Identification of G-quadruplexes in long functional RNAs using 7-deaza-Guanine RNA

Carika Weldon¹, Isabelle Behm-Ansmant², Laurence H. Hurley^{3,4}, Glenn A. Burley⁵,
Christiane Branlant², Ian C. Eperon^{1*}, Cyril Dominguez^{1*}

¹ Leicester Institute of Structural and Chemical Biology and Department of Molecular and Cell Biology, University of Leicester, Leicester, UK

² IMoPA (Ingénierie Moléculaire et Physiopathologie Articulaire), UMR 7365 CNRS-UL, Biopôle de l'Université de Lorraine, 9 Avenue de la Forêt de Haye, 54505 Vandoeuvre-lès-Nancy, France

³ College of Pharmacy and College of Pharmacy and BIO5 Institute, University of Arizona, Tucson, Arizona 85721, United States

⁴ Arizona Cancer Center, University of Arizona, Tucson, Arizona 85724, United States

⁵ Department of Pure and Applied Chemistry, University of Strathclyde, UK

* Corresponding authors: eci@le.ac.uk, cd180@le.ac.uk

RNA G-quadruplex (G4) structures are thought to affect biological processes, including translation and pre-mRNA splicing, but it is not possible at present to demonstrate that they form naturally at specific sequences within long functional RNA molecules. We have developed a novel strategy, footprinting of long 7-deazaguanine substituted RNAs (FOLDeR), that allows the formation of G4s to be confirmed in long RNAs and in functional conditions.

G-rich sequences in DNA or RNA have the potential to form four-stranded structures known as G-quadruplexes (G4)¹. DNA G4 formation has been demonstrated in vivo² and implicated in a very wide range of gene expression processes, notably telomere maintenance, transcription and genome instability³. RNA G4s exist in vivo⁴ and have been implicated in various post-transcriptional processes⁵⁻¹⁰. They would be expected to form more readily in RNA than in DNA, but it is still difficult to predict whether they do form in any given RNA sequence.

G4s have been suggested previously to be of functional significance in pre-mRNA splicing. Most of these studies used similar approaches, using bioinformatic tools to identify well-characterized G4-forming sequences and biophysical methods to show that the predicted G4 can form in isolated short sections of RNA¹¹⁻¹³. However, such sequences are naturally part of longer pre-mRNAs in which the propensity of a sequence to form a G4 is determined by competition with secondary structures and protein binding. Other methods for identifying G4s include the use of ligands or antibodies that bind selectively to G4s^{4,13,14}, but a general hazard is that the binding reagent affects the equilibrium between different structures and thus might induce the formation of otherwise unstable non-functional G4s. Mutagenesis is often used for confirmation of G4 formation in vivo^{12,15}, but mutants need to be designed so as not to affect the binding of regulatory proteins or perturb likely secondary structures.

We describe here a method in which we probed the secondary structure of an entire functional unit of pre-mRNA, comparing native RNA with 7-deazaguanine (7-deaza-G) substituted RNA, which can form the same secondary structures but not G4s¹⁶. Differences highlighted regions in which G4 might form. These differences were

confirmed in functional splicing conditions by ribonuclease H digestion in nuclear extracts. This method allowed the identification of functionally relevant G4s.

The splicing reaction we have studied is that of the human B-cell lymphoma-extra large (Bcl-x) pre-mRNA, which has two alternative 5' splice sites (5'SS) in its exon 2. The major isoform, Bcl-x_L (X_L), is an anti-apoptotic factor, while the alternative isoform, Bcl-x_S (X_S), is pro-apoptotic¹⁷. There are a number of guanine-rich (G-rich) sequences in the regions around the Bcl-x alternative splice sites that could form G4s.

We designed a functional splicing transcript, Bcl-x-681, by preserving the sequence between the alternative X_S and X_L 5'SS, shortening the intron and the exon 3 and adding an additional 5'SS at the 3' end of the construct (Fig. 1a). Splicing of this transcript substantially favored the Bcl-x_L isoform, recapitulating the preference for Bcl-x_L observed in HeLa cells¹⁸ (Fig. 1b, Supplementary Fig. 1).

