Methods xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Methods

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

In-line probing of RNA G-quadruplexes

Jean-Denis Beaudoin [1], Rachel Jodoin [1], Jean-Pierre Perreault *

RNA Group/Groupe ARN, Département de Biochimie, Faculté de Médecine et des Sciences de la Santé, Pavillon de Recherche Appliquée au Cancer, Université de Sherbrooke, QC, Canada J1E 4K8

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
RNA G-quadruplex
RNA structure
In-line probing
RNA representation
RNA structure prediction

ABSTRACT

Although the majority of the initial G-quadruplex studies were performed on DNA molecules, there currently exists a rapidly growing interest in the investigation of those formed in RNA molecules that possess high potential of acting as gene expression regulatory elements. Indeed, G-quadruplexes found in the 5′-untranslated regions of mRNAs have been reported to be widespread within the human transcriptome and to act as general translational repressors. In addition to translation regulation, several other mRNA maturation steps and events, including mRNA splicing, polyadenylation and localization, have been shown to be influenced by the presence of these RNA G-quadruplexes. Bioinformatic approaches have identified thousands of potential RNA G-quadruplex sequences in the human transcriptome. Clearly there is a need for the development of rapid, simple and informative techniques and methodologies with which the ability of these sequences, and of any potential new regulatory elements, to fold into G-quadruplexes *in vitro* can be examined. This report describes an integrated methodology for monitoring RNA G-quadruplexes formation that combines bioinformatic algorithms, secondary structure prediction, in-line probing with semi-quantification analysis and structural representation software. The power of this approach is illustrated, step-by-step, with the determination of the structure adopted by a potential G-quadruplex sequence found in the 5′-untranslated region of the cAMP responsive element modulator (CREM) mRNA. The results unambiguously show that the CREM sequence folds into a G-quadruplex structure in the presence of a physiological concentration of potassium ions. This in-line probing-based method is easy to use, robust, reproducible and informative in the study of RNA G-quadruplex formation.

1. Introduction

G-rich DNA and RNA molecules can form a non-canonical tetrahelical structure called a G-quadruplex [1,2]. The primary building block of this structure is named a G-quartet and is composed of four coplanar guanines that form Hoogsteen base pairs involving a total of eight hydrogen bonds [3]. These quartets are stabilized by a central counterion, typically potassium, and stack one on top of the other forming a very stable tetrahelical G-quadruplex structure [4]. There is evidence that this structure forms *in cellulo* and that it is frequently found, at both the DNA and RNA levels, in cellular regulatory sequences such as promoters, telomeres and 5′-UTRs [5]. Many G-quadruplexes have been found to be associated with cell disorders, and, therefore, they constitute good potential therapeutic targets [6].

While most of the early G-quadruplex studies were performed on DNA molecules, more recently a rapidly growing interest has emerged in investigating those formed in RNAs. Moreover, recent research has revealed that the size of the cellular transcriptome is considerably larger than previously thought, with results showing that over 90% of the human genome is actively transcribed [7]. In this new context, the importance of post-transcriptional regulation events is now appreciated more than ever. RNA G-quadruplexes are widely found in the cell and have been shown to act as efficient post-transcriptional regulatory elements that are involved in various biological mechanisms. These include: mRNA splicing, polyadenylation, translation and localization [8–12]. Several thousand potential RNA G-quadruplex sequences have been identified within the human transcriptome [10,13,14]. In order to test the ability of this plethora of RNA sequences to fold into G-quadruplexes, the development of a simple, reliable and reproducible technique is required.

X-ray crystallography and NMR experiments have successfully produced high-resolution structures of many G-quadruplexes which provide a significant amount of information about these structures [15]. However, these techniques are time consuming and describe only one of the structures that can be formed by a given sequence, a structure which does not necessarily correspond to the most abundant one. Clearly, quicker experiments studying the entire population of structures formed in solution must be considered. Circular dichroism (CD) is extensively used to monitor G-quadruplex formation [16]. Of particular importance, CD is able

* Corresponding author.
E-mail address: Jean-Pierre.Perreault@usherbrooke.ca (J.-P. Perreault).
[1] These authors contributed equally to this study.

1046-2023/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2013.02.017

to distinguish parallel structures from antiparallel ones. Depending on its topology, the G-quadruplex structure exhibits characteristic spectral features in CD. Typically, a spectrum exhibiting a positive peak at a wavelength around 264 nm and a negative one around 240 nm is indicative of a parallel structure, whereas a spectrum showing positive peak at 295 nm and a negative one around 260 nm indicates the presence of an antiparallel structure. Since other nucleic acid structures can produce a positive peak around 260 nm, it is important to compare spectra recorded under conditions unfavorable for G-quadruplex formation (e.g. either in the absence of salt, or in the presence of Li^+ which acts inefficiently as the G-quadruplex counterion) with others recorded under favorable conditions (e.g. in the presence of either Na^+ or K^+). A transition to a characteristic G-quadruplex spectrum has to be observed between these conditions in order to suggest the formation of this particular structure. Alternatively, thermal denaturation is also commonly used to study G-quadruplex formation. It corresponds to a melting transition caused by an increase in temperature that can be monitored by either CD (e.g. at 264 nm for the parallel structure), or by the absorbance of UV light at 295 nm [17]. These values allow the determination of the melting temperature (T_m, temperature at which half of the structures are denatured). For sequences able to fold into G-quadruplexes, the calculated T_m are typically higher under favorable conditions (e.g. in the presence of either Na^+ or K^+), reflecting the prominent stability of these structures, as compare to those obtained under unfavorable conditions (e.g. either in the absence of salt or in the presence of Li^+). One of the limits of both the CD and thermal denaturation techniques is that they require relatively high concentrations of either DNA or RNA (i.e. in the low micromolar range). At these concentrations, both intra- and intermolecular G-quadruplex structures can easily be formed. As a result, neither of these two techniques can distinguish these two G-quadruplex topologies.

In-line probing is one of the simplest RNA structure chemical mapping techniques available [18]. This technique is based on the tendency of RNA to be differentially hydrolyzed according to its structure [19]. The phosphodiester bonds of the RNA backbone are susceptible to slow, non-enzymatic cleavage through the "in-line" nucleophilic attack of the 2′-oxygen of the adjacent phosphorus group. This attack occurs when the 2′-oxygen, the phosphorus and the adjacent 5′-oxygen adopt an "in-line" conformation that allows the 2'-oxygen to act as a nucleophile and to efficiently cleave the RNA linkage. Following this logic, the relative rate of spontaneous cleavage is directly related to the surrounding structural features of each RNA linkage. The flexible nucleotides, that is to say those found in single-stranded regions and at the periphery of the RNA structure, are free to adopt various conformations, including the "in-line" geometry, and, consequently, are more susceptible to cleavage. This approach has been extensively used to study both riboswitch secondary structures and the conformational changes that occur upon ligand binding [18].

