Biochimie 93 (2011) 1267–1274

Contents lists available at ScienceDirect

Biochimie

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

Research paper

Conformational studies and solvent-accessible surface area analysis of known selective DNA G-Quadruplex binders

Stefano Alcaro*, Anna Artese, Giosuè Costa, Simona Distinto, Francesco Ortuso, Lucia Parrotta

Dipartimento di Scienze Farmacobiologiche, Laboratorio di Chimica Farmaceutica, Università degli Studi "Magna Græcia", Complesso "Ninì Barbieri", Roccelletta di Borgia, 88021 (Catanzaro), Italy

ARTICLE INFO

Article history:
Received 1 March 2011
Accepted 14 June 2011
Available online 16 June 2011

Keywords:
G-quadruplex
Conformational search
Solvent-accessible surface area
Boltzmann population
TRAP

ABSTRACT

Human telomeres are comprised of *d*(TTAGGG) repeats involved in the formation of G-quadruplex DNA structures. Ligands that stabilize these G-quadruplex DNA structures are potential inhibitors of the cancer cell-associated enzyme telomerase. In human cells, telomerase adds multiple copies of the 5′-GGTTAG-3′ motif to the end of the G-strand of the telomere and in the majority of tumor cells it results over-expressed. Several structural studies have revealed a diversity of topologies for telomeric quadruplexes, as confirmed by the different conformations deposited in the Protein Data Bank. In recent years an increasing number of chemically diverse telomerase inhibitors have been identified, including both natural and synthetic compounds. Thus telomerase has been regarded as one of the most attractive targets in cancer treatment. In this manuscript, with the aim to rationalize the different experimental activities of known telomerase inhibitors, a computational study was carried out to investigate their conformational properties and the relationships between the target affinity and the ligands solvent-accessible surface area. Among the analyzed different scaffolds of G-quadruplex binders, such a descriptor provided helpful preliminary information to discriminate end-stacking ligand binding affinities, revealing itself as a useful predictive tool in drug design and lead optimization processes.

1. Introduction

The telomeres of human cells protect chromosomal ends from fusion events; they have a length from 3 kb to 15 kb and are composed of tandem repeats of the sequence 5′-GGTTAG-3′ with 3′ overhang of the G-strand essential for structural and functional roles [1]. In human somatic cells, telomere length decreases with each cell division event [2]. The infinite extension of telomeric ends is associated with an aberrant cellular proliferation [3]. Cancer is characterized by such an immortalized state, in most cases by the activation of the reverse transcriptase enzyme telomerase, which is expressed in 85–90% of cancer cells and not significantly expressed in somatic tissue [4]. Since telomere maintenance is associated with the unlimited proliferative potential of cancer cells, most telomere related antitumor strategies target the telomerase-dependent mechanism. Several studies have demonstrated telomerase as a reliable marker [5] for some cancers as well as a target of inhibitors of immortal cell growth [6].

In presence of certain cations, the eukaryotic chromosomes at the ends of telomeric DNA G-rich sequences can associate together to form particular four-stranded conformations known as quadruplexes, tetraplexes or G4 structures [7]. Quadruplexes can be formed from one, two or four separate strands of DNA (or RNA) and display a large variety of topologies, widely analyzed by means of structural studies [3]. All quadruplexes are characterized by a repeating motif known as G-quartet, also defined G-tetrad, characterized by a high tendency to produce self-stacking [8]. Quadruplexes have a requirement for metal ions, especially the alkali metals [9], placed in the interior channel formed at the center of each quadruplex. The loops observed in the different quadruplex topologies can be diagonal, lateral or chain-reversal. In particular the number of G-quartets, the loop length, sequence and sometimes the nature of the cation are related to the plausible diverse structural arrangements. By definition chain-reversal loops connect two strands in the same parallel orientation, whereas diagonal and lateral loops connect chains in opposing, antiparallel orientations [10].

In presence of Na^+ and K^+ ions, six different main conformations of the monomolecular human telomeric sequence have been identified and deposited in the Protein Data Bank (PDB) [11] respectively with the codes 143D (antiparallel, Na^+) [12], 1KF1 (parallel, K^+) [13], 2HY9 (hybrid-1, K^+) [14], 2JPZ (hybrid-2, K^+) [15], 2JSL (hybrid-1, K^+) [16], and 2JSM (hybrid-2, K^+) [16].

* Corresponding author. Tel.: +39 0961 3694197.
E-mail address: alcaro@unicz.it (S. Alcaro).

0300-9084/$ – see front matter © 2011 Published by Elsevier Masson SAS.
doi:10.1016/j.biochi.2011.06.014

Such DNA conformations have provided novel targets for designing compounds with improved specificity. G-quadruplex selective binders or stabilizers may interfere with telomere conformation and elongation.

Synthetic small molecules were the first ligands identified to fold G-quadruplex and indirectly inhibit telomerase activity [17]. Starting from this initial work, several quadruplex binding molecules have been identified as telomere blocking agents, but corresponded to weak inhibitors of telomerase extension [18,19a,b]. Structure-activity relationship studies and physico-chemical measurements highlighted large aromatic ring-systems bearing positively charged side chains as main structural features to be efficient G-quadruplex ligands. Secondary structure stabilization and thus telomerase inhibition implies π–π stacking between planar aromatic molecules and the external guanine-rich face of G-quadruplex and electrostatic interactions of positively charged functions with the phosphate backbone of DNA [20–22].

