Duplex-hairpin transitions in DNA: NMR studies on CGCGTATACGCG

D.E.Wemmer, S.H.Chou, D.R.Hare and B.R.Reid¹

Department of Chemistry, and ¹Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

Received 13 November 1984; Revised and Accepted 4 April 1985

ABSTRACT

Two dimensional NMR methods have been used to assign proton resonances in the high salt ($>100\text{mM Na}^+$), low temperature duplex form of the self-complementary DNA dodecamer d(CGCGTATACGCG). At low salt ($\leq 10\text{mM Na}^+$) and higher temperature marked changes in the two-dimensional spectrum, and in the one-dimensional spectrum reported by others, indicate that the molecule converts to an alternate conformation. Using saturation transfer methods, many of the resonances of this new conformation have been assigned, and the kinetics of the interconversion of the two forms has been studied. The linewidth, correlation time, and concentration dependence of the formation of this alternate conformation support the idea that it is a unimolecular hairpin. Observation of chemical shifts and NOEs in the hairpin conformation allow some preliminary structural characterization. Examination of the energetics of the interconversion suggests that the exchange between forms does not proceed through a single stranded intermediate, but rather through another pathway, probably involving a cruciform structure.

INTRODUCTION

The availability of relatively large quantities of specific sequence DNA oligomers through improved synthetic methods has stimulated many studies of the detailed structural properties of various sequences by both crystallographic and NMR methods. These studies have uncovered sequence dependent "fine structure" within the classical B form helix,²⁻⁴ as well as the left handed Z structure in alternating CG sequences.⁵⁻⁷ NMR studies have shown that, in some cases, the solution and crystal structures of the same molecule are not the same^{8,9} while in other cases the crystal structure seems to be well preserved in solution.¹⁰ The NMR method has the advantage of being able to examine conformations that occur in solution, but which have not yet been crystallized, for example hairpins^{11,12} or junctions.¹³ The development of systematic methods for resonance assignments in nucleic acids by a number of groups^{9,10,14-18,29} using procedures similar to those developed by Wuthrich and coworkers for proteins,¹⁹⁻²¹ has made possible more detailed interpretation of the NMR data than was previously possible.

It has been known for many years that alternating $(TA)_n$ oligomers are capable of forming many structures in solution, including normal duplexes, cruciforms and hairpins, and several detailed studies of the stabilities of these forms have been made.²²⁻²⁴ The study of these sequences is complicated by the many interconverting conformations possible in solution, so that it is difficult to examine each of these forms independently. Such behavior has not been observed in general for self complementary sequences. It has recently been found that the sequence CGCGTATACGCG takes on two conformations in solution, which can co-exist under appropriate conditions of salt, temperature and strand concentration.¹ NMR studies have shown that the structure favored at low temperatures and high salt ($\geq 100\text{mM Na}^+$) is a relatively normal duplex, but the conformation favored at higher temperatures and low salt ($\leq 10\text{mM Na}^+$) involved unpaired bases with imino protons that were not hydrogen bonded; i.e., a bulge or loop structure of some kind. At much lower strand concentration, similar behavior in the sequence CGCGAATTCGCG has been observed at much lower strand concentration using optical methods, and it was concluded that the low salt, high temperature form is a hairpin structure.

In the present work we assign the resonances of the normal duplex form of CGCGTATACGCG using two dimensional sequential NOEs.^{10,14-16,18} These assignments are then carried over to the loop conformation using a two dimensional exchange spectrum²⁵⁻²⁷ under conditions where the two conformers coexist, and interconvert at a moderate rate. With the assignments of both forms, the NOEs and chemical shifts that are observed can be interpreted in structural terms. The rates of interconversion between these two forms have been determined through the measurement of saturation transfer between resonances from the same protons in the two forms as a function of temperature. The measurement of the kinetics of interconversion between the two conformers at various temperatures may provide an estimate for the activation energy for forming cruciform structures. Recently it has been found that certain palindromic sequences in supercoiled DNAs do form cruciforms, particularly under conditions of large negative superhelical tension.³⁹ Understanding conditions under which such structures can form in linear DNAs is important to understand further the process of cruciform formation in naturally occurring DNAs, and thus ultimately their possible roles in controlling DNA expression.

