Interplay of three G-quadruplex units in the *KIT* promoter

Cosimo Ducani,*[1] Giulio Bernardinelli,[1] Björn Högberg,[1] Bernhard K. Keppler,[2] Alessio Terenzi,*[2§]

[1] Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden

[2] Institute of Inorganic Chemistry, University of Vienna, Waehringerstrasse 42, A-1090 Vienna, Austria

KEYWORDS. G-quadruplex, DNA, KIT promoter, oncogenes, cancer.

ABSTRACT: The proto-oncogene *KIT* encodes for a tyrosine kinase receptor which is a clinically validated target for treating gastrointestinal stromal tumors. The *KIT* promoter contains a G-rich domain within a relatively long sequence potentially able to form three adjacent G-quadruplex (G4) units, namely K2, SP and K1. These G4 domains have been studied mainly as single quadruplex units derived from short truncated sequences, and are currently considered promising targets for anti-cancer drugs, alternatively to the encoded protein. Nevertheless, the information reported so far do not contemplate the interplay between those neighboring G4s in the context of the whole promoter, possibly thwarting drug-discovery efforts. Here we report the structural and functional study of the *KIT* promoter core sequence, in both single and double stranded forms, which includes all three predicted G4 units. By preventing the formation of alternatively one or two G4 units and by combining biophysical techniques and biological assays, we show for the first time that these quadruplexes cannot be analyzed independently, but they are correlated to each other. Our data suggest that, while K2 and K1 G-rich sequences retain the ability to fold into parallel G4 motifs within a long sequence, SP G-rich domain contributes to G4 structure only together with K2. Remarkably, we have found that, in the context of a dynamic equilibrium between the three G4 units, the G4 formed by K1 has the most significant influence on the structure stability and on the biological role of the whole promoter.

INTRODUCTION

G-quadruplexes (G4s) are non-canonical DNA structures formed by single stranded G-rich sequences. These motifs arise from guanine tetrads stacked on top of each other, stabilized by monovalent cations and connected by looping bases.[1] In striking contrast to the uniformity of the classical Watson-Crick B-DNA structure, intramolecular G4s are highly polymorphic leading to topologies classifiable according to their strand polarities (parallel, antiparallel or hybrid).[2]

Researchers exponentially increased their interest into G4s after the localization of these structures within human cells.[3] These unconventional DNA motifs are not randomly distributed: they are enriched in telomeres and in gene regulatory regions that include splicing sites and promoters of cancer related genes.[1,2] G4 structures derived from short G-rich oligodeoxyribonucleotides (ODNs), with an average of 20 bases and which are relatively easy to handle in solution, have been resolved by X-ray or NMR and they were used by medicinal chemists for drug discovery purposes.[4] Currently, only two G4-binding drugs have entered clinical trials for human cancers.[4,5]

The vast majority of studies on G4 structures are restricted to the above mentioned short ODNs with just one G4 unit. Only very few papers, mainly focusing on telomeric motifs,[6–8] report on longer sequences containing two or three identical G4 units in the same strand. These pioneering works have led to the formulation of contrasting models for two adjacent quadruplexes: one in which G4s are independent beads on a string and another in which they stack on each other forming higher order structures (G4-G4 crosstalk).[7,8]

Besides telomeres, the human genome contains several regions with G-rich tracts compatibles with the formation of multiple adjacent G4 units such as DNA minisatellites and promoters of diverse cancer related genes, including *BCL-2, hTERT* and *PDGFR-β*.[9–18] The majority of the existing studies investigating adjacent G4s as single units in oncogenes do not provide a full picture on the role of each of these domains in a more physiological multi-G4 context, and how this can affect cellular processes.

A well-known example of a sequence containing neighbouring G4s is represented by the proto-oncogene *KIT* (Figure 1a). It is one of the most widely studied G4 containing promoter since it controls the expression of a tyrosine kinase receptor associated to a large number of malignancies, including erythroleukemia[19] and gastrointestinal stromal tumor (GIST).[20] The *KIT* promoter comprises three adjacent regions able to fold in G4 structures: *KIT1*, *KIT**, a validated G-rich Sp1 binding site, and *KIT2* which are located between -108 and -181 nucleotides upstream the ATG codon.[21] For simplicity, from now on, we will refer to these units as K1, SP and K2 respectively. The methodological approach of previous studies unveiled structural and functional insights of single units, mainly using the corresponding short G-rich sequences,[15,22–26] not including analysis of the full G-rich *KIT* regulatory core element. According to these investigations, K1 and K2 form parallel G4s, while SP, folds in an antiparallel two-tetrad quadruplex. Even if drug-candidates have already been synthesized for one or the other structure,[27–30] it is still not clear which G4

motif is more biologically relevant and whether there is interplay between the three units. Recently, an *in vitro* study focusing on K2 and SP domains showed that a G4-G4 tandem is possible,[10] demonstrating that a complete overview on this kind of multi-unit sequences is indispensable.

In this study we analyse, for the first time, the full sequence (both in single and double stranded forms) of the core *KIT* promoter, which simultaneously contains all three G4 domains. By performing a set of biophysical, biochemical and cellular experiments, we demonstrate that there is a structural and functional connection between all the G4 units within the promoter. We show that SP is functionally related to K2 and it does not fold in the previously predicted antiparallel structure. We also demonstrate that K1, the most stable quadruplex in the long sequence, heavily influences gene expression.

RESULTS AND DISCUSSION

Spectroscopic analysis of the *KIT* core promoter sequence. Circular dichroism (CD) contributed in all of the main discoveries concerning DNA secondary structures, including G4s.[31–33] Accordingly, we decided to use such a technique to preliminarily study a 78-bases long single-stranded ODN including all three wild-type G4 units (namely WK1, WK2 and WSP, Figure 1). We then adopted a mutation strategy to investigate how a unit could be affected by the presence or the absence of the other ones. In particular, we prevented the folding of, alternatively, one, two or three G4 units (MK1, MK2 and MSP) in the wild-type sequence, mutating the corresponding guanines involved in the tetrads.

At first, we ensured the effectiveness of the G4 mutations (G-to-A and G-to-T), using the corresponding short truncated sequences (Table S1). The WK1 sequence in K^+ buffered solution formed a parallel quadruplex characterized by distinctive positive and negative bands at 263 and 241 nm, respectively (black line in Figure 1b).[24] The WK2 sequence folded as well in a parallel quadruplex topology, with a small proportion of antiparallel structure leading to a weak peak at 295 nm (red line in Figure 1b).[34] Finally, WSP showed positive peaks at 291 and 249 nm and a negative signal at 263 nm (blue line in Figure 1b), consistent with an antiparallel two tetrad G4 topology.[15,26] All our mutated short sequences displayed similar CD signals typical of mostly unstructured single-stranded DNAs (Figure S1).[35,36] G-to-T mutations on K2 sequence provided higher ellipticity values when compared to its G-to-A analogue. Moreover, NUPACK analysis, [37] revealed slightly less stable secondary structures for all the G-to-A long sequences (Figure S2-S3). For these reasons we decided to continue our investigations with G-to-A mutated sequences.

