Structure and folding dynamics of a DNA hairpin with a stabilising d(GNA) trinucleotide loop: influence of base pair mis-matches and point mutations on conformational equilibria†

Graham D. Balkwill, Huw E. L. Williams and Mark S. Searle*

Received 17th November 2006, Accepted 22nd December 2006
First published as an Advance Article on the web 25th January 2007
DOI: 10.1039/b616820e

Hairpins are known to play specific roles in DNA– and RNA–protein recognition. Various disease states are thought to originate from the ill-timed formation of a hairpin loop during transcription, particularly in the context of triplet repeats which are associated with myotonic dystrophy, fragile X syndrome and other genetic disorders. An understanding of nucleic acid folding mechanisms requires a detailed appreciation of the timescales of these local folding events, a characterisation of the conformational equilibria that exist in solution and the influence of point mutations on the relative stabilities of the different species. We investigate using NMR and CD spectroscopy the structure and dynamics of a DNA hairpin containing a highly stabilising cGNAg loop. The single-stranded 13-mer 5′-d(GCTAC**GNA**GTCGC) with N = T folds to form a hairpin structure which accommodates a C–T mis-matched base pair within the double-stranded stem region. The hairpin is in equilibrium with a double-stranded duplex form with the mixture of two interconverting conformations in slow exchange on the NMR timescale (1–2 s^{-1} at 308 K). We are able to characterise the dynamics of the interconversion process by NMR magnetisation transfer and by CD stopped-flow kinetic experiments. The latter shows that the hairpin folds too rapidly to detect by this method (>500 s^{-1}) and forms in a "kinetic overshoot" followed by a much slower equilibration to a mixture of conformations (~0.13 s^{-1} at 298 K). A point mutation that converts the G**T**A to a G**A**A loop sequence destabilises the intermolecular duplex structure and enables us to unambiguously assign the various dynamic processes that are taking place.

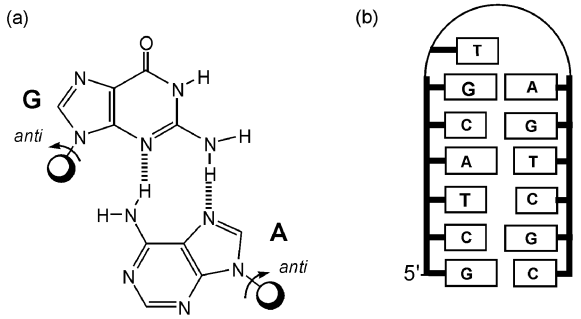
Introduction

Nucleic acids (DNA and RNA) can adopt a remarkable range of structural topologies that are closely linked to biological function. The formation of hairpin loops in single-stranded DNA (ssDNA) is known to play a key role in gene expression, DNA recombination and DNA transposition.[1–3] In the case of RNA, the formation of complex tertiary structures appears to be nucleated by local folding events involving hairpin loops that subsequently coalesce in a hierarchical assembly of secondary structure elements.[4,5] However, the kinetics are complex with evidence for parallel pathways and slow folding processes arising from meta-stable mis-folded states.[6] Although hairpins are known to play specific roles in DNA– and RNA–protein recognition, various disease states are thought to originate from the ill-timed formation of a hairpin loop during transcription. DNA hairpins are generated by triplet repeats associated with myotonic dystrophy, fragile X syndrome and other genetic disorders.[7–11] An understanding of nucleic acid folding mechanisms requires a detailed appreciation of the timescales of these local folding events, of the conformational equilibria that exist in solution and their susceptibility to point mutations. Recent investigations have shown that flexible hairpin loops are able to fold on timescales of the order of milliseconds to microseconds depending on sequence length and base composition.[12–16] Large loops result in a broad transition state ensemble characterised by a distribution of opening and closing rates that lead to non-Arrhenius folding kinetics.[12] Statistical mechanical models that embrace multiple folding pathways and mis-folded states appear to account for the complex kinetics.[13]

A number of mini-hairpin sequences with a GNA trinucleotide loop (N = G, A, C or T) have been reported with remarkably high stabilities.[17,18] These particularly stable hairpins occur in nature within the replication origin of both phage ϕX174 and herpes simplex virus, as well as being identified within the promoter region of phage N4 double-stranded DNA.[19,20] The short GNA loop folds to form a sheared G–A base pair with the unpaired nucleotide (N) stacking on the guanine base (Fig. 1a and b). The high stability appears to correlate with the presence of a CG base pair in the adjacent flanking part of the hairpin stem region, resulting in particularly favourable C–G on G–A base pair stacking within the cGNAg motif (where "c" and "g" are the first Watson–Crick paired nucleotides in the double-stranded stem region).[21,22] The stabilising effects of the cGNAg loop has enabled us to exploit this novel structural motif in the design of a number of different hairpin sequences to study drug recognition

Centre for Biomolecular Sciences, School of Chemistry, University Park, Nottingham, NG7 2RD, UK. E-mail: mark.searle@nottingham.ac.uk; Tel: +44 (0)115 951 3567

† Electronic supplementary information (ESI) available: Magnetisation transfer data demonstrating the interconversion of the GTA 13-mer d(GCTACGTAGTCGC) between the two conformational states; 1D NMR spectra of the base H6/H8 region of the 13-mer showing the concentration sensitivity of the resonances from the duplex species. See DOI: 10.1039/b616820e



(c) 5'-d($G_1C_2T_3A_4C_5\mathbf{G}_6\mathbf{T}_7\mathbf{A}_8G_9T_{10}C_{11}G_{12}C_{13}$)

5'-G C T A C **G T A** G T C G C
C G C T G **A T G** C A T C G-5'

(d) 5'-d($G_1C_2T_3A_4C_5\mathbf{G}_6\mathbf{A}_7\mathbf{A}_8G_9T_{10}C_{11}G_{12}C_{13}$)

5'-G C T A C G **A A** G T C G C
C G C T G **A A G** C A T C G-5'

