATCCTA-CTTG-TAGGAT) (curves B and D) are more stable than the hairpins formed by d(ATCCTA-GTTC-TAGGAT) (curves A and C). Note that for the latter fragment an extra transition, which is attributed to the melting of a dimeric species, appears at high concentration (curve C).

that figure, which arise from the dimeric structure, are indicated with open dots. No such salt-dependent or concentration-dependent changes are observed for the more stable hairpin d(ATCCTA-CTTG-TAGGAT). Our experiments show that seven base pairs are present in the hairpin formed by d(ATCCTA-CTTG-TAGGAT), i.e., a loop consisting of only two nucleotides is apparent. On the other hand, the results also indicate that for the sequence d(ATCCTA-GTTC-TAGGAT) a loop of four residues is formed or that the stability or lifetime of the conceivable seventh G-C base pair is too small or too short, respectively, so that the corresponding imino proton resonance is not observed in the NMR spectrum. Of course, extra base pairing in the hairpin loop formed by d(ATCCTA-CTTG-TAGGAT) is expected to contribute to the overall stability of the folded form. UV-melting experiments were performed to investigate this aspect. In Figure 2 differential melting curves are presented for d(ATCCTA-CTTG-TAGGAT) (curve B) and for d(ATCCTA-GTTC-TAGGAT) (curve A) in the presence of 200 mM NaCl. The enhanced stability of d(ATCCTA-CTTG-TAGGAT) in comparison with that of d(ATCCTA-GTTC-TAGGAT) is obvious from the observed melting temperatures being 62 and 53 °C, respectively. These melting temperatures are independent of the oligomer concentration (vide infra), demonstrating that only monomolecular transitions are involved. The corresponding ΔH values are 46 and 38 kcal/mol, respectively. The difference of 8 kcal/mol between these two values is in reasonable agreement with the value expected for the stacking enthalpy of an extra C-G base pair on top of an A-T base pair (Breslauer et al., 1986). Since the concentration of the oligonucleotides, used in the UV-melting experiments, is much lower than that we used in the NMR experiments (3 μM vs 1 mM), we also investigated the UV-melting behavior at a much higher oligonucleotide concentration. The results are given in Figure 2 where differential melting curves are presented for d(ATCCTA-CTTG-TAGGAT) (curve D) and d(ATCCTA-GTTC-TAGGAT) (curve C) at 1 mM concentration. It is clear that for d(ATCCTA-CTTG-TAGGAT) the melting temperature is the same as that at the low concentration (curve B), showing that also under "NMR conditions" hairpin molecules are present. For d(ATCCTA-GTTC-TAGGAT) two transitions are observed of which one corresponds to that of curve A, indicating that this is the hairpin-to-coil transition. The transition occurring near 28 °C is concentration dependent, which shows that in this transition a duplex with an internal loop is melting out. These results correlate perfectly with the number of resonances in the NMR spectra obtained at similar conditions (cf. Figure 1C).

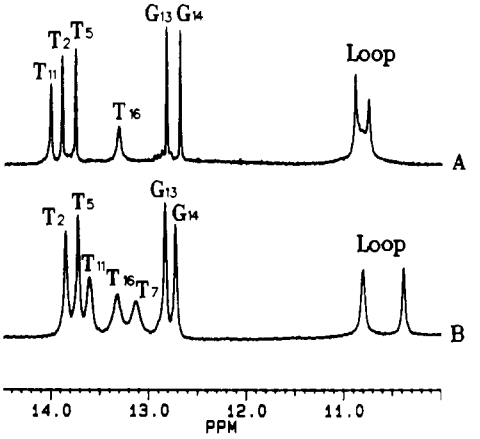

FIGURE 3: 500-MHz imino proton spectra of the hairpins formed by d(ATCCTA-TTTT-TAGGAT) (A) and d(ATCCTA-TTTA-TAGGAT) (B) at 200 mM NaCl. The imino proton resonances between 15 and 12 ppm belong to base pairs and those upfield from 11 ppm to unpaired thymidines in the loop region. The spectra were recorded at 298 K; the total strand concentration amounts to 1 mM DNA. The numbering of the resonances corresponds with that in Figure 5.

For the analogous molecules d(ATCCTA-TTTA-TAGGAT) and d(ATCCTA-ATTT-TAGGAT) similar experiments were performed. Again for the former molecule seven imino proton resonances were found (see Figure 3), where the imino proton spectrum is compared with that recorded for d(ATCCTA-TTTT-TAGGAT) showing six imino proton resonances at low field. The optical-melting experiments were conducted for solutions with 3 μM DNA and 200 mM NaCl. In analogy with the results discussed above, we find a higher melting temperature for the d(ATCCTA-TTTA-TAGGAT) sequence, i.e., 57 °C, than for the d(ATCCTA-ATTT-TAGGAT) sequence, i.e., 50 °C. The corresponding ΔH values are 45 and 39 kcal/mol, respectively. Again in analogy with the first two hexadecanucleotides, we find that only for d(ATCCTA-ATTT-TAGGAT) a dimeric species is formed at high oligomer and/or high salt concentrations. The results are summarized in Table I.