We then recorded the CD spectra of the long sequences (unit names separated by the "-" symbol) (Table S2) and we compared them with the signals induced by the mixtures of short ODNs corresponding to the single G4 domains (unit names separated by the "+" symbol). The CD spectrum of the long wild-type sequence WK2-WSP-WK1 in K^+ buffered solution showed a positive peak at 265 nm and a negative one at 245 nm (black line in Figure 1c). Interestingly, the sum of the dichroic bands produced by the solution containing a mix of the three short sequences previously annealed as single G4 units (WK2+WSP+WK1, red line in Figure 1c) does not completely overlap with the WK2-WSP-WK1 signals. In details, the positive shoulder at 293 nm, characteristic of the 2-tetrad antiparallel G4 formed by SP, is missing in the 78-bases long sequence, while it is still visible in CD signal coming from the short wild-type units' mixture. These data suggest that the parallel G4 character induced by K1 and K2 is retained, but on the other hand, prove that the SP unit is not folding into its previously reported antiparallel G4 structure in the case of the long, multiple unit, single stranded DNA sequence.

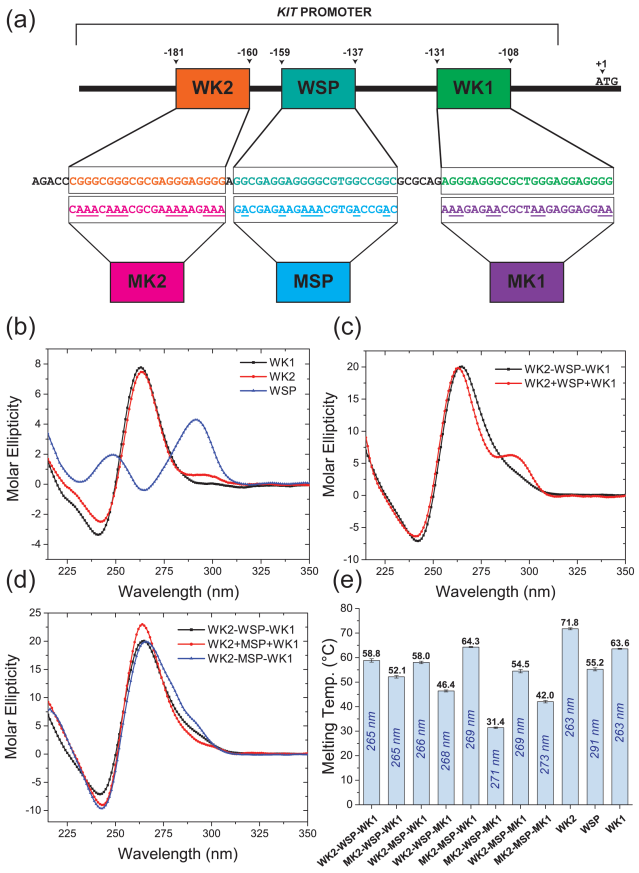


Figure 1. (a). Schematic representation of the wild-type G4 units (WK2, WSP and WK2) disposition in the *KIT* promoter, their sequences and the corresponding mutated versions (MK2, MSP and MK2) used to prevent the formation of G4 structures. (b) CD spectra of the short wild-type sequences corresponding to WK2, WSP and WK1 G4s. (c,d) CD spectra of the long sequences WK2-WSP-WK1 (black lines) and WK2-MSP-WK1 (d, blue line) compared with the spectrum produced by the mixture of single unit G4s formed by the corresponding pre-folded short sequences (red lines). (e) ODNs melting temperatures derived from the analysis of CD melting curves at the reported wavelengths. For double mutated ODNs, only the melting temperature corresponding to a G4 is reported. Units for molar ellipticity in CD spectra are 10^6 mdeg M^{-1} cm^{-1}.

Confirming this scenario, we obtained major overlap between the CD of WK2-WSP-WK1 and the one recorded for the mixture of the single units, where WSP was replaced by

its mutated version MSP (red line in Figure 1d). Strikingly, the situation was preserved when the mutation was performed within the longer sequence WK2-MSP-WK1 (blue line in Figure 1d). In this case, we observed an additional dichroic contribution in the region around 290 nm, possibly indicating a supplementary secondary structure of the long ODN.

We repeated the same experiment for the long sequences with both single and double mutations on the G4 units (Figure S4-S5). When K2 and K1 were mutated within the long sequence, the quadruplex in the SP portion did not adopt its typical antiparallel single unit conformation, characterized by a positive shoulder at 293 nm. Furthermore, we obtained a closer resemblance to the wild-type sequence spectrum when K2 was mutated within the long sequence (Figure S5c), when compared to the mutation on K1, which instead led to a general red shift of the peaks. This would suggest that K1 is the G4 which contributes the most to the parallel character of the wild-type long sequence. Among the ODNs with mutations in two units, the one involving both K1 and K2 showed, as expected, the worst overlap with WK2-WSP-WK1 (Figure S5g).

We recorded the CD spectra of all the oligonucleotides at different temperatures to investigate their thermal stability (Figure S6). Melting temperature data (Figure 1e) perfectly matched the conclusions of the previous CD experiments. In fact, the stability of the long wild-type sequence resulted to be basically unaffected by the single mutation on the SP portion, while single mutation on the K1 domain was the one producing more effects, indicating again that K1 G4 unit contributes more than the others to the general features of the long sequence. Melting curves of double mutated ODNs (Figure S6b, right panel) were characterized by two transitions, one at low temperatures (~ 35 °C) and one at higher temperatures (≥ 50 °C), most probably corresponding to a secondary non-G4 and to a G4 structure, respectively. Analysis of the multiple transitions of MK2-WSP-MK1 melting curve (Figure S6b) revealed that only the 12% is associated to a G4 melting, confirming that the quadruplex in the SP sequence does not have big influence on the thermal stability of the long ODNs.

Overall, these results clearly indicated that the tested G-rich units behave differently when observed in the context of a long sequence than when treated as single and isolated units. The main outcomes are that SP, within the long sequence, does not fold in an antiparallel fashion and that K1 plays a key role in the final structure of the full-length single stranded DNA.

