Biochimie 105 (2014) 22–35

Contents lists available at ScienceDirect

Biochimie

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

Research paper

Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations

Petr Stadlbauer [a], Lukáš Trantírek [b], Thomas E. Cheatham III [c], Jaroslav Koča [b], Jiří Šponer [a, b, *]

[a] Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic
[b] CEITEC – Central European Institute of Technology, Masaryk University, Campus Bohunice, Kamenice 5, 625 00 Brno, Czech Republic
[c] Department of Medicinal Chemistry, College of Pharmacy, University of Utah, Salt Lake City, UT 84124, USA

ARTICLE INFO

Article history:
Received 20 April 2014
Accepted 1 July 2014
Available online 17 July 2014

Keywords:
G-DNA folding
Quadruplex
Triplex
Telomere
Molecular dynamics

ABSTRACT

We have carried out extended set of μs-scale explicit solvent MD simulations of all possible G-triplexes which can participate in folding pathways of the human telomeric quadruplex. Our study accumulates almost 60 μs of simulation data, which is by about three orders of magnitude larger sampling compared to the earlier simulations of human telomeric G-DNA triplexes. Starting structures were obtained from experimental quadruplex structures by deleting either the first or the last strand. The life-times of antiparallel triplexes with lateral and diagonal loops are at least on μs-scale, which should be sufficient to contribute to the folding pathways. However, the triplex states may involve structures with various local deviations from the ideal triplexes, such as strand tilting and various alternative and incomplete triads. The simulations reveal easy rearrangements between lateral and diagonal loop triplex topologies. Propeller loops of antiparallel triplexes may to certain extent interfere with the G-triplexes but these structures are still viable candidates to participate in the folding. In contrast, all-parallel all-anti triplexes are very unstable and are unlikely to contribute to the folding. Although our simulations demonstrate that antiparallel G-triplexes, if folded, would have life-times sufficient to participate in the quadruplex folding, the results do not rule out the possibility that the G-triplexes are out-competed by other structures not included in our study. Among them, numerous possible misfolded structures containing guanine quartets can act as off-path intermediates with longer life-times than the triplexes. Besides analyzing the structural dynamics of a diverse set of G-DNA triplexes, we also provide a brief discussion of the limitations of the simulation methodology, which is necessary for proper understanding of the simulation data.

1. Introduction

Telomeres are specialized, functional DNA-protein structures found at the ends of all eukaryotic linear chromosomes. They help to protect the ends of chromosomes from being treated like damaged DNA needing repair, and they also facilitate complete replication of the chromosome. In vertebrates, telomeres comprise of double-stranded DNA of simple repetitive sequence, d(TTAGGG).d(AATCCC), terminating in a single-stranded G-rich 3′-overhang. Inherent property of the single-stranded G-rich 3′-overhang sequences is their ability to adopt non-canonical DNA structure, namely G-quadruplex. The G-quadruplex formation in telomeric G-rich DNA has been demonstrated both in vitro (reviewed in [1]) and in vivo [2–4]. Recent studies have indicated that telomeric G-quadruplex plays active roles in telomere assembly and integrity and replication/transcription of the telomeric DNA [5–9]. Moreover, as the stabilization of telomeric G-DNA by small molecular weight ligands inhibits activity of telomerase, enzyme responsible for elongation of telomeric DNA and being overexpressed in more than 80% of cancers, the stabilization of the telomeric G-quadruplexes by the ligands has emerged as a potential strategy in the anticancer therapy [10,11].

The most salient structural feature of G-quadruplexes is formation of planar guanine tetrads mediated by Hoogsteen-type guanine–guanine base-pairing. Consecutive tetrads are stacked and stabilized by cations in the central channel. Inherent property

* Corresponding author. Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic. Tel.: +420 541 517 133; fax: +420 541 212 179.
E-mail address: sponer@ncbr.muni.cz (J. Šponer).

http://dx.doi.org/10.1016/j.biochi.2014.07.009

of the telomeric G-quadruplexes is their structural polymorphism. Biologically relevant monomolecular quadruplexes can adopt various topologies depending on their sequence and surrounding conditions [12–18]. G-strands can be either parallel or antiparallel, and are connected via loops, classified into lateral, diagonal or double-chain reversal (propeller) types, which gives rise to parallel, hybrid and antiparallel topologies [10,15,19–28]. Moreover, additional diversity is brought by *syn* or *anti* orientation of the glycosidic torsion angle χ of guanosines in the G-stem. The possible *syn*/*anti* patterns are constrained by strand orientations and affected also by presence or absence of flanking nucleotides at the beginning/end of a sequence [29,30]. Thus far, six distinct folding topologies of telomeric G-quadruplex were reported [12–16,27].

