WHAT DO G-QUADRUPLEXES IN VIRAL GENOMES TELL US?

Václav Brázda

Faculty of Chemistry, Brno University of Technology
Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czech Republic

G-quadruplexes are structural elements that can form in both DNA and RNA molecules. They are formed by guanine-rich sequences that fold into very stable four-stranded structures. G-quadruplexes have been identified in various biological contexts, including genomes of viruses. In viral genomes, G-quadruplexes are proposed as important elements which play a role in viral genome stability (by protecting viral DNA from degradation and recombination events), regulation of gene expression (by affecting transcription and translation processes), in viral genomes, G-quadruplexes can be found in promoter regions and other regulatory sequences, but they are suggested also as antiviral targets (targeting G-quadruplexes can disrupt critical viral processes, such as replication and gene expression, leading to inhibition of viral proliferation). Therefore, the development of small molecules that specifically interact with G-quadruplex structures in viral genomes is a broad research area with therapeutic potential. Using the G4Hunter tool, we have analyzed H1N1 influenza genomes, all Nidovirales genomes, including SARSCoV2, and all genomes accessible in the NCBI Genome database. Our result shows that G-quadruplex-prone sequences are not distributed randomly in viral genomes, but that they are associated with function regions; moreover, we have shown that the presence of the G-quadruplexes is correlated with the type of viral infection. The presence of G-quadruplex sequences is related to viral life cycles and corresponds to the type of viral infection rather than the type of nucleic acid; while viruses that cause persistent infections are enriched for G-quadruplex-prone sequences, viruses that cause acute infections have a significantly reduced number of G-quadruplexes. Furthermore, we found a significant positive correlation between the frequencies of G-quadruplex-prone sequences in dsDNA viruses and the corresponding hosts of archaea, bacteria and eukaryotes. The strong relationship between the G-quadruplex-prone sequences in the virus and the host indicates a close coevolution and an evolutionarily reciprocal imitation of the genome organization.