Probing G-quadruplex structures. We performed biochemical and spectroscopic assays based on Thioflavin T (ThT) (Figure 2a), to better characterize the formation of quadruplexes in the *KIT* promoter. ThT is a benzothiazole-based molecule known for its switch-on fluorescence properties when bound to G4s and it has been extensively used as molecular probe for G4 detection.[38–40]

We first performed electro-mobility shift assays (EMSA) (Figure 2b) of all the pre-folded ODNs already analyzed by CD. Gels were initially stained with ThT to specifically detect G4 motifs, followed by incubation in SybrGold to reveal all the DNA bands. If compared to the corresponding mutated versions, short ODNs WK1, WK2 and WSP, folded in K^+ buffer and ran in a native PAGE, showed different electro-mobility and much higher fluorescence (Figure 2b, top), clearly confirming the G4 formation previously observed in our CD experiments. This result was further confirmed by spectrofluorimetric assays (Figure 2c, left) where the addition of ThT aliquots to wild-type 2 μM short ODN solutions led to higher emissions if compared to the same experiments performed with the mutated sequences (λ_{exc} = 425 nm). It is worth mentioning that ThT is known to have different affinity depending on the G4 arrangement.[38,40] This is also true for the selected short sequences as both EMSA and the spectrofluorimetric assay showed a higher fluorescence for WK1 compared to the other two G4s. Moreover, ThT stained also higher molecular weight bands in the wild type short ODNs only, suggesting that short G-rich ODNs can also fold into higher order G4 structures, as observed in previous studies. [41,42]

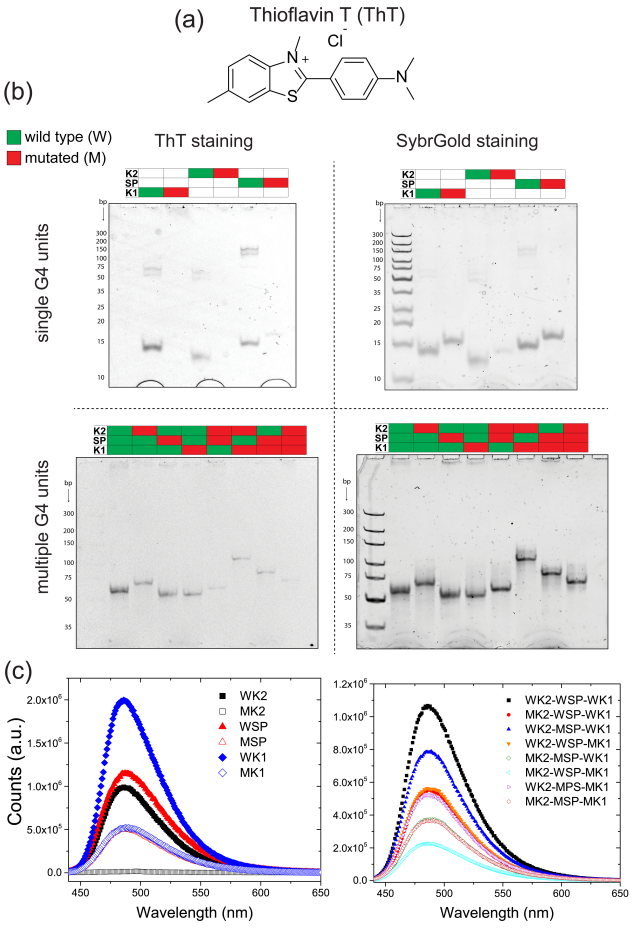


Figure 2. (a) Chemical structure of ThT. (b) EMSA in native PAGE gel of short ODNs representing the single G4 units (top) and of the long ODNs which include all the units (bottom). Wild-type sequences are represented in green, mutated in red. Gel were stained with ThT (left) and with SybrGold (right). Wild-type short ODNs are partially folded in higher ordered structures, as already reported elsewhere.[41,42] (c) Fluorescence spectra of short (left) and long (right) ODNs in the presence of 0.8 and 8.0 eq. of ThT, respectively.

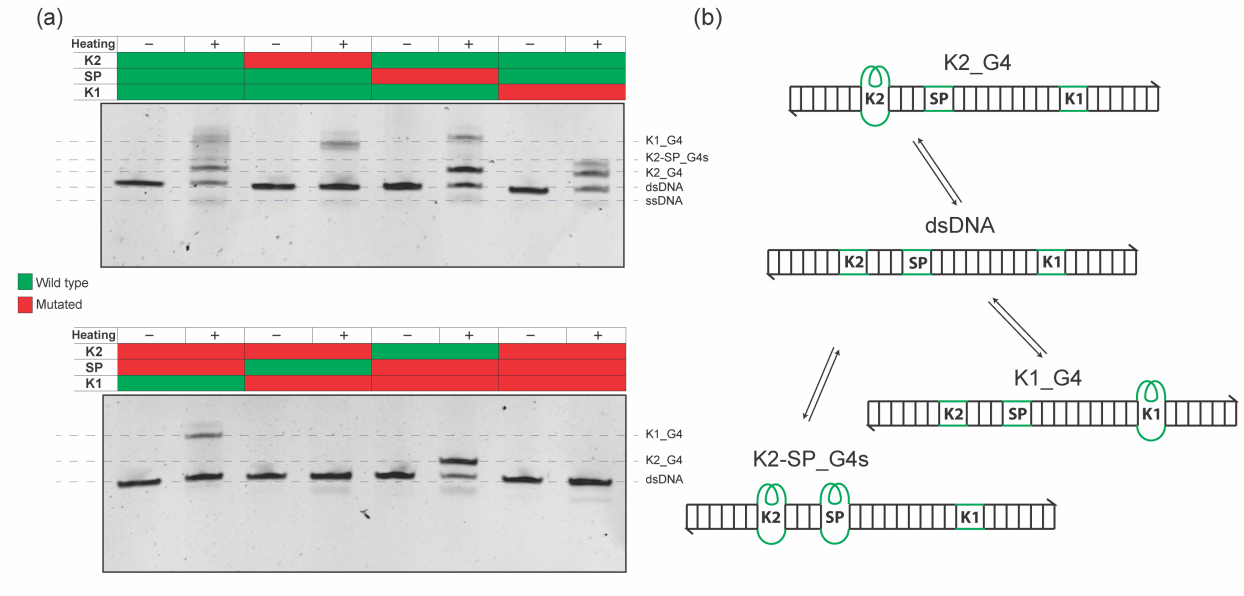


Figure 3. (a) Molecular crowding EMSA performed on dsDNA sequences before (-) or after (+) heating-renaturing step, with all the *KIT* G4 units in the wild-type and/or mutated versions; (b) schematic representation of double stranded wild-type core promoter sequence (WK2-WSP-WK1) and possible quadruplex formation sites; (c) representation of equilibrium among the different G-quadruplex structures according to the dsDNA wild-type sequence after molecular crowding EMSA.

EMSA of the long ODNs folded in K^+ buffer revealed, after ThT staining, significant differences in terms of fluorescence, but also of electro-mobility (Figure 2b, bottom), with MK2-WSP-MK1 band running slower, possibly due to its weaker sequence stability as observed in the CD assay (Figure 1e).

The band corresponding to the wild-type WK2-WSP-WK1 long sequence had the highest fluorescence and the replacement of any of the G4-units, with their corresponding mutated versions, led to a reduction of ThT associated fluorescence. By analyzing the intensity of each band taking into account the background fluorescence due to ThT unspecific binding (Figure S7), we found that WK2-WSP-WK1 ThT associated intensity was 2 or 3 times higher than ODNs with single G4 unit mutation and up to 6 times higher than the double and triple unit mutation ODNs. These results can be ascribed to a less G4-structured sequence after every mutation. Besides, in the case of the wild type and single mutated sequences, ThT may be "sandwiched" by two proximal G4s causing a further restriction in its rotational motion and a corresponding higher emission. Interestingly, the replacement of WSP with MSP produced a fluorescence decrease, indicating that the guanines in the SP unit, although not contributing with an antiparallel G-quadruplex structure as verified by CD, they definitively contribute to quadruplex formation in the long sequence.