MATERIALS AND METHODS

The DNA used in these studies was synthesized using the solid-phase phosphite triester method, as described previously.¹⁰ The synthesis was carried out on a 10 μ M scale, yielding approximately 100-200 OD₂₆₀ units (5-10mg) of DNA after purification. NMR samples were generated by lyophilizing the purified DNA from 10mM phosphate buffer, then redissolving in an equal volume of D₂O (99.996% D, Stohler Isotope Chemicals). The final concentration of DNA for most studies (other dilutions given below) was 3.8mM in strands, measured by absorbance at 260 nm. NMR spectra were taken on a Bruker WM-500 spectrometer using an Aspect-2000 computer. Two-dimensional data sets for the COSY and NOESY spectra were collected in the phase-sensitive mode using the method of States *et al.*⁴¹ The data sets were transferred to magnetic tape and processed on a VAX/11-780 using our own software (D.R. Hare, unpublished). Typically 1024 complex t_2 data points were collected for each of about 350 t_1 values, using a spectral width of 4400 Hz, giving a digital resolution of 4.3 Hz/pt. after zero filling in t_1 . The mixing time used in the NOE experiments was 300ms at 35°C and, in the exchange measurements, 400ms at 40°C. Saturation transfer measurements were made by alternating selective irradiation on and off the peak of interest using a cycle of 8 scans. The radio frequency power was adjusted to fully saturate the resonance of interest. The irradiation time was 10 seconds; i.e., more than 4 times $*T_1$ for the observed resonance. Difference spectra were generated by subtraction of the spectra after Fourier transformation. $*T_1$ values were measured using standard inversion recovery methods with continuous selective irradiation of the exchange partner. $*T_1$ values were determined using the standard three parameter fit provided in the Bruker DISNMRP software package. Selective T_1 measurements were made using a decoupler pulse of 10ms to invert only the resonances of interest, without affecting resonances beyond 50-100 Hz (50 Hz or more) from the pulse frequency. Selective recovery rates were determined using the initial slope of the recovery curve. For measurement of resonance intensities, a 60° pulse was used with a delay of 2.84 seconds between experiments. Since the kinetics of the exchange are fairly slow at low temperature, the sample was held at a particular temperature for several lifetimes before spectra were taken (the details of the kinetics will be discussed further below). The transition enthalpy between the duplex and hairpin states was calculated from

the population vs. temperature data using the equation

$$\Delta H = 6RT_m^2 \left(\frac{\partial f}{\partial T} \right)_{T_m}$$

where T_m is the temperature at which the populations of the two forms are equal.

RESULTS AND DISCUSSION

The NMR spectrum of the methyl resonances in the sequence dCGCGTATACGCG is shown at several concentrations of NaCl in Figure 1. The clear doubling of lines indicates the coexistence of two conformations of this DNA sequence, with the intensity of the lines reflecting the relative populations of the two forms.

Patel *et al.*¹ have carried out one-dimensional studies on the imino protons and adenine C2 protons of the altered conformation from which they conclude that it is either a bulged duplex or a hairpin based on the fact

Fig. 1 The upfield region of the proton NMR spectrum of CGCGTATACGCG in D_2O solution as a function of added NaCl (1mM phosphate buffer, strand concentration 0.38mM, 35°C, NaCl given in molar concentration). The doubling of the methyl peaks from coexistence of two conformers is clear at intermediate salt concentrations.

that the central residues are not base paired. In the following sections we present evidence based on cross relaxation rates and concentration dependence that indicates that the new conformer is a hairpin rather than a bulged duplex. The conversion from duplex to hairpin does not proceed via complete separation of the two strands but probably via a propagated cruciform intermediate which may be of biological significance. Following the intensities of the lines from each conformer as a function of added salt and temperature allows the "phase diagram" for these two conformations to be mapped. The fraction of DNA in the "alternate" conformation is shown as a function of temperature at the lowest salt concentration in Figure 2. The populations follow a typical sigmoidal curve as expected for a simple exchange process between two distinct conformations. Analysis of the slope of this curve at the inflection point gives a value of 30 kcal/mol for the van't Hoff enthalpy change between the duplex and hairpin states. With increasing salt concentration the duplex form of the helix is stabilized, as has been observed for $(TA)_n$ oligomers in general.²²⁻²⁴ From optical studies of the related sequence CGCGAATTCGCG at low strand concentration it was suggested that a hairpin structure was most likely for the "alternate" conformation, based on strand concentration dependence and salt dependence

Fig. 2 The partitioning of DNA between duplex and hairpin forms (left curve); and between hairpin and random coil forms (right curve) as a function of temperature for 3.8mM strands in 10mM phosphate buffer pH7. The left curve is derived from intensities of resonance lines of duplex and hairpin forms coexisting at all temperatures. The right curve is derived using the chemical shifts of aromatic resonances, which are in fast exchange between hairpin and random coil forms.