Fig. 1 (a) Structure of a G–A base pair. (b) Schematic of the 13-mer hairpin sequence containing a GTA trinucleotide loop with a T–C mis-matched base pair in the stem region. (c) Number scheme within the nucleotide sequence and the alternative intermolecular duplex form showing the contiguous alignment of 8 Watson–Crick base pairs with overhanging 3′-ends. (d) Nucleotide sequence of the GAA hairpin and the duplex alignment that introduces destabilising A–A base pairs into the stem region of the duplex.

of bulged bases and mis-matched base pairs which are otherwise highly destabilising in short double-stranded duplex DNA.[23–25] We describe NMR and CD studies of a single-stranded 13-mer, 5′-d(GCTAC**GNA**GTCGC), which is able to accommodate a C–T mis-matched base pair within the double-stranded stem region of a hairpin motif. However, a single substitution of N = A for N = T within the **GNA** loop sequence produces a structurally heterogeneous mixture of two interconverting conformations which are in slow exchange on the NMR timescale. We are able to characterise the conformation of these species and the dynamics of the interconversion process by NMR magnetisation transfer and by CD stopped-flow kinetic experiments.

Results

Conformational heterogeneity in the folding of a single-stranded 13-mer

The conformation and dynamics of a mis-matched pyrimidine C–T base pair was examined in the context of the 13-mer 5′-d(GCTAC**GTA**GTCGC) in which the C–T pair is accommodated within the double-stranded stem region of a hairpin containing the GNA loop, with N = T. ^{1}H NMR at 298 K reveals a complex set of aromatic signals consisting of one subset of resonances with relatively narrow line widths which are insensitive to temperature and a second set that demonstrate considerable temperature-dependent line broadening. The latter are poorly resolved at 288 K but are considerably sharper at 308 K, where the resonances from the two species have similar intensities and line widths (Fig. 2a). In 2D NOESY and ROESY spectra we observe strong chemical exchange cross-peaks that demonstrate that the two species are interconverting on a relatively slow timescale.

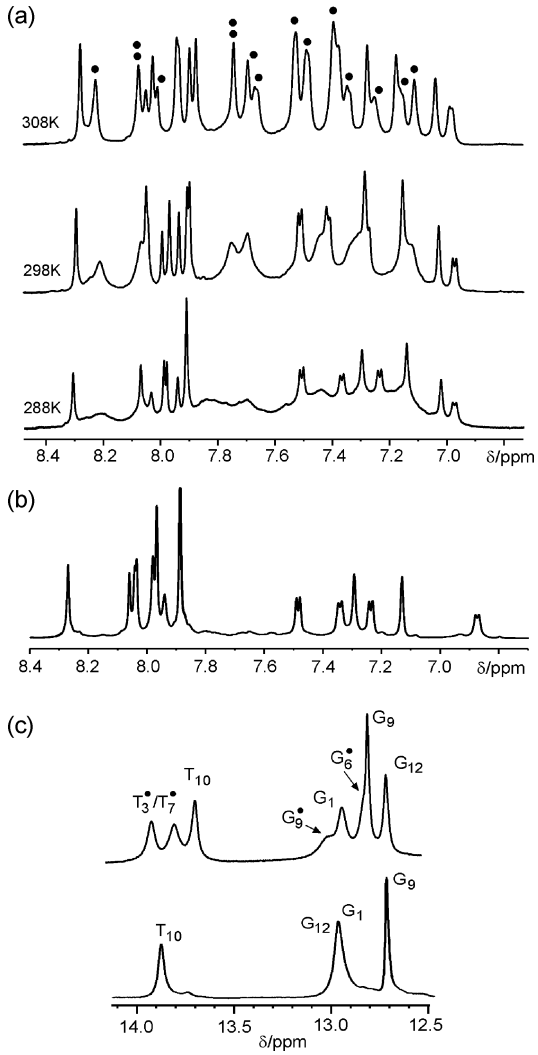


Fig. 2 (a) 1D NMR spectra of the base H6/H8 region of the GTA 13-mer showing the temperature-dependent changes in line widths between 288 K and 308 K. At low temperature the resonances of the duplex form become increasingly broadened due to chemical exchange; however, at 308 K the hairpin and duplex forms appear to have similar linewidths, and resonance intensities equating to a ~2 : 1 ratio of hairpin to duplex. (b) 1D NMR spectra of the base H6/H8 region of the GAA 13-mer at 298 K showing a single species in solution corresponding to the hairpin conformer. (c) Portion of the ^{1}H NMR spectrum of the GTA 13-mer at 288 K showing exchangeable NH resonances corresponding to the two conformers (hairpin and duplex), but only a single hairpin species is observed for the GAA hairpin (lower part of (c)). Throughout, resonances for the duplex form are indicated with a bullet point (•).

The set of resonances that are sensitive to temperature also show a sensitivity to the concentration of oligonucleotide. Although significant exchange broadening occurs at low temperatures and high solute concentrations (>2 mM), the broadening effects are not evident at lower oligonucleotide concentrations (≤1 mM),

indicative of an intermolecular association process. In contrast, the signals from the other species remain sharp and well-resolved at all concentrations, consistent with the folding of the monomeric 13-mer around the GTA loop to form a stable hairpin motif, as proposed in Fig. 1b. It is evident that the same sequence is also fortuitously able to form a two-fold symmetrical intermolecular duplex which is stabilised by the possibility of eight contiguous Watson–Crick base pairs with 3′-overhangs (Fig. 1c). Consistent with this hypothesis, at 288 K these two distinct species readily account for the number of observed NH resonances between 12.5 and 14.0 ppm (Fig. 2c). Three clearly resolved thymine N3-H resonances are apparent between 13.6 and 13.9 ppm, which we assign on the basis of temperature-dependent line broadening effects and NOE data. Two of these, which show equal intensity, arise from the duplex form of the 13-mer (T_3* and T_7*) and the third from the single A–T base pair (T_{10}) in the stem of the hairpin. We are unable to resolve hairpin resonances for T_3 and T_7, which appear not to adopt stable hydrogen-bonded base pairs. This is consistent with the exposure of T_7 in the GTA loop and with T_3 forming at best a weakly hydrogen-bonded structure at the T–C mis-match site. Despite the different sensitivities of the line widths of the two species to temperature, the relative populations of hairpin and duplex (~2 : 1 on the basis of signal intensities in Fig. 2a) show little temperature variation, indicating that both the Gibbs free energy is temperature-independent and that the entropic cost of intermolecular and intramolecular assembly is remarkably similar ($\Delta\Delta S \sim 0$).