As for the short ODNs, we investigated the ability of ThT to bind to the long sequences by spectrofluorimetry. WK2-WSP-WK1, when mixed with ThT, gave rise to the highest signal (Figure 2c, right and Figure S8-S9), confirming the EMSA outcome. Remarkably, among those sequences with a single unit mutation, the one with MSP has the closest fluorescence to the wild-type, while the double mutation involving K1 and K2 has the lowest emission. This suggests, once again, a higher contribution by K1 and K2 to the stability of the G4s compared to SP.

dsDNA EMSA in molecular crowding environment. A previous study has demonstrated how molecular crowding, which mimics experimental settings closer to physiological conditions, creates an essential environment for the stable formation of G4 motifs in dsDNA.[43] Based on this information, we tested the folding of *KIT* G4 domains in the wild-type and mutated double stranded DNA sequences (constructs available in Addgene, Table S3).

After cloning the full *KIT* promoter into eight plasmids, we digested and gel-purified a 186 base pair long dsDNA (Figure S10), which included the eight different versions of our 78 base pair *KIT* core promoter. The digestion products were incubated in 150 mM KCl and 10 mM Tris-HCl, together with 35% PEG200 as molecular crowding agent. Samples were loaded in a native PAGE gel either directly (as negative control) or after a heat denaturation-renaturation step (to induce the G4 formation) and visualized by SybrGold staining. Additional negative controls on the wild-type sequence, performed with a heat denaturation-renaturation step in the absence of alternatively or simultaneously KCl and PEG200 (Figure S11), confirmed that the newly observed bands in Figure 3a were effectively derived from G4 structures.

As expected, the fully mutated dsDNA sequence (MK2-MSP-MK1) did not show any shift corresponding to G4 formation (Figure 3a, bottom gel). With regards to the double mutated products (Figure 3a, bottom gel), it was possible to visualize the formation of both K1 (MK2-MSP-WK1), and K2 (WK2-MSP-MK1) quadruplexes, but no band corresponding to the SP G4 unit (MK2-WSP-MK1) was

detected. In addition, it is interesting how K1 and K2 G4s showed different mobility even when folded from a dsDNA. These two bands were also present in the wild-type dsDNA sequence (WK2-WSP-WK1) (Figure 3a, top gel), together with another fainter band which we identified as G4 tetrads involving K2 and SP. This last conclusion is corroborated by the fact that the same band was absent in the constructs MK2-WSP-WK1 and WK2-MSP-WK1, but it reappeared in WK2-WSP-MK1. This interplay between K2 and SP and the absence of the contribution of SP alone are in accordance with our observation in the CD experiments, where we detected a loss of the typical antiparallel SP contribution in the single stranded wild-type sequence (compare with Figure 1), but also confirm our ThT assay data, where we observed higher fluorescence in the wild-type sequence, where SP is able in any case to contribute to the total G-quadruplex structures.

We also observed a very faint band at the bottom of the dsDNA products after the denaturation-renaturation step, more visible in the constructs WK2-WSP-WK2 and WK2-MSP-WK1. We assume this is ssDNA products, not re-annealed to its complementary strand.

From this EMSA we can conclude that, in this experimental condition, the wild-type sequence promoter is in equilibrium between three different main G-quadruplex formations: K1, K2 and K2-SP (Figure 3b), and that the most represented form is the one with a single G4 in the K1 unit (Figure S12), with about 38% of the total DNA, followed by a single G4 in K2 with about 21% of the total DNA. This data overlaps with our melting temperature experiments on single stranded oligonucleotides (compare with Figure 1e), confirming a higher stability of K1 over the other G4 units.

Dual luciferase assay. To untangle the biological function of the *KIT* promoter G4 units and to better understand how the inactivation of one unit could affect the activity of the others on the gene expression, we performed a dual luciferase assay by using the eight promoter constructs already used for the dsDNA EMSA (Figure 4a, Table S3). Differently from previous investigations to study *KIT* promoter activity where sequences were partially removed,[44] we only mutated the guanines in the core promoter sequence, leaving the rest of the promoter unaltered. It is worth mentioning that the high number of G-to-A mutations could also affect the binding of transcription factors or repressors, also independently from the disruption of G4 motifs.

We performed luciferase expression assay on HEL92.1.7 cell line, which has high endogenous expression of *KIT* transcripts. Single, double and triple unit mutations significantly affected the expression of luciferase (Figure 4b). In particular, we observed that by mutating K2 only (construct MK2-WSP-WK1) there was a slightly lower expression of the luciferase compared to the wild-type (construct WK2-WSP-WK1). Much more drastic was the expression decrease when SP was mutated (construct WK2-MSP-WK1). In fact, this sequence is the binding recognition motif of transcription factors and the introduction of mutations could affect the transcription, regardless of the G4 formation.[44] Although K1 and K2 have both similar parallel G-quadruplex structures, preventing K1 G4 formation (construct WK2-WSP-MK1), surprisingly led to a gene expression enhancement, and the variation in expression in this case was much more significant compared to the one observed by mutating K2 only. The higher significance when K1 unit is mutated, compared to K2 mutation, and its resulting positive effect on expression could explain evidences found in previous studies where synthetic molecules, able to disrupt both K1 and K2, led to a general increase of gene expression.[45] At the same time, the key role of K1 over K2 can also explain the effect of molecules which, by stabilizing K1 and K2 G4s, lead to a down-regulation of *KIT* expression in carcinoma cell lines.[46,47] Our results suggest that the disruption or stabilization of the G4 formed by K1 might be more biologically and pharmacologically relevant than the one formed in the K2 region.

The first double mutation tested (construct MK2-MSP-WK1) led to a drastic decrease of gene expression. By comparing this value with the ones obtained using the corresponding single unit mutation (WK2-MSP-WK1 and MK2-WSP-WK1), we observed that, in the case of this double mutation, the gene expression is only slightly lower than the one obtained with only SP mutated, while it is significantly lower than the one where only K2 is mutated. These data suggest how, once SP is mutated and it cannot exploit its function, an extra mutation on K2 does not decrease the gene expression much more, while when only K2 is mutated there is still activity of the SP unit.

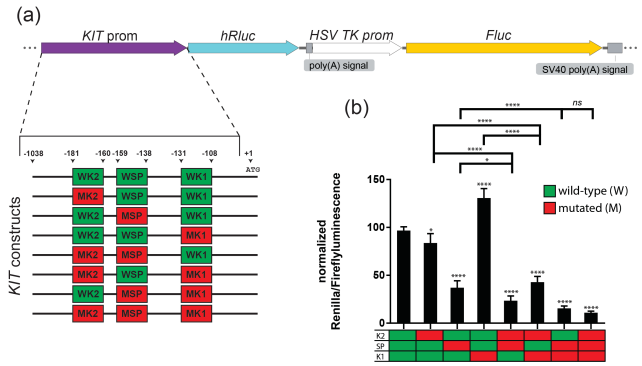


Figure 4. (a) Schematic representation of the constructs used for the dual luciferase assays, with specified all the combinations of wild-type and mutated units in the core *KIT* promoter sequence. (b) Luciferase activity in HEL92.1.7 cells transfected with the different *KIT* promoter constructs, and their relative significance based on p values.

