COMMUNICATION

OBC
www.rsc.org/obc

Solution conformation of d($C_4ACAC_4TGT)_2$; an intramolecularly folded i-motif from the insulin minisatellite†

Vandana V. Jolad, Fatima K. Murad, John R. P. Arnold and Julie Fisher*
School of Chemistry, University of Leeds, Leeds, UK LS2 9JT.
E-mail: j.fisher@chem.leeds.ac.uk; Fax: +44 (0)113 34 36565; Tel: +44 (0)113 34 36577

Received 4th April 2005, Accepted 28th April 2005
First published as an Advance Article on the web 5th May 2005

A 28-mer sequence taken from the insulin minisatellite is shown, through NMR and UV thermal melting studies, to form an intramolecular i-motif with two ACA and one TGT loop that persists to near neutral pH and room temperature.

The insulin minisatellite region, or insulin-linked polymorphic region (ILPR), is found upstream of the human insulin gene (Fig. 1).[1,2] There is considerable variability in sequence and length of the ILPR,[3] and each of these factors has implications for expression of the insulin gene. Indeed, such polymorphism has been linked with genetic susceptibility for insulin-dependent diabetes mellitus (IDDM).[4] The normal alleles for insulin are long (n = *ca.* 140, Fig. 1) but in IDDM alleles are short (n = *ca.* 40). It is still not clear why the decrease in repeat length should bring about IDDM susceptibility. However, a possible explanation has been offered for the sequence and length polymorphism in the formation of unusual DNA structures by the repeating sequences.[5] The sequence ($G_4TGTG_4ACAG_4TGTG_4$) taken from the G-rich strand of the ILPR has been shown, as a result of CD and 1D and 2D ^{1}H NMR studies, to form an intramolecularly folded hairpin G-quartet structure.[6] NMR studies of a fragment from the complementary strand, have indicated that the sequence (C_4TGTC_4) forms a hairpin that self-associates to form an intermolecular i-motif.[7] DNA i-motifs are formed when two parallel-stranded C-rich duplexes associate through the formation of C·C^+ base pairs, where C^+ is a protonated cytosine. The base pairs are fully intercalated and the typical relative orientation of the two hairpins is antiparallel.[8]

$$\left(\begin{matrix}ACAG_4TGTG_4\\TGTC_4ACAC_4\end{matrix}\right)_n$$

Fig. 1 The insulin linked polymorphic region, ILPR.

D ^{1}H spectra and molecular modelling investigations have been reported for ($C_4TGTC_4ACAC_4TGTC_4$),[7] another fragment of the ILPR. The spectra were too complicated for analysis although imino proton resonances for both loop and stem residues, characteristic of i-motifs, were observed at around pH 5.7 Two different molecular models were built for the 25-mer, one an intramolecular i-motif, the other a dimer. Subsequently, *in vitro* replication assays on an 84-mer from the ILPR were found to form a structure that blocks the progression of replication, even in the presence of its complementary G-rich strand.[7]

Clearly, the folding of the C-rich strand is important in the function of the ILPR. Although this sequence is expected to form an intramolecular i-motif, to our knowledge the solution conformation of the 'intact-repeat' has not yet been established experimentally. This is the purpose of these investigations. UV thermal melting and ^{1}H NMR experiments have been employed to determine a model for the folding of a 28-mer sequence (Fig. 2) that comprises two complete repeating units from the start of the ILPR.

5' $C_4ACAC_4TGTC_4ACAC_4TGT$ 3'
1 28

Fig. 2 The 28-mer under investigation.

The change in absorbance at 260 nm was measured as a function of temperature for the 28-mer, over a concentration range of 1–30 μM and pH range 5–7.2, for the sample in 20 mM sodium phosphate buffer. The melting temperature for the sequence was found to be strongly concentration-dependent at high pH but essentially concentration-independent from pH 6 down (refer to the electronic supplementary information). From these data $\Delta S°$, $\Delta H°$ and $\Delta G°$ values could be determined through a van't Hoff analysis[9] (also see the supplementary information). $\Delta G°$ becomes increasingly negative with decreasing pH, consistent with the formation of a stable conformation at lower pH. The absorbance *versus* temperature profiles for heating and cooling runs were found to be non-superimposable irrespective of pH. This hysteresis shows that the unfolding/folding transitions of the 28-mer follow slow kinetics.[10] Heating and cooling studies as a function of concentration, at pH 6 and below, showed that the hysteresis was concentration-independent, as expected for an intramolecular process. Varying the heating and cooling rate from 2 to 0.2 °C min^{-1} over the whole temperature range reduced this hysteresis; resulting in a 2 °C difference in the melting temperature at pH 5.0. These results supported the formation of a unimolecular folded entity under mildly acidic conditions and established conditions for subsequent NMR investigations. At the sample concentration used for the NMR studies (0.47 mM) the melting temperature (T_m) was determined to be 42.8 °C at pH 6.1 and 54.2 °C at pH 5.0 (refer to the supplementary information).

