10132–10142 *Nucleic Acids Research, 2017, Vol. 45, No. 17* *Published online 29 July 2017*
doi: 10.1093/nar/gkx678

Coexistence of two main folded G-quadruplexes within a single G-rich domain in the *EGFR* promoter

Maria L. Greco[1], Anita Kotar[2], Riccardo Rigo[1], Camilla Cristofari[1], Janez Plavec[2,3,4,*] and Claudia Sissi[1,*]

[1]Department of Pharmaceutical and Pharmacological Sciences, University of Padova, v. Marzolo 5, Padova 35131, Italy, [2]Slovenian NMR Center, National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia, [3]EN-FIST Center of Excellence, Trg OF 13, 1000 Ljubljana, Slovenia and [4]Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, Ljubljana, Slovenia

Received May 25, 2017; Revised July 19, 2017; Editorial Decision July 21, 2017; Accepted July 22, 2017

ABSTRACT

***EGFR* is an oncogene which codifies for a tyrosine kinase receptor that represents an important target for anticancer therapy. Indeed, several human cancers showed an upregulation of the activity of this protein. The promoter of this gene contains some G-rich domains, thus representing a yet unexplored point of intervention to potentially silence this gene. Here, we explore the conformational equilibria of a 30-nt long sequence located at position −272 (EGFR-272). By merging spectroscopic and electrophoretic analysis performed on the wild-type sequence as well as on a wide panel of related mutants, we were able to prove that in potassium ion containing solution this sequence folds into two main G-quadruplex structures, one parallel and one hybrid. They show comparable thermal stabilities and affinities for the metal ion and, indeed, they are always co-present in solution. The folding process is driven by a hairpin occurring in the domain corresponding to the terminal loop which works as an important stabilizing element for both the identified G-quadruplex arrangements.**

INTRODUCTION

The cell surface receptor Epidermal Growth Factor Receptor (EGFR) is an important factor in the pathogenesis and progression of several human cancers (1). Along with ErbB2/HER-2, ErbB3/HER-3 and ErbB4/HER-4, it belongs to the ErbB family of protein kinase receptors. Up to seven physiological ligands (i. e. EGF, TGF-α, etc.) have been identified that are able to activate EGFR by driving its homo- or hetero-dimerization. Physiologically, this activity is required in normal cell growth and proliferation comprising the maintenance of normal intestinal functions and homeostasis (2). However, EGFR overexpression or mutations that constitutively activate this receptor are known oncogenic drivers (3). Relevant examples of cancers connected to an aberrant activation of EGFR are non-small cell lung cancer, breast cancer and glioblastoma (4). Currently available therapeutic agents used to counteract the upregulation of EGFR are tyrosine kinases inhibitors and humanized monoclonal antibodies against the receptor extracellular domain. Both treatments are designed to switch off the kinase activity and, consequently, the signal transduction. Unfortunately, their efficacy is severely impaired by either intrinsic or acquired resistance, often deriving from a selection of pre-existing subclones (5). Several mechanisms can contribute to this phenomenon i. e. amplification of alternative pathways to activate common downstream factors able to promote cell proliferation or the expression of mutated forms of EGFR. Among them, worth of mention are the gatekeeper T790M mutation which increases the affinity of the receptor for ATP in its binding pocket (6).

The gene that codifies for EGFR is located on chromosome 7p12–13. Using *in silico* analytical tools it was discovered that sequences in *EGFR* gene at positions −37 and −272 from the transcriptional start site (TSS) potentially form G-quadruplex (G4) structures (7,8). G4s are four-stranded helical structures that can be formed by single-stranded guanine-rich DNA (and RNA) oligonucleotides. G4s arise from Hoogsteen hydrogen bonding of four guanines arranged in a planar G-quartet (9). Stacking of two or more G-quartets leads to formation of a G4 that is further stabilized by monovalent cations. The considerable structural diversity is characteristic for DNA G4s. It was shown that factors such as primary sequence, especially number of G-tracts and their length, number of assembling DNA molecules, length of loops, orientation of strands, type of cations and other external factors importantly influence the structure of G4 (10–15). G4s can adopt parallel, antiparallel and (3+1) hybrid topologies characterized by different

*To whom correspondence should be addressed. Tel: +39 049 827 5711; Fax: +39 049 827 5366; Email: claudia.sissi@unipd.it
Correspondence may also be addressed to Janez Plavec. Tel: +386 1 4760 353; Fax: +386 1 4760 300; Email: janez.plavec@ki.si

orientation of the four strands. It was demonstrated that proteins like DNA and RNA helicases can selectivity recognize the topology of G4 (16,17). The transition between different topologies could be induced by changing the type of cations, adding different cosolutes and crowding agents (11,18,19). Loops that connect guanine residues involved in G-quartets have an important role in overall folding and stability of G4s (12). Three basic types of loops are characteristic for G4s; propeller, lateral and diagonal loops (20). Orientation of loop depends on number and nature of nucleotides in loops as well strand orientation and number of G-quartets they traverse (12). Some long(er) loops have been found to adopt a well-defined structures. It has been shown that hairpin-like loop structures increase thermodynamic stability of G4s (12). It is noteworthy, that G4s are not the only structures that can be formed by G-rich sequences. According to recent reports they can adopt other non-canonical structures such as G-hairpin and AGCGA-quadruplexes (21–23).