The proposed equilibrium between hairpin and duplex could be significantly perturbed if one or other species could be selectively destabilised by a single point mutation in the 13-mer. This can be achieved through a T → A substitution that converts the GTA hairpin loop to the equally stable GAA loop while introducing destabilising A–A mismatches in the duplex when the two strands are aligned as shown in Fig. 1d. Alternative strand alignments are not able to increase the number of Watson–Crick base pairs, with the prediction that this mutation should significantly destabilise any double-stranded conformations. This mutation has the desired effect and totally eliminates the species attributed to intermolecular duplex formation (Fig. 2b), leaving the hairpin form as the only stable folded species over a wide range of concentrations and temperatures. This considerably simplifies the 1D and 2D NMR analysis and confirms the earlier assignments.

Structure of the DNA hairpin with a C–T mismatch

The hairpin conformation has been characterised in structural detail by 2D NMR using the methodologies previously described.[23–25] In NOESY spectra at 308 K, the presence of two similarly populated interconverting species gives rise to a large number of NOE cross-peaks further complicated by chemical exchange effects. However, by selecting conditions where the signals from the duplex form are significantly attenuated due to line broadening effects and where the rate of interconversion is much reduced (for example at ≤298 K), we are able to collect a complete set of NOESY spectra essentially of only the hairpin conformer. The base H6/H8 to deoxyribose H2′/H2″ region of the NOESY spectra collected at 308 K and 298 K illustrate the degree of spectral editing that is possible, with the duplex form becoming essentially "silent" at 298 K (Fig. 3a). The improved resolution has enabled us to obtain a near-complete assignment and quantitative analysis of NOE intensities over a range of mixing times to generate distance and torsion angle restraints. Similar spectra are obtained on dilute samples where the duplex form is not significantly populated; however, the signal-to-noise ratio in this case does not permit the same detailed level of analysis.

The overall fold of the hairpin is evident at a more qualitative level on the basis of key patterns of NOE connectivities. The conformation of the 5′-GTA loop is characterised by a number of features previously described in our work and that of others.[17,18,23,25] The deoxyribose H4′ of the loop thymine (T_7) undergoes a ~2 ppm upfield shift as a consequence of the stacking of the thymine base on the sheared G–A base pair. The pattern of NOEs readily identifies this resonance in this unusual environment. The H2′ of C5 also undergoes a ~0.6 ppm upfield shift through stacking interactions associated with the G–A base pair. The observation of *anti*-glycosidic torsion angles is consistent with the formation of a sheared G–A base pair, as originally proposed by Hirao *et al.*[17,18] The guanine imino resonance, which is not utilised in hydrogen bonding in the purine–purine pair, is visible at ~10.7 ppm and broadens considerably above 303 K. The turn geometry results in the thymine base stacking primarily with the guanine of the G–A pair. This is evident from either very weak, or from the absence, of sequential NOE connectivities between the T and A bases/sugars within the GTA loop; however, the corresponding cross-peaks are visible on the 5′-side across the GpT step, defining the geometry and base stacking arrangement within the GTA loop. The T–C mismatch site shows that both pyrimidine bases appear to be accommodated within the double-stranded stem region of the hairpin. We observe no deviation of the pattern of sequential NOE intensities from those expected for normal double-stranded DNA (Fig. 3a). NOE cross-peaks from the pyrimidine base protons to neighbouring base and sugar protons suggest that the two bases are accommodated within the base stack rather than displaced into more solvent- exposed positions. A cross-strand NOE from A4H2 to C11H1′ further confirms the intrahelical stacking of C11. The absence of a clearly resolved T_3 NH seems to preclude the formation of a stable hydrogen-bonded T–C pair. A broadened resonance is evident at 11.5 ppm, which we tentatively assign to the T3NH in a weakly hydrogen-bonded environment.

NOESY spectra collected in both H_2O and D_2O solutions allowed us to assimilate 213 NMR-derived distance and torsion angle restraints as a basis for generating an ensemble of structures using NOE-restrained molecular dynamics (MD) simulations (Fig. 4). Following previously reported equilibration protocols,[23,25] the hairpin was subjected to 1 ns of fully restrained dynamics at 300 K using an explicit solvent model using the AMBER package (v6). The NOE data, and indeed the calculated structures, show that the mis-matched T–C base pair is readily accommodated within the base stack of the hairpin stem region (Fig. 4). The structure was also subjected to a further 1 ns of unrestrained MD using the same protocol to investigate the stability and dynamics of the T–C base pair. Pseudo hydrogen-bonding force constants were calculated for each base pair, as described by Hirst and Brooks (Fig. 4c).[26] The T–C pair shows a considerably reduced force constant in unrestrained simulations, compared with other Watson–Crick base pairs in the stem region of the hairpin, demonstrating large deviations from an ideal stable hydrogen-bonded structure that are consistent with the absence