A recent study, demonstrated the potential of in-line probing in monitoring the formation of intramolecular RNA G-quadruplex structures [10]. It appears to be a very simple, reproducible and informative technique with which to study this motif. Since intramolecular RNA G-quadruplexes are forced to fold into parallel topologies due to their 2′-hydroxyl, C3′-*endo* sugar pucker and *anti* glycosidic bond geometry, they are typically composed of three external loops connecting the guanosine tracks [1]. The nucleotides located in these loops characteristically become highly flexible and are thus more susceptible to spontaneous cleavage upon G-quadruplex formation. This article describes a detailed integrated approach to the study of RNA G-quadruplex formation based on in-line probing analysis. This methodology takes advantage of bioinformatic algorithms for the identification of potential G-quadruplex (PG4) structures, a secondary structure prediction program, in-line probing and both quantification and structural representation software. In order to illustrate the procedure, the PG4 sequence found in the 5′-UTR of the cAMP (cyclic adenosine monophosphate) responsive element modulator (CREM) mRNA was analyzed. This gene encodes a bZIP transcription factor that binds to the cAMP responsive element found in many promoters [20].

2. Materials and methods

2.1. Designing PG4s

Initially, *in vitro* PG4 versions are designed according to a potential G-quadruplex (PG4) sequence identified by a typical bioinformatic approach using the algorithm G_x-N_{1-7}-G_x-N_{1-7}-G_x-N_{1-7}-G_x, where $x \geqslant 3$ and N is any nucleotide (A, C, G or U) [21]. Extra sequences of about 15 nucleotides are added to both the 5′ and the 3′ sides of the PG4 motif (see Fig. 1). The natures of these sequences are identical to those found in the genomic regions flanking the PG4 in question. The main purpose of using extended *in vitro* PG4 versions is to render the analysis more biologically relevant. A previous study reported evidence that both the primary and secondary RNA structure contexts in the vicinity of the G-quadruplex structure were critical to RNA G-quadruplex formation both *in vitro* and *in cellulo* [10].

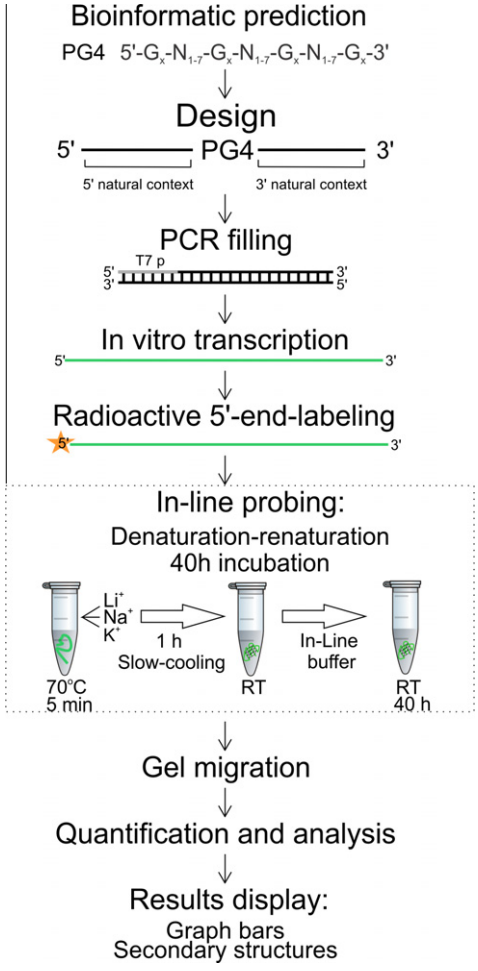

Fig. 1. Organigram of the integrative approach to the study of RNA G-quadruplex formation using in-line probing.

In addition to the wild-type (wt) PG4 version, a mutated version in which some key guanines are substituted for by adenines (G/A-mut) must also be synthesized. It is important to disrupt most of the guanosine tracks of the PG4, as well as to consider the presence of any supplemental guanosine tracks located in either the 5′ or 3′ side (for example see Fig. 2A). The G/A-mutant is a good negative control for G-quadruplex formation as it possesses only minor changes in its RNA sequence as compared to that of the wild type.

2.2. Secondary structure prediction

RNA secondary structure prediction software can be useful for both comparing and analyzing the in-line probing results. The predicted secondary structures of the candidate's *in vitro* PG4 version are retrieved using the RNAstructure software version 5.4 with the default settings [22]. For the CREM PG4 wt version, the two predicted secondary structures with the lower energy values were then manually transposed into dot-and-bracket notations and pictured using the VARNA visualization applet (Fig. 2B and C). A second secondary structure prediction was performed in order to determine the predicted structures adopted by the added 5′ and 3′ flanking sequences (i.e. by using an input sequence in which the potential G-quadruplex was substituted for by multiple adenines, thereby forcing it to form a large loop). The result of this second prediction is shown in Fig. 2D, except that the large adenine loop was replaced for by a representation of the unimolecular parallel G-quadruplex structure predicted by the algorithm seeking G-quadruplex structures (i.e. by taking into account both the length of the G-tracks and the compositions of loops 1, 2 and 3). These various representations of the predicted secondary structures are based on either strictly Watson–Crick base pairs (Fig 2B and C), or on structures that also include the formation of a G-quadruplex structure (Fig. 2D), and can be used as an aid in analyzing the results obtained from the *in vitro* experiments.

2.3. RNA synthesis

After the selection of a candidate, and the design and analysis of the *in vitro* PG4 extended version, the next step is the production of the proper RNA molecules (Fig. 1). Transcripts are synthesized by *in vitro* transcription using T7 RNA polymerase. First, two partially complementary DNA oligonucleotides (2 μM each, Invitrogen) are annealed and double-stranded DNA is obtained by filling the gaps using purified *Pfu* DNA polymerase in the presence of 5% dimethyl sulfoxide (DMSO, Fisher Scientific). One oligonucleotide corresponds to the reverse complementary sequence of the *in vitro* PG4 version with the addition of the 17 nucleotide (nt) reverse sequence of the T7 RNA polymerase promoter at the 3′ end, while the other corresponds to the 17 nt sequence of the T7 RNA polymerase promoter extended by two or more guanines at the 3′ end. In order to obtain good transcription efficiency, the T7 RNA polymerase requires the presence of a minimum of two guanines immediately 5′ of the transcript. If these guanosines are not present within the natural PG4 sequence, the minimal number of guanines must be added in order to fulfill this requirement. This point should be taken into consideration in the designing of the *in vitro* PG4 versions. The DNA duplex containing the T7 RNA polymerase promoter sequence followed by the PG4 sequence is then ethanol-precipitated, ethanol-washed and dissolved in ultrapure water (Barnstead Nanopure). Run-off *in vitro* transcription reactions are then performed in a final volume of 100 μL using purified T7 RNA polymerase (10 μg) in the presence of RNase OUT (20 U, Invitrogen), pyrophosphatase (0,01 U, Roche Diagnostics) and 5 mM NTP (Sigma–Aldrich) in a buffer containing 80 mM HEPES-KOH, pH 7.5, 24 mM $MgCl_2$, 40 mM DTT (Fisher Scientific) and 2 mM spermidine (BioShop). The reactions are incubated for 2 h at 37 °C, and are then treated with DNase RQ1 (Promega) at 37 °C for 15 min (Fig. 1). The resulting RNAs are then purified by phenol–chloroform extraction followed by ethanol precipitation. The RNA products are fractionated by denaturing (8 M urea) 10% polyacrylamide gel electrophoresis (PAGE; 19:1 ratio acrylamide to bisacrylamide) using 45 mM Tris–borate pH 7.5 and 1 mM EDTA (BioShop) solution as running buffer. The transcripts are detected by UV shadowing, and the gel slices containing those corresponding the correct sizes of the *in vitro* PG4s are excised. These slices are then incubated overnight at 4 °C on a rotating wheel in a buffer containing 1 mM EDTA (BioShop), 0.1% SDS (BioShop) and 0.5 M ammonium acetate (Fisher Scientific). The eluted RNAs are then ethanol-precipitated, dried, dissolved in ultrapure water and analyzed by spectrometry at 260 nm in order to determine their concentrations.