together with thermodynamic comparisons with previously studied hairpin structures.³⁶ We have examined the relative concentrations of the two conformers as a function of total strand concentration in 10mM phosphate buffer and 37°C. Over the range of concentrations examined (3.8mM to 0.38mM) the fraction of hairpin conformer observed is in close agreement with predictions based on a two-strand duplex dissociating to form two single strand hairpins. The observed hairpin populations are $f_h = 0.40$ at 3.8mM, $f_h = 0.50$ at 1.9mM, $f_h = 0.64$ at 0.95mM, $f_h = 0.83$ at 0.38mM strands; the predicted values at 1.9mM, 0.95mM, and 0.38mM strands are 0.52, 0.68 and 0.79, respectively. Further evidence is provided by a comparison of selective inversion recovery measurements of the C6H protons of the cytidine in position 1. These resonances from both the duplex and alternate forms are resolved from other resonances and can be inverted together. The initial rate of recovery observed for the duplex C1 H6 is almost exactly a factor of two faster than for the alternate conformation (4.2 sec^{-1} vs. 2.0 sec^{-1}). These initial slopes are dominated by cross relaxation, and thus indicate a rotational correlation time for the alternate conformation of about one half that of the duplex, again completely consistent with the new conformer being a hairpin.

Assignment of Resonances

Assignment of the imino and adenine C2H resonances for the "TATA" sequence has been carried out by Patel *et al.* using one dimensional interbase-pair NOEs.¹ We have now extended the assignments to the nonexchangeable hydrogens of this sequence using two-dimensional COSY and NOESY spectra. The use of the sequential NOE method for the assignment of non-exchangeable resonances in nucleic acids has been described in detail elsewhere,^{10,14-16,18} and will not be repeated here. It suffices to note that base and sugar protons on neighboring residues are sufficiently close in space that NOEs can be observed between some of them. Following the primary sequence, and using the identity of the base type at each step, established from the COSY data, one can assign unambiguously all of the resolved protons of the sequence. These generally include all of the base protons, and sugar protons except the C5'H and C5"H, which occur in such crowded spectral regions that they cannot be assigned except in short, simple sequences (unpublished observations). To assign the protons of the duplex form, COSY and NOESY spectra were collected on a sample at fairly high salt and moderate temperature (3.8mM strand concentration, 200mM NaCl, 30°C) - conditions under which the duplex is highly favored. To verify that chemical shifts do not

Fig. 3 Expansion of the aromatic to C1' proton region of the NOESY spectrum of CGCGTATACGCG at 35°C, 3.8mM strands in 10mM phosphate buffer pH7. The sequential assignment path in the duplex form is traced with a solid line, with cross peaks between aromatic and sugar protons of the same residue labeled with that residue. NOEs between cytosine C5 protons and neighboring aromatics are connected by vertical lines. Peaks that are not connected or labeled are NOEs in the hairpin form.

change significantly at low salt, the assignments were also carried out at low salt and moderate temperatures where equal populations of the two conformers are observed. An expansion of the aromatic to C1' proton cross-peak region of the phase sensitive spectrum under such conditions is shown in Figure 3. The sequential connectivities for the B-form duplex are indicated by a solid line. Peaks not connected by lines correspond to the hairpin conformation. The assignments obtained using this spectral region were confirmed, and extended to the C2' and C2'' protons, using the aromatic to C2' proton cross-peak region. The C3' and C4' sugar protons were then assigned using both the NOESY and COSY spectral regions corresponding to C2' to C3',

Table 1. Chemical shifts of assigned resonances in the duplex and hairpin forms of CGGTATACGG.