The formation of G4 structures in human cells has been demonstrated clearly (24–29). G-rich regions are found in telomeric regions. In addition, bioinformatic genome analysis showed that they frequently cluster upstream of the TSS of many oncogenes (30). Compared to telomeres, G4-forming sequences found in promotor regions are more diverse with varying number and length of G-tracts and intersecting residues resulting in potential formation of multiple G4s. It was demonstrated that formation of G4 in promoter regions can be involved in regulation of gene transcription. Investigations into structural features of G4s enable structure based design of ligands that would bind to and stabilize G4 structures. Indeed, in promotor regions, G4 stabilization by small-molecule ligands frequently results in the suppression of gene expression (31–34). The silencing of *EGFR* transcription by promoting G4 formation in its promotor could thus represent a powerful complementary therapeutic strategy to the currently available treatments.

With the aim to explore the formation of G4 in the promotor region of the *EGFR* gene, we focused on the 30 nt sequence named EGFR-272 d[GGGGACCGGGTCCAGAGGGGCAGTGCTGGG] that starts at positions −272 from the TSS. With the use of spectroscopic and electrophoretic techniques we describe its structural features in ionic conditions comparable to those found in the intracellular environment. The experimental data clearly established formation of G4 with additional Watson–Crick GC base pairs within the loop regions. The effect of individual guanine and cytosine residues on the folding of EGFR-272 in the presence of K^+ ions has been tested rigorously with several spectroscopic methods including nuclear magnetic resonance (NMR). We identified two main G4 forms of EGFR-272, a kinetically favored (3+1) hybrid and a slowly forming parallel one, both comprised of three stacked G-quartets.

MATERIALS AND METHODS

Oligonucleotides

Oligonucleotides (Table 1) were purchased from Metabion International AG (German) and resuspended in milliQ water and then purified by electrophoretic technique (EGRF-272: 5′-GGG GAC CGG GTC CAG AGG GGC AGT GCT GGG-3′. The EGFR-272, ΔG1, G4T, G4,17T and G4,20T were also synthesized on K&A Laborgeraete GbR DNA/RNA Synthesizer. In all cases, the standard phosphoramidite chemistry was used. Deprotection and cleavage from the solid support was done with the use of aqueous ammonia at 55°C for 12 h. The crude oligonucleotides were then purified by RP-HPLC and desalted, before use.

Electrophoretic mobility shift assay (EMSA)

^{32}P end-labeled single-stranded oligonucleotides were obtained by incubating the oligonucleotides with T4 polynucleotide Kinase (M-Medical S.r.l., Italy) and [γ-^{32}P] ATP (Perkin Elmer S.p.a., Italy) for 30 min at 37°C. The enzyme was then removed by extraction with phenol/chloroform/isoamylic alcohol (25:24:1). A mixture of purified labelled and unlabelled oligonucleotides (total final concentration 1 μM) was heated to 95°C for 5 min in 10 mM Tris, 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0 buffer at increasing KCl concentrations and let to cool overnight at room temperature. The folding of the starting material was monitored by native 20% polyacrylamide gel electrophoresis in 0.5× TBE (44.5 mM Tris base, 44.5 mM boric acid and 1 mM Na_2EDTA) added of 10 mM KCl. Resolved bands were visualized and quantified on a Phosphor Imager (STORM 840, Pharmacia Biotech Amersham).

Thermal differential spectrum (TDS)

The thermal difference spectrum was obtained by subtracting of the oligonucleotide UV-spectra at 25°C from the one recorded at 95°C (below and above the oligo melting temperature respectively). The experiments were performed in 10 mM NaCacodilate, 150 mM KCl pH 7.0. The resulting thermal difference spectra have been normalized to the value of 1 at the maximal intensity (35). Before the thermal difference spectra, the CD signal of the oligo in the same buffer has been recorded.

Circular dichroism (CD) measurements

Circular dichroism (CD) spectra were acquired on a Jasco J 810 spectropolarimeter equipped with a Peltier temperature controller using 10 mm path length cells. Before data acquisition, oligo solutions (ca. 4 μM in 10 mM Tris, pH 7.5) were heated at 95°C for 5 min and cooled overnight at room temperature. The reported spectrum of each sample represents the average of three scans recorded with 1-nm step resolution. For kinetic analyses, a single CD spectrum was acquired every 2 min. Thermal denaturation experiments were performed by heating the sample by 2°C and allowing sample equilibration before spectra acquisition. Observed ellipticities were converted to mean residue ellipticity [Θ] = deg x cm^2 x $dmol^{-1}$ (Mol. Ellip.).