Telomerase inhibitors of natural and synthetic origins are chemically diverse ligands such as telomestatin [23], ethidium derivatives [17], bisamido-anthraquinones [24], fluorenones [25a,b,c], acridones [26], acridines [27], perylene diimides [28], fluoroquinolones [29a,b], indoloquinolines [30], cryptolepines [31a,b,c], quindolines [32], porphyrins [33], phenanthrolines [34], triazines [35] and carbazole derivatives [36]. Among all these numerous heterocyclic systems, some substituted acridines excelled by their efficiency. Thus, the 3,6-bis(aminoalkyl) side-chain substituted acridine BRACO displayed potent G-quadruplex stabilization ability and Telomere Repeat Amplification Protocol (TRAP) inhibitory potential [37–39]. Replacement of the C-9 dimethylaminoaniline moiety by a difluorobenzyl group allowed pharmacological profile improvements [40].

Another class of compounds lately reported included perylene derivatives characterized by side chains ending with linear or cyclic amines. These molecules were capable of strong and selective recognition to the G-quadruplex showing a remarkable biological and pharmacological anticancer effects in living systems [41].

Even though experiments with the parent compound, ethidium bromide, had proven low selectivity toward G-quadruplexes, several ethidium derivatives, characterized by a high affinity and selectivity toward quadruplex DNA, were synthesized. Such an improvement was probably due to the aromatic surface of the derivatives, along with the positive charge inside the ring system, which could interact with negative charges in the G-quadruplex and also promote binding. Actually it has been shown that these ethidium derivatives not only stabilize G-quadruplex, but they promote its formation [42].

The binding of a ligand to its receptor is dependent on several chemical-physical factors. Many descriptors based approaches have been successfully applied to drug discovery process [43]. These methods correlate calculated properties derived from the chemical structure of examined compounds to experimentally determined biological activity and are generally indicated as Quantitative Structure–Activity Relationship (QSAR) methods. Molecular properties can be determined through experiments, but more often computational methods have been devised to calculate them from the topology of any given molecule. Although 3D-descriptors are conformation-dependent and require more computational resources, 3D QSAR approaches are the most powerful and applied ones [44]. Among the alignment dependent methods, CoMFA is the most popular [45], however there are other 3D approaches which are independent from alignment. Some examples reported in literature consider distributions of molecular surfaces (MaP) [46], comparative molecular moment analysis (CoMMA) [47], and comparative spectra analysis (CoSA) [48] that apply molecular moments and molecular spectra as descriptors.

$$p_i = e^{-E_i/RT} \Big/ \sum_1^n e^{-E_i/RT}$$

Scheme 1. The Boltzmann probability is computed as the ratio of the single microstate exponential and the partition equation. The value of p_i is between 0 and 1.

In the present study we report the application of the well documented Averaged Solvent Accessible Surface Area (ASASA) descriptor as a tool to give direct insights into the role of the solvent in the binding event of known G-quadruplex binders exhibiting different conformational profiles.

The concept of Solvent Accessible Surface Area (SASA) is of great importance in biological field. It was first introduced to evaluate the solvation effects in proteins and their complexes [49a,b,c,d]. It also became an important descriptor of the protein structure itself and has been used for the identification of binding sites and protein–protein interfaces [50]. Moreover DNA ability to form double helix or unwind strongly depends on hydration [51]. In a recent work [52] the exploration of amino acid solvent accessibility was applied to DNA in order to give additional insights of DNA-water-protein interactions. Furthermore the important role of the ASASA descriptor was also reported for the interaction of DNA-small molecules [53,54], while only recently it was applied to investigate its significance for G-quadruplex binders [55–57].

2. Materials and methods

2.1. Monte Carlo (MC) conformational search

The conformational search was carried out by molecular mechanics techniques coupled to the analysis of the Solvent Accessible Surface Area (SASA) computed onto a defined planar aromatic system depending on the chemical nature of the analyzed ligands. In details, after building the 3-D structures of each compound, we proceeded with the exploration of their internal degrees of freedom by Monte Carlo (MC) randomization of any rotatable bond. 10,000 conformations were generated and submitted to 5000 iterations of energy minimization using the Polak–Ribiere Conjugated Gradient (PRCG) algorithm [58], AMBER* as force field [59] with the united atoms notation and the implicit model of solvation GB/SA water [60] as implemented in Macro Model ver. 7.2 [61a,b].

For all the studied G-binders the convergence in the conformational search was evaluated using the averaged number of duplicate conformers that resulted always higher than 2, thus indicating a good conformational space exploration.

Among all the generated conformations, those within 3 kcal/mol above the energy global minimum were selected and submitted to the SASA analysis.

2.2. BASASA (Bolzmann averaged solvent accessible surface area) descriptor

The SASA calculations were computed on defined aromatic portions of the selected compounds. In particular such a descriptor allowed the estimation of the degree of exposition of the planar rings of three different classes of ligands (acridine, perylene and ethidium scaffolds) to the solvent, providing an indirect measure of the stacking potential property of our compounds.

$$BASASA = \sum_1^n p_i \cdot SASA_i$$

Scheme 2. The BASASA is a weighted descriptor, calculated as the integral of the Boltzmann population (p_i) and surface area values of each conformer ($SASA_i$).