Exciting NMR Structures of G-rich DNA Repeats

Janez Plavec[a,b,c]

[a] *Slovenian NMR center, National Institute of Chemistry, Ljubljana, Slovenia*
[b] *University of Ljubljana, Faculty of Chemistry and Chemical Technology, Slovenia*
[c] *EN-FIST Centre of Excellence, Trg OF 13, Ljubljana, Slovenia*
** e-mail: janez.plavec@ki.si*

Structural insights into DNA including its repetitive elements contribute to understanding of evolution and function as well as complex relations with the genetics of human diseases. A large portion of DNA repeats lies within genes and their regulatory regions. 1.5 million short tandem repeats consist of repeating units of two to six base-pair motifs of DNA. Among the various types of tandem repeat diseases two major neurodegenerative diseases, frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS), were found to be caused by hexanucleotide GGGGCC repeats expansions. GGGAGCG repeats occur in the regulatory regions of genes responsible for neurological disorders, cancer and abnormalities in bone and cartilage development.

DNA with its canonical Watson-Crick paired duplex plays a major role in inheritance of genetic material and gene expression. Alternative secondary structures including quadruplexes and i-motifs have been associated with many different biological functions of DNA. The most well studied alternative DNA structures are G-quadruplexes. They are formed by G-rich sequences and consist of four-stranded columnar structures. G-quadruplexes are stabilized by the stacking of multiple Hoogsteen-hydrogen-bonded G-quartets. Notably, cations residing in the center of G-quartets contribute additional stabilizing factor through electrostatic interactions. Our laboratory has been using NMR spectroscopy to uncover structural details of G-quadruplexes in relation to sequence details, presence of cosolutes, interaction with ligands and even expand the structure and sequence complexity of DNA four-stranded architectures by discovering AGCGA-quadruplexes.

Selected references: *Chem. Eur. J.* **2020**, 26, 814. *Molecules* **2020**, *25*: 434. *Nucleic Acids Res.* **2019**, *47*, 11057. *Hum. Mol. Genet.* **2019**, *28*, 3163. *Molecules* **2019**, *24*: 1294. *Nucleic Acids Res.* **2019**, *47*, 2641. *Angew. Chem. Int. Ed.* **2019**, *58*, 2387. *Angew. Chem. Int. Ed.* **2018**, *57*, 15395. *J. Am. Chem. Soc.* **2019**, *141*, 2594. *J. Am. Chem. Soc.* **2018**, *140*, 5774. *Nat. Commun.* **2017**, 8: 15355.