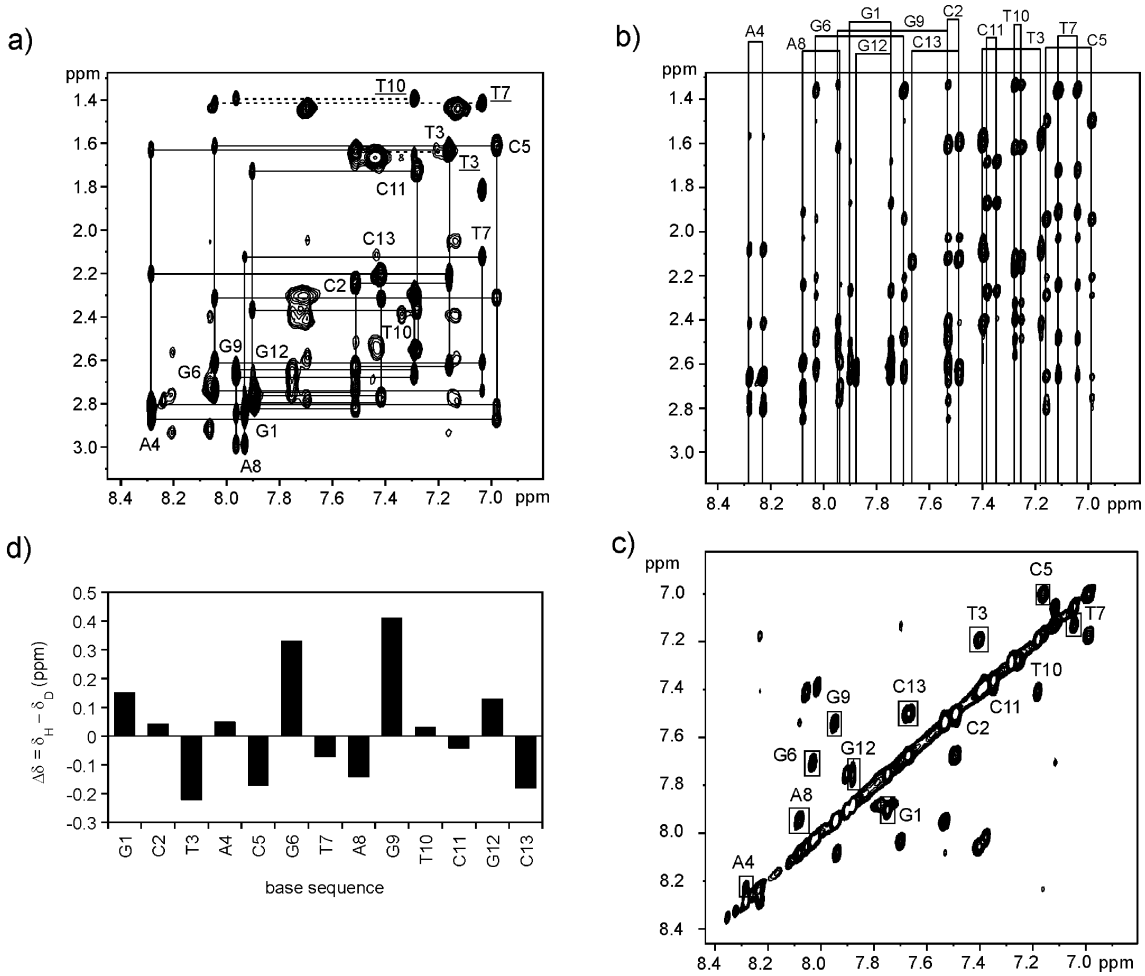


Fig. 3 2D NOESY data collected for the GTA 13-mer at (a) 298 K and (b) 308 K illustrating the same base H6/H8 to deoxyribose H2′/H2″ region. In (a), signals from the duplex are sufficiently broadened that only NOE cross-peaks for the hairpin are visible. Labels indicate the positions of the base H6/H8 signal and cross-peaks arising from thymine methyl groups are underlined. Sequential connectivities linking G_1 through to C_{13} are highlighted. In (b), assignments are indicated for the base protons (H6/H8) of both the hairpin and duplex forms, with the signals connected along the upper axis of the plot. In (c), the portion of the NOESY spectrum of the H6/H8 region at 308 K shows strong chemical exchange cross-peaks between the hairpin and duplex forms (base labelling shown). In (d), chemical shift differences ($\Delta\delta = \delta_H - \delta_D$) for base H6/H8 are shown *versus* sequence as a histogram to illustrate those protons that show the biggest change in structural environment.

of a clearly resolved T_3 NH resonance in the experimental NMR data.

Assignment of the duplex form of the 13-mer

We are unable to identify a unique set of experimental conditions under which only the duplex form is populated which would permit a detailed structural analysis. The NOESY spectra at 308 K (Fig. 3b) show both species to be equally populated. A combination of the pattern of NOE connectivities and chemical exchange effects (Fig. 3c), which enable us to transfer the assignments from the hairpin form to that of the duplex, have enabled us to assign the majority of base H6/H8 and deoxyribose H1′ and H2′/H2″ resonances in the duplex form. Comparison of base proton H6/H8 chemical shifts shows that there are some significant differences between the two structural environments ($\Delta\delta = 0.3$–0.4 ppm), particularly for nucleotides close to the GTA hairpin loop (notably G6 and G9; Fig. 3d).

Folding and unfolding kinetics from magnetisation transfer studies

The slow switching that we observe by NMR between the hairpin and duplex forms has enabled us to determine the rate of interconversion (folding/unfolding kinetics) using magnetisation transfer methods through selective resonance inversion using a 180° Gaussian-shaped on-resonance pulse.[27,28] We are unable to observe signals for the single-stranded form, demonstrating that it is significantly higher in energy (meta-stable state) than either the duplex or hairpin, approximating the kinetics to a two-state model. At 308 K several pairs of well-resolved resonances corresponding to the two structured environments were selected for study (A4 H8 and T7 H6). The data for A4 H8 are shown in Fig. 5, where the recovery of the intensity of the inverted resonance for the duplex form at 8.23 ppm is shown together with magnetisation transfer effects on the corresponding signal for the hairpin conformer. In each case the data fit well to a double exponential with a random distribution of residuals which enables us to determine the respective rate constants for hairpin → duplex