We have also considered the influence of the base pair sequence in the common stem region of the molecules. To this end three hexadecanucleotides were synthesized with the sequence d(ATTACG-XTTY-CGTAAT), in which the sequence of the intervening residues was -TTTT-, CTTG-, and -GTTC-, respectively. Again the same trends were observed as for the series of partly self-complementary molecules discussed above. This is demonstrated in Figure 4 where the imino proton spectra of the hairpins formed by d(ATTACG-

Table I: Melting Temperatures and Enthalpies of the Hairpin-to-Coil Transition of the Examined DNA Hairpins[a]

oligonucleotide	T_m (°C)	ΔH (kcal/mol)
d(ATCCTA-CTTG-TAGGAT)	62	46
d(ATCCTA-GTTC-TAGGAT	53	38
d(ATCCTA-TTTA-TAGGAT)	57	45
d(ATCCTA-ATTT-TAGGAT)	50	39
d(ATTACG-CTTG-CGTAAT	67	40[b]
d(ATTACG-GTTC-CGTAAT)	53	38
d(ATTACG-TTTT-CGTAAT)	52	37
d(ATCCTA-TAAA-TAGGAT)	49	36
d(ATCCTA-TTTT-TAGGAT)	53	41
d(ATCCTA-AAAA-TAGGAT)	43	27
d(ATCCTA-AAAC-TAGGAT)	47	35
d(ATCCTA-AACA-TAGGAT)	48	34
d(ATCCT-AAAT-AGGAT)	47	30
d(ATCCTA-TTTTTT-TAGGAT)	47	41
d(ATCCTA-AAAAAA-TAGGAT	38	24

[a] All data are obtained by means of a least-squares fit procedure of the UV absorbance–temperature curve (see Materials and Methods). [b] This value is somewhat less reliable than the other melting enthalpies.

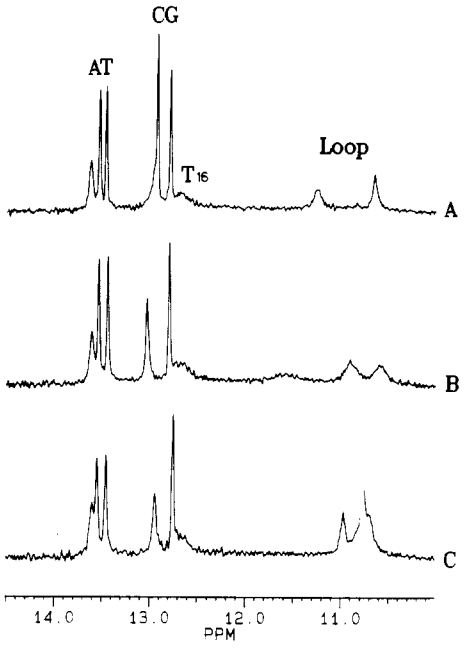

FIGURE 4: 500-MHz imino proton spectra of the hairpins formed by d(ATTACG-CTTG-CGTAAT) (A), d(ATTACG-GTTC-CGTAAT) (B), and d(ATTACG-TTTT-CGTAAT) (C) at 5 mM NaCl. The imino proton resonances between 15 and 12 ppm belong to base pairs and those near 11 ppm to unpaired thymidines in the loop region. The spectra were recorded at 298 K; the total strand concentration amounts to 1 mM DNA.

XTTY-CGTAAT) are shown. For the intervening sequence -CTTG- an extra imino proton resonance is observed at 12.9 ppm, demonstrating the formation of a seventh base pair in the stem. This is not found for the intervening sequence -GTTC-, but it is noted that a broad resonance is visible at 11.5 ppm for the unpaired (or partially paired) G residue. Also the difference in stability of these hairpins was as expected, i.e., the stability of the hairpin with the -CTTG- sequence was significantly higher than that of the other two hairpins (see Table I).

Since earlier NMR experiments (Haasnoot et al., 1983a; Blommers et al., 1987) and molecular mechanics calculations (Haasnoot et al., 1986, 1987) indicated that the loop folding in the hairpins considered might be rather tight, we investigated the influence of the introduction of more bulky bases in the loop sequence. Hairpins where adenines are introduced in the loop region, i.e., -AAAA-, -AAAC-, -AACA-, and -AAAAAA-, show a lowered stability with respect to hairpins where T residues are located at homologous positions (cf. Table I). A study of the hairpin formation by d(ATCCTA-TAAA-TAGGAT) shows that the imino proton spectrum of this molecule corresponds with the formation of a hairpin with six base pairs in contrast to that by d(ATCCTA-TTTA-TAGGAT) (vide supra); the imino proton resonance of T7 is observed in d(ATCCTA-TAAA-TAGGAT) as a broad line near 11 ppm (not shown). In line with this finding is the observation that the melting temperature drops from 57 °C for d(ATCCTA-TTTA-TAGGAT) to 49 °C for d-(ATCCTA-TAAA-TAGGAT).