By re-establishing the wild-type sequence of SP and instead mutating K1 (construct MK2-WSP-MK1), we observed an upregulation of the expression compared to the previous double unit mutation case. Nevertheless, if we compare this expression level with the one obtained using the corresponding single unit mutations on K2 or K1, it is significantly lower. Finally, we noticed an even lower expression with mutation on both SP and K1 (WK2-MSP-MK1) and with all the units mutated (MK2-MSP-MK1). These results might indicate that wild-type SP is not able to completely perform its function if K2 is mutated but, on

the other hand, we also show, for the first time, that a mutation in K1 is able to increase the gene expression over the wild-type levels, only when both K2 and SP are wild-type (construct WK2-WSP-MK1). This denotes that there is a functional link between all the *KIT* promoter quadruplex units meaning that the regulative role of a single unit is definitively affected by the formation or disruption of the adjacent ones.

CONCLUSIONS

The emerging importance of G-quadruplexes as anticancer target requires calls for a deeper understanding of their structure and functionality in a context as close as possible to physiological conditions. Considering the relevant number of G-rich tracts in human genome compatible with the formation of multiple adjacent G4 units, one of the main challenges is to clarify whether these units can be considered as separate beads in a string or if the folding of one G4 influences the others.

The pioneering studies conducted mainly by the groups of Balasubramanian and Neidle revealed that, among multiple G4 systems, the quadruplexes within *KIT* have the potential to be concrete selective targets for anticancer drug candidates.[24,48]

Following such suggestions, we expanded the study of the *KIT* promoter to a sequence which simultaneously includes the three predicted G4 units, an approach never attempted so far.

Through a combination of spectroscopic techniques and gel electrophoresis analysis, we have found that, when analyzed within the full-length promoter sequence, the SP is not able to fold in the antiparallel quadruplex recently described in the literature.[15] According to our data, SP G-rich tract exists in equilibrium between an unfolded state and a folded one which includes simultaneously K2. This supports a recent study, restricted to a single stranded sequence without the K1 domain, according to which K2 and SP form a higher order cross-talk structure.[10] At the same time, we highlight that the quadruplex corresponding to the K1 G-rich sequence is the most stable both in the single and double stranded form. This last observation confirms and strengthen previous findings where it was demonstrated that K1 structure is highly conserved, stable, not environmentally sensitive, and with a dynamic much reduced compared to other G4s (e.g. the telomeric one).[48,49] Importantly, in the context of double stranded DNA we never observed a simultaneous formation of the three G4s but rather an equilibrium between K1, K2 and K2-SP.

Extending this study with a functional cell-based assay, we show that all the above considerations have a direct consequence on the *KIT* promoter activity in the leukemia cell line HEL92.1.7. We demonstrate that K1 and K2 G4s have opposite roles in gene regulation, one reducing and the other increasing the expression levels, respectively. Furthermore, K1 quadruplex demonstrated to have a major effect compared to K2 G4, whose function is closely related to SP, probably with a supportive role in the recruitment of transcription factors. Remarkably, the function of each unit is strongly affected by the formation or disruption of the other ones. This outcome paves the way for future studies to evaluate how different mutations in the *KIT* promoter affect the binding of transcription factors, depending or not on the formation of G-quadruplex structures.

Overall, in this study we demonstrate that there is a clear link between all the G4 units within the *KIT* promoter and we suggest that this link must be taken in consideration when, for example, these DNA motifs (or the ones of other promoters containing multiple G4s) are treated as target for anticancer drugs. Our data suggest that a good strategy for *KIT* G4 targeting is to synthesize molecules able, not only to distinguish between the double helix and the G4s, but specifically between K1 and K2, avoiding that the stabilization or the disruption of both G4s, at the same time, could lead to antagonistic effects. In particular we indicate K1 as a more reliable target due to its stability and reduced dynamic. At the same time, we propose that our mutation strategy and experimental settings could help in understanding how transcription factors might affect or being affected by the equilibrium between K1, K2 and SP G4s.

EXPERIMENTAL SECTION

Single stranded ODNs. Oligonucleotides (see Table S1 and S2) were purchased from IDT (Integrated DNA Technologies, Belgium) in PAGE purity grade. We firstly diluted the lyophilized strands in IDTE buffer (10 mM Tris, pH 7.5, 0.1 mM EDTA) to obtain 100 μM stock solutions. We then prepared working solutions diluting the stock solutions to the required concentrations with 10 mM tris-hydroxymethyl-aminomethane buffer (Tris-HCl, pH=7.5) in the presence of 100 mM KCl. In order to perform the folding of the diluted oligonucleotide solutions we heated them to 95 °C for 5 min and we let cool to room temperature overnight.

We checked the exact strand concentration of the oligonucleotides by measuring the absorbance at 260 nm of the corresponding diluted solutions and using the extinction coefficient values provided by the manufacturer.

Spectroscopic analysis of single stranded ODNs. Circular dichroism spectra were recorded on a Chirascan™ CD Spectrometer (by Applied Photophisics) equipped with a single cell Peltier temperature controller, using 1 cm path-length quartz cuvettes, at 25 °C with the following parameters: range 500-200 nm, bandwidth: 1.0 nm, time per point: 0.5 s, repeats: 4. Concentration of ODNs were adjusted so that the HV (HT) voltage could be properly controlled, giving reliable ellipticity values over the investigated wavelength range. Concentration of both short and long ODNs were in the range 0.5-1.0 μM (strand concentration). CD raw data were divided for the actual ODN concentration, measured by UV-Vis, and represented as molar ellipticity. CD at variable temperature were analyzed plotting the CD signal at the specified wavelength against the temperature and fitting the corresponding plot with a Boltzmann sigmoidal curve implemented in the Origin 9.5 software package (OriginLab Corp.) (see Figure S6, left panel). Melting curves of double mutated ODNs (Figure S6b, right panel), characterized by two transitions, were

fitted with Double-Boltzmann sigmoidal curves which give as a result the two melting points and the % of the curve corresponding to the melted structure. CD at different temperatures of MK2 and MK1 ODNs showed no significant changes and, accordingly, no melting points were reported, while MSP non-G4 secondary structure melted at 34.1 °C.

Fluorescence spectra were recorded on Horiba FluoroMax-4 spectrofluorometer. ThT assay was performed adding aliquots of Thioflavin T (Sigma Aldrich) stock solution to the corresponding ODN (short or long) solutions at the indicated concentrations and measuring after 2 min incubation time (λ_{exc} = 425 nm, λ_{em} = 440-750 nm). Scans were run at room temperature with excitation and emission slit widths of 1/5 nm and 2.5/5 nm for short and long ODNs experiments, respectively. Spectrofluorimetric data were normalized according to the ODNs concentration before being analysed and represented.