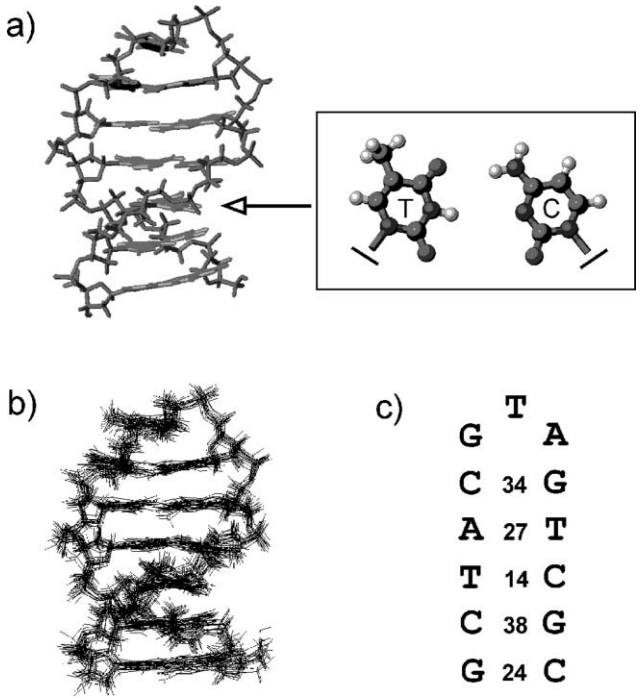


Fig. 4 Structure of the hairpin conformation of the GTA 13-mer d(GCTACGTAGTCGC) generated using NMR-restrained molecular dynamics simulations. (a) Energy-minimised mean structure of the hairpin showing the geometry of the GTA loop and the internal stacking of the mis-matched T–C base pair; the geometry of the T–C base pair is shown in the inset. (b) Ensemble of 10 NMR restrained structures of the hairpin derived from molecular dynamics simulations (snap shots taken during the final 100 ps of a 1 ns simulation; pairwise RMSD 0.89 Å). (c) Calculated average hydrogen-bonding force constants within each base pair.

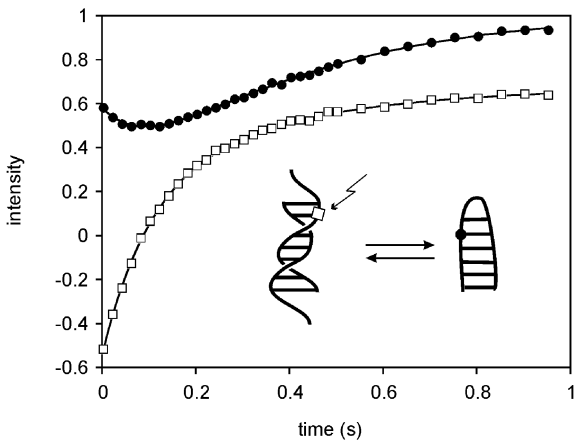


Fig. 5 Magnetisation transfer data at 308 K demonstrating the interconversion of the GTA 13-mer d(GCTACGTAGTCGC) between the two conformational states (see inset schematic) following the inversion of A4 H8 for the duplex form at 8.23 ppm. Open symbols represent the recovery of the inverted peak (duplex) and closed symbols the intensity of the signal for A4 H8 in the hairpin conformer (8.28 ppm). The data show in each case the best fit to a double exponential with good agreement between relaxation constants for each curve ($\lambda_1 = 2.79$ s^{-1} and $\lambda_2 = 9.89$ s^{-1} (□); $\lambda_1 = 2.57$ s^{-1} and $\lambda_2 = 10.01$ s^{-1} (●)), and with data for T7 H6 (not shown).

(k_{HD}) and duplex → hairpin (k_{DH}) at 308 K as being 1.1 ± 0.3 s^{-1} and 2.4 ± 0.6 s^{-1} respectively (±standard deviation from four data sets), giving an equilibrium constant $K_{eq} = k_{DH}/k_{HD} \sim 2.2$. The rate of interconversion is at least three orders of magnitude slower than reported for the folding of isolated hairpins from a single strand,[12–16,29] but consistent with rates of conformational interconversion between two hairpin forms.[30]

Thermal, equilibrium and stopped-flow CD studies of hairpin folding

Previous kinetic studies of the fast dynamics of hairpin folding (≤ 1 ms) have required T-jump and fluorescence correlation spectroscopic methods to provide the necessary fast time resolution.[12–16] The stopped-flow rapid mixing approach using far-UV CD detection is typically limited by instrumental dead times (> 2 ms) that reflect the mechanics of rapid dilution of buffer solutions. Initially we examined the hairpin unfolding transition by CD on a 6.0 μM sample in 10 mM phosphate buffer at pH 7.0. The thermal unfolding transition was monitored at 277 nm and gave a sigmoidal unfolding curve with a T_m of 52 °C and an enthalpy of unfolding of 104 (±5) kJ mol^{-1}. Parallel melting studies with the GAA hairpin analogue suggests that the loop substitution of T for an A increases the stability by 3 °C. Chemical denaturation with GdmCl (0–8 M) produced similarly large changes in ellipticity at 250 and 280 nm; however, denaturation was less clearly associated with a sigmoidal transition. Thus, under equilibrium conditions, we observe significant changes in ellipticity as a function of denaturant concentration. The time-dependence of these changes in ellipticity upon rapid dilution from highly denaturing conditions (8 M GdmCl) was investigated by stopped-flow CD measurements.

Although we expected the rate of hairpin folding to be fast, surprisingly, we were able to detect time-dependent changes in the CD spectrum using stopped-flow methods.[31] Rapid 10 : 1 dilution of denatured 13-mer in 8 M GdmCl into different refolding buffers enabled refolding data to be collected over a range of final denaturant concentrations between 0.7 and 5.3 M GdmCl at 298 K. The data in Fig. 6b show the change in ellipticity for refolding into 1.6 M GdmCl. The ellipticity initially increases, as predicted by the equilibrium unfolding studies, but with a rate constant too fast to measure (>500 s^{-1}) on our instrument (dead-time ~2–3 ms). However, a subsequent decay (relaxation) process occurs which is much slower and fits to a single exponential with a rate constant of 1.8 ± 0.04 s^{-1}. Kinetic traces for refolding at higher denaturant concentrations (up to 4 M GdmCl) also give rise to the same single exponential decay following the initial rapid increase in ellipticity; however, the amplitude of the slower process decreases and is not detectable at [GdmCl] > 4 M. The data at 5.3 M (Fig. 7a) demonstrate that all spectroscopic changes are over within the dead-time of the instrument. The logarithm of the rate constant varies linearly with [GdmCl] < 4 M, enabling an extrapolated rate constant to be determined in water alone of 0.13 s^{-1} under the conditions and final concentrations employed (Fig. 7b).