The presence of G4s in this functional RNA was assessed by an electrophoretic mobility shift assay (EMSA) using a previously described G4-specific antibody, BG4^{2,4}, that induced a shift in the migration of the Bcl-x-681 RNA (Fig. 1c, left). This transcript contains 6 G-rich sequences predicted to potentially form a G4¹⁹, one upstream of the X_S splice site (Q1), three between the X_S and the X_L splice sites (Q2, Q3 and Q4), and one downstream of the 3' splice site (3'SS) (Q6). All of these short sequences displayed a circular dichroism spectrum that is typical for parallel G4s, with a positive signal at 265 nm and a negative signal at 240 nm (Supplementary Fig. 2). The stability of these G4s ranged from 37°C (Q4) to >70°C (Q2). However, these

experiments did not provide information on whether these sequences do form G4 in the face of competing secondary structures in functional Bcl-x pre-mRNA.

Therefore, our strategy for identifying the location of G4 elements in this RNA was to (i) map structured regions using ribonucleases (RNases), and then (ii) repeat the assay with 7-deaza-G RNA, in which G4s would not form¹⁶. For (i), we performed RNA footprinting using RNases T1, T2 and V1, providing structural restraints at nucleotide resolution (Supplementary Figs. 3 and 4). In total, we defined 163 and 91 nucleotides as single-stranded and double-stranded, respectively (Supplementary Fig. 5, Supplementary Dataset 1). We used these restraints with a secondary structure prediction software (Mfold²⁰) to derive the best fit model (Supplementary Fig. 6). This model predicted that the Bcl-x-681 RNA is highly structured and suggested the presence of four structurally independent domains, denoted as X_S, X_L, intron and 3'ss domains according to the splicing elements they contained (Supplementary Fig. 6). We transcribed each structural domain separately and RNA footprints of each domain showed a very similar pattern to that of the full-length RNA (Supplementary Fig. 7), supporting the structural model.

To identify potential G4s within this structure, we replaced all guanines by 7-deaza-G during *in vitro* transcription. The absence of the N7 in the 7-deaza-G was shown previously to abolish the possibility of Hoogsteen base pairing and thus of G4 formation, whilst still permitting Watson-Crick base-pairing and the formation of stem-loops¹⁶. We confirmed the absence of G4s in the 7-deaza-G Bcl-x-681 RNA by EMSA. As expected, the antibody BG4^{2,4} did not induce a shift in the migration of the 7-deaza-G RNA (Fig. 1c, right).

We performed RNA footprinting on the 7-deaza-G RNA (Supplementary Fig. 4) and a comparison between the RNA footprinting of the 7-deaza-G-substituted transcripts and the native transcript was assessed by calculating an index factor for each nucleotide corresponding to $(V1 \text{ 7-deaza-G} / V1 \text{ native}) / (T2 \text{ 7-deaza-G} / T2 \text{ native})$ (Supplementary Dataset 2). The logarithms of the index were distributed as a Gaussian with a mean of -0.0416 and a standard deviation of 0.3051 (Supplementary Fig. 8). We considered the nucleotides whose index value had changed by more than two standard deviations from the mean (corresponding in this case to a difference of approximately three-fold or more in structural probing between the native and the 7-deaza-G RNA) to be disproportionately affected by 7-deaza-G substitution. Most of these were located in the X_S and the X_L domains, and many of these nucleotides coincided with two of the possible G4 regions (Q2, and Q5) (Fig. 1d). We therefore concluded that the Bcl-x-681 transcript contains G4s located near the X_S and X_L splice sites. Previous reports identified G4s by comparing structural probing of RNAs in the presence of either potassium (K^+) or Lithium (Li^+) ions⁸, although on much shorter sequences than Bcl-x-681. For comparison, we performed an RNA footprinting experiment of Bcl-x-681 in the presence of Li^+ or K^+ ions (Supplementary Dataset 3) and calculated an index as $(V1 K^+ / V1 Li^+) / (T2 K^+ / T2 Li^+)$. By analysing the distribution of the logarithms of these scores as above, we found a mean of -0.02 and a standard deviation of 0.17. Only 19 nucleotides showed a change beyond two standard deviations, and these were not concentrated around potential G4s (Supplementary Fig. 9). These results showed that the use of 7-deaza-G substitution is more sensitive than the comparison of structural probing in K^+ and Li^+ containing buffers.

A major benefit of using 7-deaza-G RNA in structural probing is that, in contrast to the comparison between Li^+ and K^+ buffers, it can be done in functional conditions, such as in nuclear extracts. Therefore, we tested the accessibility of native and 7-deaza-G modified Bcl-x-681 RNAs using RNase H cleavage directed by 10-mer DNA oligonucleotides in functional splicing conditions in nuclear extracts (Supplementary Fig. 10). We focused our analysis on the X_S and X_L domains. We assigned protected or accessible regions on the basis of either $< 40\%$ or $> 60\%$ cleavage, respectively (Supplementary Fig. 10). The RNase H cleavage patterns of native RNA in nuclear extract were in good agreement with the structure predicted in solution (Fig. 2a). Putative stems were protected in both the X_S and X_L domains, while several of the predicted loops were accessible. To compare the RNase H patterns of the native and the 7-deaza-G RNA, each transcript was analysed three times and a Student's t test was done for each oligonucleotide to identify whether there were significant differences between the native and deaza RNAs. The result (Fig. 2b) showed that most of the significant differences were found near the putative G4s, especially Q2 and Q5, consistent with the footprinting results. These data demonstrated that the use of 7-deaza-G RNA can identify G4s in functional splicing conditions.

