Aggregating distryrylpyridinium dye as a bimodal structural probe for G-quadruplex DNA

<table>
<thead>
<tr>
<th>Journal:</th>
<th>New Journal of Chemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>NJ-LET-05-2015-001325.R1</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Letter</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>17-Jun-2015</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Xie, Xiao; Institut Curie, Centre de Recherche, CNRS UMR9187 / Inserm U1196  
                      Renvoisé, Aurélie; AP-HP, Hôpital Pitié-Salpêtrière, Bactériologie-Hygiène,  
                      Centre National de Référence des Mycobactéries et de la Résistance des  
                      Mycobactéries aux Antituberculeux; INSERM, U1135, Centre  
                      d'Immunologie et des Maladies Infectieuses, Team E13 (Bacteriology);  
                      Sorbonne Universités, UPMC University Paris 6, CR7, Centre d'Immunologie  
                      et des Maladies Infectieuses, Team E13 (Bacteriology)  
                      Granzhan, Anton; Institut Curie, Centre de Recherche, CNRS UMR9187 /  
                      Inserm U1196  
                      Teulade-Fichou, Marie-Paule; Institut Curie, Centre de Recherche, CNRS  
                      UMR9187 / Inserm U1196 |
A colorimetric and fluorimetric probe for detection of G-quadruplex (G4) DNA was designed by introduction of coumarin units into the distyrylpyridinium chromophore. It has been successfully applied for optical discrimination of G4 vs. non-G4 DNA and validated using two previously unknown G4-forming sequences, identified in the genome of Mycobacterium tuberculosis.

Guanine-rich nucleic acid sequences are capable of folding into four-stranded structures called G-quadruplexes (G4), composed of several layers of guanine quartets held by hydrogen bonds and coordination to metal cations. G4 structures attract considerable interest because of their putative roles as regulatory elements in DNA- and RNA-related processes such as recombination, replication, transcription and telomere maintenance. Thus, a number of putative G4-forming sequences (PQS) have been identified in human genome, particularly in promoter regions of oncogenes and in telomere overhangs, as well as in genomes of bacteria and viruses. The PQS are typically identified through a search of sequence patterns containing (usually) four repeats of at least three consecutive guanines separated by one to seven nucleotides, using bioinformatics algorithms such as Quadfinder or QGRS Mapper. However, the growing body of structural data on G-quadruplexes gives evidence that these structures do not necessarily follow the consensus motif; for instance, snap-back G4 structures feature isolated guanine bases which participate in formation of guanine quartets, leading to complex folding topologies which currently can hardly be predicted by bioinformatics algorithms. On the other hand, sequences fulfilling the generally accepted consensus may fail to adopt G4 structures in vitro. In this context, experimental verification of formation of G4 structures is essential both for their discovery in genetic screens and for optimization of bioinformatics tools.

Besides high-resolution structural methods such as high-field NMR or X-ray crystallography, which can provide detailed information on secondary structures of nucleic acids, a number of biophysical techniques can deliver sufficient proof of G4 formation in vitro. Among those, thermal denaturation experiments, thermal difference spectra (TDS), circular dichroism, chromatographic and gel electrophoresis-based methods have been more or less widely employed, but require specialized equipment, remain rather time-consuming, and are poorly suited for high-throughput identification of PQS in genomic samples. Therefore, there is a need for fast, inexpensive and reliable methods for detection of G4 structures formed by DNA sequences.

With increasing availability of multimode microplate readers and the even more performant DNA microarray technology, optical methods (absorption and fluorescence) have become the technique of choice for high-throughput analysis. Along these lines, several groups undertook attempts to design optical probes for sensitive and selective detection of G4 structures. Fluorescent probes selective for particular G4 topologies, such as parallel or anti-parallel folds, have been also developed. However, there is still a need of optical probes capable of robust detection of G4-DNA structures irrespective of their topology, but with a sufficient selectivity over other (non-G4) DNA forms. Recently, N-methylmesoporphyrin IX and Thioflavin T (ThT) have been proposed as universal G4-selective fluorescent probes due to their high fluorescence enhancement in the presence of a large panel of G4 structures; nevertheless, in the case of ThT, the fluorimetric response delivers a significant overlap between G4 and non-G4 structures, which limits its applicability as a discriminating probe.

Colorimetric probes, especially ratiometric ones, i.e. those operating via absorption changes at two wavelengths, represent a valuable alternative to fluorescent probes due to the simplicity of the experimental set-up and robustness of the absorbance signal. In this context, the isaindigotone–thiazole
orange hybrid (ISTO) and the isandigotone−coumarin hybrid (ISCH-1) have been developed as efficient colorimetric probes for discrimination of G4 and non-G4 DNA structures due to the G4-induced disaggregation of the probe and the accompanying colour change. In addition to colour changes, ISCH-1 undergoes a significant fluorescence enhancement (15 to nearly 100-fold) upon interaction with G4 structures, allowing their discrimination from other DNA forms.