	C6H	C6H	CSH	C1'H	C2"H	C2'H	C3'H	C4'H	CH ₃	C2H
1C		7.48a) 7.50b) (0.02)c)	5.74 5.64 (-0.10)	5.61 Δd)	2.23	1.76	X	3.89		
2G	7.78 7.83 (0.05)			5.73 Δ	2.56	2.48	4.79	4.16		
3C		7.15 7.23 (0.08)	5.19 5.27 (0.08)	5.55 5.65 (0.10)	2.23 2.16	1.87 1.64	4.67	4.02		
4G	7.71 7.70 (Δ0)			5.79 Δ	2.60	2.42	4.78	4.18		
5T		7.01 7.18 (0.17)		5.54 5.84 (0.30)	2.29 2.23	1.91 1.90	4.68	4.00	1.32 1.46 (0.14)	
6A	8.09 8.06 (0.03)			6.03 Δ	2.72	2.44	4.80	4.22		7.14 7.92 (0.78)
7T		6.96 7.06 (0.10)		5.43 5.53 (0.10)	2.21 1.93	1.81 1.70	4.66	3.98	1.23 1.18 (-0.05)	
8A	8.02 8.10 (0.08)			5.96 6.05 (0.09)	2.68 2.77	2.45 2.22	4.82	4.20		7.21 7.90 (0.69)
9C		7.06 7.27 (0.21)	5.05 5.33 (0.28)	5.40 5.36 (-0.04)	2.11 2.21	1.74 1.89	4.61	3.95		
10G	7.66 7.77 (0.11)			5.67 5.60 (-0.07)	2.50	2.40	4.77	4.15		
11C		7.13 7.14 (Δ0)	5.22 5.32 (0.10)	5.60 Δ	2.15	1.71	4.62	3.96		
12G	7.75 7.80 (0.05)			5.97 Δ	2.42	2.17	4.57	X		

a) is the chemical shift of B form
b) is the chemical shift of hairpin
c) () is the chemical shift difference (δ hairpin - δ B form)
d) Δ too close to the diagonal to measure X not assigned

and C3' to C4', proton coupling, as well as the C1' to C3' and C4' NOEs. There are no degeneracies in the peak positions and no ambiguities in the assignments. The chemical shifts do not change significantly upon addition of salt and the assignments obtained under different conditions agree completely. The chemical shifts of the assigned protons are listed in Table 1.

The related sequence CGCGAATTCGCG has also been studied by NMR,¹⁰ and it is interesting to compare the results obtained for these two sequences. It has been noted by Dickerson and coworkers,²⁻⁴ and described theoretically by Calladine,²⁸ that compared to uniform B DNA, there is a sequence dependent "fine structure" which appears to be determined by optimization of base overlap, tempered by avoidance of cross-strand steric clashes. The structural predictions, which agree fairly well with the limited crystallographic data available, are that this local helix distortion will be determined by the nature of the base type in the sequence (purine or pyrimidine), rather than by the actual sequence itself. The intensity of NOE cross peaks reflect the distances between the protons involved, provided a sufficiently short mixing time is used so that NOEs are dominated by first-order effects. The aromatic to C1' proton cross peak region seems to be particularly useful for structural analysis, since the distances between these protons are sensitive to the local structural distortions. Some caution must be used since indirect NOE pathways (spin diffusion) may become important at longer NOE mixing times, particularly for proton pairs separated by 3.5Å or more. It appears that, at moderate mixing times, the cross peak intensities still reflect primarily the direct cross relaxation rates and hence the distances between spins. Examination of the NOE spectrum, Figure 3, shows that NOE intensities are nearly the same for each of the purine-pyrimidine steps. Similarly each of the pyrimidine-purine steps has a roughly equal intensity that is obviously different from the intensities of the purine-pyrimidine steps. These observations are in agreement with Calladine's rules,²⁸ which predict that such an alternating sequence, although it may vary from classical B-form DNA, should have a uniform structure with dinucleotide repeats. Considerable non-uniformities have been observed in CGCGAATTCGCG,¹⁰ and these will be analyzed in more detail in a separate publication.³⁰

Assignment of the Alternate Conformation

It is well known that transfer of saturation can be observed for systems in which spins are chemically exchanging from one environment to another if

Fig. 4 Expansions of the aromatic (left) and C1' (right) proton regions of the two-dimensional "exchange" spectrum at 40°C, same sample as Figure 3. Cross peaks represent saturation transfer between the resonances of protons in the duplex and hairpin forms. Assignments shown are derived from sequential connectivities in the duplex form, such as shown in Fig. 3.

