Biochemistry and Biophysics Reports 5 (2016) 388–395

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

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

A bouquet of DNA structures: Emerging diversity

Mahima Kaushik [a,b,*], Shikha Kaushik [b], Kapil Roy [b], Anju Singh [b], Swati Mahendru [b], Mohan Kumar [b], Swati Chaudhary [b], Saami Ahmed [b], Shrikant Kukreti [b]

[a] Cluster Innovation Centre, University of Delhi, Delhi, India
[b] Nucleic Acids Research Laboratory, Department of Chemistry, University of Delhi, Delhi, India

ARTICLE INFO

Article history:
Received 28 October 2015
Received in revised form
28 December 2015
Accepted 22 January 2016
Available online 28 January 2016

Keywords:
Hairpin
Triplex
Quadruplex
i-motif
Structural polymorphism
Alternate DNA structures

ABSTRACT

Structural polymorphism of DNA has constantly been evolving from the time of illustration of the double helical model of DNA by Watson and Crick. A variety of non-canonical DNA structures have constantly been documented across the globe. DNA attracted worldwide attention as a carrier of genetic information. In addition to the classical Watson–Crick duplex, DNA can actually adopt diverse structures during its active participation in cellular processes like replication, transcription, recombination and repair. Structures like hairpin, cruciform, triplex, G-triplex, quadruplex, i-motif and other alternative non-canonical DNA structures have been studied at length and have also shown their *in vivo* occurrence. This review mainly focuses on non-canonical structures adopted by DNA oligonucleotides which have certain prerequisites for their formation in terms of sequence, its length, number and orientation of strands along with varied solution conditions. This conformational polymorphism of DNA might be the basis of different functional properties of a specific set of DNA sequences, further giving some insights for various extremely complicated biological phenomena. Many of these structures have already shown their linkages with diseases like cancer and genetic disorders, hence making them an extremely striking target for structure-specific drug designing and therapeutic applications.

Contents

1. Introduction

In eukaryotic cells, deoxyribose nucleic acid (DNA) is present in its supercoiled form which is stabilized by ancillary proteins. However, during the biological processes such as replication and

* Corresponding author. Tel.: +91 11 27667702.
E-mail addresses: mkaushik@cic.du.ac.in, kaushikmahima@yahoo.com (M. Kaushik).

http://dx.doi.org/10.1016/j.bbrep.2016.01.013

transcription, DNA gets unwind and may form structures that differ from the Watson–Crick B-form of DNA. This biological phenomenon of adopting various conformations by this marvelous biomolecule is known as structural polymorphism and it depends upon a number of factors like oligonucleotide sequence, solution condition, hydration, ions, proteins, ligands and superhelical stress. B-form of DNA double helix, proposed by Watson and Crick, generally accounts for most of the DNA behavior in the cell [1]. Widely studied DNA conformations are A, B and Z forms, while DNA structures like bulge, hairpin, cruciform, parallel-stranded DNA, triplex, quadruplex, and i-motif are also well documented. These alternative DNA structures might not only be important for interactions with proteins involved in replication, gene expression and recombination but these would also have an impact on DNA damage, repair and genetic stability [2]. They also play different roles in the formation of nucleosomes and other supramolecular structures involving DNA [3].

With the exception of the A-form, which is usually adopted by RNA duplexes and is capable of accommodating any sequence, all known DNA structures are sequence-dependent. The sequence requirement for Z-DNA is an alternating purine–pyrimidine/GC-rich sequence and such Z-DNA forming sequences have shown their occurrence near chromosomal breakpoints involving the *c-MYC* and *BCL-2* genes [4]. DNA sequences containing $(CGG)_n$ repeats have been shown to form Z-DNA at high salt and millimolar concentrations of Ni^{2+} ions [5]. Further, their stability can be enhanced with methylation of CGG repeats. The conformational properties of DNA sequences containing $(CCA)_n$ and $(TGG)_n$ repeats were reported by Zemánek et al. using CD spectroscopy, polyacrylamide gel electrophoresis, and UV absorption spectroscopy. This study revealed that $(CCA)_n$ repeats associate to form i-motif structure at acidic pH, while it existed as single strand at alkaline or neutral pH. DNA sequences containing $(TGG)_n$ stretches form antiparallel homoduplex or hairpin structure at low salt concentrations, whereas these sequences adopt G-quadruplex structure in the presence of potassium ions at physiological pH [6]. The inverted repeat DNA sequences are capable of forming cruciform structures, while the G-quadruplex and i-motif formation require a contiguous stretch of guanines and cytosines respectively [7,8]. Some of the sequence requirements for the formation of various non-canonical DNA structures are summarized in Table 1. Most of these non-canonical DNA structures have already shown their *in vivo* existence, making them biologically very significant. Some of these fascinating DNA structures (Fig. 1) are discussed in the following sections.