At higher concentrations of denaturant the duplex could be preferentially destabilised such that this slow equilibration process is no longer visible. To test this hypothesis we also examined the stopped-flow kinetics of the GAA hairpin, which is unable to form significant populations of the duplex form even at millimolar concentrations. The data were collected under identical

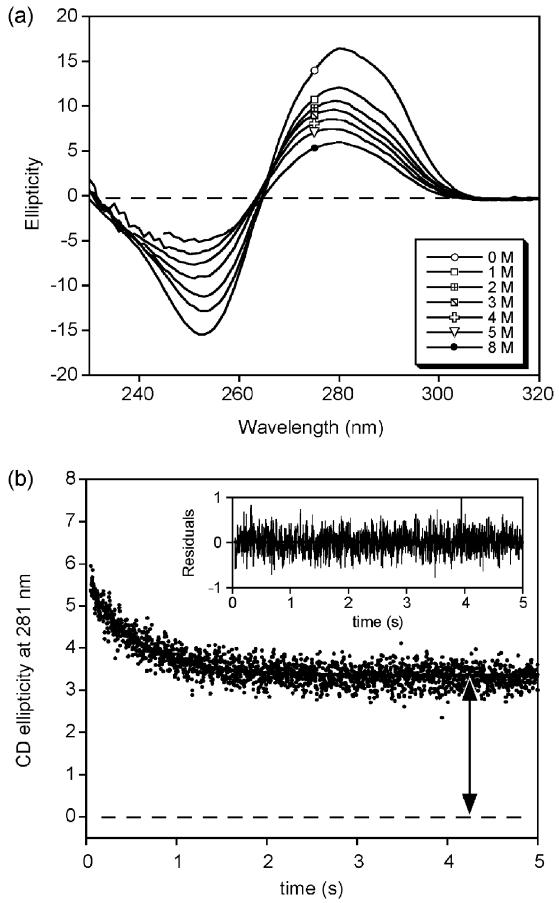


Fig. 6 (a) CD spectra of the GTA 13-mer at 298 K and 100 mM NaCl/10 mM phosphate buffer at pH 7.0 and oligonucleotide concentration of 6.5 μM showing the change in ellipticity as a function of the concentration of the denaturant guanadinium chloride (GdmCl). The minimum at 255 nm and maximum at 280 nm are characteristic of double-stranded DNA. The intensity at both wavelengths decreases uniformly with denaturant concentration as the single-stranded form becomes populated. (b) Stopped-flow kinetic data for the folding of the hairpin monitored by circular dichroism spectroscopy. Data were collected at 298 K following 1 : 10 mixing, resulting in dilution of denaturant from 8 M GdmCl to 1.6 M. The ellipticity increases rapidly within the dead-time of the experiment (~2–3 ms), as predicted by the equilibrium unfolding studies, but with a rate constant too fast to measure. The subsequent decay is much slower and fits to a single exponential process with a rate constant of 4 s^{-1}. The residuals from the subtraction of the single exponential fit to the data points are shown in the inset. The double-headed arrow represents the net change in ellipticity, which is in good agreement with the difference in equilibrium intensities shown in (a).

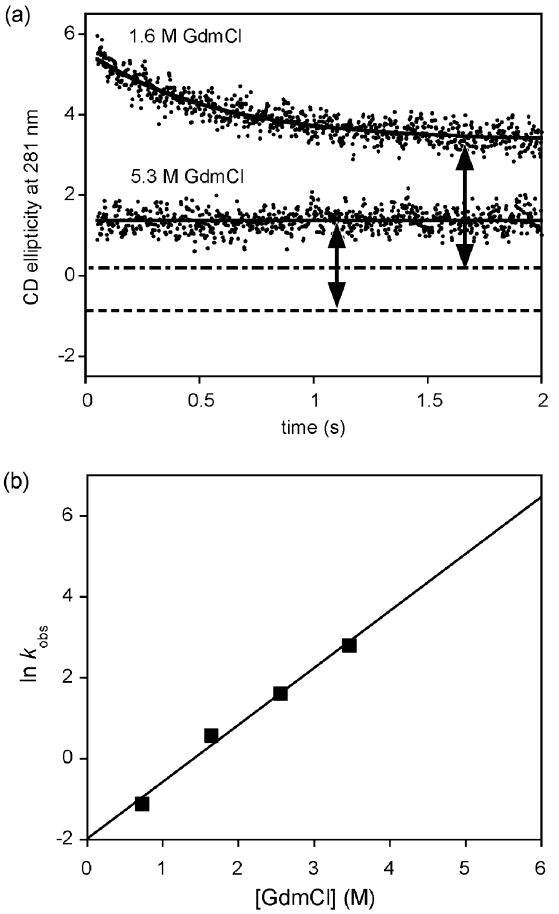


Fig. 7 (a) Stopped-flow kinetic data at final denaturant concentrations of 1.6 M and 5.3 M GdmCl. The ellipticity in the latter case is arbitrarily shifted to improve visibility; however, although the ellipticity increases rapidly within the dead-time of the experiment, we no longer detect the slow decay process. (b) Dependence of the rate constants for the decay process on the denaturant concentration derived from a single exponential fit to kinetic data collected at [GdmCl] < 4 M. The extrapolated rate of duplex formation in water at 298 K and at 6.5 μM concentration is 0.13 s^{-1}. Data collected under the same conditions for the GAA 13-mer do not show the slow equilibration phase at any final concentration of denaturant.

conditions and suggest the same rapid formation of hairpin in the dead-time of the instrument; however, now there is no evidence for the subsequent slow equilibration process. We are led to the conclusion that GdmCl is much less effective at denaturing oligonucleotides than it is proteins. The lack of a well-defined sigmoidal denaturation curve supports this conclusion, suggesting that we have a mixture of species in solution even at low solute concentrations. The high ionic strength of GdmCl appears to balance the effects of chemical denaturation with cationic stabilisation, resulting in the duplex being populated at solute concentrations as low as 6 μM. High salt concentrations are known to lead to faster rates of renaturation because nucleation is more likely to occur when charged phosphate groups are shielded. Our observations show that the rate constants associated with the slow folding phase increase with GdmCl concentration, demonstrating the balance between the renaturing effects associated with the guanidinium cation and ultimately its denaturing capacity at high concentrations that results in the elimination of the slow kinetic phase at >4 M GdmCl.

