Abstract

The *AUTS2* gene has been shown to influence neurodevelopment by controlling the number of neurons, influencing neuronal migration and promoting axon and dendrite growth. The AUTS2 protein can be expressed either as full-length or C-terminal isoform. Precise regulation of expression of both isoforms is crucial for brain development. The molecular mechanism responsible for the expression regulation of the two isoforms is unknown and no mechanism has been proposed.

In this work, we have identified a CGAG-rich region in the promoter of *AUTS2* gene. The CGAG-rich region spans 60 nucleotides and contains a putative protein binding site (PPBS), d(AGCGAAAGCACGAA). Using NMR, UV-Vis spectroscopy and MD simulations, we show that oligonucleotides derived from this region adopt thermally stable non-canonical hairpin structures. We uncover a mechanism, which we refer to as a four-nucleotide register shift, that enables the formation of a variety of hairpins depending on the number of CGAG repeats in the oligonucleotide sequence. The hairpins are stabilized by G:C and sheared G:A base pairs that are arranged in structural motifs that we term CGAG blocks. We observe significant structural differences between the loop regions of these non-canonical hairpin structures. We show that epigenetic modifications 5-methylcytosine and 5-hydroxymethylcytosine incorporated in CGAG-rich oligonucleotide result in marginal increase and minor decrease in thermal stability, respectively. The modifications cause only local structural changes. We also examine an oligonucleotide that is complementary to the CGAG-rich strand and show that the oligonucleotide $d(CGCT)_4$ adopts a thermally rather labile hairpin structure stabilized by G:C and C:T base pairs. We conclude that a transient formation of non-canonical hairpin structures by CGAG-rich oligonucleotides is possible *in vivo*. Therefore, our work provides an insight into a plausible mechanism responsible for regulation of expression of the two AUTS2 isoforms.