In conclusion, we have described FOLDeR, a new method that allowed us to demonstrate the presence of G4s within long RNAs by using 7-deaza-G RNA modifications in combination with secondary structure probing (Fig. 1). By coupling footprinting with ribonuclease H cleavage assays on the two different types of transcript, it was also possible to validate the findings in functional conditions, such as nuclear extracts (Fig. 2). Although we demonstrated the power of this method

using nuclease digestions, its validity relies on the modification of the RNA and not on the structural probing method. Therefore, the use of 7-deaza-G RNAs to probe G4s could also be done in combination with other structural probing methods such as selective 2'-hydroxyl acylation analysed by primer extension (SHAPE) chemistry²¹ or dimethyl sulphate (DMS) footprinting²². It is now becoming clear that RNA G4s have a higher propensity to form than their DNA counterparts. Recently, structural data revealed the presence of RNA G4s that could not be predicted by bioinformatics tools^{13,23,24,25}. The development of an experimental method for identifying and locating G4s is therefore a major step forward. The existence of G4s in a functional RNA can be tested using the BG4 antibody⁴ and their location established using 7-deaza-G-RNA as described here.

Acknowledgments

This work was supported by a Medical Research Council Career Development Award (G1000526) to C.D. and a Sir Dudley Spurling Post Graduate Scholarship from the Bank of Butterfield Foundation in Bermuda to C.W. This work was also funded by CNRS and Lorraine University (UMR 7365 and previously 7214) and the European Alternative Splicing Network of Excellence (EURASNET, FP6 life sciences, genomics and biotechnology for health; LSHG-CT-2005-518238).

Author Contributions

C.W. performed all experiments under the guidance of I.C.E. and C.D. Footprinting experiments and structural model calculation were performed under the guidance of I.B.A. and C.B. G.A.B. and L.H.H. contributed to the development and the validation

of the strategy. C.W., I.C.E., and C.D. interpreted the results and wrote the manuscript.

The authors declare no competing financial interests

References

1. Huppert, J. L. *FEBS J* **277**, 3452–3458 (2010).
2. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. *Nat Chem* **5**, 182–186 (2013).
3. Rhodes, D. & Lipps, H. J. *Nucleic Acids Res* **43**, 8627–8637 (2015).
4. Biffi, G., Di Antonio, M., Tannahill, D. & Balasubramanian, S. *Nat Chem* **6**, 75–80 (2014).
5. Agarwala, P., Pandey, S. & Maiti, S. *Org. Biomol. Chem.* **13**, 5570–5585 (2015).
6. Millevoi, S., Moine, H. & Vagner, S. *WIREs RNA* **3**, 495–507 (2012).
7. Kumari, S., Bugaut, A., Huppert, J. L. & Balasubramanian, S. *Nat Chem Biol* **3**, 218–221 (2007).
8. Darnell, J. C. *et al. Cell* **107**, 489–499 (2001).
9. Decorsiere, A., Cayrel, A., Vagner, S. & Millevoi, S. *Genes Dev.* **25**, 220–225 (2011).
10. Wolfe, A. L. *et al. Nature* **513**, 65–70 (2014).
11. Marcel, V. *et al. Carcinogenesis* **32**, 271–278 (2011).
12. Ribeiro, M. M. *et al. Hum Genet* **134**, 37–44 (2015).
13. Smith, L. D. *et al. Cell Rep* **9**, 193–205 (2014).
14. Zizza, P. *et al. Nucleic Acids Res* **44**, 1579–1590 (2015).
15. Munroe, S. H., Morales, C. H., Duyck, T. H. & Waters, P. D. *PLoS ONE* **10**, e0137893 (2015).
16. Murchie, A. I. & Lilley, D. M. *Nucleic Acids Res* **20**, 49–53 (1992).
17. Boise, L. H. *et al. Cell* **74**, 597–608 (1993).
18. Garneau, D., Revil, T., Fisette, J.-F. & Chabot, B. *J Biol Chem* **280**, 22641–22650 (2005).