The hairpin formed by d(ATCCTA-TTTA-TAGGAT) was further examined by means of 2D-NMR. Part of its NOESY spectrum is presented in Figure 5; in this region the H1′ and H8/H6 crosspeaks are found. This part of the spectrum was used to assign the ring proton resonances and the sugar H1′ resonances in the standard manner (Haasnoot et al., 1983b; Scheek et al., 1983; Hare et al., 1983). It can be seen that when one starts at the crosspeak generated by the 5′-terminal residue (A1), the sequential assignment can be continued to the H1′ and H6 resonances of residue T8. On the other hand, when one starts at the 3′-end (T16), the assignment can be continued to T9. On the basis of these assignments the J coupling patterns of the H1′ protons were analyzed. It was found for all of the sugars that the sum $J(1'2') + J(1'2'') >$ 13.3 Hz, which indicates that they predominantly adopt an S-type conformation (Rinkel & Altona, 1987). Scrutiny of the H8/H6–H1′ crosspeak intensities shows that the intra-residue crosspeak intensity of residue A10 is abnormally high. This can be seen more clearly by considering the cross sections presented in Figure 6, showing the Overhauser effects of three different ring protons, namely, H8 of A10 and H6 of T7 and of C4, to the H1′ and H2′/H2″ resonances. It is obvious that the intensity of the H8–H1′ crosspeak of adenine A10 is much higher than those of thymidine T7 and cytidine C4 but comparable to the H5–H6 crosspeak of cytidine C4. That this effect is not caused by spin diffusion follows from a NOESY experiment performed with a mixing time of 75 ms. Then, in the spectral region considered, only the H8–H1′ crosspeak of A10 and the H5–H6 crosspeaks of C4 are retained with about the same intensity, indicating that the distance between the adenosine H8 and H1′ spins is near 2.5 Å. This strongly suggests that the adenine base is in a syn conformation. However, the UV–melting experiments and the imino proton spectra indicate that this base (A10) is involved in base pairing (vide supra). This is only possible when the base pair is of the Hoogsteen type (cf. Figure 7). Confirmation of the occurrence of such a base pair can be obtained by consideration of intra base pair proton–proton distances. In a Hoogsteen pair the distance between the thymidine imino proton and H8 of the adenine residue is 2.6 Å. Thus one expects to observe a crosspeak between the resonances of these two protons in a NOESY spectrum recorded for a H_2O solution, as indeed is borne out by experiment (cf. Figure 7). In addition, a spin diffusion induced crosspeak between the imino proton resonance and the adenine H1′ sugar resonance is observed. To avoid any possible confusion with crosspeaks between imino proton resonances and adenine H2 resonances, the hairpin was deuterated at the purine C8 position. It followed that all resonances at and near 8.2 ppm (the position of the adenine H8 resonance) were reduced to 10% of their original intensity. In the NOESY spectrum no connectivity between the imino