the rate constant for the exchange is comparable to the spin lattice relaxation rates of the spins observed.^{31,32} The two-dimensional "NOE" experiment is known to be sensitive to chemical exchange in addition to direct magnetic cross relaxation (in fact this was its first application²⁵), and hence provides a method for assigning the resonances of the alternate conformation using the assignments obtained above for the duplex form. This method has also been applied to the assignment of the Z form of the DNA sequence (dm⁵CdG)₃ in methanolic solution.³³ Figure 4 shows portions of the two-dimensional "exchange" spectrum taken of CGCGTATACGCG at low salt and 40°C. Under these conditions chemical exchange between the conformers occurs at a moderate rate, and the populations of the two conformers are almost equal. Extra peaks, in spectral regions where NOEs are not expected, are observed which correspond to chemical exchange between conformers. Since the duplex form of the sequence is already assigned, these new cross peaks immediately assign the alternate conformation. A majority of the aromatic and C1' protons change sufficiently in chemical shift that resolved cross peaks away from the diagonal are observed. In other regions of the spectrum, the exchange cross peaks are not as clearly observable; however using a phase-sensitive COSY spectrum taken under identical conditions, the assignments of the C1' protons in the alternate form could be extended to the C2' and C2" protons without difficulty. Table 1 shows the change in chemical shift upon going from the duplex to hairpin form; the hairpin chemical shift is shown below the chemical shift in the duplex form, with the change in parentheses. We observe that there are relatively small changes observed for the ends of the sequence (residues 1, 2, 3 and 10, 11 and 12), which correspond to the double helical stem of the hairpin. However much larger changes are seen in the central residues, the largest of which are the C2H protons of A6 and A8, as has been noted by Patel.¹ By decreasing the temperature, NOE spectra could be obtained for the two conformers in equilibrium, but without the interference of exchange and exchange-transferred NOE cross peaks. The aromatic to C1' and the aromatic to C2" proton regions of the NOE spectrum are shown in Figures 3 and 5. Since the hairpin conformation is half of the molecular weight of the duplex form, the NOE buildup rate for this conformation is slower, and peaks from the duplex form dominate the spectrum, although a number of clear peaks from the hairpin conformation are observed. Some of these can be used to obtain direct sequential connectivities in the hairpin conformation which confirm the assignments obtained using the exchange cross peaks. Examination of the NOE

Fig. 5 Expansion of the aromatic to H2', H2'' and methyl NOE region of the NOESY described in Fig. 3. The H2' and H2'' to aromatic proton cross peaks of the duplex form are labeled with their residue. The vertical lines indicate methyl to neighboring aromatic cross peak connectivities in the duplex form.

and chemical shift data reveals several interesting features of the hairpin conformation. From Table 1, and the NOE region of Figure 4, it is apparent that the chemical shift of G4 does not change significantly upon going from the duplex to the hairpin form. In addition, a "normal" NOE is observed between the C6H of T5 and the C1'H of G4. These facts together suggest that, in the hairpin conformation, T5 is still stacked "normally" on G4. However the large change in chemical shift of the T5 protons, and the lack of NOEs from T5 to A6, suggests that these two bases are not stacked in the loop. The stacking of T5 on G4 is consistent with the observation of an NOE between one of the loop T imino protons and the imino proton of the base pair at the top of the hairpin stem in TTTT loops as reported by Haasnoot *et al.*^{11,12} On the other side of the stem however, the protons of C9 do undergo large changes in chemical shift upon transfer to the hairpin form, and NOEs are not observed between protons in residues A8 and C9, indicating the absence of strong stacking. It is interesting to note that changes in chemical shift

between the duplex and hairpin forms are observed even for C1 and G12, indicating some change in structure even for base pairs far from the loop, although these changes are probably small. Similar features have been observed for the sequence CGCGTAATCGCG, which is apparently stable only in the hairpin conformation, and which will be described in more detail elsewhere.³⁴

Kinetics of Exchange Between Conformations

The saturation transfer method also provides a quantitative method for measuring the unidirectional rate constants for the exchange process between the duplex and hairpin conformations. It is well known that, as the rate of chemical exchange between two sites increases, the resonances from the spins affected by the exchange process broaden, then eventually coalesce at a weighted average chemical shift. When the rate of exchange is less than the separation between resonances in frequency units, but is large compared to the relaxation rates of the spin, then the exchange process leads to linebroadening which can be used to determine the exchange rate. This effect has been applied recently to the study of the exchange among the B, Z and single strand forms of (m⁵CGCGm⁵CG).³⁵ However in the present application, the concentrations of the two forms are more highly temperature dependent, and hence when the temperature is high enough so that exchange rates could lead to significant linebroadening, the sample has converted almost totally into the hairpin form. Hence we have used direct measurement of saturation transfer to derive the kinetic parameters. In this method the saturation of a particular resonance leads to partial saturation of the resonance with which it exchanges. The fractional saturation, f_s , observed in such a case is determined by the spin lattice relaxation time of the observed spin, and the unidirectional rate of exchange, k_f , between the two species:

$$f_s = k_f {}^*T_1 \text{ and } {}^*T_1 = T_1 / (1 + k_f T_1)$$

where *T_1 is the spin lattice relaxation time of the observed spin in the presence of saturation of the exchange partner, and T_1 is the relaxation time which would be observed in the absence of exchange. *T_1 is measured directly so that the rate constant for exchange can be calculated from the fractional saturation observed under the same conditions. It is important that the fractional saturation observed be derived only from the exchange process and not from NOEs, or the direct effect of the applied rf field. In the TATA sequence the C2H resonance of A6 in the duplex form was saturated, and the effect upon the well separated resonance of the same proton in the hairpin