19. Kikin, O., D'Antonio, L. & Bagga, P. S. *Nucleic Acids Res* **34**, W676–W682 (2006).
20. Zuker, M. *Nucleic Acids Res* **31**, 3406–3415 (2003).
21. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).
22. Kwok, C. K., Ding, Y., Tang, Y., Assmann, S. M. & Bevilacqua, P. C. *Nature Comms* **4**, 1–12 (2013).
23. Phan, A. T. *et al. Nat Struct Mol Biol* **18**, 796–804 (2011).
24. Huang, H. *et al. Nat Chem Biol* **10**, 686–691 (2014).
25. Warner, K. D. *et al. Nat Struct Mol Biol* **21**, 658–663 (2014).

Figure Legends:

Figure 1: The functional Bcl-x-681 RNA contains G4s *in vitro*

(a) Schematic representation of native Bcl-x (top) and Bcl-x-681 transcript (bottom) used in this study. The size of introns and exons are indicated in nucleotides below the diagrams. X_L and X_S 5' splice sites are indicated above the diagrams. (b) In vitro splicing assay of Bcl-x-681. Bands corresponding to the unspliced transcript, the X_L and the X_S products are labelled. The full-length gel is displayed in Supplementary Fig. 11 (c) Electrophoretic-mobility shift assay of native (left) and 7-deaza-G substituted (right) Bcl-x-681 in the absence or presence of increasing amount of BG4 antibody (0, 150 pM, 1.5 nM, 15 nM, 150 nM and 1.5 μ M). The full-length gel is displayed in Supplementary Fig. 11. (d) Mapping of RNA footprinting of Bcl-x-681. Nucleotides having a different footprinting pattern upon 7-deaza-GTP substitution are circled and are located mainly to the X_S and X_L domains. Putative G4s are labelled Q1 to Q6.

Figure 2: The Bcl-x-681 RNA contains G4s *in functional conditions*

(a) Schematic drawing of the RNase H cleavage of native Bcl-x-681 on the secondary structure model of Bcl-x-681. Accessible regions (average >60% cleavage) are in black, and protected regions (average <40% cleavage) are in grey. RNase cleavage experiments were done in triplicate and are presented in supplementary Figure 9A. (b) Significance of changes in RNase H cleavage between 7-deaza-G and native RNA in nuclear extract. Three experiments were done on each transcript and, for each oligonucleotide, a Student's t test was done to establish the probability that the two sets of results might come from the same population. Values of *p* below 0.05

indicate that the cleavage of the two transcripts was significantly affected by deaza substitution. Representative experiments on the two transcripts are shown in Supplementary Figure 10.

Online Methods:

***In Vitro* Transcription of RNA**

A transcription reaction containing 40 mM Tris (pH 7.5), 20 mM MgCl₂, 10 mM NaCl, 2 mM spermidine HCl, 10 mM DTT, 4 mM NTPs, 1 µg DNA template (PCR product), 5% RNaseOut (Invitrogen), and 5% T7 polymerase (1:20 dilution) was incubated at 37 °C for 4 h. DNA fragments were removed using an S-300 column (GE Healthcare). The product was extracted by phenol-chloroform and precipitated by ethanol. The pellet was dissolved in the desired amount of TE buffer and stored at –20 °C. For radiolabelled RNA, A transcription reaction containing 40 mM Tris (pH 7.5), 6 mM MgCl₂, 10 mM NaCl, 2 mM spermidine HCl, 0.5 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 0.05 mM GTP, 50 ng DNA template (PCR product), 4% RNaseOut, 10% T7 (1:20 dilution), 1 mM Ribo m⁷G Cap analogue (Promega) and 0.33 µM [α -³²P] GTP (10 mCi/ml, 3000 Ci/mmol) (Pelkin Elmer) was incubated at 37 °C for 1.5 h. Samples were then purified via a denaturing gel.

