

Non-canonical DNA structures and their interactions with small molecule ligands

L. A. Yatsunyk, L. Y. Lin, D. Beseiso, D. Beseiso, D. A. Jordan, B. M. Powell, S. McCarthy

Swarthmore College, 500 College Ave, Swarthmore PA, 19081, USA

lyatsun1@swarthmore.edu

Non-canonical DNA structures, notably G-quadruplexes and i-motifs, draw significant attention because biological evidence suggests that they play crucial roles in a variety of disease-related biological processes. G-quadruplex DNA is composed of planar guanine tetrads which are engaged in efficient π - π stacking and are further stabilized by monovalent central cation (e.g. K^+ or Na^+). Sequences with G-quadruplex forming potential are present in telomeres and in oncogene promoters, according to bioinformatics studies. I-motifs are intercalated hemi-protonated cytosine-rich structures formed in the C-rich sequences. Naturally, such sequences are present in the regions complimentary to the G-rich parts of genome.

In this work we have investigated nine variants of telomeric DNA with the repeat (TTGGGG) $_n$ from the *organism Tetrahymena thermophila* using biophysical and x-ray crystallographic studies. Biophysical characterization showed that all sequences folded into stable GQs which adopted a variety of conformations, most commonly parallel and hybrid. Native PAGE suggested that most of the sequences form multiple species in the presence of potassium. All species, but one, are monomolecular. We successfully crystallized two variants, TET25 (resolution 1.56 Å) and TET26 (three crystal forms with resolution 1.99 and 1.97 Å) and solved the structures via molecular replacement. TET25 adopted a hybrid (3+1) conformation with a four G-tetrad core, three lateral loops, one propeller loop, and 5' snapback. TET26 fold into a parallel GQ conformation with a four G-tetrad core and three TT propeller loops. We have also investigated binding of N-methylmesoporphyrin IX (NMM) to all sequences and crystallized three variants with NMM. NMM induces parallel fold in all sequences. Both crystal structures display 5'-5' dimers of parallel GQs with NMM bonded to the 3' G-tetrad. NMM binds GQ with one of its faces and another NMM molecule with another. Our structural data demonstrate great plasticity of the telomeric sequence from the telomeric region of *T. thermophila* where small variation in the overhang length and composition leads to drastically distinct GQ structures.

I will also share our progress toward the structure an i-motif DNA from the HRAS oncogene promoter as well as the structure of repetitive DNA (CAGAGG) $_n$ from difficult-to-replicate regions of the mouse genome implicated in replication stress. Our findings have potential to contribute to the development of new and efficient anticancer therapies.

Keywords: G-quadruplex, i-motif, X-ray crystallography, ligands

This work was supported by NIH grant (grant No. 1R15CA208676-01A1) and Camille and Henry Dreyfus Teacher-Scholar Award