The other intriguing aspect of the data shown in Fig. 6b is that the slow phase is associated with a decay process (decrease in ellipticity) after the initial increase in ellipticity following fast formation of the hairpin. Thus, we appear to observe an initial "over-shoot" in which rapid over-population of the hairpin structure occurs under kinetic control. This is followed by a second slower equilibration process to establish thermodynamic equilibrium (see Fig. 8). This initial over-shoot has been identified in the context of the formation of single-stranded DNA quadruplex structures where either the parallel or anti-parallel fold could become over

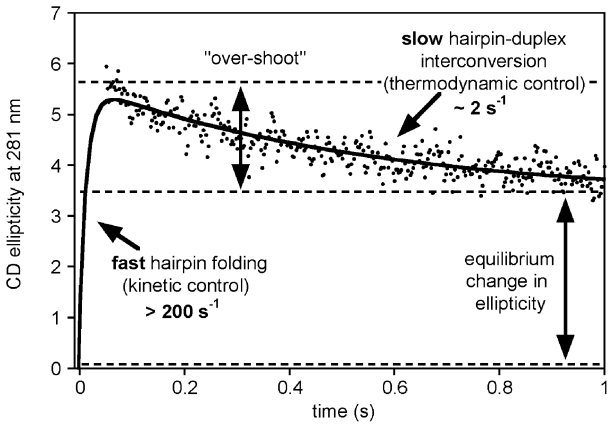


Fig. 8 Summary of the kinetic processes taking place in stopped-flow CD experiments on the GTA 13-mer. Fast hairpin folding occurs (>500 s^{-1}) under kinetic control, which is too fast to characterise by stopped-flow methods, followed by slow equilibration in which hairpin–duplex interconversion enables thermodynamic equilibrium to be established. This slow decay phase represents a kinetic "over-shoot" where the hairpin is overpopulated in the first folding event.

populated in T-jump experiments, with the system subsequently evolving to reach the new equilibrium.[32] Similar effects have been observed in the formation of helical peptides where rapid propagation appears to follow nucleation leading to an initial ellipticity higher than that observed at equilibrium, followed by a relaxation to a more dynamic structure with flexible N- and C-termini.[33]

Conclusions

Mini-hairpin motifs of the type cGNAg have been shown to impart remarkably high hairpin stabilities to relatively short oligonucleotide sequences with a T_m of 72 °C reported for sequences as short as a heptamer.[17,18] We have exploited this stabilising motif in the design of sequences to probe the structure and thermodynamic stability of hairpins carrying bulged-bases of mis-matched base pairs, and have investigated the interaction of drug molecules with these sequences.[23–25] In this work we have investigated the conformational dynamics of introducing a T–C mis-match within the stem region and have shown that the choice of loop sequence (GTA *versus* GAA) can result in the conformational equilibrium between competing hairpin and intermolecular duplex formation. The rate of interconversion of these conformers has been investigated by magnetisation transfer NMR experiments and by stopped-flow CD spectroscopy. In the latter case, hairpin formation appears to be too rapid to detect directly due to instrumental limitations, in agreement with the measurements of others.[12–16] However, we observe a novel initial over-population of the hairpin in a kinetically controlled process, and subsequent slow re-equilibration to thermodynamic equilibrium where some double-stranded duplex appears to be populated. Although hairpins are known to play specific roles in DNA– and RNA–protein recognition, various disease states are thought to originate from the ill-timed formation of a hairpin loop during transcription, particularly in the context of triplet repeats associated with myotonic dystrophy, fragile X syndrome and other genetic disorders.[7–11] An understanding of nucleic acid folding mechanisms requires a detailed appreciation of the timescales of these local folding events, of the conformational equilibria that exist in solution and their susceptibility to point mutations.

Experimental

Materials

Oligonucleotides were synthesised by standard solid-phase methods and purified by reverse-phase HPLC. 2D NMR data were collected at 600 MHz using standard pulse sequences and oligonucleotide samples in the concentration range 0.5–2.0 mM in 10 mM phosphate buffer at pH 7.0 and at various [NaCl] in the range 0–150 mM.

NMR experiments and molecular modelling

Magnetisation transfer experiments followed reported methodologies[27,28] and employed a selective inversion pulse with an excitation band width of 20–30 Hz optimised to minimise spill-over effects to adjacent resonances. For structural modelling, a total of 213 restraints (137 NOE distances, 65 torsion angle restraints and 11 hydrogen-bonding restraints) were employed. Derivation of the interproton NOE distance restraints have been described previously.[23] Torsion angle restraints were estimated using high-resolution DQF-COSY data, analysis of coupling constants from 1D spectra and examination of NOESY-determined distances incorporating sugar protons.[34] Hydrogen-bonding restraints between bases were included when imino resonances showed slow exchange rates. Molecular dynamics simulations were performed using the AMBER 6.0 suite of programs (http://amber.scripps.edu/doc6/install.html) and explicit solvation models, as previously described.[23,25] The starting structure was generated by combining an idealised duplex structure, obtained from the NUCGEN program within AMBER 6.0, with an existing GTA loop motif. This structure was subsequently edited using the LEaP module within AMBER and was allowed to fully equilibrate before restrained molecular dynamics simulations were performed. The restraints were gradually applied in the form of square well potentials with a force constant of 20 kcal mol^{-1} $Å^{-2}$ over a period of 150 ps as the temperature was raised from 100 to 300 K. Restrained molecular dynamics, incorporating all 213 restraints, was then performed for 1 ns. This was followed by the removal of the restraints and a further 1 ns of unrestrained dynamics. Analysis of the autocorrelation function revealed that the structure was judged to be stable during the last 200 ps of the run. As a result, 1 ps snapshots from the last 200 ps of restrained dynamics were taken to generate an ensemble of hairpin conformations displaying a pairwise RMSD of 0.89 Å for all heavy atoms. The final energy-minimised structure only violated 3 restraints >0.5 Å (but less than 0.9 Å). These violated restraints corresponded to weak cross-strand NOEs. Structural analysis was performed using the CARNAL module of AMBER 6.0, and the structures were displayed using MOLMOL.[35]

Hydrogen bonding was analysed in terms of pseudo force constants, as described by Hirst and Brooks.[26] This method of analysis uses a quasiharmonic approximation to examine relative stabilities. Large deviations from ideal hydrogen-bonding

distances give rise to small weak force constants, whereas small deviations give rise to large stable force constants. The calculated value for the T–C pair from unrestrained MD simulations shows that the hydrogen-bonded T–C conformation is not particularly stable (see estimated force constants in Fig. 4).