***In vitro* transcription of 7-deaza RNA**

A transcription reaction containing 40 mM Tris (pH 7.5), 20 mM MgCl₂, 10 mM NaCl, 2 mM spermidine HCl, 10 mM DTT, 4 mM NTPs (c⁷GTP instead of GTP), 1 µg DNA template (PCR product), 5% RNaseOut, 2 mM Ribo m⁷G Cap analogue (Promega) and 5% T7 (1:20 dilution) was incubated at 37 °C for 4 h. DNA fragments were removed using an S-300 column (GE Healthcare). The product was extracted by phenol-chloroform and precipitated by ethanol. The pellet was dissolved in the desired amount of TE buffer and stored at –20 °C. For radiolabelled deaza RNA, A transcription reaction containing 40 mM Tris (pH 7.5), 6 mM MgCl₂, 10 mM NaCl,

2 mM spermidine HCl, 0.5 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 0.05 mM GTP, 50 ng DNA template (PCR product), 4% RNaseOut, 10% T7 (1:20 dilution), 1 mM Ribo m7G Cap analogue (Promega) and 0.33 μ M [α -³²P] GTP (10 mCi/ml, 3000 Ci/mmol) (Pelkin Elmer) was incubated at 37 °C for 1.5 h. Samples were then purified via a denaturing gel.

***In Vitro* Splicing of RNA**

Samples containing 1.53 mM ATP, 20 mM CrPi, 3.2 mM MgCl₂, 20 mM Hepes (pH 7.5), 50 mM KGlu, 0.125% NP-40, 50% nuclear extract (Cilbiotech) and 10–20 cps of radiolabelled RNA transcript were set up in 10 μ l reactions on ice then incubated at 30 °C for 2 h in a pre-heated waterbath. Aliquots of 2 μ l were taken into a microtitre plate on dry ice at timepoints 0, 30, 60 and 120 min. Samples were then treated by addition of 50 μ l of 0.4 mg/ml proteinase K and incubation for 10 mins at 37 °C. Samples were precipitated twice with 100% ethanol, and dissolved in 10 μ l formamide dyes. To prepare for loading on an 8% denaturing polyacrylamide gel, samples were placed in boiling water for 30 sec and run until bromophenol blue has just run out of the gel. Gels were then fixed, dried and exposed to a phosphorimaging screen overnight. Quantification of pre-mRNA and both mRNA products was done using OptiQuant software, and intensities were normalised to account for the number of guanines in each RNA species.

RNase probing

Samples containing 400 ng RNA (either native or 7-deaza-substituted), 0.5 mg/ml yeast tRNA, and either 150 mM KCl, 1.5 mM MgCl₂ and 20 mM Hepes pH 7.6 (K⁺ containing buffer) or 150 mM KCl, 1.5 mM MgCl₂ and 20 mM Hepes pH 7.6 (Li⁺

containing buffer) in a total volume of 9 μ l, were heated at 65 °C in a water bath for 10 min then slow cooled to 37 °C. 1 μ l of either Tris buffer (0.5 mM Tris pH 7.6, 1 M KCl or 1M LiCl, and 25 mM MgCl₂), RNase T1 (Roche Diagnostics), RNase T2 (Ambion) or RNase V1 (Mobitech) was added. Reactions were incubated at ambient temperature for 6 min then stopped with 101.8 μ l of stop buffer (3.9 mM EDTA and 0.2 μ g yeast tRNA) and 101.8 μ l phenol-chloroform then spun at 13,500 rpm for 10 min. The top layer was removed and added to 700 μ l ethanol (96%), 12 ng glycogen and 22 mM NaAc (pH 3.0). Samples were then ethanol-precipitated and dissolved in 4 μ l potassium borate or sterile water.

Reverse transcription of probed RNA

A hybridization mix was made containing 0.25 μ l hybridization buffer (reverse transcription buffer without MgCl₂), 1 μ l of 100 cps/ μ L radiolabelled primer and 0.25 μ l water. 1 μ l of probed RNA was incubated with 1.5 μ l of hybridization mix at 65 °C for 10 min then cooled on ice. 2.5 μ l of extension mix (0.1 μ l 5 mM dNTPs, 0.25 μ l reverse transcriptase buffer, 0.25 μ l 1/10 diluted AMV reverse transcriptase (MP Biomedical) and 1.9 μ l water) was added to each probed RNA mix and incubated at 42 °C for 30 min. 3 μ l formamide dye was added and the sample was heated to 96 °C for 2 min, and then put on ice for 2 min. 2 μ l was loaded onto 7% DNA sequencing gel and run for 2 h at 100 W. Gels were dried and exposed to a phosphorimaging screen overnight and quantified using SAFA. The data were analysed using a cleavage index to determine the change in accessibility of each nucleotide in the RNA, and the mean and standard deviation of this index were used as described in the Results to identify nucleotides that showed a significantly greater change.

Footprinting of the native RNA was done twice and the structural model was confirmed by further footprinting experiments on the isolated structural domains.