Initially 1D ^{1}H NMR spectra were recorded as a function of pH and temperature to confirm the formation of an 'ordered' structure. Imino proton resonances were detected at chemical shifts greater than 15 ppm at pH 7.0; at 15 °C and lower temperatures, and at pH 6.0 and 5.0, at room temperature and below (Fig. 3). This is evidence of the presence of an i-motif.[8,11] A weak signal is detected at around 12.7 ppm at room temperature and pH 7.0. This is typical of G imino protons in G·C base pairs indicating the presence of some other folded form at this pH. 2D spectra, recorded at pH 7.0 and 20 °C, indicated a significant population of folded (subsequently established as i-motif folded) molecule. Indeed the average ^{1}H chemical shift variation from pH 7.0 to 6.0 was found to be less than 0.03 ppm (refer to the supplementary information). However, the best quality 2D data were recorded for the 28-mer at pH 6. The proton NMR assignment process followed established procedures[12] for nucleic acids, utilising COSY,

† Electronic supplementary information (ESI) available: thermal data, chemical shifts, NOESY data and spectra. See http://www.rsc.org/suppdata/ob/b5/b504606h/

DOI: 10.1039/b504606h

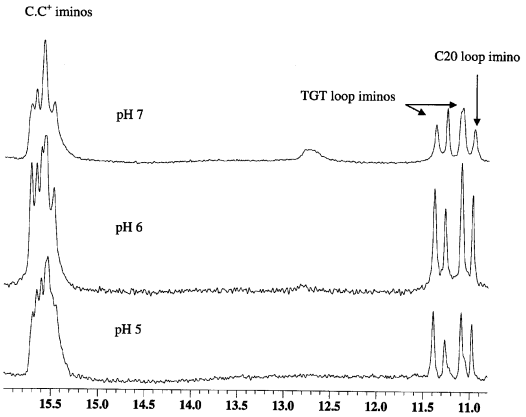

Fig. 3 Low field region of the ^{1}H NMR spectrum of the 28-mer in 90% H_2O–10% D_2O, at pH, 5, 6, and 7 and at 5 °C. A comparison of the integrals for the C·C$^+$ iminos and TGT loop iminos indicates than in excess of 90% of the sample is in i-motif form at pH 6, compared with 53% at pH 7.

TOCSY AND NOESY experiments. Almost all of the non-exchangeable protons were assigned, based on the assumption of B-type helical parameters for the individual nucleotides; the majority of the H5′/5″ resonances remain unassigned (see the supporting information). With extensive resonance assignment, it quickly became apparent that the sequence was not adopting a simple helix. C2, C15, C18 and C24 all displayed H5 to n-1 H2′/H2″ nuclear Overhauser enhancements (NOEs) that were much greater than those between corresponding H6 and H2′ protons. Also, the methyl groups of T14 and T28 displayed NOEs to the H2′/H2″ protons of G13 and G27 respectively. Comparison of the volumes of NOE cross-peaks connecting H6/H8 to n-1 H2′/H2″, H1′ to H2′/H2″, and H5/H1′ to n-1 H2′/H2″ revealed that the majority of the residues adopted C3′-*endo* sugar puckers. In a B-type helix the sugar pucker is typically C2′-*endo* however the C3′-*endo* conformation is a feature of i-motifs.[8,14] Perhaps of greater significance were the long-range NOEs. Residues C22 and C18, C11 and T14, T12 and C25, C8 and A19, T26 and T12, and C23 and C4, are pairs connected by H5 and/or H1′ inter-residue connections to H3′ and/or H4′ protons. H1′-H1′ NOE connections between C1 and C23, C10 and C17, C4 and C22, and C8 and C17, are indicative of the presence of two intercalating anti-parallel strands.[8,11,13] As the 28-mer is a single strand system, the H1′ interactions must be across two narrow grooves. All of these observations are strongly supportive of the formation of an i-motif.

Resonance assignments were initially made for spectra recorded at 20 °C and subsequently transferred to those recorded at 5 °C, at which temperature the best data for the exchangeable protons were recorded. Assignment of the exchangeable protons progressed by first connecting the amino protons of cytosine to their respective H5 or H6. Where both aminos could be distinguished, the *cis* proton was 1.5–2.0 ppm downfield of the *trans* proton.[14] This then enabled connections to iminos of protonated cytosines to be established. When a set of imino, aminos and H6 or H5 protons displayed an NOE to another amino proton, a C$^+$·C base pair was identified.[15] In this way base-pairing interactions were found between C18 and C4, C3 and C17, C2 and C16, C1 and C15, C25 and C11, C23 and C9 and C22 and C8; the base-pairing of C24 and C10 has not been established. Imino–imino NOE connections between C23 and C9 and C16 and C2 are suggestive of these base-pairs being stacked one above the other. Also T12 imino, assigned due to its close spatial proximity to its methyl group, was found to be close in space to T26 imino, suggesting a T·T base-pairing interaction.[16] All of these data are consistent with the 28-mer being folded to an i-motif with the topology indicated in Fig. 4. The topology is similar to that found for other sequences.[17]