In our previous work, we identified 2,4-distyrylpyridinium dyes as another promising scaffold for G4-DNA-selective fluorescent probes. Upon further investigation of this series, we discovered that a coumarin-substituted distyrylpyridinium derivative BCVP (2,4-bis[(7-diethylamino-3-coumarinyl)-vinyl]-1-methylpyridinium, Fig. 1a), displays both colorimetric and fluorescent signals in the presence of a wide panel G4-DNA structures. We report herein our investigation of this probe and its use for in vitro detection of G4-DNA structures.

The distyryl dye BCVP was synthesized via a Knoevenagel condensation of 1,2,4-trimethylpyridinium iodide with two equivalents of 7-diethylamino-3-formylcoumarin (ESI†). BCVP displays photophysical properties typical for distyryl-type donor−acceptor dyes, i.e. a long-wavelength absorption spectrum (λ_{max} = 527 nm in MeOH) and a weak fluorescence with a large Stokes shift in the red region of the spectrum (ϕ = 0.003, λ_{max} = 650 nm in MeOH, Fig. S2, ESI†). Despite the presence of intrinsically fluorescent coumarin units in the dye molecule, they do not manifest in absorption or emission spectra and behave simply as electron-donating groups within the D−π−A−π−D scaffold of a distyryl dye (D = electron donor, A = electron acceptor units). This behaviour is similar to that of other styryl dyes containing conjugated coumarin groups.

The absorption spectrum of a salt-free aqueous solution of BCVP is almost identical to that in MeOH; however, even at relatively low ionic force (10 mM lithium cacodylate buffer) the absorption band of BCVP undergoes a strong blue shift (λ_{max} = 459 nm), which we ascribe to formation of H-aggregates of the dye (Fig. S3, ESI†). Formation of aggregates in buffer solution was further confirmed by changes of absorption spectra to the monomeric form (λ_{max} = 525 nm) upon increase of the temperature (Fig. 1b) as well as upon lowering dye concentration at a constant temperature (Fig. 1c).

Interaction of BCVP with G4 and non-G4 DNA structures was initially studied by spectrophotometric and spectrofluorometric titrations with 22AG (human telomeric G4, cf. Table S1, ESI†) and calf thymus DNA (ct DNA) as representative G4 and double-stranded structures. Addition of 22AG as well as several other G4 structures (c-kit2, hras-1) resulted in a decrease of absorption band of H-aggregates and formation of the red-shifted (λ_{max} = 555 nm) monomeric band of the dye (Fig. 2a). This was accompanied by a ~2.5-fold increase and a strong blue shift (from 705 to 640 nm) of the fluorescence emission band of the dye (Fig. 2c). These changes were attributed to binding of BCVP monomers to G4-DNA, leading to disaggregation of H-aggregates, as in the case of other similar systems.

A very different behaviour was observed upon titration with ct DNA: the absorption band of H-aggregates of the dye was slightly decreased without any significant shift or formation of the monomer band, while the fluorescence band was quenched by about 70%, again without any significant shift (Fig. 2b and d). In circular dichroism (CD) spectra, we observed formation of weak positive induced CD bands of BCVP at λ = 460 nm (i.e. corresponding to H-aggregates) in the presence of ct DNA (Fig. S4, ESI†). Altogether, these observations point out to interaction of H-aggregates of BCVP with double-stranded DNA, without disaggregation of the dye.
investigate the bimodal (absorption and fluorescence) optical response of this dye towards a larger panel of DNA structures, including 17 well-characterized G4 structures, nine random, presumably single-stranded sequences, and six double-stranded structures (sequences and structural details, cf. Table S1, ES1†). The secondary structures of selected DNA sequences were additionally confirmed by their TDS and UV melting profiles (Fig. S5–S9, ES1†). The absorption spectra and emission intensity at 670 nm in the presence of five molar equivalents of DNA oligonucleotides (or ct DNA at an equivalent nucleotide concentration) were monitored using a microplate reader. The data shown at Fig. 3 demonstrate that the absorption ratio at 450 and 555 nm ($A_{555 \text{nm}}/A_{450 \text{nm}}$) can be used as a discriminating factor for G4 vs. non-G4 structures: thus, for all G4 structures from our test set (blue bars on Fig. 3a) the $A_{555 \text{nm}}/A_{450 \text{nm}}$ value exceeds 1.5, whereas in the case of single-stranded (except for ss 8) and duplex structures it does not exceed this threshold value. A similar discriminating profile was observed using the fluorimetric response of BCVP, i.e. the relative increase of fluorescence intensity ($S$). Thus, the fluorescence intensity of BCVP was enhanced from 2.1-fold to 11-fold in the presence of various G4-DNA structures, with the exception of thrombin-binding aptamer (TBA) which caused a slight decrease of fluorescence intensity ($S = 0.9$). Remarkably, TBA represents a “hard case” in fluorescent sensing of G4-DNA,16,21,25 which is probably the reason why it is often omitted from the test sets of G4 structures in other studies.20