RNase H assays

RNase H cleavage was done in splicing conditions by adding each of 44 DNA oligonucleotides at 10 μ M after incubation for 30 min and continuing incubation for 5 min. Reactions were processed as for splicing reactions. Three independent sets of experiments were done for each RNA. The extent of cleavage was measured using a phosphorimager a Student's t-test was performed to evaluate the significance of the changes between the 7-deaza-G and the native RNA triplicate experiments.

Electromobility Shift assays (EMSA)

The BG4 antibody was purchased from Absolute Antibody (Ab00174-1.1). Radiolabelled RNA at 10 count per seconds was incubated with various concentrations (0, 150 pM, 1.5 nM, 15 nM, 150 nM and 1.5 μ M) of BG4 antibody in a microtitre plate. 3 μ l of native gel dye was added, and 4 μ l was run on a small 5% native polyacrylamide gel at 4°C. The gel was run at 5 W for 1 h. Gels were dried and exposed to a phosphorimager screen overnight.

Circular Dichroism

RNA sequences corresponding to Q1, Q2, Q3, Q4, Q5, and Q6 were purchased from Dharmacon, GE Healthcare, deprotected according to the manufacturer instructions, lyophilised, and resuspended in 20mM KH_2PO_4 , pH7, 100mM KCl at a final concentration of 20 μ M. Circular dichroism spectra were acquired on a Chirascan spectrometer at 20°C and data were collected between 350 and 220nm in triplicate

and averaged. Buffer baseline was subtracted for each spectrum. CD Melting curves were obtained by measuring the CD signal at 265nm from 20 to 80°C at 1°C heating rate.

Figure 1

Figure 2

SUPPLEMENTARY INFORMATION

Identification of G-quadruplexes in functional RNAs using 7-deaza-Guanine RNA

Carika Weldon¹, Isabelle Behm-Ansmant², Laurence H. Hurley^{3,4}, Glenn A. Burley⁵, Christiane Branlant², Ian C. Eperon^{1*}, Cyril Dominguez^{1*}

¹ Leicester Institute of Structural and Chemical Biology and Department of Molecular and Cell Biology, University of Leicester, Leicester, UK

² IMoPA (Ingénierie Moléculaire et Physiopathologie Articulaire), UMR 7365 CNRS-UL, Biopôle de l'Université de Lorraine, 9 Avenue de la Forêt de Haye, 54505 Vandoeuvre-lès-Nancy, France

³ College of Pharmacy and College of Pharmacy and BIO5 Institute, University of Arizona, Tucson, Arizona 85721, United States

⁴ Arizona Cancer Center, University of Arizona, Tucson, Arizona 85724, United States

⁵ Department of Pure and Applied Chemistry, University of Strathclyde, UK

* Corresponding authors: eci@le.ac.uk, cd180@le.ac.uk

SUPPLEMENTARY RESULTS

Supplementary Figure 1: Time courses of splicing of Bcl-X 681 pre-mRNA in vitro and the effects of inhibition of U1 snRNA.

Supplementary Figure 2: Circular dichroism spectra and CD melting curves of the six putative G4 of Bcl-x-681

Supplementary Figure 3: Sequence of Bcl-x-681

Supplementary Figure 4: Footprinting experiments showing the effects of deazaG substitution on cleavage by RNase T1, T2 and V1

Supplementary Figure 5: Single-stranded and double-stranded nucleotides in Bcl-x-681 wild type transcript

Supplementary Figure 6: Secondary structure model of Bcl-x-681

Supplementary Figure 7: Comparison of RNA footprinting of Bclx-681 full-length (FL) and fragments (Frag) X_S, X_L, intron and 3'SS

Supplementary Figure 8: Distribution of the logarithm of the index from comparison of nuclease footprinting on native and 7-deaza-G RNA.

Supplementary Figure 9: Mapping of RNA footprinting of Bcl-x-681 in Li⁺ and K⁺ containing buffers

Supplementary Figure 10: RNase H cleavage and quantification of the X_S and X_L fragments of Bcl-x-681 WT (Top) and 7-deaza (bottom) in nuclear extract

Supplementary Figure 11: Electrophoretic-mobility shift assay of native (left) and 7-deaza-G substituted (right) Bcl-x-681 in the absence or presence of increasing amount of BG4 antibody

Supplementary Dataset 1: SAFA quantification of Bcl-x-681 native footprinting

Supplementary Dataset 2: Comparison of footprinting on Bcl-x-681 native and 7-deaza-G-substituted transcripts

Supplementary Dataset 3: Comparison of footprinting on Bcl-x-681 in either potassium or Lithium containing buffers