Table 2.
Observed saturation transfer parameters for CGCGTATACGCG.

T	f_s	*T_1	k_f
310	0.163	1.47	0.11
315	0.455	0.87	0.52
320	0.732	0.43	1.69

(T = temperature in K; f_s = fractional saturation observed;
 *T_1 = relaxation time of hairpin conformation in the
presence of saturation; k_f = rate of conversion of duplex to
hairpin)

form was observed. These protons undergo a large change in chemical shift upon exchange, and thus can be easily saturated without directly affecting the observed resonance. In addition the C2H protons are well separated in space from other protons in the sequence, and hence do not have significant NOEs to other protons in either conformation. The rate of exchange and the populations of the two conformations are such that rates can be measured accurately with the saturation transfer method over the temperature range 42-57°C. Three measurements were made in this range, and the results are summarized in Table 2. At the highest temperature the random coil state is significantly populated; see Figure 1. Since the hairpin and random coil forms are in fast exchange, saturation transfer from duplex to hairpin and from duplex to random coil will be indistinguishable, their effects being additive. If the rate of conversion from duplex to random coil were substantial, the rate of saturation transfer to the hairpin would increase anomalously via the random coil route, being too high at higher temperatures. Since this is not observed (though our kinetic data are limited) the duplex-coil rate is probably not yet significant. The temperature dependence of these rates can be used in the usual fashion to derive an activation energy for the exchange process. The Arrhenius plot derived from these rates is shown in Figure 6, from which an activation energy of 55 kcal/mol is derived. This seems a reasonable value considering the van't Hoff ΔH value of 30 kcal/mol for this transition derived above.

Hairpin to Random Coil Transition

At temperatures above about 40°C the hairpin is the predominant form at low salt. Upon further increase in temperature, the lines of this form shift continuously and, over the range 50-60°C, some of the resonances broaden

Fig. 6 Arrhenius plot for the rate of dissociation of the duplex form of CGCGTATACGCG into two unimolecular hairpins, measured by using saturation transfer for the same sample as Figure 3. The activation energy derived is 55 kcal/mol.

significantly. This temperature range corresponds to the transition from hairpin to random coil form. The linebroadening observed, in the absence of splitting of the resonances, indicates that this process occurs on a time scale near the fast exchange limit. This is similar to the behavior observed for the duplex to coil transition in a number of different duplex sequences (unpublished results). This transition appears to be comparable in width to the duplex to hairpin described above, Figure 2. The van't Hoff enthalpy of transition derived is approximately 30 kcal/mol of hairpin. This value, together with the 30 kcal/mol observed for the duplex to hairpin conversion, gives a duplex to random coil enthalpy of 90 kcal/mol, in good agreement with the expected value.

Energetics of the Transitions

The enthalpy of transition between the duplex and hairpin forms found for the TATA sequence (30 kcal/mol at 10mM sodium phosphate) is surprisingly low in comparison to the value for the related AATT sequence³⁶ (90 kcal/mol at 10mM NaCl) using optical methods and much lower strand concentrations.

However, the optical results are somewhat surprising in themselves, since the enthalpy of the transition from duplex to hairpin at low salt (10mM NaCl) is 90 kcal/mol, but the calorimetrically measured enthalpy of the duplex to single strand transition, measured at higher strand concentration, is also 90 kcal/mol. In fact, at lower salt the apparent enthalpy of transition for the duplex to hairpin increases to 140 kcal/mol, while the extrapolated calorimetric value for the duplex to single strand transition decreases to less than 90 kcal/mol. It appears that the van't Hoff enthalpy derived from these measurements cannot reflect the true energetics of the system.