In our case, the combination of two outputs (absorption and fluorescence) improves the predicting power of the probe, as TBA is correctly assigned to G4 structures on the basis of its absorbance signal. Single-stranded (except for ss 8) and all duplex DNA structures induced decrease of fluorescence intensity of BCVP ($S < 1$), similarly to what was observed with ct DNA. Thus, a threshold value of $S = 1$ may be used for fluorimetric discrimination of G4 and non-G4 structures, with only two exceptions (TBA and ss 8)16 among the 32 DNA samples in our set (Fig. 3c). The statistical analysis of optical responses shows that, except for the cases noted above, there is essentially no overlap between the groups of G4 and non-G4 structures (Fig. 3b and d), which represents a significant improvement with respect to the ThT assay.19

In order to formally validate BCVP as a G4-specific probe, we attempted its use for characterization of previously unknown G4-forming sequences. We turned our attention to 31-base guanine-rich sequences repeatedly found within the polymorphic GC-rich repetitive sequence (PGRS) of Mycobacterium tuberculosis H37Rv27 after in silico parsing by qGRS algorithm7 using default parameters.28 Five closely situated (<1500 bp distant) sequences found in the PE_PGRS13 gene (Fig. S10, ES1†), each giving a qGRS score of 42, were aligned to highlight the conservation of multiple runs of two to four consecutive guanines within the 31-base sequence (Fig. 4a). As the corresponding conserved sequence could potentially give rise to multiple G4 structures of complex topology, for the initial studies it was split into two 23-mer sequences, arbitrarily named MTB2.1 and MTB2.2 (Fig. 4b), both including multiple runs of two to four consecutive guanines and thus potentially able to form simpler G4 structures. In the BCVP assay, both sequences are unambiguously classified as G4-DNA structures on the basis of their colorimetric and fluorimetric outputs (Fig. 3a and c, blue-hatched bars) which are close to group medians. To confirm that MTB2.1 and MTB2.2 form G4 structures in vitro, we recorded their TDS and CD spectra as well as UV-melting profiles in K⁺-rich conditions (100 mM KCl), identical to those used in the BCVP assay. The TDS of MTB2.1 and MTB2.2 are characterized by negative peaks at 295 nm and positive peaks at 275 and 245 nm (Fig. 4c), typical for G4 structures. Both sequences show characteristic negative melting transitions at 295 nm, with melting temperatures at 71 and 68 °C for MTB2.1 and MTB2.2, respectively. CD spectra of both oligonucleotides show negative peaks at 240–245 nm and strong positive peaks at 260–265 nm, as well as shoulders at 290–295 nm (Fig. 4d). The latter feature speaks rather in favour of antiparallel G4 structures in both cases, although the present data are not sufficient for an unambiguous assignment of quadruplex topology. Further structural studies will be necessary to characterize the G4 structures adopted by MTB2.1 and MTB2.2, as well as by the 31-mer repeat sequence.

In summary, we have developed a bimodal (colorimetric and fluorimetric) optical probe BCVP, combining coumarin and distyrylpyridinium fragments, for detection of G4 structures in vitro. The combination of colorimetric and absorption outputs leads to a better discrimination between G4 and non-G4 structures compared to the previously used methods, resulting
in only one false positive (ss8). A particularly interesting feature of BCVP is quenching of its fluorescence in the presence of most single- and double-stranded DNA samples, in contrast to the fluorescence enhancement observed in the presence of G4 structures. This represents a clear advantage in the presence of double-stranded DNA. Finally, the validity of the probe was demonstrated by characterization of two previously unknown G4-forming sequences from the PGRS of M. tuberculosis, whose structural details are under investigation. The BCVP protocol is suitable for identification of G4-forming DNA using microplate technology and can, in future, be adapted to high-throughput screening through the use of oligonucleotide microarrays.

Notes and reference
26. Although our data do not indicate that ss8 (5′-GGGACAGAGAGCTGGTGTGGG-3′) adopts a well-defined secondary structure in our conditions (cf. Figure S7, ES1†), we cannot exclude induction of an (unknown) secondary structure in the presence of the probe, as indicated by the changes of the melting profile of this oligonucleotide (Figure S10, ES1†).
28. A detailed analysis of G4-forming motifs in the genome of M. tuberculosis will be published elsewhere.