Supplementary Figure 1: Time courses of splicing of Bcl-X 681 pre-mRNA in vitro and the effects of inhibition of U1 snRNA. 32P-labelled pre-mRNA was added to aliquots of HeLa cell nuclear extract that had been pre-incubated with a 2'-O-methyl oligonucleotide complementary to U1 snRNA (anti-U1) at the concentrations shown, and the mixtures were incubated for 0, 30, 60 or 120 mins, as shown. Products were analysed by electrophoresis on a denaturing polyacrylamide gel. The shift towards use of the upstream (Bcl-Xs) 5' splice site has been seen in other examples with alternative 5' splice sites (Hudson et al., NAR, 40, 6850-62, 2014 ; Eperon et al, EMBO J., 12, 3607-17, 1993) and is predicted by models for selection by U1 snRNA (Eperon et al, Mol. Cell Biol., 20, 8303-18, 2000).

Q1: GAGGGAGGCAGGCGACGAGUUUGAACUGCGGUA (QGRS score: 6)

Q2: CCGGGAUAGGGGUAAACUGGGGUCGCAUUGUGGCC (QGRS score: 17)

Q3: UCGGCGGGGCACUGUGCGUGGAA (QGRS score: 11)

Q4: AAGGAGAUGCAGGUUAUUGGUGAGUCGGAU (QGRS score: 18)

Q5: UUGGAUCCAGGAGAACGGCGGCUGGGUA (QGRS score: 21)

Q6: CUGGUUCCUGACGGGCAUGACUGUGGCCAGGUA (QGRS score: 15)

Supplementary Figure 2: Circular dichroism spectra and CD melting curves of the six putative G4 of Bcl-x-681. The RNA sequence used is indicated above the spectra.

Bcl-X-681 RNA

Key

- Exon 2
- Exon 2.1
- Intron
- Exon 3
- hnRNP F/H binding site
- G tracts

Supplementary Figure 3: Sequence of Bcl-x-681. Primers used for footprinting analysis by reverse transcription are shown below the sequence.

SL XI

Q6

Supplementary Figure 4: Footprinting experiments showing the effects of deazaG substitution on cleavage by RNase T1, T2 and V1. The primers used are shown in Supplementary Figure 2. All putative stem-loops and G-quadruplexes are indicated. Each reaction was done as part of a set with a range of concentrations of a quadruplex-binding reagent (triangles above lanes), GQC-05. The presence of GQC-05 affected the pattern in the vicinity of the predicted G4s (manuscript in preparation).