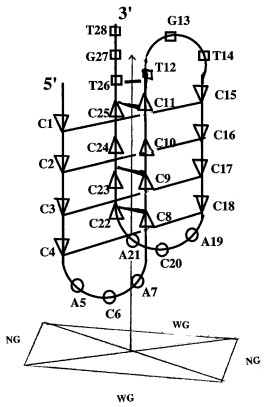

Fig. 4 Folding topology of the 28-mer. NG and WG mean narrow and wide grooves respectively. The lines between non-adjacent residues indicate observed base-pairing patterns.

Previous studies of a 25-mer sequence, related to the 28-mer investigated here, have suggested that the 25-mer is capable of forming an i-motif,[6] although experimental data could not rule out the possibility of this being as a result of two 25-mers interacting; *i.e.*, an intermolecular i-motif was a possibility. The dimer system was modelled with a protonated cytosine (from what would otherwise be an ACA loop) involved in the stem of the hairpin formed by the 25-mer. The results reported here do not support such a complex for the 28-mer. UV thermal melting analyses are unambiguous regarding the molecularity of the ordered system at slightly acidic pH. Also, the single loop cytosine, although protonated, is not hydrogen bonded and displays no close spatial relationship to protons other than those of its own nucleotide. The 1D spectra presented for the 25-mer[6] displayed imino proton signals for the protonated and hydrogen bonded cytosines, but imino protons for the loop thymines were not observed at pH 7.0. In contrast, under the same conditions of temperature and pH, imino protons for the loop thymines of the 28-mer are clearly visible (refer to the electronic supplementary information). It is possible, therefore, that the shifted 25-mer sequence is less stable than the 28-mer, leading to the complexity of NMR spectra commented on by the authors.[6] This comparison in itself may shed some light on the mechanistic detail of i-motif formation at slightly acidic pH. An understanding of this is desirable as the biological relevance of this C-tetraplex is becoming more apparent.[18]

Acknowledgements

This work was supported by the provision of an Overseas Research Scholarship to V.V.J.

References

1 G. I. Bell, M. J. Selby and W. J. Rutter, *Nature*, 1982, **295**, 31–35.
2 D. Owerbach and L. Aagaard, *Gene*, 1984, **32**, 475–479.
3 A. M. Lucassen, G. R. Screaton, C. Julier, T. J. Elliot, M. Lathrop and J. I. Bell, *Hum. Mol. Genet.*, 1995, **4**, 501–506.
4 G. C. Kennedy, M. S. German and W. J. Rutter, *Nat. Genet.*, 1995, **9**, 293–298.
5 P. Catasti, X. Chen, S. V. S. Mariappan, E. M. Bradbury and G. Gupta, *Genetica*, 1999, **106**, 15–36.
6 P. Catasti, X. Chen, R. K. Mooyzis, E. M. Bradbury and G. Gupta, *J. Mol. Biol.*, 1996, **264**, 534–545.
7 P. Catasti, X. Chen, L. L. Deaven, R. K. Moyzis, E. M. Bradbury and G. Gupta, *J. Mol. Biol.*, 1997, **272**, 369–382.
8 K. Gehring, J.-L. Leroy and M. Gueron, *Nature*, 1993, **363**, 561–565.
9 J. D. Puglisi and I. Tinoco, Jr., *Methods Enzymol.*, 1989, **180**, 304–325.
10 J.-L. Mergny and L. Lacroix, *Nucleic Acids Res.*, 1998, **26**, 4797–4803.
11 J.-L. Leroy, K. Gehring, A. Kettani and M. Gueron, *Biochemistry*, 1993, **23**, 6019–6031.

12 K. Wuthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
13 X. Han, A. T. Phan, and M. Gueron, in *Structure, Motion, Interaction and Expression of Biological Macromolecules, vol. 1*, ed. R. H. Sarma and M. H. Sarma, 1998, 49–62.
14 M. Gueron and J.-L. Leroy, *Structure*, 1995, **3**, 101–120.
15 P. Catasti and K. Robert, *J. Mol. Biol.*, 1997, **272**, 369–382.
16 J.-L. Leroy, *J. Mol. Biol.*, 2003, **333**, 125–139.
17 M. Gueron and J.-L. Leroy, *Curr. Opin. Struct. Biol.*, 2000, **10**, 326–331.
18 (*a*) K. Halder, V. Mathur, D. Chugh, A. Verma and S. Chowdhury, *Biochem. Biophys. Res. Commun.*, 2005, **327**, 49–56; (*b*) T. Simonsson, M. Pribylova and M. Vorlickova, *Biochem. Biophys. Res. Commun.*, 2000, **278**, 158–166.