The measured activation energy for the duplex to hairpin transition is also of interest. Since this value is derived from the temperature dependence of the rate of interchange between these forms, it puts an upper bound on the enthalpy difference between these two forms. The activation energy of 55 kcal/mol, which we have derived for CGCGTATACGCG, is consistent with the van't Hoff enthalpy for this sequence, but not with the enthalpy for the CGCGAATTCGCG sequence. It is also interesting to compare this activation energy with the enthalpy of transition between the duplex and single strand form. The enthalpy for a number of sequences with the same number of base pairs has been measured,^{36,37} and can also be estimated from base stacking energies. Using either extrapolations from other sequences, correcting for the differences in stacking energy using experimentally derived energy values,³⁷ or using the purely theoretical value,³⁸ the enthalpy of melting the duplex to single strands should be on the order of 80-90 kcal/mol. If the transition between the duplex and hairpin forms were going through a single stranded intermediate, then its activation energy should be at least this value. Since the experimental value is approximately half of this, it is clear that the strands cannot fully dissociate in forming the hairpin conformation. Although with the data presently available it is not possible to rigorously say more about the intermediate formed in this transition, the results are completely consistent with the hypothesis that a cruciform structure is formed. Once an initial bulge is created at the center of the duplex and a single base pair is formed between bases on the same strand, little energy should be required to propagate the "mobile junction" which has been formed and to complete the separation of the two hairpins. It seems reasonable that the "bulged" intermediate initially formed would need to have approximately six base pairs disrupted, requiring about half of the enthalpy of melting of the sequence, or about 40-45 kcal/mol in this case. This is entirely in line with the observed value of 55 kcal/mol. Such a pathway is clearly a special case of the "indirect slippage" mechanism proposed for the

interconversion of duplex and hairpin forms of $(TA)_n$ oligomers by Scheffler *et al.*²² In their pioneering work Scheffler *et al.* measured the kinetics of the conversion of hairpins to duplexes (the opposite direction to that measured in the present work). The activation energy found for this reaction was about 50 kcal/mol, regardless of the length of the DNA, indicating a somewhat higher barrier than we have found for the duplex to hairpin transition. The analysis of sequences such as CGCGTATACGCG should be simpler than that for strictly repeating polymers since many of the "indirect slippage" paths will be energetically unfavorable (requiring unpairing of more bases than the "symmetric" cruciform described above). Further studies of the length and sequence dependence of these transitions should further clarify the pathway.

Using an intrastrand cross-linking assay it has been found that a 76 base pair perfect inverted repeat derived from the lac operator can form stable cruciforms.³⁹ The activation energy for formation of this cruciform from the linear DNA is found to be about 33 kcal/mol.⁴⁰ This is somewhat lower than the value we have found for the transition between the linear and hairpin conformation; however the differences in sequence, and the superhelical twist strain, will certainly influence this value, if indeed the intermediate in the formation of hairpins is also a cruciform structure. Further experiments are underway to further illuminate the structures involved in these transitions.

ACKNOWLEDGEMENTS

We would like to thank Dr. A.S. Benight for helpful discussions and we gratefully acknowledge support through NIH grant P01 GM 32681 and instrumentation grants from the Murdock Foundation, NIH (GM 2874-0151) and NSF (PCM80-18053).

REFERENCES

1. Patel, D.J., Kozlowski, S.A., Ikuta, S., Itakura, K., Bhatt, R. and Hare, D.R. (1983) Cold Spring Harbor Symp. Quant. Bio. **47**, 197-206.
2. Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.F. (1981) Proc. Natl. Acad. Sci. USA **78**, 2179-2183.
3. Dickerson, R.E. and Drew, H.R. (1981) J. Mol. Bio. **149**, 761-786.
4. Dickerson, R.E. (1983) J. Mol. Bio. **166**, 419-441.
5. Drew, H.R. and Dickerson, R.E. (1981) J. Mol. Bio. **152**, 723-736.
6. Wang, A.H.-J., Quigley, G.J., Kolpak, F.J., van der Marel, G., van Boom, J.H. and Rich, A. (1981) Science **211**, 171-126.
7. Wang, A.H.-J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G. and Rich, A. (1979) Nature **282**, 680-686.
8. Shakked, Z., Rabinovitch, D., Cruse, W.B.T., Egert, E., Kennard, O., Sala, G., Salisbury, S.A. and Viswamitra, M.A. (1981) Proc. Roy. Soc. London Ser. B **213**, 479.

