Abstract

During this Ph. D. study we focused on d(GGGAGCG) repeats connected by adenine residues found in the regulatory region of the PLEKHG3 gene in the fourteenth human chromosome. The gene is involved in many key cellular processes such as regulating the development and function of neurons. Additionally, *de novo* deletions of this gene have been linked to autism. The simplest oligonucleotides found in the d(GGGAGCG) repeat region are oligonucleotides d(GGGAGCGAGGGAGCG), VK1 and d(GCGAGGGAGCGAGGG), VK34. From the structural point of view the oligonucleotides are interesting since they could, based on the data from the literature, fold into G-quadruplexes comprised out of two G-quartets. These two G-quartets could be linked by bulges or could sandwich in GCGC-quartets. In addition, researchers have detected a lot of folds in guanine-rich regions of the human genome for which it is impossible to conclude with certainty that they adopt G-quadruplex folds. Similarly, we concluded that VK1 and VK34 oligonucleotides could also adopt folds different from G-quadruplexes. All new structural motifs found in the biologically relevant regions are potentially of great importance as targets for development of new drugs.

We have discovered that VK1 and VK34 oligonucleotides fold into structures that belong to a new tetrahelical structural family that we named AGCGA-quadruplexes. Only two tetrahelical structural families called G-quadruplexes and i-motifs have been known so far. Two VK1 oligonucleotides fold into a symmetric dimeric structure that is stabilized by four G-C base pairs in Watson-Crick geometry that form the core of the structure, two G-A base pairs in N1-N7 carbonyl amino geometry that are stacked on each side of the G-C core and six G-G base pairs in N1-carbonyl symmetric geometry. Three G-G base pairs on each side of the AGCGA core of the VK1 structure stabilize two fold-back GGG lops.

VK34 oligonucleotides can fold into dimeric or tetrameric structures that are both highly symmetric. The transition from dimeric to tetrameric fold can occur spontaneously or is accelerated at higher oligonucleotide concentrations and in presence of Na^+, K^+ and NH_4^+ cations. In both structures two G-A base pairs in N1-N7 carbonyl amino geometry associate into GAGA-quartets that form GAGA-cores, where two GAGA-quartets are stacked on one another. Major (tetramer) and minor (dimer) groove GCGC-quartets which are rare structural elements are stacked on GAGA-cores. Additionally, the VK34 folds contain G-G base pairs in N1-carbonyl symmetric geometry that adopt criss-cross topologies. We have confirmed that AGCGA-quadruplexes are formed by oligonucleotides containing several VK1 or VK34 repeats (VK2, 2VK34, 4VK34). We were able to fold a mutant of VK34 (G11 to I11) named VK34_I11 into a structure that is very similar to VK1. Two oligonucleotides with different sequences (VK1 and VK34_I11) can form similar structures since they fold by forming a pre-folded duplex. Specifically, two VK34_I11 oligonucleotides align to form a duplex that encompasses a segment with the same composition and distribution of base pairs as found in the VK1 structure. We have concluded that pre-folded states are not unique to VK1, VK34 and VK34_I11 oligonucleotides bat are typical for all oligonucleotides that fit the AGCGA(N_{1-20})AGCGA(N_{1-20})AGCGA(N_{1-20})AGCGA folding motif.

A human genome-wide search revealed 146 sequences with 41 (in addition to VK1, VK2, VK34, 2VK34 and 4VK34) of them found in promoter regions and transcription start sites of known genes as well as CTCF (CCCTC-binding factor) binding sites and CNV (copy number variations) regions. Together 46 oligonucleotides are found in regulatory regions of 38 different human genes in addition to PLEKHG3 connected to neurodevelopment and neurological disorders, abnormal cartilage and bone formations, cancer and regulation of basic cellular processes. With the help of 1D ^1H NMR and CD spectroscopy we have confirmed that all of the 46 potentially biologically relevant oligonucleotides fold into AGCGA-quadruplexes. Unique structural features of AGCGA-quadruplexes together with their lower sensitivity to cations and pH variation, compared to G-quadruplexes, support the biological relevance of structures formed by AGCGA-repeat sequences.

Keywords: PLEKHG3 gen**,** AGCGA-quadruplexes, G-quadruplexes, i-motifs, VK1 oligonucleotide, VK34 oligonucleotide, pre-folded duplex