SVD analysis

Multiple wavelength CD experiments were analyzed using Singular Value Decomposition (SVD) analysis (36,37). The

Table 1. Sequences of oligonucleotides used in this work

Oligonucleotide	Sequence										
EGFR-272	5′-	GGGG	ACC	GGG	TCC	AGA	GGGG	CAG	TGCT	GGG	–3′
EGFR-272-p	**5'-phos**	GGGG	ACC	GGG	TCC	AGA	GGGG	CAG	TGCT	GGG	–3′
ΔG1	5′-	/GGG	ACC	GGG	TCC	AGA	GGGG	CAG	TGCT	GGG	–3′
G1T	5′-	**T**GGG	ACC	GGG	TCC	AGA	GGGG	CAG	TGCT	GGG	–3′
G2T	5′-	G**T**GG	ACC	GGG	TCC	AGA	GGGG	CAG	TGCT	GGG	–3′
G3T	5′-	GG**T**G	ACC	GGG	TCC	AGA	GGGG	CAG	TGCT	GGG	–3′
G4T	5′-	GGG**T**	ACC	GGG	TCC	AGA	GGGG	CAG	TGCT	GGG	–3′
G17T	5′-	GGGG	ACC	GGG	TCC	AGA	**T**GGG	CAG	TGCT	GGG	–3′
G18T	5′-	GGGG	ACC	GGG	TCC	AGA	G**T**GG	CAG	TGCT	GGG	–3′
G19T	5′-	GGGG	ACC	GGG	TCC	AGA	GG**T**G	CAG	TGCT	GGG	–3′
G20T	5′-	GGGG	ACC	GGG	TCC	AGA	GGG**T**	CAG	TGCT	GGG	–3′
ΔG1-G20T	5′-	/GGG	ACC	GGG	TCC	AGA	GGG**T**	CAG	TGCT	GGG	–3′
G1,20T	5′-	**T**GGG	ACC	GGG	TCC	AGA	GGG**T**	CAG	TGCT	GGG	–3′
G4,17T	5′-	GGG**T**	ACC	GGG	TCC	AGA	**T**GGG	CAG	TGCT	GGG	–3′
G4,18T	5′-	GGG**T**	ACC	GGG	TCC	AGA	G**T**GG	CAG	TGCT	GGG	–3′
G4,19T	5′-	GGG**T**	ACC	GGG	TCC	AGA	GG**T**G	CAG	TGCT	GGG	–3′
G4,20T	5′-	GGG**T**	ACC	GGG	TCC	AGA	GGG**T**	CAG	TGCT	GGG	–3′
C12,13T	5′-	GGGG	ACC	GGG	T**TT**	AGA	GGGG	CAG	TGCT	GGG	–3′
G25T-C26T	5′-	GGGG	ACC	GGG	TCC	AGA	GGGG	CAG	T**TT**T	GGG	–3′
ΔT27	5′-	GGGG	ACC	GGG	TCC	AGA	GGGG	CAG	TGC/	GGG	–3′
M_{22}	5′-	GGAT	GTG	AGT	GTG	AGT	GTGA	GG			–3′
M_{30}	5′-	GTTG	ACC	GTG	TCC	AGA	GTGG	CAG	TGCT	GGG	–3′

In bold, mutated residues.

entire dataset forms a matrix, called D matrix, in which each row corresponds to a single wavelength and each column refers to a acquisition time. The applied analytical tool broke up the D matrix into three different submatrices according to the relation: D = U × S × V; where S matrix keeps information about the importance of every species contributing to the dichroic signal; U matrix contains information related to the spectral shapes of the significant species; V matrix indicates how the spectral changes occur over time. The S values and U and V autocorrelation coefficients allow determining the number of species that significantly contribute to the dichroic changes, thus selecting the significant V eingenvectors.

For kinetic CD experiment the significant V eingenvectors were globally fitted by applying different kinetic models and the best fitting was obtained using a first order monoexponential kinetic model,

$$\theta_{t,\lambda} = \theta_{\infty,\lambda} + \theta_{\lambda} \cdot e^{(-k \cdot t)} \quad (1)$$

where, $\theta_{t,\lambda}$ is the value of signal at time t, $\theta_{\infty,\lambda}$ is the final value of the signal, θ_{λ} is amplitude factor for the exponential, k is the kinetic constant. The fitting parameters formed the so-called H matrix that allowed us to determine the spectral shapes of the species contributing to the dichroic signal.

Dimethyl sulfate (DMS) footprinting

For each reaction, 150 000 cpm; ca. 300 ng of 32P-labeled DNA were annealed 10 mM Tris, 1 mM EDTA, pH 8.0 in the presence/absence of 200 mM KCl. The samples were loaded onto a 20% polyacrylamide gel in 1×TBE and the solved bands were extracted by crushed and soak in the same buffer. The recovered DNA solutions were added of 0.4% dimethyl sulfate (DMS) (25 μl EtOH, 5 μl DMS, 20 μl milliQ water) and 1 μM ctDNA (final concentration) in 50 μl total volume. After 5 min incubation at r.t. the reaction was stopped with 3.5 μl of β-mercaptoethanol and 10 μl of 40% glycerol. DNA samples were ethanol precipitated and further incubated for 30 min at 90°C in 100 μl of diluted piperidine (1:10 in milliQ water). Finally, the samples were dried, washed two times with 20 μl of milliQ water and loaded on a 20% denaturing polyacrylamide sequencing gel along with Maxam and Gilbert purine marker.

