Gene 572 (2015) 175–183

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Role of long purine stretches in controlling the expression of genes associated with neurological disorders

Himanshu Narayan Singh, Moganty R. Rajeswari *

Department of Biochemistry, All India Institute of Medical Sciences, New Delhi 110029, India

ARTICLE INFO

Article history:
Received 18 April 2015
Received in revised form 17 June 2015
Accepted 2 July 2015
Available online 3 July 2015

Keywords:
Purine repeats
Human genome
Neurological disorder
Genetics
DNA-triple helix
Neuroinformatics

ABSTRACT

Purine repeat sequences present in the human genome are known to act as hotspots for mutations leading to chromosomal imbalances. It is established that large purine repeats (PRs) form stable DNA triplex structure which can inhibit gene expression. Friedreich's ataxia (FRDA), the autosomal neurodegenerative disorder is the only human disease known so far, where a large purine (*GAA*) repeat in the *FXN* gene is known to inhibit the expression of frataxin protein. We explored the hidden purine repeats (PR_n with $n \geq 200$) if any, in the human genome to find out how they are associated with neurological disorders. The results showed 28 PRs, which are mostly restricted to the intronic regions. Interestingly, the transcriptome expression analysis of PR-carrying genes (PR-genes) revealed that most of them are down-regulated in neurological disorders (autism, Alzheimer's disease, schizophrenia, epilepsy, mental retardation, Parkinson's disease, brain tumor) as compared to that in healthy controls. The altered gene expression in brain disorders can be interpreted in terms of a possible expansion of purine repeats leading to formation of very stable DNA-triplex and/or alleviation of the repair enzymes and/or other unknown cellular factors. Interactome analysis identified four PR-genes in signaling pathways whose dysregulation is correlated directly with pathogenesis: *GRK5* and *KLK6* in Alzheimer's disease; *FGF14* in craniosynostosis, mental retardation and *FLT1* in neuroferritinopathy. By virtue of being mutational hotspots and their ability to form DNA-triplex, purine repeats in genome disturb the genome integrity and interfere with the transcriptional regulation. However, validation of the disease linkage of PR-genes can be validated using knock-out techniques.

1. Introduction

Announcement of the first draft of the complete human genome in 2004 generated a lot of enthusiasm and expectations of understanding pathology at the genetic level (International Human Genome Sequencing Consortium, 2004). A number of pharmaceutical industries came forward to exploit the data for therapeutic developments. Health professionals were keen to use it for personalized medicine, diagnosis and therapy. There was an initial burst of attempts to discover the hidden secrets lying in the human genome sequence. However, a lot of information available on the public domain remained as mere data. There are several reasons for this and one of them is that the genes were of hypothetical origin or the expression data of the known genes in any disease were not readily available. In recent years, there has been a tremendous progress in genomic technologies (Ansorge, 2009), which helped in our understanding of complex biochemical pathways and molecular/genetic basis of human disease. Further, the advances in technologies like microarrays helped in generating large amounts of data of expression of genes at the DNA/mRNA level (Frank-Kamenetskii and Mirkin, 1995). These reports are a huge boost in interpreting a variety of patterns that are seen in the human genome. In this paper we made an attempt to understand the existence of non-overlapping purine repeat patterns in the human genome and their role in genomic instability which can lead to neurological disorders. Molecular mechanisms that underlie neurological diseases are relatively difficult to understand because of the limited availability of the pathological tissue. The majority of our understanding of neurological diseases comes from basic science using animal models. Further, neurological disorders are multifactorial and pose challenges to scientists in diagnosis and therapy. Therefore, the present study is expected to give a new dimension and genetic basis of neurological diseases.

It is interesting to note that the purine repeats (PRs) in DNA have high propensity of generating triple helical structures or DNA-triplex (Arya, 2011; Frank-Kamenetskii and Mirkin, 1995; Rajeswari, 2012). It is now established that PRs present in gene can induce mutations, expansions, translocations and chromosomal rearrangement and play a role in transcriptional regulation (Mirkin, 2007; Wells et al., 2005).

Abbreviations: G, guanine; A, adenine; PRs, purine repeats; FRDA, Friedreich's ataxia; PR-genes, PR-carrying genes; MR, mirror repeat; nBMST, non-B DNA Motif Search Tool; AD, Alzheimer's disease; PD, Parkinson's disease; DS, Down Syndrome.

* Corresponding author.
E-mail addresses: himanshu720@gmail.com (H.N. Singh), rajeswari3011@hotmail.com (M.R. Rajeswari).