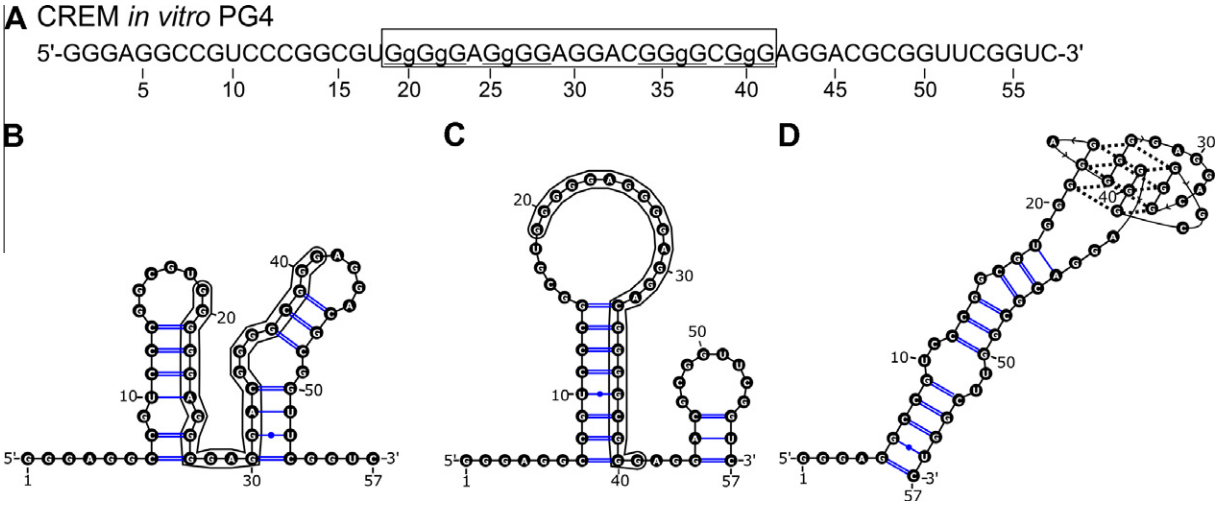

Fig. 2. CREM PG4 sequence and its predicted secondary structures. (A) Nucleotide sequence of the characterized CREM wt transcript. The boxed sequence denotes the predicted PG4. The four G-tracks are underlined. The lowercase guanines (g) correspond to those substituted for by adenines in the G/A-mutant. (B–C) The two secondary structures predicted by the RNAstructure software (version 5.4) for the wt CREM. (D) Possible secondary structures of the additional 5′ and 3′ regions of the PG4, predicted using the RNA structure software and combined with the representation of the unimolecular G-quadruplex structure predicted using the algorithm seeking potential PG4 sequences.

Please cite this article in press as: J.-D. Beaudoin et al., Methods (2013), http://dx.doi.org/10.1016/j.ymeth.2013.02.017

2.4. Radioactive 5′-end-labeling

The next step is to radioactively label the RNA transcripts (Fig. 1). In order to produce 5′-end-labeled RNA molecules, 50 pmol of purified transcripts are dephosphorylated at 37 °C for 30 min in the presence of 1 U of antartic phosphatase (New England BioLabs) in a final reaction volume of 10 μL containing 50 mM Bis-propane (pH 6.0), 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$ and RNase OUT (20 U, Invitrogen). The enzyme is then inactivated by incubating for 7 min at 65 °C. The dephosphorylated RNAs (10 pmol) are then 5′-end-radiolabeled using 3 U of T4 polynucleotide kinase (Promega) for 1 h at 37 °C in the presence of 3.2 pmol of [γ-^{32}P]ATP (6000 Ci/mmol; New England Nuclear). The reactions are stopped by the addition of 10 μL formamide dye buffer (95% formamide, 10 mM EDTA, 0.025% bromophenol blue and 0.025% xylene cyanol). Finally, the samples are purified by 10% polyacrylamide 8 M urea denaturing gel electrophoresis. The bands corresponding to the 5′-end-labeled RNAs are detected by autoradiography, and the gel slices containing those of the correct sizes are excised and recovered as described in the RNA synthesis (Section 2.3). The eluted and precipitated 5′-end-labeled transcripts are then dissolved in 30 μL ultrapure water, and the final radioactivity is calculated using a scintillation counter (Bioscan QC-2000).

2.5. In-line probing experiment

Prior to performing the in-line probing experiment, all 5′-end-labeled RNAs (both wt and G/A-mut PG4 versions) are heat-denatured and then allowed to slowly renature (Fig. 1). More specifically, trace amounts of 5′-end-labeled transcripts (50 000 cpm, <1 nM) are heated at 70 °C for 5 min, and are then slow-cooled to room temperature over 1 h in a buffer containing 20 mM lithium cacodylate pH 7.5 and 100 mM of LiCl, NaCl or KCl, depending on the conditions tested, in a final volume of 10 μL. Following the initial slow-cooling step, the volume of each sample is adjusted to 100 μL such that the final concentrations are 20 mM lithium cacodylate pH 8.5, 20 mM $MgCl_2$ and 100 mM of LiCl, NaCl or KCl. The reactions are then incubated for 40 h at room temperature, at which point the samples are ethanol-precipitated in presence of glycogen, ethanol-washed and dissolved in 10 μL ice-cold formamide loading buffer (95% formamide and 10 mM EDTA, 0.025% xylene cyanol).

Two ladders should be used for this kind of in-line probing experiment, an alkaline hydrolysis (permits the mapping of each nucleotide of the sequence) and an RNase T1 digestion of the transcripts (permits the mapping of the guanines). For the alkaline hydrolysis ladder, 50 000 cpm of the 5′-end-labeled wt transcripts (<1 nM) are dissolved in 5 μL of water, 1 μL of 1 N NaOH is added and the reaction is incubated for 1 min at room temperature prior to being quenched by the addition of 3 μL of 1 M Tris–HCl (pH 7.5). The RNA molecules are then ethanol-precipitated, and the RNA pellet dissolved in 10 μL formamide dye loading buffer (95% formamide, 10 mM EDTA and 0.025% xylene cyanol). For the RNase T1 ladder, 50 000 cpm of 5′-end-labeled wt transcript (<1 nM) are dissolved in 9 μL of buffer containing 20 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$ and 100 mM LiCl. The reaction mixture is incubated for 2 min at 37 °C in the presence of 0.6 U of RNase T1 (Roche Diagnostic). The reaction is then quenched by the addition of 20 μL of formamide loading buffer (95% formamide, 10 mM EDTA and 0.025% xylene cyanol). All of the samples and ladders are then transferred into new microcentrifuge tubes, and the radioactive content of the in-line probing samples and both ladders are then quantified using a scintillation counter (Bioscan QC-2000). Equal amounts, in terms of cpm, of all samples (Li^+, Na^+, K^+), and approximately two-thirds of this amount of the ladders, are then fractionated on 10% polyacrylamide 8 M urea denaturing gels. The resulting gels are subsequently dried and visualized by exposure to phosphorscreen (GE Healthcare) using a Typhoon Trio instrument (GE Healthcare)(see Fig. 3 for an example).