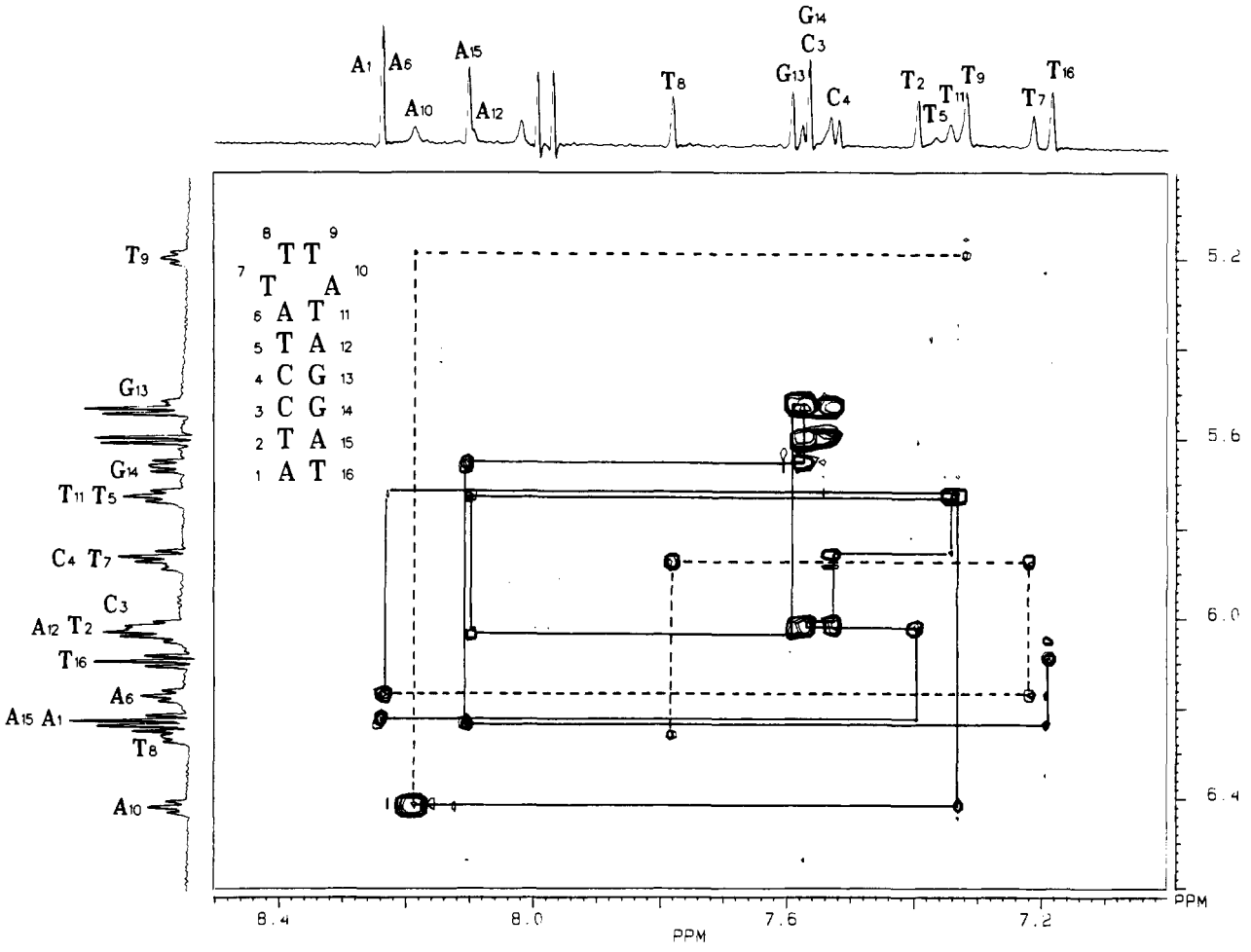

FIGURE 5: Part of the 600-MHz phase-sensitive NOESY spectrum of d(ATCCTA-TTTA-TAGGAT), dissolved in D_2O, exhibiting crosspeaks between aromatic H6/H8 ring protons and sugar H1′ protons. The spectrum was obtained for a 1 mM DNA sample at 1 mM NaCl, pH 7, and at a temperature of 298 K. The mixing time was 0.3 s. The sequential analysis of the crosspeaks is indicated. The drawn lines represent the connectivity diagram for residues in the stem; the dashed lines represent the connectivity diagram of the residues in the loop. The assignment of the base H6/H8 protons is indicated along the horizontal axis and the assignment of the H1′ sugar protons along the vertical axis.

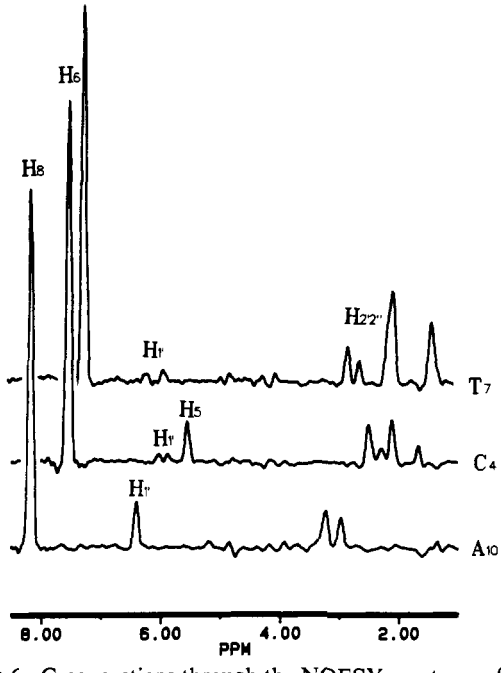

FIGURE 6: Cross sections through the NOESY spectrum of Figure 5 taken along F_1 for three different F_2 values so that crosspeaks from the H6/H8 ring proton resonances to other proton resonances are visible for residues T7, C4, and A10. The assignments of various crosspeaks are indicated.

proton resonance of thymine and the ring proton spectral region nor the H1′ proton spectral region was found. All these results demonstrate that T7 and A10 in d(ATCCTA-TTTA-TAGGAT) form a T-A Hoogsteen base pair with the adenine base in the syn conformation.

Similar experiments were performed for the hexadecanucleotide d(ATCCTA-CTTG-TAGGAT) which also forms a hairpin with an additional base pair (vide supra). Here we did not find any indications that the seventh base pair, C7-G10, forms a Hoogsteen pair. Instead, the intensity pattern of the NOE crosspeaks is in accordance with the presence of a Watson–Crick-type base pair.