http://dx.doi.org/10.1016/j.gene.2015.07.007
0378-1119/

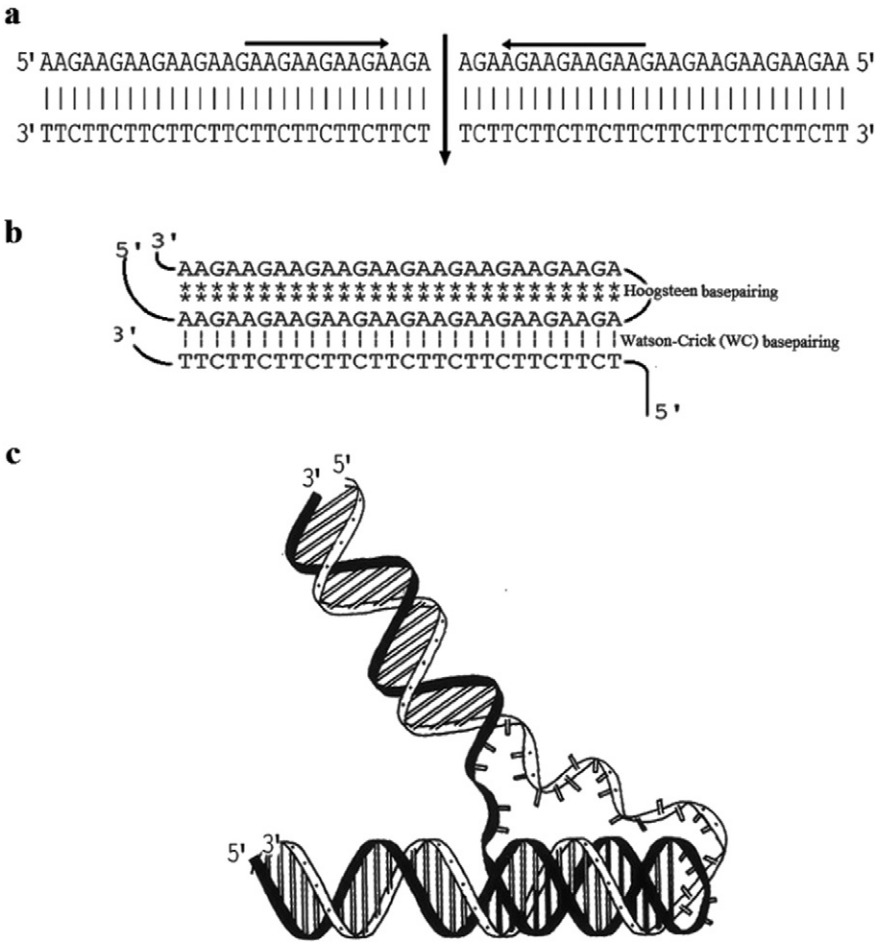

Fig. 1. (a) Schematic representation of the DNA triplex, constructed using purine repeat sequence with repeat number, n, 56 (AGAAGAAGAAGAAGAAGAAGAAGAAGAAAGAA GAAGAAGA AGAAGAAGAAGAAGA), which is part of the PR-gene and ALK gene identified in the present study. Plane of symmetry is shown as a dashed line and direct mirror repeat of 28 purines are highlighted with arrows. (b) The intramolecular triplex formed by the fold back of GAA strand onto itself is shown in the centre. The Hoogsteen pairing between GG and AA is shown with 'asterisk' (*) and the Watson–Crick base-pairing between the sharing strand and the complementary strand is shown by dash (–). (c) The schematic diagram of the intramolecular triplex is given in the bottom.

A schematic representation of a typical DNA triplex structure generated by a segment containing 56 purines of the 411 long PR is shown in Fig. 1. In fact the 411 PR obtained from the present study is part of the *ALK* gene. Interestingly, the PR with a length of 56 is a perfect mirror repeat of 28 purines (Fig. 1a). The intramolecular triplex formed by the fold-back of the GAA strand onto itself by forming the Hoogsteen pairing between GG and AA and the Watson–Crick base pairing between the sharing strand and the complementary strand is shown in Fig. 1b. A schematic representation of an intramolecular triplex is shown in Fig. 1c. Generally, DNA base triads can be formed by reverse Hoogsteen base pairing as in G*G:C, T*A:T and A*A:T or Hoogsteen base pairing as in C^{+}*G:C, T*A:T and G*G:C (Frank-Kamenetskii and Mirkin, 1995; Goñi et al., 2004).

The presence of DNA-triplex structures in the gene impedes gene transcription by any or both mechanisms: (i) promoter occlusion, where triplex formation at promoter does not allow transcription

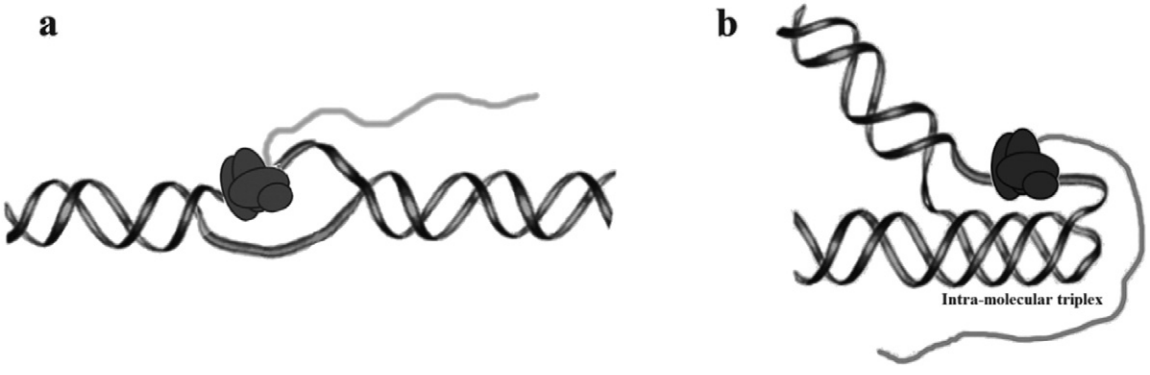

Fig. 2. Transcriptional regulation (a) under normal conditions; and its inhibition by (b) intramolecular triplex mediated elongation arrest.