ThT concentration was calculated recording the UV-Vis spectrum of its diluted solutions (on a PerkinElmer LAMBDA 35 double beam spectrophotometer) and using the molar extinction coefficient at 412 nm in water of 36 000 M^{-1} cm^{-1}.[40]

All the spectroscopy-based experiments were performed in Tris-HCl 50 mM buffer (pH = 7.4) supplemented with KCl 100 mM and analyzed by Origin 9.5 (OriginLab Corp.)

Electromobility Shift Assay on single stranded DNA oligonucleotides. We resuspended single units short and multiple units long oligonucleotides to a final concentration of 2 μM together with 10 mM Tris-HCl and 100 mM KCl. We incubated all the mixtures at 95 °C for 5 min and we left them cooling down at room temperature overnight. We prepared 15% native polyacrylamide gels with 1× TBE (made by 89 mM Tris-borate, 2 mM Na_2EDTA dissolved in deionized water), and 25 mM KCl. We loaded 1 μl and 2 μl of the long and short oligonucleotides mixture respectively, and we ran the electrophoresis at 17 V/cm (2:35 hr for the long ODNs and 1:30 hr for the shorts ODNs). We stained gels into a 0.5 μM Thioflavin T solution (Sigma Aldrich), we quickly washed them with 1x TBE and acquired images, before performing SybrGold staining (1×; Invitrogen). All images were acquired by LAS 4000 (GE) imager and analyzed by ImageJ.

EMSA of dsDNA in molecular crowding conditions. The double stranded DNA sequences (186 bp) of the eight plasmids enclosing the *KIT* promoter were enzymatically digested (BsaI/NheI) and purified by gel extraction (Nucleospin gel and PCR clean-up, Macherey-Nagel). For each sequence variant, 200 ng were resuspended in the folding buffer (150 mM KCl, 10 mM Tris-HCl pH 7.4, PEG200 35% w/v), and heat denatured and renatured in a thermocycler (95 °C for 5 min followed by a ramp to 25 °C at a rate of 0.1 °C /second). For comparison we resuspended the same dsDNA sequences in a control buffer (final concentration: 150 mM LiCl, 10 mM Tris-HCl pH 7.4, PEG200 35% w/v) and we kept this solution at room temperature. Further negative controls were performed in absence of KCl and PEG200 after a denaturation-renaturation step.

Samples were assessed on a PAGE gel as previously described.[43] Briefly, polyacrylamide mixture (29:1), final concentration 8% in 1x TBE buffer supplemented with 150 mM KCl and 35% w/v PEG200 were polymerized with 1% APS and 0.1% TEMED and run with 1x TBE buffer with 150 mM KCl at 11 V/cm in ice cold water for 6 hours. The images were acquired, after SybrGold staining, by LAS 4000 (GE) imager and analyzed by ImageJ.

Luciferase assay. The wild-type human *KIT* promoter sequence (GeneBank: S67773.1) and its seven mutants (insert sequences in supplementary information Table S2) were synthesized (BioCat GmbH) and cloned (BglII/NheI) in the psiCHECK-2 plasmid in order to replace the SV40 promoter upstream the *hRluc* reporter gene. Plasmids were transformed in DH5α competent cells (Thermo Fisher Scientific) and harvested from an overnight culture by using the maxiprep kit Genejet (Thermo Fisher Scientific) and the sequences were confirmed by Sanger sequencing. The erythroleukemia cell line HEL92.1.7 was obtained from the European Collection of Authenticated Cell Culture and was cultured in RPMI 1640 Glutamax (GIBCO) supplemented with 10% fetal bovine serum (FBS, Sigma Aldrich), 1 mM Sodium citrate (Sigma Aldrich) and grown in suspension at 37 °C, 5 % CO_2 atmosphere. Cells were typically transfected with 20 μg of plasmid DNA with the Neon electroporator (Thermo Fisher Scientific) using the 100 μl tips (electroporation settings: 2 pulses, 1400 Volts, 20 ms pulse width). 36 hours after transfection, the expression of the luciferase reporter gene was measured using the Dual-Glo kit (Promega), following the manufacturer instruction, in a Veriscan lux luminometer (ThermoFisher Scientific). The normalized ratio was obtained from three independent biological replicates and graphed by Prism 7 software.

AUTHOR INFORMATION

Corresponding Author

Cosimo Ducani; cosimo.ducani@ki.se
Alessio Terenzi; aterenzi@dipc.org

Present Addresses

§ Donostia International Physics Center, Paseo Manuel de Lardizabal 4, Donostia, 20018, Spain

Author Contributions

The manuscript was written through contributions of all authors. / All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

We thank the David and Astrid Hagelén Foundation and Svenska Sällskapet för Medicinsk Forskning (SSMF) for funding to CD, the EU FP7 through the EScoDNA ITN to G.B., the Knut and Alice Wallenberg Foundation and the Swedish Foundation for Strategic Research (grant FFL12-0219) for funding to BH, and the Mahlke-Obermann Stiftung and the EU FP7 Programme (grant agreement no. 609431) for funding to A.T. Furthermore, we thank Dr. G. Miglietta for the experimental sugges-

tions, Prof. L. Salassa, Dr. C. De Visu and F. Bobbio for the fruitful discussions.

ABBREVIATIONS

PAGE, Polyacrylamide Gel Electrophoresis; dsDNA (double stranded DNA); ssDNA (single stranded DNA); EMSA, Electromobility Shift Assay; APS, Ammonium Persulfate; TEMED, Tetramethylethylenediamine; PEG, Polyethylene Glycol; Tris, Tris(hydroxymethyl)aminomethane; PCR, Polymerase Chain Reaction.

REFERENCES

(1) Hänsel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S. DNA G-Quadruplexes in the Human Genome: Detection, Functions and Therapeutic Potential. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (5), 279–284.

(2) Rhodes, D.; Lipps, H. J. Survey and Summary G-Quadruplexes and Their Regulatory Roles in Biology. *Nucleic Acids Res.* **2015**, *43* (18), 8627–8637.

(3) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, *5* (3), 182–186.

(4) Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J. Med. Chem.* **2016**, *59* (13), 5987–6011.

(5) Xu, H.; Di Antonio, M.; McKinney, S.; Mathew, V.; Ho, B.; O'Neil, N. J.; Santos, N. Dos; Silvester, J.; Wei, V.; Garcia, J.; et al. CX-5461 Is a DNA G-Quadruplex Stabilizer with Selective Lethality in BRCA1/2 Deficient Tumours. *Nat. Commun.* **2017**, *8* (205), 14432.

(6) Collie, G. W.; Parkinson, G. N.; Neidle, S.; Rosu, F.; De Pauw, E.; Gabelica, V. Electrospray Mass Spectrometry of Telomeric RNA (TERRA) Reveals the Formation of Stable Multimeric G-Quadruplex Structures. *J. Am. Chem. Soc.* **2010**, *132* (27), 9328–9334.