2.6. Data analysis

Several types of data can be extracted from in-line probing gels (Fig. 1). Initially, each gel is analyzed using the Semi-Automated Footprinting Analysis (SAFA) software [23]. The RNase T1 ladder lane is used as the "anchor" line, using the guanines as cleavage sites for the sequence references in SAFA. The intensity of each band in each condition is determined and is exported in a text format file. This file can be opened with the Excel program in order to produce an easily usable table.

First, the intensity of the bands under the Li^+ conditions are used to examine the secondary structure adopted under conditions unfavorable to G-quadruplex formation. The intensities are normalized with a method commonly used for SHAPE structure probing [24]. Briefly, the intensities of the bands having the next highest 10% intensities after the highest 2%, which corresponds to positions that are highly prone to cleavage, are averaged and each band's intensity is divided by this number, giving a ratio ranging between 0 and ~1. Low ratios correspond to constrained positions (i.e. mainly base-paired positions), while higher ratios indicate positions of greater flexibility such as single-stranded nucleotides. Normalization is performed using the data from 3

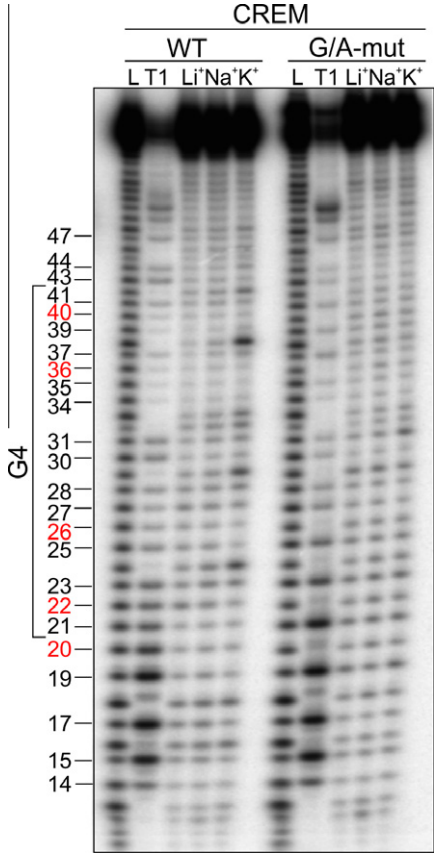

Fig. 3. In-line probing results. Autoradiogram of a 10% denaturing (8 M urea) polyacrylamide gel of the in-line probing of both the 5′-labelled CREM wt and the G/A-mutant PG4 versions performed in the presence of 100 mM of either LiCl, NaCl or KCl. The L and T1 lanes indicate the alkaline hydrolysis and ribonuclease T1 mapping lanes, respectively. The positions of the guanines are indicated at the left. The numbers in red represent guanines converted to adenines in the G/A-mutant version. The bracket at the left indicates the nucleotides involved in the formation of the G-quadruplex in the wt version.

independent experiments, and the results are presented as a color map on the best predicted secondary structures (see Fig. 4A and B for an example of the results obtained with the CREM candidate; blue indicates constrained nucleotides and red highly flexible ones).

Secondly, both the similarities and the discrepancies of the RNA structures under conditions both favorable and unfavorable for the formation of G-quadruplex structures are determined and examined. In order to do this, the raw intensity of each band from the lanes representing favorable G-quadruplex conditions (i.e. in the presence of either Na^+ or K^+) is divided by the intensity of the corresponding band from the Li^+ lane (i.e. the unfavorable condition). The in-line probing experiments are performed in triplicate (Materials and Methods point 2.5), and are then analyzed for each sequence (i.e. both the wt and the G/A-mut PG4s). The averages and standard deviations are calculated for the Na^+/Li^+ and K^+/Li^+ ratios for each nucleotide. These values are then used to generate bar graphs (with the intensity ratios of Na^+/Li^+ or K^+/Li^+ on the *y* axis and the sequence on the *x* axis) which permit an easier analysis of the data (see Fig. 5A for the wt and Fig. 5B for the G/A-mut sequences). Another way to represent the probing data is to show them directly on the predicted secondary structure. The same set of values can be used to create a color code in which the color of each nucleotide represents its cleavage susceptibility under conditions favorable for the formation of G-quadruplex structures relatively to that found under conditions that are unfavorable (K^+/Li^+) (Fig. 5C and D).

These various representations (see Figs. 4 and 5) facilitate the identification of the secondary structure that most likely fits the in-line probing data obtained under conditions either unfavorable or favorable for the formation of the G-quadruplex.

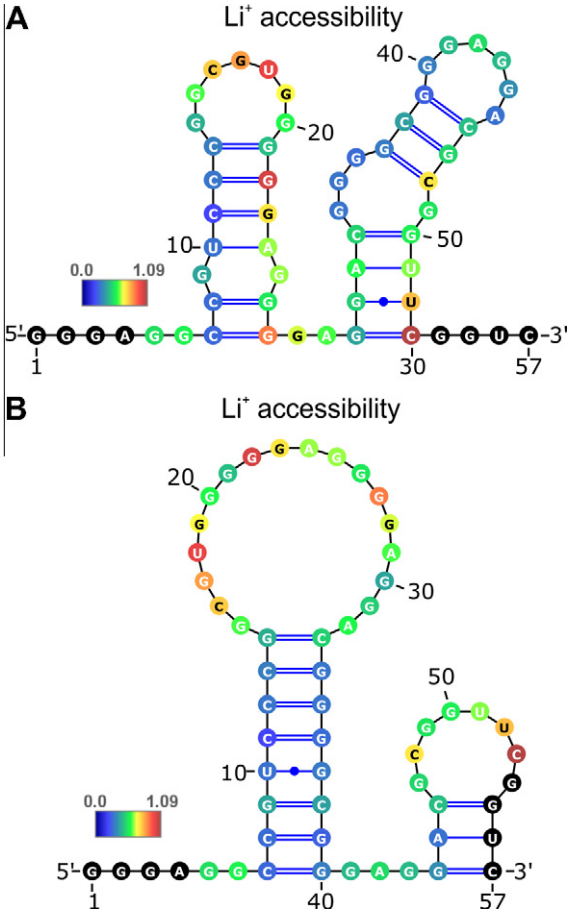

Fig. 4. Nucleotide accessibility in the presence of Li^+. (A) The first and (B) the second predicted secondary structures of the CREM wt. The color map illustrates each nucleotides' accessibility based on the normalization of the intensity of the band corresponding to each nucleotide, and is obtained by dividing by the average intensity of the 10% most intense bands. Ratios of ~0 (blue) show constrained regions, while ratios of ~1 (red) show flexible regions. The nucleotides in blacks are those for which no significant ratio could be calculated because their representative bands either migrated off of the gel, or were not sufficiently resolved. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results and discussion