Preparation of NMR samples

Oligonucleotides for NMR samples were dissolved in H_2O with 10% of ^2H_2O and titrated with aliquots 3 M KCl solutions to a finale concentration ranging from 50 to 200 mM K^+ ions. pH values of samples were adjusted to 7.0 with the 10 mM potassium phosphate buffer and to 8.0 with the 10 mM Tris buffer. Strand concentration in the samples was from 2.6 μM to 0.4 mM and was determined by UV absorption at 260 nm using UV/VIS Spectrophotometer Varian CARY-100 BIO UV-VIS.

NMR spectroscopy

NMR spectra were recorded on Agilent (Varian) NMR System 300, 600 and 800 MHz spectrometers at 25°C in 90%/10% $H_2O/^2H_2O$. ^1H NMR spectra were recorded with the use of the DPFGSE solvent suppression method. NMR spectra were processed and analyzed by using VNMRJ 4.2 (Varian Inc.) and the Sparky (UCSF) software. All NMR spectra were referenced to the TMSP.

RESULTS

EGFR-272 shows distinct conformational features in the presence/absence of KCl

To assess the potential of EGFR-272 to fold into a de-

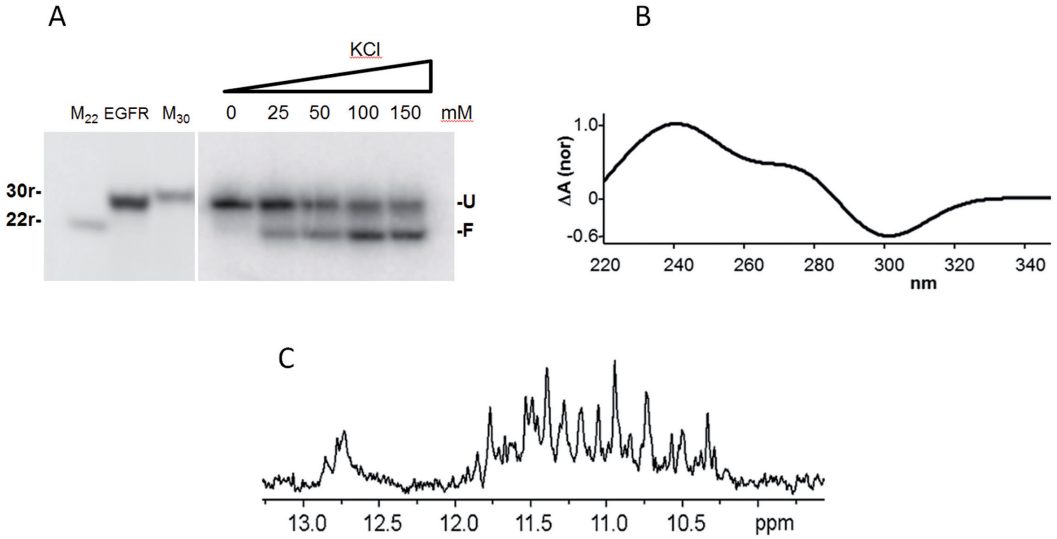

Figure 1. (**A**) Electrophoretic mobility shift assay (EMSA) of EGFR-272 annealed in the presencee of increasing KCl concentrations. Lane 0 and EGFR refer to the oligonucleotide annealed in the absence of the metal ion; lanes M_{22} and M_{30} refer to random sequences 22 and 30 residues long; U and F to the unfolded and folded form of EGFR-272, respectively. (**B**) Thermal differential spectrum of 4 μM EGFR-272 previously annealed in 150 mM KCl. (**C**) Imino region of ^{1}H NMR spectrum of 0.1 mM EGFR-272 recorded at 100 mM KCl, pH 7.0, 25°C on a 800 MHz spectrometer.

fined G4 secondary structure, we investigated the effect of KCl as a G4 stabilizer/inducer. When loaded on a native polyacrylamide gel, EGFR-272 runs essentially as a single band comprised between a 22 and a 30-mers (Figure 1A). Conversely, samples containing increasing concentrations of KCl showed a second band characterized by a higher electrophoretic mobility. Due to its reduced hydrodynamic volume, this form can be easily referred to an intramolecular folded form. Interestingly, a thermal differential spectrum analysis on the KCl-containing sample showed a negative pick at about 295–300 nm and of two positive bands at about 245 and 270 nm (Figure 1B) (35). These evidences sustain the folding of EGFR-272 into a G4 conformation in KCl concentrations (150 mM) comparable to those find in the intracellular environment. In full agreement, ^{1}H NMR spectrum of EGFR-272 revealed signals between δ 10.35 and 12.00 ppm which are characteristic of imino protons of guanine residues involved in G-quartets as building blocks of G4 structures (Figure 1C). Additionally, the signals observed between δ 12.70 and 12.90 ppm indicated formation of GC base pairs (38,39).