(7) Yu, H.-Q.; Miyoshi, D.; Sugimoto, N. Characterization of Structure and Stability of Long Telomeric DNA G-Quadruplexes. *J. Am. Chem. Soc.* **2006**, *128* (48), 15461–15468.

(8) Payet, L.; Huppert, J. L. Stability and Structure of Long Intramolecular G-Quadruplexes. *Biochemistry* **2012**, *51* (15), 3154–3161.

(9) Palumbo, S. M. L.; Ebbinghaus, S. W.; Hurley, L. H. Formation of a Unique End-to-End Stacked Pair of G-Quadruplexes in the HTERT Core Promoter with Implications for Inhibition of Telomerase by G-Quadruplex-Interactive Ligands. *J. Am. Chem. Soc.* **2009**, *131* (31), 10878–10891.

(10) Rigo, R.; Sissi, C. Characterization of G4–G4 Crosstalk in the c-KIT Promoter Region. *Biochemistry* **2017**, *56* (33), 4309–4312.

(11) Morgan, R. K.; Batra, H.; Gaerig, V. C.; Hockings, J.; Brooks, T. A. Identification and Characterization of a New G-Quadruplex Forming Region within the KRAS Promoter as a Transcriptional Regulator. *Biochim. Biophys. Acta - Gene Regul. Mech.* **2016**, *1859* (2), 235–245.

(12) Agrawal, P.; Lin, C.; Mathad, R. I.; Carver, M.; Yang, D. The Major G-Quadruplex Formed in the Human BCL-2 Proximal Promoter Adopts a Parallel Structure with a 13-Nt Loop in K + Solution. *J. Am. Chem. Soc.* **2014**, *136* (5), 1750–1753.

(13) Chaires, J. B.; Trent, J. O.; Gray, R. D.; Dean, W. L.; Buscaglia, R.; Thomas, S. D.; Miller, D. M. An Improved Model for the HTERT Promoter Quadruplex. *PLoS One* **2014**, *9* (12), e115580.

(14) Cogoi, S.; Shchekotikhin, A. E.; Xodo, L. E. HRAS Is Silenced by Two Neighboring G-Quadruplexes and Activated by MAZ, a Zinc-Finger Transcription Factor with DNA Unfolding Property. *Nucleic Acids Res.* **2014**, *42* (13), 8379–8388.

(15) Raiber, E. A.; Kranaster, R.; Lam, E.; Nikan, M.; Balasubramanian, S. A Non-Canonical DNA Structure Is a Binding Motif for the Transcription Factor SP1 in Vitro. *Nucleic Acids Res.* **2012**, *40* (4), 1499–1508.

(16) Amrane, S.; Adrian, M.; Heddi, B.; Serero, A.; Nicolas, A.; Mergny, J. L.; Phan, A. T. Formation of Pearl-Necklace Monomorphic G-Quadruplexes in the Human CEB25 Minisatellite. *J. Am. Chem. Soc.* **2012**, *134* (13), 5807–5816.

(17) Kang, H. J.; Cui, Y.; Yin, H.; Scheid, A.; Hendricks, W. P. D.; Schmidt, J.; Sekulic, A.; Kong, D.; Trent, J. M.; Gokhale, V.; et al. A Pharmacological Chaperone Molecule Induces Cancer Cell Death by Restoring Tertiary DNA Structures in Mutant HTERT Promoters. *J. Am. Chem. Soc.* **2016**, *138* (41), 13673–13692.

(18) Brown, R. V.; Wang, T.; Chappeta, V. R.; Wu, G.; Onel, B.; Chawla, R.; Quijada, H.; Camp, S. M.; Chiang, E. T.; Lassiter, Q. R.; et al. The Consequences of Overlapping G-Quadruplexes and i-Motifs in the Platelet-Derived Growth Factor Receptor β Core Promoter Nuclease Hypersensitive Element Can Explain the Unexpected Effects of Mutations and Provide Opportunities for Selective Targeting of Both Structures by Small Molecules to Downregulate Gene Expression. *J. Am. Chem. Soc.* **2017**, *139* (22), 7456–7475.

(19) Kubota, A.; Okamura, S.; Shimoda, K.; Ikematsu, W.; Otsuka, T.; Niho, Y. Analysis of C-Kit Expression of Human Erythroleukemia Cell Line, HEL: Clonal Variation and Relationship with Erythroid and Megakaryocytic Phenotype. *Leuk. Res.* **1995**, *19* (4), 283–290.

(20) Lennartsson, J.; Ronnstrand, L. Stem Cell Factor Receptor/c-Kit: From Basic Science to Clinical Implications. *Physiol. Rev.* **2012**, *92* (4), 1619–1649.

(21) Yamamoto, K.; Tojo, A.; Aoki, N.; Shibuya, M. Characterization of the Promoter Region of the Human C-Kit Proto-Oncogene. *Jpn. J. Cancer Res.* **1993**, *84* (11), 1136–1144.

(22) Kuryavyi, V.; Phan, A. T.; Patel, D. J. Solution Structures of All Parallel-Stranded Monomeric and Dimeric G-Quadruplex Scaffolds of the Human c-Kit2 Promoter. *Nucleic Acids Res.* **2010**, *38* (19), 6757–6773.

(23) Hsu, S. T. D.; Varnai, P.; Bugaut, A.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. A G-Rich Sequence within the c-Kit Oncogene Promoter Forms a Parallel G-Quadruplex Having Asymmetric G-Tetrad Dynamics. *J. Am. Chem. Soc.* **2009**, *131* (37), 13399–13409.

(24) Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D. J. Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-Kit Promoter. *J. Am. Chem. Soc.* **2007**, *129* (14), 4386–4392.

(25) Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. Putative DNA Quadruplex Formation within the Human C-Kit Oncogene. *J. Am. Chem. Soc.* **2005**, *127* (30), 10584–10589.

(26) Kotar, A.; Rigo, R.; Sissi, C.; Plavec, J. Two-Quartet Kit* G-Quadruplex Is Formed via Double-Stranded Pre-Folded Structure. *Nucleic Acids Res.* **2019**, *47* (5), 2641–2653.

(27) Bejugam, M.; Gunaratnam, M.; Müller, S.; Sanders, D. A.; Sewitz, S.; Fletcher, J. A.; Neidle, S.; Balasubramanian, S. Targeting the C-Kit Promoter G-Quadruplexes with 6-Substituted Indenoisoquinolines. *ACS Med. Chem. Lett.* **2010**, *1* (7), 306–310.

(28) McLuckie, K. I. E.; Waller, Z. A. E.; Sanders, D. A.; Alves, D.; Rodriguez, R.; Dash, J.; McKenzie, G. J.; Venkitaraman, A. R.; Balasubramanian, S. G-Quadruplex-Binding Benzo[a]Phenoxazines down-Regulate c-KIT Expression in Human Gastric Carcinoma Cells. *J. Am. Chem. Soc.* **2011**, *133* (8), 2658–2663.