3.1. Molecular design

In the last few years, RNA G-quadruplexes found in the 5′-UTRs of mRNAs acting as translational repressors have attracted a lot of attention (for a review see [11]). In this vein, a potential PG4 sequence found in the 5′-UTR of the human CREM mRNA was chosen with which to illustrate, step-by-step, an in-line probing protocol that analyzes the ability of this candidate to fold *in vitro* into a G-quadruplex structure. Multiple tools permit the prediction of G-quadruplex formation [21,25–27], and various databases of the PG4 sequences found in pre-mRNAs, mature mRNAs and both 5′- and 3′-UTRs are publically available [10,13,14]. The content of most of these databases was generated using the algorithm for predicting PG4 motifs mentioned in Section 2.1 (Materials and Methods). The CREM PG4 was chosen from a database built in our laboratory, and is located in the 5′-UTR of the human CREM transcript variant 19 mRNA (NM_183013) [10]. This 5′-UTR is 407 bp long and the PG4 sequence starts at position 110. The CREM PG4 is predicted to be composed of 23 nt (Fig. 2A), to possess several relatively short loops and to not contain any important cytosine stretches in its flanking sequences. These characteristics strongly increase its probability of folding into a G-quadruplex structure both *in vitro* and *in cellulo*. Indeed, these two criteria have been demonstrated to greatly influence not only the ability of an RNA sequence to fold into a G-quadruplex structure, but also its stability [10,28]. Clearly, the CREM PG4 represented an ideal candidate for this study. A previous study showed that it was more representative of the actual cellular context to probe, *in vitro*, a slightly extended version of the PG4 sequence in question in order to obtain more accurate data [10]. Consequently, the sequence probed included an additional 18 nt at the 5′-end and 16 nt at the 3′-end of the PG4 sequence and is referred to as *in vitro* PG4 (Fig 2A). The added nucleotides were identical to those found in the natural 5′-UTR. Finally, a mutant version in which several guanines were substituted for by adenines was also synthesized. These substitutions have the effect of disrupting the G-tracks and, consequently, abolishing the ability of the RNA to fold into a G-quadruplex structure (see the lower case "g" in Fig. 2A corresponding to the G/A-mutations). We suggest to use this handbook for the mutation of G-tracks in order to make sure to disrupt them adequately: GGG/GaG; GGGG/GaGG or GGGG/GGaG; GGGGG/GaGaG; GGGGGG/GGaGaG or GGGGGG/GaGGaG; and so on.

3.2. Secondary structure predictions

Previous characterizations of many PG4 sequences revealed that some do not in fact adopt a G-quadruplex structure because, instead, they fold into stable secondary structures that are formed by Watson–Crick base pairs ([10] and unpublished data). Because the latter structures are rapidly formed, this significantly impairs the folding into a G-quadruplex structure, a process that requires more time both *in vitro* and *in vivo*. As a result of these observations, the protocol was adapted so as to consider the predicted secondary structures based on both Watson–Crick base pairs and on