While the diverse telomeric G-DNA structures have been rather well characterized experimentally, much less is known about folding kinetics and folding pathways for the individual G-quadruplexes. Importance of the kinetic information has emerged recently along with the experimental data suggesting that number of physiologically relevant processes involving telomeric G-DNA folding/unfolding such as chromatin remodeling, replication, transcription, protein binding or sequestration of non-coding RNA might be under kinetic rather than thermodynamic control [31,32]. However, the experimentally observed folding rates and the number of folding phases may be affected by the experimental setups involving the cation and buffer conditions, the methods used to interrogate the folding, or whether the G-DNA construct used for the investigations has flanking nucleotides [32–36]. Yet, regardless of the discrepancies, these studies have collectively indicated that telomeric G-DNA folds on millisecond to second time scale via multiphasic processes involving intermediates [32–34,37–39]. Recently, the pathway for unimolecular quadruplex folding of human telomeric sequence was simultaneously monitored by several kinetics experiments, providing the so far most comprehensive insight into the G-DNA folding process [38]. This study suggested that the folding process consisted of phases with times scales ranging from ms to 1000 s.

The triplex structure was suggested as one of the most plausible intermediates in folding pathway of human telomeric quadruplex [37,40–45]. Yet, due to the limited temporal and structural resolution of the experimental techniques used to monitor G-DNA folding as well as due to the model-dependent nature of kinetic data interpretation, the structure of the intermediate (triplex) has remained speculative and the data have not provided any deeper insight into molecular events taking place along the folding pathway. Actually, while some studies suggested triplexes acting as on-pathway intermediates, other studies suggest triplexes acting as long-living off-pathway intermediates [46].

In principle, although limited by force field accuracy and affordable time scale, computer simulations (molecular dynamics, MD) can complement experimental data and provide insights into selected aspects of G-DNA folding [47–54]. For example, standard simulations suggested participation of straightforward vertical strand-slippage movements in late stages of formation of parallel tetrameric all-*anti* G-DNA stems [47,48]. In contrast, antiparallel and hybrid G-stems are unable to undergo direct vertical slippage of their strands due to steric conflicts between *syn* and *anti* oriented bases [48]. Thus, their rearrangements rather occur via strand binding-unbinding processes coupled with *syn* − *anti* dynamics of the individual nucleotides. This observation would indirectly support presence of the triplex intermediates.

Standard (unbiased) simulations allow observing only those motions that a real molecule would sample on the simulation time scale, which is presently in the microsecond range. Standard simulations, therefore, are too short to visualize full folding of G-quadruplexes. This limitation can be partially alleviated by investigation of a diverse spectrum of starting structures that could occur along the folding pathway [47,48]. The alternative option is utilization of enhanced sampling simulations which use various approaches to overcome the energy barriers. Steered dynamics, temperature replica exchange molecular dynamics (T-REMD) and metadynamics have been used to investigate G-DNA unfolding [50–53]. However, enhanced sampling simulations are not fully equivalent to standard simulations and their outcome may be affected by the specific method of enhanced sampling [55]. T-REMD accelerates crossing over enthalpic barriers but cannot increase sampling of entropy-driven processes and therefore changes the entropy/enthalpy balance. Metadynamics assumes that the unfolding process can be sufficiently completely described by only a few collective variables providing a coarse-grained description of all the important slow modes. If the real process includes additional slow motions that are orthogonal to the defined collective variables (which may happen for complex systems), the metadynamics description becomes inaccurate and incomplete. Metadynamics can be unsatisfactory when different stages of the folding path or different routes of a multi-pathway process are dominated by different types of slow motions, requiring different collective variables. The so far most advanced enhanced sampling G-DNA study reported unfolding of the human telomeric hybrid-1 quadruplex using four-dimensional bias-exchange metadynamics using four collective variables [54]. The study also illustrates (as discussed by its authors) limitations of the metadynamics, such as the lack of acceleration of *syn*/*anti* guanine reorientations not included in the collective variable set. This precludes sampling of alternative folds and of misfolded structures. Another study combined metadynamics simulations with NMR experiments which provided evidence for formation of the triplex structure by truncated sequence of the thrombin binding aptamer (15-TBA) DNA [50]. Note that none of the currently available studies could simulate true folding of G-DNA. All studies have been initiated from folded structures and thus capture unfolding. The accelerated disruption of a folded molecule does not necessarily capture the folding pathway and, in addition, the results may be influenced by the specific enhanced sampling method. For example, the T-REMD [53] and metadynamics [50] models of 15-TBA G-DNA unfolding differ visibly. As demonstrated for very simple RNA tetraloops, achieving folding from the unfolded state would be considerably more difficult [56]. Therefore, despite some successes, we remain far from an exhaustive computer description of true G-DNA folding processes.

An important suggestion emerging from recent standard simulations is the possibility that folding pathways of G-DNA molecules include at the atomistic level numerous intermediates which cannot be resolved by the existing experimental techniques, resulting in extremely multi-pathway process [48]. Even a capture of a single ion by a G-DNA stem likely proceeds via a number of distinct micro-routes [57]. High complexity has been explicitly revealed by standard simulations in folding studies of the smallest DNA hairpins that fold on microsecond time scale [58]. Thus, the human telomeric G-quadruplex folding pathways suggested by the recent experiments could potentially be oversimplified by the resolution of available methodologies. G-quadruplex folding may resemble folding funnels suggested for proteins and RNA [59–62].

Polymorphism of folded G-quadruplex structures in the thermodynamics equilibrium is well established [12–18,21,25,27]. However, diversity of the structures occurring during folding processes may be much larger. A single G-DNA folding strand can in principle adopt 2^N different combinations of *syn* and *anti* G orientations (N is the number of guanines involved in the G-DNA