Biochimie 93 (2011) 1219–1230

Contents lists available at ScienceDirect

Biochimie

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

Research paper

The evolving world of protein-G-quadruplex recognition: A medicinal chemist's perspective

Claudia Sissi, Barbara Gatto, Manlio Palumbo*

Department of Pharmaceutical Sciences, University of Padova, Via Marzolo 5, 35131 Padova, Italy

ARTICLE INFO

Article history:
Received 26 February 2011
Accepted 20 April 2011
Available online 29 April 2011

Keywords:
G-quadruplex
Drug binding
Nucleic acids
Proteins

ABSTRACT

The physiological and pharmacological role of nucleic acids structures folded into the non canonical G-quadruplex conformation have recently emerged. Their activities are targeted at vital cellular processes including telomere maintenance, regulation of transcription and processing of the pre-messenger or telomeric RNA. In addition, severe conditions like cancer, fragile X syndrome, Bloom syndrome, Werner syndrome and Fanconi anemia J are related to genomic defects that involve G-quadruplex forming sequences. In this connection G-quadruplex recognition and processing by nucleic acid directed proteins and enzymes represents a key event to activate or deactivate physiological or pathological pathways. In this review we examine protein-G-quadruplex recognition in physiologically significant conditions and discuss how to possibly exploit the interactions' selectivity for targeted therapeutic intervention.

1. Introduction

G-Rich DNA or RNA sequences can fold into quadruplexes (G4s) consisting in non canonical structures stabilized by the stacking of G-quartets, in which four guanines are assembled in a planar arrangement by Hoogsteen hydrogen bonding [1,2].

G4s are characterized by the relative direction (parallel vs. antiparallel) of the strands connecting the guanines, by the syn vs. anti glycosyl conformation, by the nature and length of G–G connecting loops, by the intra- vs. inter-molecular nature of the structure and by the number of stacking tetrads. In addition, the conformational properties of G4 are largely influenced by the environmental conditions, in particular by the presence of monovalent ions such as K^+ or Na^+. These features grant high degree of polymorphism to G4 arrangements and make them suitable to differential recognition. In addition, relatively mild changes in the experimental setting or addition of specific low molecular weight ligands may lead a G-rich sequence to fold/unfold, hence conferring G4s the characteristics of a molecular switch.

Due to these properties, G-quadruplex structures do not only represent novel nucleic acid arrangements worth of scientific investigation, they also emerge as biologically significant due to the presence of G-rich sequences in specific regions of the genome. In particular, guanines are over-represented in the terminal repeating sequences of chromosomes (telomeres) and in promoter regions of genes, especially proto-oncogenes, such as c-myc, c-kit, bcl-2, VEGF, H-ras and N-ras, as well as in other human genes. In addition, G4s can be selectively formed at the RNA level further contributing to a modulation of the information flow leading to proteins [3,4]. These findings suggest a role of G4 in controlling biological events including chromosome protection and gene expression [5–8] and foresee several potential biophysical, diagnostic and therapeutic applications for G4. Recent reviews cover this matter thoroughly [9–11].

G4s effect on regulation of physiological (or pathological) processes can be considered in two ways. They can be exploited as targets for protein intervention, thus modulating their basal activity, or, alternatively, they can be potentially used as non-physiologic players to produce desired cellular effects. This latter concept is evidenced by the activity exhibited by some G4 folded sequences (aptamers) toward selected targets. Their strong affinity and specificity make them a sort of nucleic acid-based antibody. For recent reviews see references [7,12–14].

In this review, we will consider the most recent information available on the role played by G4s interactions with proteins, both to unravel naturally occurring recognition and regulation pathways with physiological and pathological relevance and to help identifying yet undisclosed non-physiologic nucleic acid structures (aptamer) able to interfere with biological processes.

In approaching this subject, we wish to remind the following important issues, which are relevant to our discussion, i.e.:

- a specific interaction occurring *in vitro* does not represent a safe proof for the existence of such an interaction *in vivo*;
- G4s are very plastic structures and can easily undergo conformational changes during recognition processes or coexist in

* Corresponding author. Tel./fax: +39 049 8275699.
E-mail address: manlio.palumbo@unipd.it (M. Palumbo).

0300-9084/$ – see front matter © 2011 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.biochi.2011.04.018

different structures with diverse binding affinities to a given target;
- at present very few three-dimensional structures of G4-protein complexes are available, which renders molecular modeling and rational targeting a difficult task.

2. In vitro selected protein-G4 recognition

2.1. Protein-directed G4 aptamers

Aptamers are nucleic acids sequences able to appropriately fold to specifically recognize target molecules with specificities reminiscent those exhibited by antibodies. Aptamers include potential applications as anticoagulants, antivirals, antirheumatics, antiproliferartive agents and biosensors. For recent reviews see reference [15].