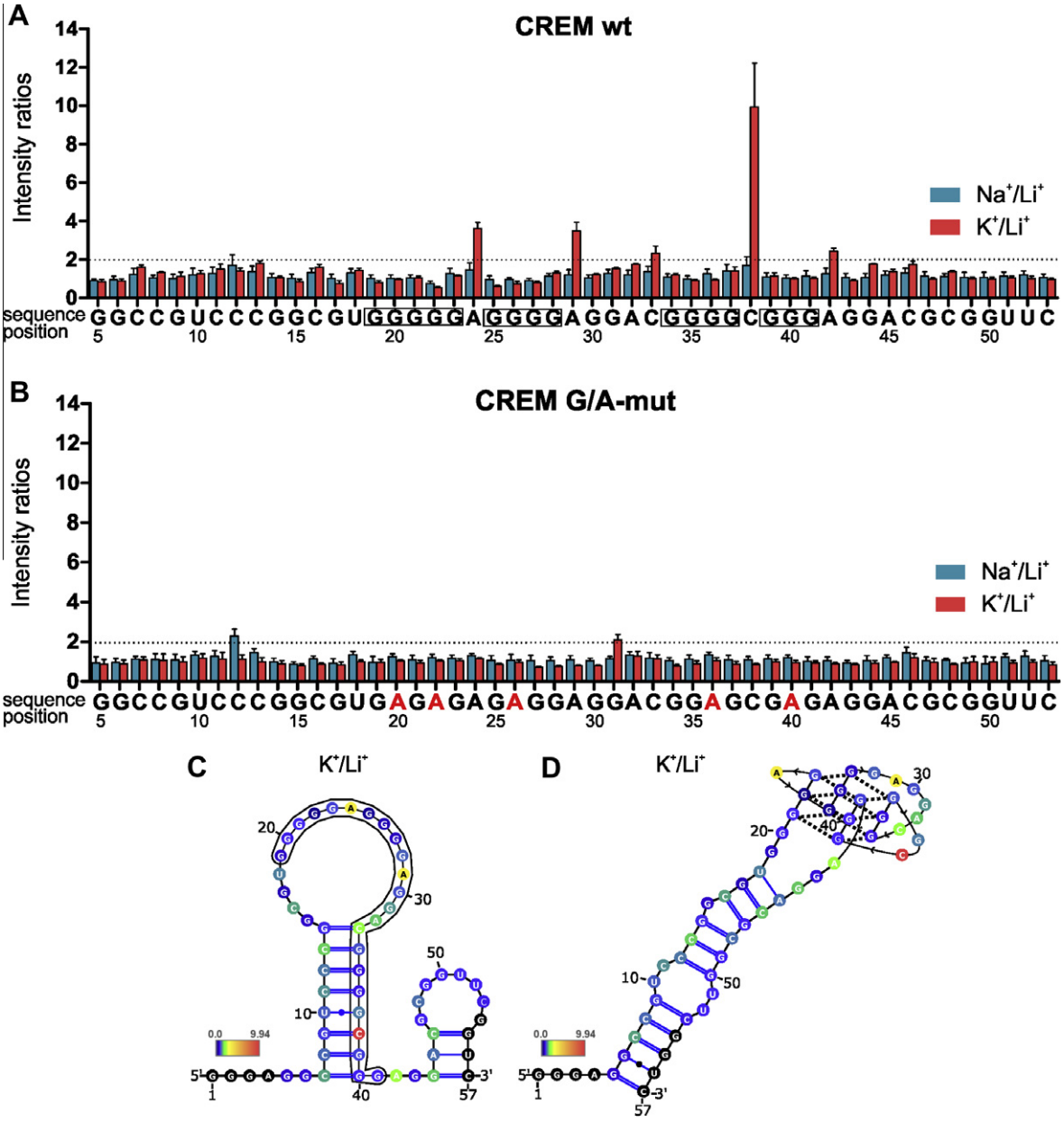

Fig. 5. Semi-quantitative analysis of the in-line probing experiments and interpretation of the secondary structures. (A–B) Ratios of the bands' intensity of the CREM wt (A) and G/A-mut (B) *in vitro* PG4 versions for each nucleotide. The Na^+/Li^+ ratios are shown in blue and K^+/Li^+ ones are in red. The dotted lines represent the 2-fold threshold that denotes a significant gain in flexibility. Both the sequence and the positions of the nucleotides are indicated on the *x*-axis. The boxed guanines represent the G-tracks involved in the G-quadruplex formation. The adenines shown in red are those replacing the guanines in the CREM G/A-mutant. Each bar represents the average of three independent experiments, and the error bars represent the standard deviations. (C–D) The K^+/Li^+ ratios of bands' intensity transposed as a color map on the best predicted secondary structure of the CREM wt PG4 either *in vitro* (C), or on the predicted structure containing the additional 5′ and 3′ regions flanking the PG4 with the putative CREM G-quadruplex folded at the top of the stem (D). The flexibilities of the nucleotides are proportional to their ratio of bands' intensity. Low ratios are shown in blue and high ones in red. The boxed nucleotides represent the predicted PG4 region. The nucleotides shown in black are those for which no significant ratio could be calculated because their representative bands either migrated off of the gel, or were not sufficiently resolved in the electrophoresis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the one including the G-quadruplex for each of the *in vitro* PG4 candidates studied. The secondary structures of the designed wt CREM *in vitro* PG4 version was predicted using the RNAstructure software [22]. Two potential structures were obtained (Fig. 2B and C). Briefly, the first one includes two hairpins, of 6 and 7 base pairs, that are linked by two single-stranded nucleotides, and harbours medium sized loops (Fig. 2B). The second is also composed of two hairpins, which are different from those of the previous structure, and are linked by three single-stranded nucleotides. Here, the first hairpin is composed of an 8 base pair stem that is capped by a large 18 nt loop, while the second is a small one composed of a 3 base pair stem and harbouring a 6 nt loop (Fig. 2C). One possible way to differentiate both of these predicted structures based on in-line probing should be the position of the single-stranded regions in both structures as they are distinct. Importantly, both of the predicted secondary structures showed a limited stability which was estimated to be between −18.8 and −17 kcal/mol (Fig. 2B and C, respectively).

When looking at the probability of a given sequence to form a G-quadruplex structure, an intrinsic parameter should be the number of nucleotides of the potential PG4 sequence that might be involved in Watson–Crick base pairs according to the RNA structure

prediction. In the case of the CREM PG4 sequence, 10 of the 23 nt appeared to be in single-stranded regions in the first predicted structure as compared to 15 nt in the second structure (Fig. 2B and C). Considering both the lack of highly stable predicted secondary structures, and the relative abundance of single-stranded nucleotides, the CREM PG4 appeared to be a suitable candidate to fold into a G-quadruplex.

RNAstructure folding software cannot predict the presence of a G-quadruplex motif. Nonetheless, the folding of the nucleotides on either side of the PG4 was predicted by preventing the PG4 region from being involved in the folding. In order to do so, the predicted unimolecular parallel G-quadruplex was considered as being already folded and was removed from the equation (Fig. 2D). The sequences surrounding the PG4 (i.e. the 5′- and 3′-extensions) were then folded together, if possible as a helical region (Fig. 2D). For the CREM wt sequence, this permitted the formation of an additional stem of 5 base pairs.

3.3. RNA synthesis and in-line probing

Subsequent to the designing of the sequence and the analysis of their predicted computer-based structures, RNA transcripts have to be synthesized. Double-stranded DNA templates for both the wt and G/A-mut versions of the CREM candidate were synthesized by the filling of two partially complementary oligonucleotides (Fig. 1). Upon performing the experiment it was noticed that DMSO was generally essential for this step. It creates slightly denaturing conditions that impair stable secondary structure formation, and thus permit the polymerase to read through the entire sequence. DMSO is known to increase the PCR amplification efficiency of GC-rich sequences [29]. Once the DNA templates were ready, they were *in vitro* transcribed using purified T7 RNA polymerase (see Section 2.3). The resulting reaction mixtures were treated with DNase to remove the DNA template. Phenol chloroform extraction was then performed in order to remove the proteins, and, lastly the RNA transcripts were fractionated by denaturing (8 M urea) 10% polyacrylamide gel electrophoresis. The RNAs in the gel bands of the appropriate sizes were recovered, dephosphorylated and 5′-end labelled with P^{32} using standard procedures.

Prior to the in-line probing experiment, trace amounts of all RNA samples were denatured at 70 °C for 5 min, followed by slow-cooling to room temperature in the presence of 100 mM of monovalent cation (i.e. either Li^+, Na^+ or K^+). In principle, this step should favor the prefolding of G-quadruplexes or other RNA structures. After the addition of the in-line probing buffer, all RNA samples were subjected to in-line probing reactions at room temperature for 40 h. The length of the incubation should be sufficient for the G-quadruplex structure to be formed and to reach equilibrium. Hydrolysis of the phosphodiester bonds was observed to occur in the most flexible regions.

It is important to note that only a trace amount of RNA (50000 cpm,<1 nM) is characteristically used in the in-line experiment. Therefore, most likely only intramolecular G-quadruplex formation is possible. This is an important difference as compared to other biophysical methods that are commonly used to study G-quadruplex formation, as methods such as circular dichroism and thermal denaturation require RNA concentrations in the low micromolar range which permit the formation of intermolecular G-quadruplexes. In our opinion, limiting the analysis to solely unimolecular topologies by using trace amounts of RNA is more biologically relevant, and is therefore essential in order to be able to properly evaluate both the potential of G-quadruplex formation and the role of these structures *in cellulo*. Although, even if it has never been observed in our hand, the relatively high concentration of magnesium ions (20 mM) could potentially affect the RNA structure equilibrium (e.g. between G-quadruplexes and alternative secondary structures) represents a limit of the technique. To get over this possible limitation, we suggest to confirm in-line probing results with a short experiment using one of the complementary biophysical methods mentioned above in condition corresponding to physiological concentration of magnesium ions (~1 mM).