9. Reid, D.G., Salisbury, S.A., Bellard, S., Shakked, Z. and Williams, D.H. (1983) *Biochem.* 27, 2019-2025.
10. Hare, D.R., Wemmer, D.E., Chou, S.-H., Drobny, G. and Reid, B.R. (1983) *J. Mol. Bio.* 171, 319-336.
11. 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. and van Boom, J.H. (1983) *J. Biomolec. Stereodyn.* 1, 115-129.
12. Haasnoot, C.A.G., den Hartog, J.H.J., de Rooij, J.F.M., van Boom, J.H. and Altona, C. (1980) *Nuc. Acids Res.* 8, 169-181.
13. Kallenbach, N.R., Ma, R.-I., Wand, A.J., Veeneman, G.H., van Boom, J.H., and Seeman, N.C. (1983) *J. Biomolec. Stereodyn.* 1, 159-168.
14. Scheek, R.M., Russo, N., Boelens, R., Kaptein, R. and van Boom, J.H. (1983) *J. Am. Chem. Soc.* 105, 2914-2916.
15. Scheek, R.M., Boelens, R., Russo, N., van Boom, J.H., and Kaptein, R. (1984) *Biochem.* 23, 1371-1376.
16. Feigon, J., Leupin, W., Denny, W.A. and Kearns, D.R. (1983) *Biochem.* 22, 5943-5951.
17. Sanderson, M.R., Mellema, J.-R., van der Marel, G.A., Wille, G., van Boom, J.H. and Altona, C. (1983) *Nucl. Acids Res.* 11, 3333-3347.
18. Weiss, M.A., Patel, D.J., Saver, R.T. and Karplus, M. (1984) *Proc. Natl. Acad. Sci. USA* 81, 130-134.
19. Wuthrich, K., Wider, G., Wagner, G. and Braun, W. (1982) *J. Mol. Bio.* 155, 311-319.
20. Billeter, M., Braun, W. and Wuthrich, K. (1982) *J. Mol. Bio.* 155, 321-346.
21. Wagner, G. and Wuthrich, K. (1982) *J. Mol. Bio.* 155, 347-366.
22. Scheffler, I.E., Elson, E.L. and Baldwin, R.L. (1968) *J. Mol. Bio.* 36, 291-304.
23. Scheffler, I.E., Elson, E.L. and Baldwin, R.L. (1970) *J. Mol. Bio.* 48, 145-171.
24. Baldwin, R.L. (1971) *Accts. of Chem. Res.* 4, 265.
25. Meier, B.H. and Ernst, R.R. (1979) *J. Am. Chem. Soc.* 101, 6441.
26. Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) *J. Chem. Phys.* 71, 4546-4553.
27. Macura, S., Huang, Y., Suter, D. and Ernst, R.R. (1981) *J. Mag. Res.* 43, 259-281.
28. Calladine, C.R. (1982) *J. Mol. Bio.* 161, 343-352.
29. Wemmer, D.E., Chou, S.-H. and Reid, B.R. (1984) *J. Mol. Bio.* 180, 41-60.
30. Wemmer, D.E., Hare, D.R., Chou, S.-H. and Reid, B.R. (To be published).
31. Forsen, S. and Hoffman, R.A. (1963) *J. Chem. Phys.* 39, 2892-2901.
32. Forsen, S. and Hoffman, R.A. (1963) *Acta. Chem. Scand.* 17, 1787.
33. Feigon, J., Wang, A.H.-J., van der Marel, G.A., van Boom, J.H. and Rich, A. (1984) *Nucl. Acids Res.* 12, 1243-1263.
34. Klevit, R.E., Wemmer, D.E., Chou, S.-H. and Reid, B. R. (To be published).
35. Tran-Dinh, S., Tabovry, J., Neumann, J.-M., Huyn-Dinh, T., Genissel, B., d'Estaintot, B.L. and Igolen, J. (1984) *Biochem.* 23, 1362-1371.
36. Marky, L.A., Blumenfeld, K.S., Kozlowski, S. and Breslauer, K.J. (1983) *Biopolymers* 22, 1247-1257.
37. Marky, L.A. and Breslauer, K.J. (1982) *Biopolymers* 21, 2185-2194.
38. Ornstein, R.L. and Fresco, J.R. (1983) *Biopolymers* 22, 1979-2000.
39. Sinden, R.R., Broyles, S.S. and Pettijohn, D.E. (1983) *Proc. Natl. Acad. Sci. USA* 80, 1797-1801.
40. Sinden, R.R. and Pettijohn, D.E. (1984) *J. Biol. Chem.* 259, 6593-6600.
41. States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) *J. Mag. Res.* 48, 286-292.