Kinetic experiments by CD

CD stopped-flow experiments were performed on an Applied Photophysics PiStar spectrophotometer and employed rapid mixing methods to trigger the folding of the 13-mer. The temperature was regulated using a Neslab RTE-300 circulating programmable water bath. Stocks of the 13-mer oligonucleotide were prepared (50 μM in 10 mM phosphate buffer, 100 mM NaCl and 8 M GdmCl at pH 7.2) and rapidly diluted 10-fold with the same buffer containing variable concentrations of denaturant. Rapid dilution lead to a final oligonucleotide concentration of 4.55 μM and denaturant concentrations in the range 0.73–5.27 M. Kinetics were monitored at $\lambda_{max} = 281$ nm, and data were recorded for up to 10 s with 3–5 traces averaged.

Acknowledgements

GDB was supported by the Engineering and Physical Sciences Research Council of the UK, and HELW by Cancer Research UK and the University of Nottingham. We thank Emma Simpson for advice in setting up stopped-flow CD experiments.

References

1 D. B. Roth, J. P. Menetski, P. B. Nakajima, M. J. Bosma and M. Gellert, *Cell*, 1992, **70**, 983.
2 A. Bhasin, I. Y. Goryshin and W. S. Reznikoff, *J. Biol. Chem.*, 1999, **274**, 37021.
3 A. K. Kennedy, A. Guhathakurta, N. Kleckner and D. B. Haniford, *Cell*, 1998, **95**, 125.
4 I. Tinoco, Jr. and C. Bustamante, *J. Mol. Biol.*, 1999, **293**, 271.
5 D. J. Williams and K. B. Hall, *J. Mol. Biol.*, 2000, **297**, 1045.
6 J. Pan, D. Thirumalai and S. Woodson, *J. Mol. Biol.*, 1997, **273**, 7.
7 C. T. Ashley, Jr. and S. T. Warren, *Annu. Rev. Genet.*, 1995, **29**, 703.
8 G. R. Sutherland and R. J. Richards, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 3636.
9 T. A. Kunkel, *Nature*, 1993, **365**, 207.
10 J.-L. Mandel, *Nat. Genet.*, 1994, **7**, 453.
11 J. Petruska, N. Arnheim and M. F. Goodman, *Nucleic Acids Res.*, 1996, **24**, 1992.
12 M. I. Wallace, L. Ying, S. Balasubramanian and D. Klenerman, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 5584.
13 A. Ansari Kuznetsov and Y. Shen, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 7771.
14 G. Bonnet, O. Krichevsky and A. Libchaber, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 8602.
15 Y. Shen, S. V. Kuznetsov and A. Ansari, *J. Phys. Chem. B*, 2001, **105**, 12202.
16 D. J. Proctor, H. Ma, E. Kierzek, R. Kierzek, M. Gruebele and P. C. Bevilacqua, *Biochemistry*, 2004, **43**, 14004.
17 S. Yoshizawa, K. Kawai, K. Watanabe, K. Miura and I. Hirao, *Biochemistry*, 1997, **36**, 4761.
18 I. Hirao, G. Kawai, S. Yoshizawa, Y. Nishimura, K. Ishido, K. Watanabe and K. Miura, *Nucleic Acids Res.*, 1994, **22**, 576.
19 K. Arai, R. Low, J. Kobori, J. Shlomai and A. Kornberg, *J. Biol. Chem.*, 1981, **256**, 5273.
20 M. A. Glucksmann, P. Markiewicz, C. Malone and L. B. Rothman-Denes, *Cell*, 1992, **70**, 491.
21 E. Moody and P. C. Bevilacqua, *J. Am. Chem. Soc.*, 2003, **125**, 2032.
22 E. Moody and P. C. Bevilacqua, *J. Am. Chem. Soc.*, 2003, **125**, 16285.
23 M. L. Colgrave, H. E. L. Williams and M. S. Searle, *Angew. Chem., Int. Ed.*, 2002, **41**, 4754.
24 C. T. Gallagher and M. S. Searle, *Chem. Commun.*, 2003, 1824.
25 H. E. L. Williams, M. L. Colgrave and M. S. Searle, *Eur. J. Biochem.*, 2002, **269**, 1726.
26 J. D. Hirst and C. L. Brooks, *Biochemistry*, 1995, **34**, 7614.
27 J. J. Led, H. Gesmar and F. Abildgaard, *Methods Enzymol.*, 1989, **176**, 311.
28 L.-Y. Lian and G. C. K. Roberts, *NMR of Macromolecules*, ed. G. C. K. Roberts, IRL Press, Oxford, 1993.
29 D. Thirumalai, N. Lee, S. A. Woodson and D. K. Klimov, *Annu. Rev. Phys. Chem.*, 2001, **52**, 751–762.
30 P. Wenter, B. Furtig, A. Hainard, H. Schwalbe and S. Pitsch, *ChemBioChem*, 2006, **7**, 417–420.
31 J. Buchner and T. Kiefhaber, *Protein Folding Handbook*, Wiley-VCH, Weinheim, Germany, 2005.
32 A. T. Phan and D. J. Parel, *J. Am. Chem. Soc.*, 2003, **125**, 15021–15027.
33 D. T. Clarke, A. J. Doig, B. J. Stapley and G. R. Jones, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 7232.
34 L. J. Rinkel and C. Altona, *J. Biomol. Struct. Dyn.*, 1987, **4**, 621.
35 R. Koradi, M. Billeter and K. Wuthrich, *J. Mol. Graphics*, 1996, **14**, 51.