More than double number of imino signals in ^{1}H NMR spectrum than expected for a three G-quartet G4 indicated the formation of multiple G4 forms. This is not surprising since EGFR-272 sequence contains four guanine (G) tracts in which the guanine residues are repeated 4-3-4-3 times, respectively, that can differently pair thus producing variable G4 arrangements. To get better insights on EGFR-272 conformational features, we performed CD studies (Figure 2). The CD spectrum of the oligonucleotide in the absence of KCl presents three main bands thus indicating the occurrence of a pre-folded state: this would actually explain the slightly higher electrophoretic mobility rate of EGFR-272 in comparison to a random 30-mer. In line, a theoretical prediction indicates the possible formation of different hairpins with melting temperatures comprised between 35 and 39°C and repeated melting/annealing cycles monitored by recording the CD signal at 260 nm showed a reversible melting process within this temperature range although with a not well-defined thermal transition (Supplementary Figure S1).

Addition of the monovalent cation extensively altered the oligonucleotide CD spectrum according to a process that, in agreement with polyacrylamide gel electrophoresis (PAGE) data, was associated to a metal ion half maximal effective concentration $EC_{50} = 38 \pm 2$ μM (Figure 2A and B). The observed spectral variations were not extremely fast and required about 30 min to reach equilibrium (Figure 2C and D). In these conditions, the metal ion induced a positive signal at about 265 nm and a negative one close to 240 nm that is reminiscent of a parallel G4. However, a significant shoulder at 290 nm is preserved and, even upon careful equilibration of the sample after each KCl addition, not well resolved isodichroic points are present. Noteworthy, the formation rate of the 290 nm contribution was much faster than the one at 265 nm (Figure 2C and D). This picture anticipates the formation of at least two G4 conformations comprising both parallel and antiparallel components.

This model was further supported by an SVD analysis of the oligonucleotide CD spectra acquired at different incubation time after addition of 200 mM KCl. As anticipated, the very first step was extremely fast and this prevents us to kinetically analyze it according to the applied protocol. Thus, we moved to consider the slower part that comprises only spectra acquired after the addition of the metal ion thus removing from the analysis the contribution of the metal-free oligonucleotide (spectra ranging from the dashed-dotted line up to the dotted one in Figure 2C). Data analysis indicated that two main species are sufficient to properly describe this second folding process (Supplementary Figure S2A and B). Accordingly, the two significant derived V eingenvectors were satisfactorily described