Discussion

The experiments performed in the present study were inspired by two earlier observations. In our laboratory it was found that the hexadecanucleotide d(ATCCTA-TTTT-TAGGAT) folds into a hairpin loop in which in the expected loop region a wobble T-T pair is formed (Bommers et al., 1987; Hilbers et al., 1987) so that apparently a loop of two residues is present. For the octanucleotide d(C*GC*GTGC*G) Orbons et al. observed that a mini-hairpin can be generated also leading to a loop of two bases (Orbons et al., 1986, 1987). In our earlier paper (Blommers et al., 1987)), we concluded the existence of a wobble T-T pair on the basis of sequential NOE measurements and NOE measurements between imino protons. It followed from these experiments that the first and

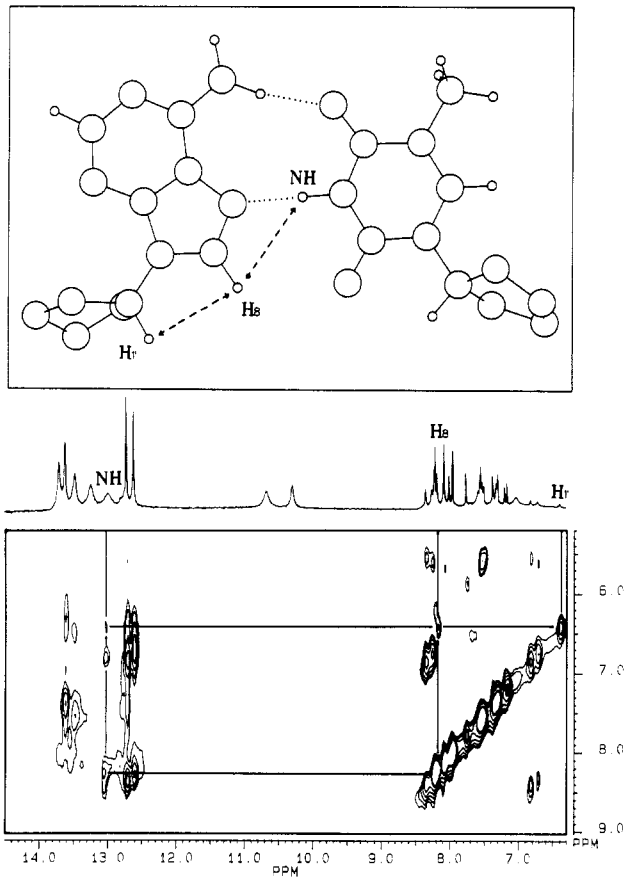

FIGURE 7: Part of the 600-MHz NOESY spectrum of d(ATCCTA-TTTA-TAGGAT) dissolved in H_2O. The spectrum was obtained for a 1 mM DNA sample containing 1 mM NaCl, pH 6, at a temperature of 298 K. The mixing time was 0.3 s. The crosspeaks between the imino proton resonance of residue T7 and the H8 and H1′ resonances of residue A10 are indicated. The geometry of the Hoogsteen base pair displayed on top corresponds with the observed intensities of the crosspeaks.

last thymine bases of the loop are stacked upon the stem of the hairpin and that their imino protons are arranged in such a way that these bases can form a pair. That actual hydrogen bonding occurs follows from NMR–melting experiments which show that the mentioned imino protons are bleached out from the spectrum at the same temperature as the majority of the imino protons in the stem. This indicates that the T-T pair has a lifetime similar to that of the other base pairs in the molecule. The resonance position of the imino protons in the T-T pair is somewhat higher than is normally found for uridine or thymidine imino protons involved in N–H···O hydrogen bonds. This we do not completely understand at this moment. In contrast, for the hairpins studied in this paper the existence of a seventh base pair is only accepted when an extra resonance is observed between 12.5 and 14 ppm and this needs some explanation. Contrary to the observations described above, we found that for molecules with -PuTTPy- loop sequences the exchange of the imino protons in the loop region is enhanced compared to that of those of the stem region of the hairpin and also compared to that of those of the loop region of molecules with -TTTT- and PyTTPu- loop sequences (cf. Figure 1, 3, and 4). The imino proton resonances of the -PuTTPy- loop sequences are strongly broadened. In the case of d(ATTACG-GTTC-CGTAAT) we were fortunate to be able to see the resonance of G7, as a very broad line but separated from the resonances at 11 ppm. For the other molecules with -PuTTPy- loop sequences, the corresponding resonances are bleached out completely or hidden beneath the broadened resonances near 11 ppm. This can only mean that for the -PuTTPy- sequences the base pair is not at all or transiently formed, in contrast to the -TTTT- loop sequence. This conclusion is further corroborated by the occurrence of a seventh resonance in the 12.5–14-ppm region for the molecules with -PyTTPu- loop sequences, which is diagnostic for the existence of a seventh base pair, and by the comparative UV–melting experiments.