After the incubation period, equivalent amounts of radioactivity (cpm) from each reaction were analyzed on a denaturing polyacrylamide gel. The bands were visualized by exposure of the dried gel to a phosphorscreen. A typical autoradiogram for both the wt and G/A-mut versions of the CREM candidate probed in the presence of either Li^+, Na^+ or K^+ is shown in Fig. 3. A change in the banding patterns was observed solely for the wt sequence. More precisely, specific nucleotides appeared to become more susceptible to hydrolysis in the presence of K^+. It is noteworthy that the Li^+ cation is an excellent negative control in the study the formation of G-quadruplexes as it maintains the same ionic strength in solution, but is unable to stabilize the stacking of the G-quartets due, primarily, to its smaller size. In other words, it favors the formation of the Watson–Crick base pair based secondary structure. Conversely, the presence of K^+ may stabilize both the G-quartet motifs and their stacking, which, therefore, favours the formation of a G-quadruplex structure. The bands showing an increased intensity in the presence of K^+ correspond to those nucleotides located within the predicted loops that are intercalated between the guanosine tracks, as well as those located immediately 3′ of the PG4 sequence (Fig. 3; nucleotides A_{24}, A_{29}, C_{33}, C_{38} and A_{42}). All of these regions were predicted to be single-stranded and therefore are probably more flexible upon formation of the G-quadruplex, thereby supporting the folding into this structure. Contrastingly, the susceptibility to hydrolysis of the corresponding nucleotides in the G/A-mut version remained unchanged when probed in the presence of K^+ instead of Li^+. Finally, the same probing pattern was observed in the presence of Li^+ and Na+, suggesting that no G-quadruplex structure was formed by this sequence in presence of Na^+.

3.4. Semi-quantitative analysis of the in-line probing

In order to achieve a more robust analysis, and to provide a quantitative aspect to the probing, triplicate experiments of in-line probing reactions were performed for each RNA sequence. The resulting band intensities were then quantified for each band using the SAFA Software [23]. The K^+/Li^+ intensity ratio was calculated for each position for both the wt and the G/A-mutant versions. The average and standard deviation (SD) of these ratios for each position and sequence were used to build bar graphs. Examples for the CREM candidate are illustrated in Fig. 5A and B. In order to determine if a specific nucleotide was truly more accessible in the presence of K^+ as compared Li^+, the K^+/Li^+ ratio was compared to that of the G/A-mutant. The reproducibility of the results is illustrated by the analysis of the G/A-mut sequence, which should exhibit no structural difference between these two ionic conditions, and thus permits the establishment of a threshold. In fact, no significant variation was observed in the bands' intensities between the three ionic conditions for the G/A-mutant version (Fig. 5B). The ratios over this threshold value should represent those nucleotides that are significantly more flexible. The study of more then twenty G-quadruplexes ([10] and unpublished data) indicated that a threshold of 2-fold was an accurate indication of a nucleotide that shows a significantly higher flexibility. For the CREM wt sequence, five nucleotides showed K^+/Li^+ ratio over 2 (A_{24}, 3.61; A_{29}, 3.48; C_{33}, 2.32; C_{38}, 9.94; and A_{42}, 2.43) (Fig. 5A). Four of these are located in the predicted loops of the folded G-quadruplex. More specifically, these nucleotides are situated immediately either 5′ or 3′ of the G-tracks. The last of the fare is located at the 3′ end of the last G-track (i.e. in position 42). According to our other probings of

RNA G-quadruplexes, this is typical. Depending on the particular G-quadruplex studied, the sequences on both the 5′ and 3′ sides of the PG4 region can also be affected by the G-quadruplex's formation ([10] and unpublished data). Moreover, it was observed that pyrimidine residues (i.e. C and U) are more susceptible to exhibiting significant hydrolysis in the G-quadruplex structure, in good agreement with a previous demonstration that pyrimidines are more prone to non-enzymatic spontaneous hydrolysis than are purines [30]. This might explain why some nucleotides in the loop, -GGA- in loop 2 of CREM for instance, showed superior cleavage levels (K^+/Li^+ ratio of 1.21, 1.52 and 1.74 respectively), but remain under the fixed 2-fold threshold (Fig. 5A). In summary, a clear modification in RNA structure driven by the presence of K^+ was observed. Moreover, the new structural features seemed to support the folding into a G-quadruplex structure. Finally, this procedure brings a semi- quantitative aspect to the analysis; however, it should always be considered with precaution and the appropriate controls must always be performed.

Several additional controls were required in order to validate the method, more specifically to verify that the quantification and the ratio calculations were accurate. Firstly, the amount of radioactivity of all of the samples was determined and equivalent amounts of cpm were loaded onto the gel for each of the samples. After migration and visualization by phosphorimaging, the total amounts of cpm in each of the lanes containing the in-line probing samples were quantified (using the ImageQuant software version 7.0; GE Healthcare Life Sciences) and compared (Supplementary Fig. 1). For each gel, the average radioactivity, in terms of cpm, for all of the lanes was calculated. If the standard deviation was too high (±15%), the results of a specific lane, or the complete gel, were rejected. This event in fact occurred very unfrequently. Secondly, equal amounts of cpm of a specific CREM wt PG4 sample were loaded into two distinct wells in order to assess any possible bias arising from either the loading step or the position of the samples on the gel. No significant variation was observed, in terms of cpm, between the intensities of the bands in the two lanes, nor in the banding patterns (Supplementary Fig. 2). Thirdly, since the K^+/Li^+ ratios used for building the bar graphs represent the averages of three distinct experiments, standard deviations (SD) were determined and are illustrated using error bars (see Fig. 5A and B). Clearly, the standard deviations were relatively small. Finally, this method was applied to several other candidates in order to ensure that it worked for candidates other than CREM. Specifically, more than twenty transcripts including potential G4 structures have been probed to date and in all cases conclusive data were obtained ([10] and unpublished data).

3.5. Comparing structure predictions and in vitro probing results

With the results of the in-line probing experiments in hand, it is of interest to take a closer look at the secondary structure adopted by the transcripts, starting with the one found in presence of Li^+ (i.e. under conditions unfavorable to G-quadruplex formation). In order to do so, the raw intensity values of the Li^+ conditions were specifically normalized with the help of a methodology used for the analysis of SHAPE results (see Section 2.6). The results of this normalization correspond to ratios that represent the levels of cleavage for each nucleotide under the same conditions (in the presence of Li^+ here). An initial color code can be created with these ratios, and the values can be superposed on the two initial predicted secondary structures (Fig 2B and C). Blue indicates constrained (base-paired) nucleotides, and colors from green to yellow to red indicate regions of increasing flexibility and accessibility, that is to say residues that are most likely single-stranded (or are less stable). Clearly, the best fit was with the second predicted structure that is composed of two stem loops with loops of 18 and 8 nt (compared Fig. 4B to A). Specifically, the in-line probing results showed that the most accessible nucleotides were found to be located in the hairpin loops, and the long stem was confirmed to contain constrained nucleotides (Fig. 4B). Thus, the experimental data support the second predicted structure for the CREM wt sequence under conditions unfavorable for G-quadruplex formation (i.e. in the presence of Li^+). A second color code can be produced using the averaged K^+/Li^+ ratios presented above (Section 3.4). With this new code, blue represents a ratio near 1, and colors from yellow to red show increasing ratios up to 9.94, the maximum ratio observed for the CREM PG4. The results of this second color code were transposed onto the secondary structure suggested to be adopted in the presence of Li^+, as well as on that obtained with the predicted unimolecular parallel G-quadruplex structure (Fig. 5C and D). With these representations, it appeared obvious that the differences in accessibility in the presence of the Li^+ versus in the presence of K^+ preferentially occurred for the nucleotides located in the loops and in those located 3′ of the PG4 region. Clearly, the K^+/Li^+ values have a better fit on the predicted structure that includes the G-quadruplex (Fig. 5D), as several discrepancies are observed when the structure including solely Watson–Crick base pairs is considered (Fig. 5C). In summary, the in-line probing of the CREM wt transcript unambiguously demonstrated the transition from a secondary structure composed of two stem loops to an unimolecular G-quadruplex structure is due to the presence of KCl.