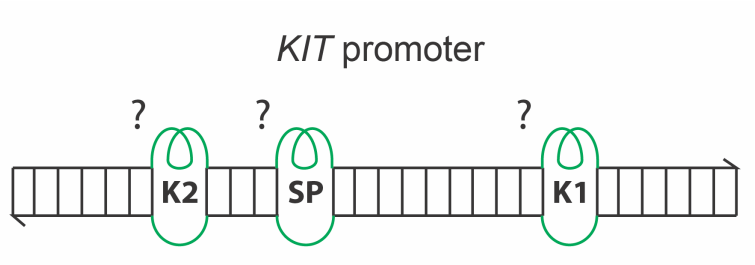
(29) Zorzan, E.; Da Ros, S.; Musetti, C.; Shahidian, L. Z.; Coelho, N. F. R.; Bonsembiante, F.; Létard, S.; Gelain, M. E.; Palumbo, M.; Dubreuil, P.; et al. Screening of Candidate G-Quadruplex Ligands for the Human c-KIT Promotorial Region and Their Effects in Multiple in-Vitro Models. *Oncotarget* **2016**, *7* (16), 21658–21675.

(30) Terenzi, A.; Lötsch, D.; van Schoonhoven, S.; Roller, A.; Kowol, C. R.; Berger, W.; Keppler, B. K.; Barone, G. Another Step toward DNA Selective Targeting: Ni II and Cu II Complexes of a Schiff Base Ligand Able to Bind Gene Promoter G-Quadruplexes. *Dalton Trans.* **2016**, *3* (18), 7758–7767.

(31) Kypr, J.; Kejnovska, I.; Renciuk, D.; Vorlickova, M. Circular

Dichroism and Conformational Polymorphism of DNA. *Nucleic Acids Res.* **2009**, *37* (6), 1713–1725.
(32) Randazzo, A.; Spada, G. P.; da Silva, M. W. Circular Dichroism of Quadruplex Structures BT - Quadruplex Nucleic Acids; Chaires, J. B., Graves, D., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 67–86.
(33) del Villar-Guerra, R.; Trent, J. O.; Chaires, J. B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chemie Int. Ed.* **2018**, *57* (24), 7171–7175.
(34) Fernando, H.; Reszka, A. P.; Huppert, J.; Ladame, S.; Rankin, S.; Venkitaraman, A. R.; Neidle, S.; Balasubramanian, S. A Conserved Quadruplex Motif Located in a Transcription Activation Site of the Human C-Kit Oncogene. *Biochemistry* **2006**, *45* (25), 7854–7860.
(35) Zhang, W.; Chen, M.; Ling Wu, Y.; Tanaka, Y.; Juan Ji, Y.; Lin Zhang, S.; He Wei, C.; Xu, Y. Formation and Stabilization of the Telomeric Antiparallel G-Quadruplex and Inhibition of Telomerase by Novel Benzothioxanthene Derivatives with Anti-Tumor Activity. *Sci. Rep.* **2015**, *5* (1), 13693.
(36) Choi, J.; Kim, S.; Tachikawa, T.; Fujitsuka, M.; Majima, T. PH-Induced Intramolecular Folding Dynamics of i-Motif DNA. *J. Am. Chem. Soc.* **2011**, *133* (40), 16146–16153.
(37) Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. NUPACK: Analysis and Design of Nucleic Acid Systems. *J. Comput. Chem.* **2011**, *32* (1), 170–173.
(38) Renaud de la Faverie, A.; Guédin, A.; Bedrat, A.; Yatsunyk, L. A.; Mergny, J.-L. Thioflavin T as a Fluorescence Light-up Probe for G4 Formation. *Nucleic Acids Res.* **2014**, *42* (8), e65–e65.
(39) Zhao, D.; Dong, X.; Jiang, N.; Zhang, D.; Liu, C. Selective Recognition of Parallel and Anti-Parallel Thrombin-Binding Aptamer G-Quadruplexes by Different Fluorescent Dyes. *Nucleic Acids Res.* **2014**, *42* (18), 11612–11621.
(40) Mohanty, J.; Barooah, N.; Dhamodharan, V.; Harikrishna, S.; Pradeepkumar, P. I.; Bhasikuttan, A. C. Thioflavin T as an Efficient Inducer and Selective Fluorescent Sensor for the Human Telomeric G-Quadruplex DNA. *J. Am. Chem. Soc.* **2013**, *135* (1), 367–376.
(41) Dexheimer, T. S.; Sun, D.; Hurley, L. H. Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the *Bcl-2* P1 Promoter. *J. Am. Chem. Soc.* **2006**, *128* (16), 5404–5415.
(42) Marsh, T. C.; Henderson, E. G-Wires: Self-Assembly of a Telomeric Oligonucleotide, d(GGGGTTGGGG), into Large Superstructures. *Biochemistry* **1994**, *33* (35), 10718–10724.
(43) Zheng, K.; Chen, Z.; Hao, Y.; Tan, Z. Molecular Crowding Creates an Essential Environment for the Formation of Stable G-Quadruplexes in Long Double-Stranded DNA. *Nucleic Acids Res.* **2010**, *38* (1), 327–338.
(44) Park, G. H.; Plummer, H. K.; Krystal, G. W. Selective Sp1 Binding Is Critical for Maximal Activity of the Human C-Kit Promoter. *Blood* **1998**, *92* (11), 4138–4149.
(45) Waller, Z. A. E.; Sewitz, S. A.; Hsu, S.-T. D.; Balasubramanian, S. A Small Molecule That Disrupts G-Quadruplex DNA Structure and Enhances Gene Expression. *J. Am. Chem. Soc.* **2009**, *131* (35), 12628–12633.
(46) McLuckie, K. I. E.; Waller, Z. A. E.; Sanders, D. A.; Alves, D.; Rodriguez, R.; Dash, J.; McKenzie, G. J.; Venkitaraman, A. R.; Balasubramanian, S. G-Quadruplex-Binding Benzo[*a*]Phenoxazines Down-Regulate *c-KIT* Expression in Human Gastric Carcinoma Cells. *J. Am. Chem. Soc.* **2011**, *133* (8), 2658–2663.
(47) Domarco, O.; Lötsch, D.; Schreiber, J.; Dinhof, C.; Van Schoonhoven, S.; García, M. D.; Peinador, C.; Keppler, B. K.; Berger, W.; Terenzi, A. Self-Assembled Pt2L2 Boxes Strongly Bind G-Quadruplex DNA and Influence Gene Expression in Cancer Cells. *Dalton Trans.* **2017**, *46* (2), 329–332.
(48) Wei, D.; Husby, J.; Neidle, S. Flexibility and Structural Conservation in a C-KIT G-Quadruplex. *Nucleic Acids Res.* **2015**, *43* (1), 629–644.
(49) Shirude, P. S.; Okumus, B.; Ying, L.; Ha, T.; Balasubramanian, S. Single-Molecule Conformational Analysis of G-Quadruplex Formation in the Promoter DNA Duplex of the Proto-Oncogene C-Kit. *J. Am. Chem. Soc.* **2007**, *129* (24), 7484–7485.

Insert Table of Contents artwork here



Graphical abstract: we unveil structural and functional interplay of G-rich domain in the *KIT* promoter