The observation of the two-membered loops thus obtained came originally as a surprise because in a regular B-type (or A-type) helix the distance between the two terminal phosphates at one end of the stem is about 17.5 Å and bridging this distance by two nucleotides was thought to be sterically impossible (Tinoco et al., 1971). For aforementioned hexadecanucleotide d(ATCCTA-TTTT-TAGGAT) and octanucleotide d(C*GC*GTGC*G) it was observed that the torsion angle γ of the residue directly following the two-membered loop adopts a trans conformation instead of the standard gauche$^+$ conformation found in normal B-DNA (Orbons et al., 1987; Blommers et al., 1987). This adjustment of the backbone conformation decreases the distance between the 5′-end phosphate of the stem and the second loop residue, thus facilitating the formation of a two-membered loop. Another interesting feature of this type of loop closure is displayed by modeling studies (Haasnoot et al., 1987; van Beuzekom et al., 1989) which show that the C-G base pair closing such a two-membered loop is not flat but buckled and hence points to a strained conformation.

The results of this paper show that there is an additional mechanism by which the distance between the terminal phosphates at one end of the stem can be diminished, and this is the totally unexpected formation of a Hoogsteen base pair. The distance between the C1′ sugar carbons, which is 10.4 Å in a standard Watson–Crick base pair, is diminished to 8.6 Å in a Hoogsteen pair (Figure 7). This will concomitantly decrease the distance between the phosphates at the end of the stem where the Hoogsteen pair is situated and will additionally facilitate the formation of a two-membered loop. The occurrence of Hoogsteen base pairs as part of the end of double-helical regions in nucleic acids is relatively rare. In yeast $tRNA^{Phe}$ two reversed Hoogsteen base pairs, U8-A14 and T54-m^1A58, are observed. They facilitate the occurrence of sharp bends in the molecule and may in this respect resemble the role of the T-A Hoogsteen base pair in our hairpin. Hoogsteen T-A base pairs have also been observed in the complex of the bis-intercalator triostin A and the double helix formed by $d(CGTACG)_2$ (Wang et al., 1984). Here the reduced distance between the C1′ carbon atoms of the A-T pairs serves to stabilize the complex through a close packing of the oligonucleotide near the end of the triostin A molecule.

These observations illustrate that nucleic acids quite easily adjust their structure in order to accommodate local energetic needs so as to avoid highly localized strain situations and to optimize local interactions (Haasnoot et al., 1987). Our experiments do show that the formation of loops consisting of residues in partly self-complementary molecules depends on the sequence of the intervening residues. With a pyrimidine at the 5′-end and a purine at the 3′-end of this intervening sequence, two-membered loops are easily formed in the molecules considered, while inversion of this sequence makes the formation of such a loop much more difficult if not impossible. It is noted that replacement of the two central thymidines by more bulky bases appears to invalidate this simple rule in the sense that in that case a four-membered loop is retained.

We also studied the salt and concentration dependence of the UV-melting profiles for the oligonucleotides. At higher strand and salt concentrations the d(ATCCTA-PuTTPy-TAGGGAT) and not the d(ATCCTA-PyTTPu-TAGGAT) sequences give rise to dimer formation. The first type of fragments does not form the two-membered loop, but in the duplex form the Py and Pu are involved in base pair formation which explains the relative preference for the dimer structure compared to the second fragment where Py and Pu can pair in the hairpin form. Perusal of the literature shows that these considerations can be extended to some self-complementary sequences as well. For instance, a comparison of the melting behavior of the self-complementary sequences d(CGCG-AATT-CGCG) and d(CGCG-TATA-CGCG) shows that it is easier for the latter molecule to form hairpin structures. Thus at 0.07 mM strand concentration d(CGCG-AATT-CGCG) does not form a hairpin in 0.01 M NaCl (Marky et al., 1983) while at 0.38 mM strand concentration d(CGCG-TATA-CGCG) still is able to form hairpins in 0.1 M NaCl (Wemmer et al., 1985).

At first sight the above pieces of information seem rather unrelated: in some cases a two-membered loop closed by a distorted base pair (buckled or Hoogsteen type) is apparent; in other cases, sometimes only subtly changed sequences give rise to four-membered loops. Although it may seem a matter of semantics, we prefer to think of the above-mentioned two-membered loops as four-membered loops in which local optimization (e.g., base pairing) occurs. Such a view has its completely accepted analogies; e.g., the extra base pair T54-m^1A58 in $tRNA^{Phe}$ does not make the TΨC loop a three-membered one, nor is the m_2^2G26-A44 base pair in the same molecule normally reckoned to be part of the anticodon stem.