Although G4 represents just one of several possible options of aptamer architecture, a conspicuous number of G4-based aptamers are reported, the most celebrated one being the thrombin-binding 15mer GGTTGGTGTGGTTGG encompassing two G4 tetrads. Its complex with thrombin provides one of the very few X-ray structures of a G4 bound to a protein thus far available (Fig. 1) [16]. The aptamer is sandwiched between two thrombin molecules and binds (largely through salt bridges) the fibrinogen exosite of one and the heparin binding site in the C-terminal region of the other thrombin molecule. The TGT sequence forming one of the three loops is involved in a hydrophobic cluster near the fibrinogen exosite (Ile^{24}, His^{71}, Ile^{79}, Tyr^{117}), whereas a T belonging to one of the TT loops associates with His^{91}, Pro^{92} and Trp^{237} of the heparin binding site. Thus, both electrostatic and hydrophobic interactions explain stabilization and recognition of the G4-thrombin complex.

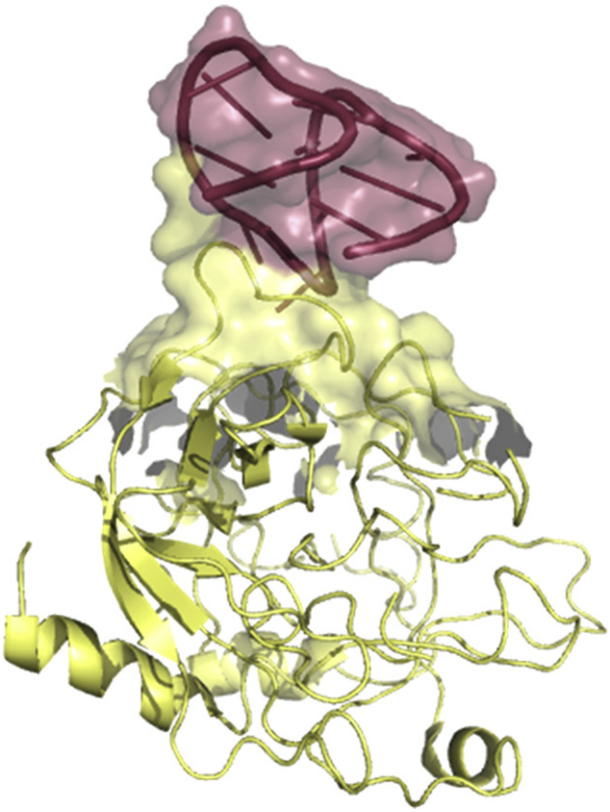

Fig. 1. Three-dimensional structure of the GGTTGGTGTGGTTGG aptamer bound to α-thrombin. PDF 1HUT. The G4 component is highlighted in red, the protein in yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Current advances refer to an improved thrombin aptamer deriving from the incorporation of an unlocked nucleotide monomer (UNA), lacking the 2′-3′ C—C bond in the D-ribose moiety, at each position of the 15mer. Substitution of D-ribose with UNA derivative may produce either destabilization or stabilization of the aptamer G4 structure depending upon its position along the DNA chain. This is likely related to a higher degree of local flexibility produced by the open ribose. Particularly effective was the oligomer modified at position 7, which was found to be more efficient than the original aptamer both in protein binding and in anticoagulant activity [17].

To target VEGF, an aptamer raised against the growth factor receptor-binding domain was identified as the 25mer TGTGGGGGTG GACGGGCCGGGTAGA showing binding constants in the nanomolar range [18]. Its pharmacological interest is related to the fact that it shares the same target with Macugen (Pegaptanib), a non G4 RNA-based aptamer clinically approved against wet macular degeneration [19].

To further emphasize the potential clinical application of G4 aptamers, a cancer chemotherapeutic G-rich oligonucleotide, the 26mer AS1411 (GGTGGTGGTGGTTGTGGTGGTGGTGG), discovered over 10 years ago, is now undergoing phase II clinical trials for the treatment of AML and renal cell carcinoma [20]. Surprisingly, G4 forming potential seems to represent a general feature leading to cancer-selective antiproliferative activity as shown by the impressive data referring to 11 synthetic G-rich oligonucleotide libraries [21].

Potentially clinically useful G4s can target also non-human proteins. As an example, tetra-end linked oligonucleotides having a cluster of four TG_4T sequences, arranged in a parallel G4 were investigated for anti-HIV activity and showed potencies in the submicromolar range. The authors speculate that Arg190, the only conserved residue in the hypervariable V3 loop of GP120 represents a key site for aptamer binding, recognition being generally driven by electrostatic forces [22].

Finally, the field of G4 aptamer recognition can be also profitably exploited for developing biosensing systems for analytical application through the use of logic gate strategies. Very recently, synthetic receptors have been generated by functionalized G4 sequences, specifically an octanucleotide with a G_5 stretch, fit for targeting protein surfaces and for sensing specific intervening contacts. The assembled DNA quadruplexes are modified with three fluorophores exhibiting individual but overlapping emission spectra. This will produce distinctive emission patterns when interacting with different proteins. The observed differences can be used both to monitor protein concentration levels and to investigate analyte combinations producing conformational changes on the probe [23,24].

A list of G4 aptamers recognizing specific protein targets is reported in Table 1.

2.2. G4-directed antibodies and peptides

Using a sort of “mirror” approach, another line of research has led to major advances in the production and development of antibodies and peptide sequences able to bind G4s very efficiently. One of the first examples concerns a high-affinity (sub-nanomolar range) single-chain antibody fragment able to recognize the G4 formed by the *Stylonychia lemnae* telomeric repeat [38]. Peculiarly, this protein is effective in discriminating between parallel or antiparallel quadruplex conformations formed by the same sequence.