nature structural biology
February

# The art of intercalation

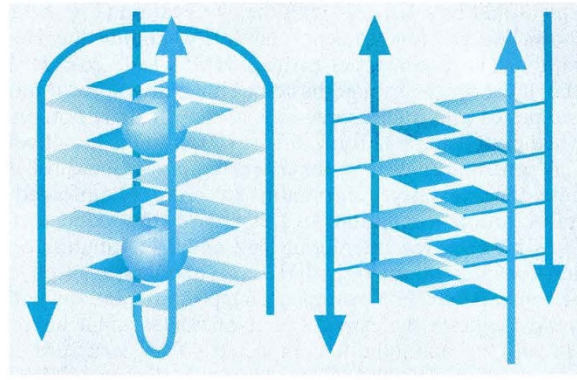
**The biological relevance of the i-motif comes under examination with the structural characterization of oligonucleotides formed of repeats of the human telomeric C-strand sequence d(CCCTAA)$_n$.**

FOR all its apparent chemical simplicity, DNA can adopt a surprising range of structures *in vitro*. One of the icons of the twentieth century, the right-handed antiparallel double helix, can exist in a number of conformations (A- and B-forms being the best known). More exotic structures include Z-DNA, a left-handed double helix, triple helices, where a third strand nestles in the major groove of a standard double helix, and the quadruple helix of the guanine tetraplex.

Although the biological importance of some of these structures is beyond doubt, the status of others is less clear. A paper from Shawn Ahmed and colleagues in this month's *Nature Structural Biology*[1] now provides the first hint that a newly characterized four-stranded structure, the cytosine-protonated cytosine (C–C$^+$) tetraplex or i-motif[2], may be involved in telomere formation and nucleic-acid self-recognition.

The best characterized of the four-stranded structures are the guanine quartets[3]. These consist of four guanine bases, in a square-planar arrangement, which cohere by Hoogsteen base-pair interactions. Stacks of these G-quartets, stabilized by monovalent cations, readily form *in vitro* under physiological conditions (figure). The complexes are polymorphic; they can be formed from one, two or four separate DNA, or RNA, strands which can be in a parallel, or antiparallel, orientation relative to one another. The connectivity of the strands can vary, as can the glycosidic conformation — *syn* or *anti* — of the guanine bases.

A variety of biologically important guanine-rich sequences are known to form G-quartet structures *in vitro*. Telomeric DNA, which caps the ends of linear eukaryotic chromosomes, consists of simple tandem repeats that are G-rich on one strand, which protrudes as a single strand by two-repeat-units at the very tip of the chromosome[4]. Indeed, the structures of a number of the telomeric sequences that form G-tetraplexes have been determined by X-ray crystallography and nuclear magnetic resonance spectroscopy (ref. 5 and references therein). Do these structures exist *in vivo*? Evidence from the study of telomere-binding proteins[6,7] and the dimerization of retroviral RNA genomes[8] indicates that they probably do.



Diagrams of a guanine tetraplex (left), the spheres representing monovalent cations, and a C–C$^+$ tetraplex (right). The DNA backbone is shown as solid lines with arrowheads which indicates the 5'→3' polarity of the strand.

For every stretch of G-rich sequence capable of forming G-quartet structures in a DNA double helix, there is a complementary strand rich in cytosine bases. What is the fate of these sequences? Last year's report[2] of an unusual structure, the i-motif, formed in solution by d(TCCCCC)$_4$ suggested that such tracts of DNA rich in cytosine may in fact have a highly unusual structure.

The i-motif, like the G-quartet, has a backbone of four strands, which may all be part of the same DNA molecule or formed as dimers or tetramers. Here the similarity ends. Hemi-protonated cytosine bases participate in non-Watson-Crick base-pairs, forming parallel-stranded C–C$^+$ duplexes. In the i-motif structure two such duplexes intertwine in an antiparallel fashion, with the C–C$^+$ base pairs of one duplex fully intercalated with those of the other duplex (figure).

Does this addition to the pantheon of nucleic-acid structures have any biological ramifications? Clearly, the first place to look would be on the C-rich strand of telomeric DNA. Ahmed *et al.*[1] have done just this and show, using native gel electrophoresis, that the human telomeric C-strand oligonucleotide d(CCCTAA)$_2$ forms a dimer and d(CCCTAA)$_4$ forms an intramolecular complex at low pH consistent with the formation of four-stranded structures. More notably, NMR spectroscopic analysis of the d(CCCTAA)$_2$ dimer reveals imino proton spectra characteristic of C$^+$ imino protons, and connectivities considered to be diagnostic of DNA molecules that form the i-motif — it is, therefore, a C–C$^+$ tetraplex.

The demonstration that human telomeric C-rich sequences adopt an i-motif conformation suggests that this structure may be of relevance *in vivo*. G-tetraplex and i-motif structures could act in concert; any sequence, such as telomeric DNA, that will form a G-tetraplex on one strand potentially has the ability to form a C–C$^+$ quadruplex on the other strand. Indeed, one structure might promote the formation of the other. G-tetraplexes have been suggested to mediate self recognition, telomere formation, meiotic chromosome pairing and recombination; the i-motif has the potential to do the same.

An obvious drawback to such speculation is the low pH, and cytosine protonation, needed for formation of the i-motif. Of course, it is possible to invoke i-motif-binding proteins, super-helical stress and so on as conditions that might yield the structure. Whether such conditions exist *in vivo* will only become evident after further experiment. Perhaps of equal interest are details of the chemistry of the i-motif (how the tetraplex folds to its final form, for example), and the answers to such questions should be just as illuminating. **Guy Riddihough**

*Guy Riddihough is Editor of* Nature Structural Biology.

Also in this month's *Nature Structural Biology*: two alternative conformations of a 2-aminofluorene DNA duplex; the binding site of GrpE on DnaK; the structure of a 'cysteine ring' — a disulphide bond linked by a *cis*-peptide bond; correlated intramolecular motions and diffuse X-ray scattering in lysozyme crystals; the common fold of the matrix metalloproteinases — structures of the catalytic domains of fibroblast and neutrophil collagenase and stromolysin-1.

1. Ahmed, S., Kintanar, A. & Henderson, E. *Nature struct. Biol.* **1**, 83–88 (1994).
2. Gehring, K., Leroy, J.L. & Guéron, M. *Nature* **363**, 561–565 (1993).
3. Williamson, J.R. *Curr. Opin. struct. Biol.* **3**, 357–362 (1993).
4. Blackburn, E.H. *Nature* **350**, 569–573 (1991).
5. Wang, Y. & Patel, D.J. *Structure* **1**, 263–282 (1993).
6. Fang, G. & Cech, T.R. *Cell* **74**, 875–885 (1993).
7. Liu, Z., Frantz, J.F., Gilbert, W. & Tye, B.-K. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3157–3161 (1993).
8. Sundquist, W.I. & Heaphy, S. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3393–3397 (1993).