Seen in this light all the above hairpins fall within the scope of a general loop folding model for which under specific conditions further architectural details can be stated. For example, the present results demonstrate that for intervening loop sequences -PyXXPu- (Py = pyrimidine, X = thymidine, Pu = purine complementary to Py) the terminal bases are involved in some kind of base pairing scheme, while reversal of this sequence, i.e., -PuXXPy-, makes base pairing much more difficult if not impossible. Bulkiness of the two central bases in the loop is another limiting factor in local optimization: when the first X base is a purine [cf. d(C*GC*GTGC*G)]; an extra base pair may be formed, but replacing both X bases by purines [cf. d(ATCCTA-TAAA-TAGGAT)] limits the hairpin to the basic four-membered loop scheme.

Although the observed effects can thus be formulated in terms of a simple rule, the available dataset is still too small to declare the model generally applicable to all DNA hairpins. A case in point is formed by the results reported for d-(CGCG-TTTT-CGCG) (Hare & Reid, 1986) and d-(CGCGCGCG-TTTT-CGCGCGCG) (Wolk et al., 1988). These authors claim a loop folding for the four central T's different from our loop folding model in the sense that stack propagation from the 3′-end of the stem into the loop is limited to two thymidines; the third T is supposed to be rotated away from the center of the loop. Although we have some doubts about the ability of the NOESY–distance geometry methods used to deduct such conformational details in nucleic acids in an ab initio way (van de Ven & Hilbers, 1988), we cannot offer an explanation for this aberration apart from the observation that these loops have in common that they are folded on top of an all C-G stem. This is also true for the results obtained by Xodo et al. (1988). These authors found that neither the melting temperatures nor the melting enthalpies measured for d(CGCGCG-TTTTT-CGCGCG) and d(CGCGCG-AAAAA-CGCGCG) differed within experimental accuracy. This is in contrast to our results; upon substitution of all intervening thymidines by adenines the melting temperature as well as the transition enthalpy is significantly lower (cf. Table I). Very recently similar results were obtained by Breslauer and collaborators (Senior et al., 1988). These authors demonstrate that the stability of the related hairpins formed by d(CGAACG-XXXX-CGTTCG) decrease in the order X = T, C, G, and A, respectively. It is noted in passing that the NOESY crosspeak patterns of the loop regions of d(CGAACG-TTTT-CGTTCG) and d(CGAACG-AAAA-CGTTCG) agree with the interresidue connectivity pattern observed for d(ATCCTA-TTTT-TAGGAT) (Blommers et al., 1987): connectivities are observed between the loop residues X7 to X9, while no connectivity is observed between the loop residues X9 and X10. These results are in perfect agreement with our loop folding principle; i.e., the base stacking pattern of the loop bases is propagated on top of the 3′-end of a B-type double-helical stem, while after the third loop base the sugar–phosphate backbone changes its direction (Blommers et al., 1987).

The thesis put forward by Xodo et al. (1988) that the self-complementary sequences d(CGCGC-GATC-GCGCG) and d(CGCGC-GTAC-GCGCG) form hairpins in which the loops consist of only the two central (A/T) bases and G6-C9 base pairs close the loop is obviously also deviating from our results, i.e., -PuXXPy- sequences forming four-membered loops only. However, we note that their thesis is based on UV-melting data recorded for a very limited set of hairpins. Their main argument is that the enthalpy change on denaturation of the hairpins is ca. 3 kcal higher than that of the "unconstrained" hexamer duplex $d(CGCGCG)_2$. That this finding in itself is not sufficient for drawing conclusions regarding the loop size is examplified by our previously reported results for d(ATCCT-ATTTT-AGGAT). In the latter molecule the hairpin denaturation enthalpy is also slightly higher (ca. 3 kcal/mol) than that of the corresponding helix formed by the unconstrained hexamer duplex d(ATCCTA)·d(TAGGAT) Hilbers et al., 1985; Haasnoot et al., 1986), but NMR experiments indicate (Haasnoot et al., 1985, 1986) that the stem of the hairpin formed by this molecule involves only five base pairs (no imino proton resonance was found for the conceivable pair A6-T10). In other words, the enthalpy contribution of the loop to the overall hairpin stability may be equal to or exceed that of one base pair (Hilbers et al., 1985; Haasnoot et al., 1985 1986).

In summary, the present results demonstrate that nucleotide sequences and composition have a marked influence on loop folding and hairpin stability. Notwithstandingly, the results can be explained by a rather simple model in which under certain conditions local optimization of the hairpin structure within the conformational space delineated by our loop folding principle may occur.

Acknowledgments

NMR spectra were recorded at the Dutch National hf-NMR facility at Nijmegen, which is supported by SON. Dr. S. S. Wijmenga and J. Joordens are thanked for expert technical assistance.

Supplementary Material Available

Two figures in which, by means of NOE difference spectra, the assignment of the imino proton resonances is given for the sequences d(ATCCTA-TTTA-TAGGAT) and d(ATCCTA-CTTG-TAGGAT) (3 pages). Ordering information is given

on any current masthead page.

References

Blommers, M. J. J., Haasnoot, C. A. G., Hilbers, C. W., van Boom, J. H., & van der Marel, G. A. (1987) Structure and Dynamics of Biopolymers, *NATO ASI Ser. E 133*, 78–91.