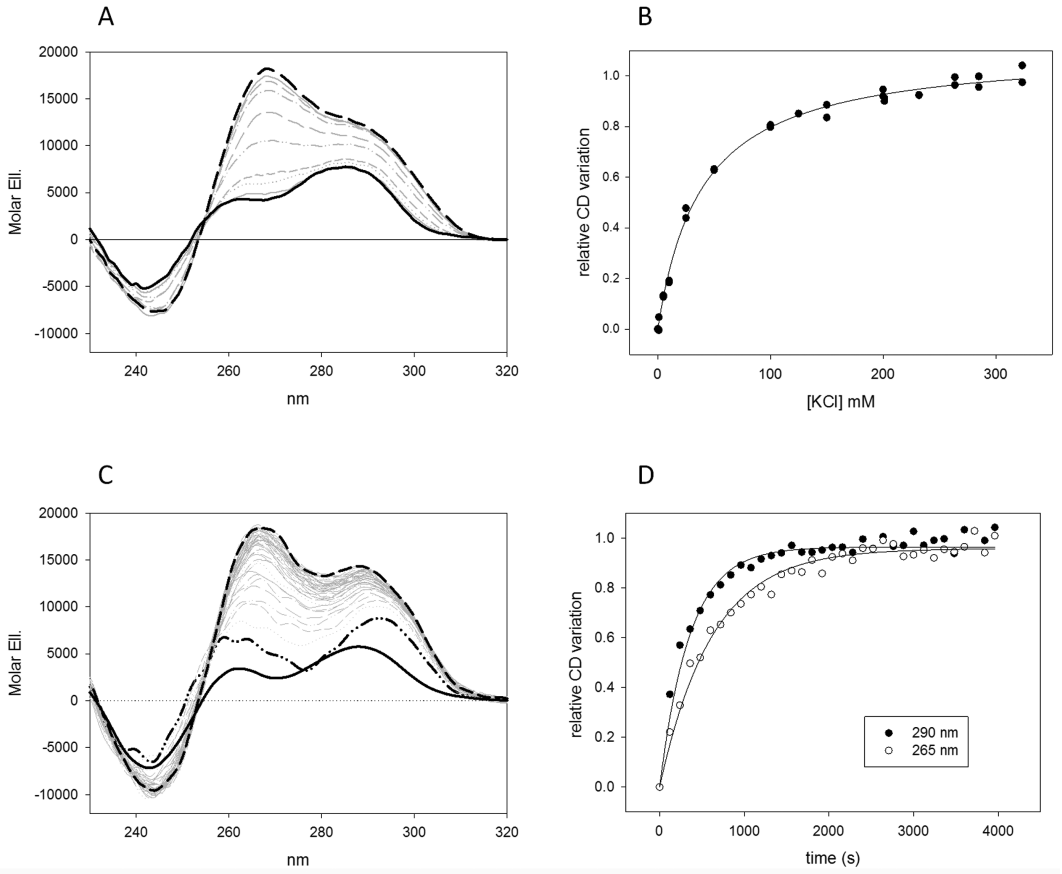

Figure 2. CD spectra of 4 μM EGFR-272 acquired in 10 mM Tris, pH 7.5, 25°C in the absence (solid black line) and in the presence of increasing concentrations of KCl (dashed black line refers to 200 mM KCl) (**A**) and corresponding relative variation of the dichroic signal recorder at 265 nm as a function of metal ion concentration (**B**). In (**C**) the time-dependent variation of CD spectrum of 4 μM EGFR-272 in 10 mM Tris, pH 7.5, 25°C upon addition of KCl is reported (1 acquisition/120 s). Black line refers to the sample in the absence of KCl, dotted-dashed line to the one acquired immediately after addition of 200 mM KCl, dotted line to the one after equilibration. The derived relative variations of the dichroic signal recorder at 265 and 290 nm as a function of incubation time are reported in (**D**).

by a global mono-exponential process (Supplementary Figure S2C and D). The derived fitting parameters allowed us to obtain the actual spectral shapes of the two folded forms in solution that are reported in Figure 3A. Among them, one can easily be referred to a parallel (Form 2) and the other one to a hybrid G4 (Form 1). The shape of the equilibrated EGFR-272 does not correspond to any of the two deconvoluted forms thus suggesting they coexist in solution.

Their stabilities were analyzed by following the melting profile of EGFR-272 after proper equilibration in 200 mM KCl (Supplementary Figure S3). In these conditions, EGFR-272 showed a fully reversible thermal denaturation profile. Analysis of the CD signal in the whole wavelength range derived from spectra acquired at increasing temperatures indicates that three species are significantly changing along the process. Comparison of the melting profile acquired at 265 versus 290 nm (the two wavelength corresponding to the more intense dichroic contribution of the two main forms) showed a modest, although significant, difference in the melting temperature (Figure 3B). In particular, a $T_{m1} = 63.4 \pm 0.2$°C and a $T_{m2} = 64.6 \pm 0.2$°C were determined at 265 and 290 nm, respectively. This points toward the presence of two main forms with, unexpectedly, very similar thermal stabilities or to a large prevalence of one form over the other.

This prompted us to investigate which were the most relevant guanines involved in the G-tetrads pairing. Thus, we isolated the high mobility band resolved by PAGE and we treated the extracted product(s) according to the DMS footprinting protocol in order to detect the guanines involved in the G4 tetrads as bases protected from the chemical modification (40). Results are summarized in Figure 4. Comparison of the DMS footprinting traces obtained in the presence and in the absence of KCl showed a clear metal-induced protection of the guanine triplets at positions 8–10 and 27–30. Within the G17-G20 tract G20 was always clearly exposed to the cleavage reaction as well as the isolated G15, G23 and G25. Conversely, it was less easy to unambiguously attribute G-quartets pairing within the G1–4 guanines repeat.

Mutation of selected guanines in EGFR-272 alters its conformational arrangement

To solve this lacking information we decided to analyze some oligonucleotides containing one or two G→T muta-

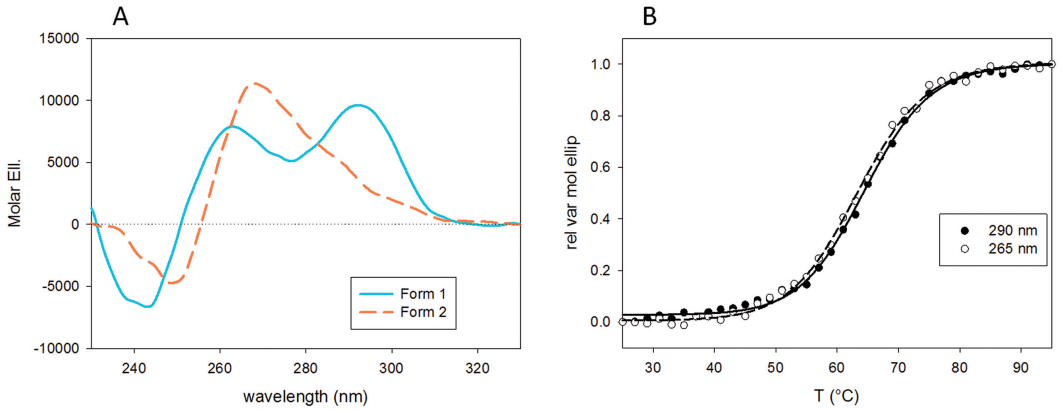

Figure 3. Generated CD spectra for the two folded species (Form 1 and Form 2 in solid and dashed lines, respectively) derived from SVD analysis of the time-dependent titration of EGFR-272 in 200 mM KCl, 25°C (**A**) and temperature-dependent relative variations of the dichroic signals of EGFR-272 recorded at 265 and 290 nm in 10 mM Tris, 200 mM KCl, pH 7.5 (**B**).