2. Cruciform DNA

DNA supercoiling and base sequences are the two prime factors which are responsible for the structural complexity of DNA molecule. A cruciform structure is formed when interstrand base pairing in duplex DNA with inverted repeats, convert to intrastrand base pairing. It is well documented that an essential requirement for cruciform DNA formation is an inverted repeat DNA sequence which should be embedded in an A+T-rich region [9]. It is of great importance in many biological processes, including the nucleosome positioning, replication and regulation of gene expression. Two mechanisms are proposed for the formation of cruciform structure which differ in salt concentration, activation energy and temperature [10]. The thermodynamic stability of cruciform is very less in comparison to normal B-DNA and showed highly retarded mobility in polyacrylamide gel. Two classes of cruciform have been found till now, one with four-fold symmetry with all arms perpendicular to each other, and another with arms at an acute angle. Cruciform structures are generally a preferential target for many proteins such as HMG proteins, H1, H5 histones, and help to target many diseases [9].

Table1
Summarizing the sequence requirements for various DNA structures.

S.No.	Structures	Sequence	Reference
1.	Parallel-Stranded DNA	5′-CCTATTAAATCC	[51]
		5′-AAAAAAAAAATAATTTTAAATATTT	[52]
2.	Hairpin	$(CAG)_n/(CTG)_n$	[35]
		5′-TGGGGA/GCCCCA (Hairpin and duplex)	[11,12]
3.	Cruciform	5′-ATGGTCTTACCTA	[92]
4.	Triplex	5′-$(AAG)_5$ (Intermolecular triplex)	[93]
		5′-$C_2TC_5TC_2T_5G_2AG_5AG_2T_5G_2AG_5AG_2$	[58]
5.	i-motif	5′-CCCTAACCCTAA (Bimolecular)	[65,66]
		5′-$(CCCTAACCCTAA)_2$ (Unimolecular)	[66]
6.	Quadruplex	5′-$AG_6AG_3AG_3TG_2$ (Dimeric parallel-stranded)	[73]
		5′-GGTTGGTGTGGTTGG (Antiparallel unimolecular)	[94]
		5′-TTAGGGTTAGGG (Antiparallel tetramer)	[70]
7.	Z-DNA	5′-$(CGCGCGCGCGCG)_2$	[95]
8.	A-DNA	5′-$(GCGGGCCCGC)_2$	[96]

3. Hairpin DNA

DNA sequences with inverted repeats (IRs) or palindromes lead to the formation of hairpin structure, which might also be in equilibrium with duplex DNA, depending upon salt or oligomer concentrations. A single nucleotide polymorphism (SNP) in an 11-mer DNA oligonucleotide (d-TGGGG(A/G)CCCCA) had been shown to exhibit equilibrium between duplex and hairpin [11,12], while its RNA counterpart (UGGGG(G/A)CCCCA) adopted only hairpin structure [13]. This polymorphism of DNA sequences of Locus Control region (LCR) of beta-globin gene had been discussed from our laboratory in a review [14]. Hairpins have a base-paired stem and a small loop of unpaired bases, which are usually formed *via* two main mechanisms. In the first mechanism, hairpin is formed in the same way as cruciform structure is formed from double stranded DNA. In the second mechanism, single stranded DNA (ssDNA), produced during various cellular processes like replication on the template for lagging-strand synthesis, DNA repair, rolling-circle replication (RCR), and infection by some viruses leads to hairpin formation [15]. Single strand DNA binding protein (SSB) facilitate the binding of RecA to ssDNA without sequence specificity and leads to hairpin formation [16]. Basically, the possibility of hairpin or cruciform formation at the protein binding site can affect the coiling state of DNA which may either facilitate or prevent the DNA–protein interactions, and alter the gene expression. Although hairpins forming long palindromes are genetically unstable, yet the hairpin structure is known to play a key role in a number of cellular processes such as gene expression, recombination, and transcription.

Poly(dG)●Poly(dC) sequences exhibit A- as well B-forms depending on the solution conditions. These sequences adopt A-form of DNA at higher (molar) salt conditions, while at low (millimolar) salt concentration, they exist in B-form [17]. The transition from B- to A-form can be induced in the presence of methanol or hexamine cobalt or spermine or with the use of high oligomer concentration [18–21]. On the basis of CD spectroscopy, NMR spectroscopy and unrestricted molecular dynamics, d(CCCCGGGG) was shown to exhibit A- as well as B-like DNA characteristics. This study suggested that stacking between bases is accountable for A-form, while backbone framework is responsible for B-DNA signatures [22].

On the contrary, the sequence with inverted guanine and cytosine tracts *i.e.* d(GGGGCCCC) showed both B- as well as A-form