4. Concluding remarks

The in-line probing method appears to be a simple, robust, reproducible and informative one with which to study RNA G-quadruplex formation. More importantly, compared to circular dichroism, thermal denature and NMR techniques, a much lower concentration of RNA is required for in-line probing (i.e. <1 nM). Specifically, only trace amounts of RNA are necessary, which permits avoiding the potential formation of intermolecular G-quadruplexes. Another advantage is that it is relatively quick to perform, as only a few days are required for both the probing and the analysis of both the wt and the mutated versions.

As presented, in-line probing permits the confirmation of whether or not a given PG4 sequence folds into a G-quadruplex structure. The corresponding G/A-mutant version does not permit this folding and is in fact important when further *in cellulo* investigations of the G-quadruplex need to be performed. Moreover, in-line probing offers the advantage of providing information on the structural modifications of the whole molecule following the G-quadruplex's formation. It is also possible to gain structural information for the nucleotides located on both sides of the G-quadruplex motif. So far, we have successfully used this technique to probe G-quadruplex sequences found in RNA molecules over 120 nt long (unpublished data). The possibility to probe relatively long molecule may be instructive in several situations, for example if the formation of a G-quadruplex is used to expose an adjacent regulatory region that was previously trapped in a hairpin, or the opposite situation, in which it is used to hide a region that was previously accessible.

In brief, the detailed methodology described here combines the use of bioinformatic algorithms to identify potential G-quadruplex sequences, a program for secondary structure prediction, in-line probing and its semi-quantification analysis and the representation of the resulting structure. Together, this represents a complete and accurate method with which to study RNA G-quadruplex formation. The results obtained are easy to interpret and provide a concrete and understandable visualization of the various structures adopted in different conditions.

Please cite this article in press as: J.-D. Beaudoin et al., Methods (2013), http://dx.doi.org/10.1016/j.ymeth.2013.02.017

Acknowledgements

This work was supported by a grant from the Canadian Institute of Health Research (CIHR, grant number MOP-44022) to J.P. Perreault. The RNA group is supported by a grant from the Université de Sherbrooke. J.D.B. was the recipient of the CIHR Frederick Banting and Charles Best Canada Graduate Scholarship Doctoral Award. R.J. was the recipient of the CIHR Frederick Banting and Charles Best Canada Graduate Scholarship Master's Award. J.P.P. held the Canada Research Chair in Genomics and Catalytic RNA and currently holds the Chaire de Recherche de l'Université de Sherbrooke en Structure et Génomique de l'ARN. He is also a member of the Centre de Recherche Clinique Étienne-Le Bel.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ymeth.2013.02.017.

References

[1] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Nucleic Acids Res. 34 (2006) 5402–5415.
[2] S. Millevoi, H. Moine, S. Vagner, Wiley Interdiscip Rev. RNA 3 (2012) 495–507.
[3] M. GELLERT, M.N. LIPSETT, D.R. DAVIES, Proc. Natl. Acad. Sci. USA 48 (1962) 2013–2018.
[4] J.L. Huppert, Chem. Soc. Rev. 37 (2008) 1375–1384.
[5] H.J. Lipps, D. Rhodes, Trends Cell Biol. 19 (2009) 414–422.
[6] G.W. Collie, G.N. Parkinson, Chem. Soc. Rev. (2011) 5867–5892.
[7] E. Birney, J.A. Stamatoyannopoulos, A. Dutta, R. Guigó, T.R. Gingeras, E.H. Margulies, et al., Nature 447 (2007) 799–816.
[8] V. Marcel, P.L.T. Tran, C. Sagne, G. Martel-Planche, L. Vaslin, M.-P. Teulade-Fichou, et al., Carcinogenesis 32 (2011) 271–278.
[9] A. Decorsière, A. Cayrel, S. Vagner, S. Millevoi, Genes Dev. 25 (2011) 220–225.
[10] J.-D. Beaudoin, J.-P. Perreault, Nucleic Acids Res. 38 (2010) 7022–7036.
[11] A. Bugaut, S. Balasubramanian, Nucleic Acids Res. (2012) 4727–4741.
[12] M. Subramanian, F. Rage, R. Tabet, E. Flatter, J.-L. Mandel, H. Moine, EMBO Rep. 12 (2011) 697–704.
[13] J.L. Huppert, A. Bugaut, S. Kumari, S. Balasubramanian, Nucleic Acids Res. 36 (2008) 6260–6268.
[14] O. Kikin, Z. Zappala, L. D'Antonio, P.S. Bagga, Nucleic Acids Res. 36 (2008) D141–D148.
[15] S. Neidle, G.N. Parkinson, Biochimie 90 (2008) 1184–1196.
[16] A. Randazzo, G.P. Spada, M.W.D. Silva, Top. Curr. Chem. (2012), http://dx.doi.org/10.1007/128_2012_331.
[17] J.-L. Mergny, L. Lacroix, Curr. Protoc. Nucleic Acid Chem. (2009). Chapter 17 (2009), Unit 17.1..
[18] E.E. Regulski, R.R. Breaker, Methods Mol. Biol. 419 (2008) 53–67.
[19] G.A. Soukup, R.R. Breaker, RNA 5 (1999) 1308–1325.
[20] M. Lamas, L. Monaco, E. Zazopoulos, E. Lalli, K. Tamai, L. Penna, et al., Philos. Trans. R. Soc. Lond. B Biol. Sci. 351 (1996) 561–567.
[21] H.M. Wong, O. Stegle, S. Rodgers, J.L. Huppert, J. Nucleic Acids 2010 (2010), http://dx.doi.org/10.4061/2010/564946.
[22] J.S. Reuter, D.H. Mathews, BMC Bioinformatics 11 (2010) 129.
[23] A. Laederach, R. Das, Q. Vicens, S.M. Pearlman, M. Brenowitz, D. Herschlag, et al., Nat. Protoc. 3 (2008) 1395–1401.
[24] J.T. Low, K.M. Weeks, Methods 52 (2010) 150–158.
[25] V. Scaria, M. Hariharan, A. Arora, S. Maiti, Nucleic Acids Res. 34 (2006) W683–W685.
[26] O. Kikin, L. D'Antonio, P.S. Bagga, Nucleic Acids Res. 34 (2006) W676–W682.
[27] R.Z. Cer, K.H. Bruce, D.E. Donohue, N.A. Temiz, U.S. Mudunuri, M. Yi, et al., Curr. Protoc. Hum. Genet. (2012). Chapter 18 (2012), Unit18.7..
[28] A.Y.Q. Zhang, A. Bugaut, S. Balasubramanian, Biochemistry (2011) 7251–7258.
[29] K. Varadaraj, D.M. Skinner, Gene 140 (1994) 1–5.
[30] Y. Li, R.R. Breaker, J. Am. Chem. Soc. 121 (1999) 5364–5372.