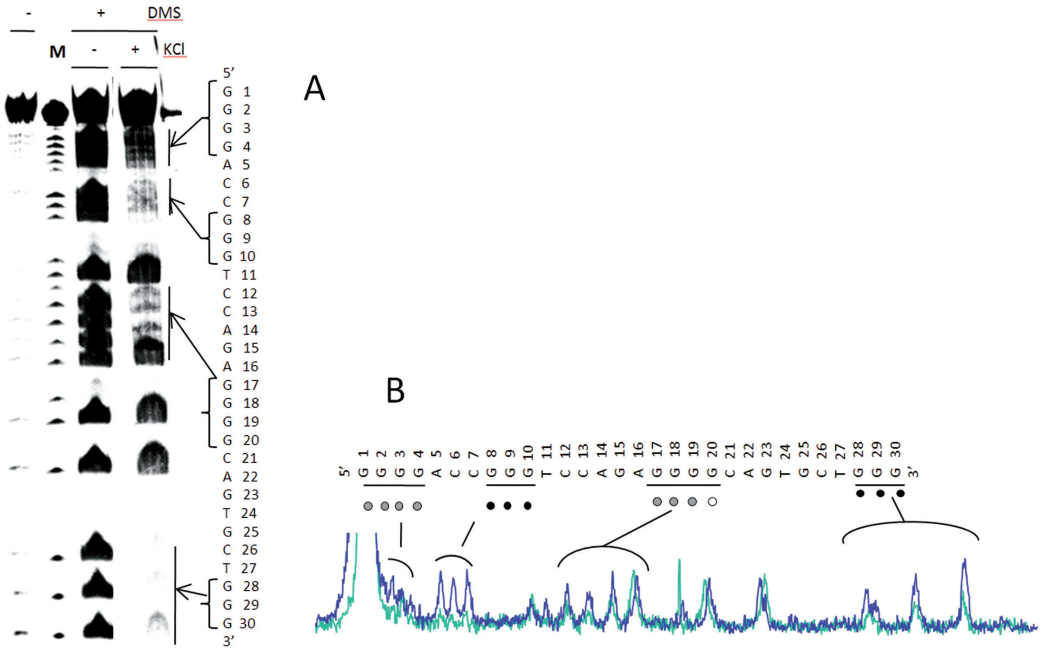

Figure 4. DMS footprinting of EGFR-272 in the absence/presence of 200 mM KCl (**A**). The densitometric analyses of the lanes corresponding to samples prepared in the absence (blue line) and in the presence of the metal ion (cyan line) are reported in (**B**).

tions within the two stretches formed by four consecutive guanines.

Electrophoretic mobility shift assay (EMSA) of the oligonucleotides annealed in 200 mM KCl showed clear differences among the tested sequences (Figure 5). Mutation of either G1 or G4 poorly affected the extent of the folded fraction in comparison to the wt sequence thus suggesting that a shuffling of the guanines forming this G4 column is actually possible. Conversely, as expected, mutation of the G17 (G17T) almost completely prevents the formation of any high electrophoretic mobility form. Thus, in this third G4 column, no recruitment of other guanines is accepted. Nevertheless, peculiar data derived upon mutation of G20. Indeed, although footprinting results indicated that for the wt sequence G20 was not involved in G-quartet pairing, the G20T sequence remains largely unfolded (≈65%). Unexpectedly, the folded fraction was split into two well resolved bands (B1 and B2). Using them as markers, it can be noticed that whereas for EGFR-272 and G1T the fastest band is almost totally absent, with G4T the metal ion induces only the more compact form although to a minor percentage.

To get clearer insights on these folded forms, we acquired the CD spectrum of all the mutated sequences both in the absence and presence of 200 mM KCl as well as the corresponding melting profiles (Figures 5 and 6; Supplementary Figure S4).

In the absence of KCl, the spectral features of the tested sequences were variable. This is likely related to the aforementioned occurrence of hairpins formations as highlighted by the variation in the oligonucleotides electrophoretic mo-