Nucleotide	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
Sequence	G	G	G	A	G	G	T	G	A	T	C	C	C	C	A	T	G	G	C	A	G	C	A	G	T	A	
Nucleotide	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	
Sequence	A	A	G	C	A	A	G	C	G	C	T	G	A	G	G	G	A	G	G	C	A	G	G	C	G	A	
Nucleotide	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	
Sequence	C	G	A	G	T	T	T	G	A	A	C	T	G	C	G	G	T	A	C	C	G	G	C	G	G	G	
Nucleotide	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	
Sequence	C	A	T	T	C	A	G	T	G	A	C	C	T	G	A	C	A	T	C	C	A	G	C	T	C	C	
Nucleotide	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	
Sequence	C	A	C	A	T	C	A	C	C	C	C	A	G	G	G	A	C	A	G	C	A	T	A	T	C	A	
Nucleotide	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	
Sequence	G	A	G	C	T	T	T	G	A	A	C	A	A	G	G	T	A	G	T	G	A	A	T	G	A	A	C
Nucleotide	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	
Sequence	T	C	T	T	C	C	G	G	G	A	T	G	G	G	G	T	A	A	A	C	T	G	G	G	G	T	
Nucleotide	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	
Sequence	C	G	C	A	T	T	G	T	G	C	C	T	T	T	T	T	T	C	T	C	C	T	T	C	G	G	
Nucleotide	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	
Sequence	C	G	G	G	G	C	A	C	T	G	T	G	C	G	T	G	G	A	A	A	G	C	G	T	A	G	
Nucleotide	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	
Sequence	A	C	A	A	G	G	A	G	A	T	G	C	A	G	G	T	A	T	T	G	G	T	G	A	G	T	
Nucleotide	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	
Sequence	C	G	G	A	T	C	G	C	A	G	C	T	T	G	G	A	T	G	G	C	C	A	C	T	T	A	
Nucleotide	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	
Sequence	C	C	T	G	A	A	T	G	A	C	C	A	C	C	T	A	G	A	G	C	C	T	T	G	G	A	
Nucleotide	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	
Sequence	T	C	C	A	G	G	A	G	A	A	C	G	G	C	G	G	C	T	G	G	g	t	a	a	a	a	
Nucleotide	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	
Sequence	a	c	c	a	a	g	c	c	c	c	t	t	g	t	g	t	g	t	c	c	c	t	t	t	t	c	
Nucleotide	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	
Sequence	t	t	t	g	g	c	c	t	c	t	g	g	t	c	a	g	a	g	a	t	c	c	c	c	a	a	
Nucleotide	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	
Sequence	c	a	g	c	c	c	t	t	c	t	t	c	t	g	t	a	t	c	t	c	t	t	t	c	t	g	
Nucleotide	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	
Sequence	t	t	g	t	g	t	t	g	g	g	t	g	a	t	t	g	t	t	g	g	a	g	a	c	g	t	
Nucleotide	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	
Sequence	t	g	a	t	a	g	t	t	g	a	g	g	a	a	a	c	c	t	g	a	c	t	g	g	c	c	
Nucleotide	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	
Sequence	t	c	a	t	t	t	c	a	c	c	a	c	a	a	g	a	g	g	t	t	a	a	c	t	c	c	
Nucleotide	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	
Sequence	t	t	g	g	a	a	a	t	a	c	a	a	a	t	g	a	c	c	c	c	c	a	c	c	c	a	
Nucleotide	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	
Sequence	c	c	t	a	c	a	t	c	a	c	t	c	t	c	t	g	a	g	g	a	c	t	a	a	a	a	
Nucleotide	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	
Sequence	c	t	g	c	a	c	c	t	t	c	c	t	c	t	c	t	c	t	c	c	c	c	t	g	c	a	
Nucleotide	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	
Sequence	g	G	A	T	A	C	T	T	T	T	G	T	G	G	A	A	C	T	C	T	A	T	G	G	G	A	
Nucleotide	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	
Sequence	A	C	A	A	T	G	C	A	G	C	A	G	C	C	G	A	G	A	G	C	C	G	A	A	A	G	
Nucleotide	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	
Sequence	G	G	C	C	A	G	G	A	A	C	G	C	C	T	T	C	A	A	C	C	G	C	T	G	G	T	T
Nucleotide	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	
Sequence	C	C	T	G	A	C	G	G	G	C	A	T	G	A	C	T	G	T	G	G	C	C	A	G	G	T	
Nucleotide	677	678	679	680	681																						
Sequence	A	A	G	T	A	Key																					

Key

Single stranded
Double stranded
Splice site

Supplementary Figure 5: Single-stranded and double-stranded nucleotides in Bcl-x-681 wild type transcript. Nucleotides that are defined as double-stranded (strong V1 cleavage) and those defined as single-stranded (strong T1 and/or T2 cleavage) are coloured blue and red, respectively.

Supplementary Figure 6: Secondary structure model of Bcl-x-681. The model was generated using Mfold and the constraints derived from footprinting experiments (see Supplementary Figures 3 and 4 and Supplementary Dataset 1). The XL 5'SS, XS 5'SS and the 3'SS are indicated in green. Stem-loops are designated I to XI. Putative G4 sequences as predicted by QGRS Mapper are designated Q1 to Q6 and coloured blue. Putative structurally independent domains are circled and labelled XS, XL, Intron and 3'SS domains.

Supplementary Figure 7: Comparison of RNA footprinting of Bclx-681 full-length (FL) and fragments (Frag) X_S , X_L , intron and 3'SS. Primers used are the same as those used on the full-length pre-mRNA.

Supplementary Figure 8 : distribution of the logarithm of the index derived from comparison of nuclease footprinting on native and 7-deaza-G RNA. Note that the interval size of the values changes from 0.1 to 0.5 for values above 0.5 or below -0.5.

Supplementary Figure 9: Mapping of RNA footprinting of Bcl-x-681 in Li⁺ and K⁺ containing buffers. Nucleotides having a significantly different footprinting pattern in Li⁺ and K⁺ containing buffers are circled.

RNAse H Cleavage of WT Bcl-x-681

RNAse H Cleavage of 7-deaza Bcl-x-681

Supplementary Figure 10: RNAse H cleavage and quantification of the XS and XL fragments of Bcl-x-681 WT (Top) and 7-deaza (bottom) in nuclear extract. Experiments were done in triplicate and data represent mean values $\pm$ standard deviation

Supplementary Figure 11: Electrophoretic-mobility shift assay of native (left) and 7-deaza substituted (right) Bcl-x-681 in the absence or presence of increasing amount of BG4 antibody (0, 150 pM, 1.5 nM, 15 nM, 150 nM, 1.5 μ M, and 15 μ M)