Breslauer, K. J., Frank, R., Blocker, H., & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746–3750.

Gralla, J., & Crothers, D. M. (1973a) *J. Mol. Biol. 73*, 497–511.

Gralla, J., & Crothers, D. M. (1973b) *J. Mol. Biol. 78*, 301–319.

Haasnoot, C. A. G., & Hilbers, C. W. (1983) *Biopolymers 22*, 1259–1266.

Haasnoot, C. A. G., de Bruin, S. H., Berendsen, R. G., Janssen, H. G. J. M., Binnendijk, T. J. J., Hilbers, C. W., van der Marel, G. A., & van Boom, J. H. (1983a) *J. Biomol. Struct. Dyn. 1*, 115–129.

Haasnoot, C. A. G., Westerink, H. P., van der Marel, G. A., & van Boom, J. H. (1983b) *Biomol. Struct. Dyn. 1*, 131–149.

Haasnoot, C. A. G., de Bruin, S. H., Hilbers, C. W., van der Marel, G. A., & van Boom, J. H. (1985) *Proc. Int. Symp. Biomol. Struct. Interact., Suppl. J. Biosci. 8*, 767–780.

Haasnoot, C. A. G., Hilbers, C. W., van der Marel, G. A., van Boom, J. H., Singh, U. C., Pattabiraman, N., & Kollman, P. A. (1986) *J. Biomol. Struct. Dyn. 3*, 843–857.

Haasnoot, C. A. G., Blommers, M. J. J., & Hilbers, C. W. (1987) Structure, Dynamics and Function of Biomolecules, *Springer Ser. Biophys. 1*, 212–216.

Hare, D. R., & Reid, B. R. (1986) *Biochemistry 25*, 5341–5350.

Hare, D. R., Wemmer, D. E., Chou, S.-H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319–336.

Hilbers, C. W., Haasnoot, C. A. G., de Bruin, S. H., Joordens, J. J. M., van der Marel, G. A., & van Boom, J. H. (1985) *Biochimie 67*, 685–695.

Hilbers, C. W., Blommers, M. J. J., Haasnoot, C. A. G., van der Marel, G. A., & van Boom, J. H. (1987) *Fresenius' Z. Anal. Chem. 327*, 70–71.

Hoogsteen, K. (1963) *Acta Crystallogr. 16*, 907–916.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.

Marky, L. A., Blumenfield, K. S., Kozlowski, S., & Breslauer, K. J. (1983) *Biopolymers 22*, 1247.

Orbons, L. P. M., van der Marel, G. A., van Boom, J. H., & Altona, C. (1986) *Nucleic Acids Res. 14*, 4187–4196.

Orbons, L. P. M., van Beuzekom, A. A., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 965–987.

Rich, A., & RajBhandary, U. L. (1976) *Annu. Rev. Biochem. 45*, 805–860.

Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621–649.

Sänger, W. (1984) *Principles of Nucleic Acid Structure*, 1st ed., Springer-Verlag, New York.

Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., & van Boom, J. H. (1983) *J. Am. Chem. Soc. 105*, 2914–2916.

Senior, M. M., Jones, R. A., & Breslauer, K. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6242–6246.

Tibanyenda, N., de Bruin, S. H., Haasnoot, C. A. G., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1984) *Eur. J. Biochem. 139*, 19–27.

Tinoco, I., Uhlenbeck, O. C., & Levine, M. D. (1971) *Nature (London) 230*, 362–370.

Uhlenbeck, O. C., Borer, P. N., Dengler, B., & Tinoco, I. (1973) *J. Mol. Biol. 73*, 483–496.

van Beuzekom, A. A. (1989) Ph.D. Thesis, State University Leiden.

van Boom, J. H., van der Marel, G. A., Westerink, H. P., van Boeckel, C. A. A., Mellema, J.-R., Altona, C., Hilbers, C. W., Haasnoot, C. A. G., de Bruin, S. H., & Berendsen, R. G. (1983) *Cold Spring Harbor Symp. Quant. Biol. 47*, 403–409.

van de Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1–38.

Wang, A. H.-J., Ughetto, G., Quigley, G. J., Hakoshima, T., van der Marel, G. A., van Boom, J. H., & Rich, A. (1984) *Science 225*, 1115–1121.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., & Weiner, A. M. (1987) *Molecular Biology of the Gene*, 4th ed., Benjamin/Cummings, Menlo Park, CA.

Wemmer, D. E., Chou, S. H., Hare, D. R., & Reid, B. R. (1985) *Nucleic Acids Res. 13*, 3755.

Wolk, S. K., Hardin, C. C., Germann, M. W., van de Sande, J. H., & Tinoco, I. (1988) *Biochemistry 27*, 6960–6967.

Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G. A., & van Boom, J. H. (1988) *Biochemistry 27*, 6321–6326.