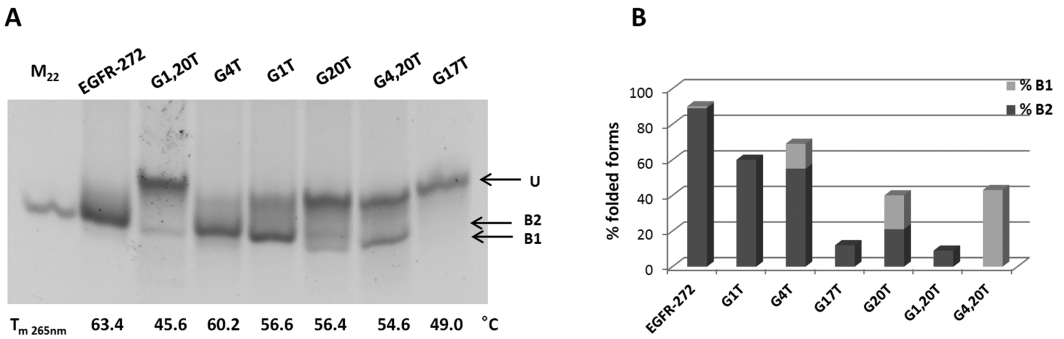

Figure 5. (**A**) EMSA of EGFR-272 and selected mutants annealed in 200 mM KCl. Lane M_{22} refers to a random sequence 22 residues long, U to the unfolded oligonucleotides and B1 and B2 to two high electrophoretic mobility forms. On the bottom, the melting temperatures derived by recording the CD signal at 265 nm in 200 mM KCl, is reported. In (**B**) the quantification of B1 and B2 with reference to the total DNA amount is reported.

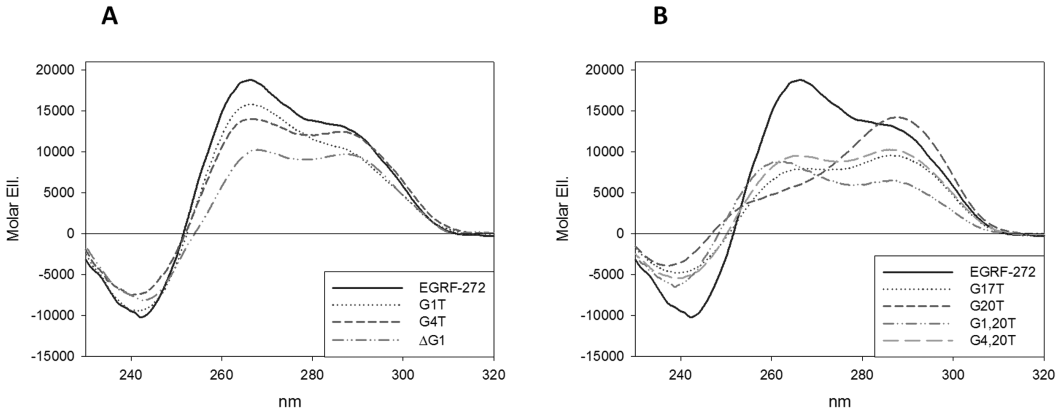

Figure 6. CD spectra of EGFR-272 single mutations in the first G-tract (**A**) or in the third one along with double mutations (**B**) acquired in 10 mM Tris, 200 mM KCl, pH 7.5, 25°C.

bility in native conditions which is actually elicited in denaturating one (Supplementary Figure S5). After annealing in KCl, the acquired CD spectra present a picture that well parallels EMSA results (Figure 6).

Among oligonucleotides containing a single mutation, G17 was confirmed to be essential for G4 formation and stability. Indeed, upon its substitution with a T (G17T), a low dichroic signal was recorded which was lost at quite low temperature.

The mutations of G1 as well as G4 did not prevent the overall folding of the nucleic acid thus further supporting that a shuffling of the guanines forming the first column within the G4 is acceptable. The slight lower stability of the folded G1T (T_m 56.6°C at 265 nm) derives from a limited reduction of the affinity for the metal ion and fits with the presence of a small fraction of unfolded DNA when resolved by PAGE. It is worth to underline that the G4T mutant showed a CD signal overlapping the one corresponding to the wt EGRF-272 at 290 nm but a lower one at 265 nm, thus suggesting a redistribution of the relative amounts of the folded forms.

Unexpectedly, the removal of the first G1 (ΔG1) did not provide a CD profile comparable to the wt EGFR-272 or to its closely related G1T. Indeed the dichroic spectrum of ΔG1 showed two well resolved maxima at 265 and 290 nm which relative ratio parallels the one obtained for G4T. Consistently, the melting temperature determined at 265 nm is notable (T_m 65.0°C).

Striking result was obtained with the G20T mutant. G20 was not expected to be involved in G-quartet pairing. However, G20T had a CD signal at 265 nm (as well as the melting temperature herein detected corresponding to 56.4°C) extremely reduced whereas the contribution at 290 nm was actually reinforced. When the G20T mutation was associated to G4T or G1T mutation, the effect was even more dramatic: both sequences with two mutations showed CD intensities comparable (G4,20T) or even lower (G1,20T) with reference to G17T. Thus, we can suggest that G20 although not directly involved in G-quartet pairing might contribute to drive G4 formation, possibly by providing an external capping or through interactions with other bases in loop(s).

Small modifications in the first G-tract of EGFR-272 leads to two well-defined G4 forms

Despite trying many different experimental conditions and annealing protocols, we were not able to reduce the structural diversity of EGFR-272 in ^{1}H NMR samples (Supplementary Figure S6). Furthermore, in order to trap only one G4 fold in solution we enlarged the mutant library up to 16 different mutated EGFR-272 oligonucleotides, which included an added phosphate group at the 5′-end (EGFR-

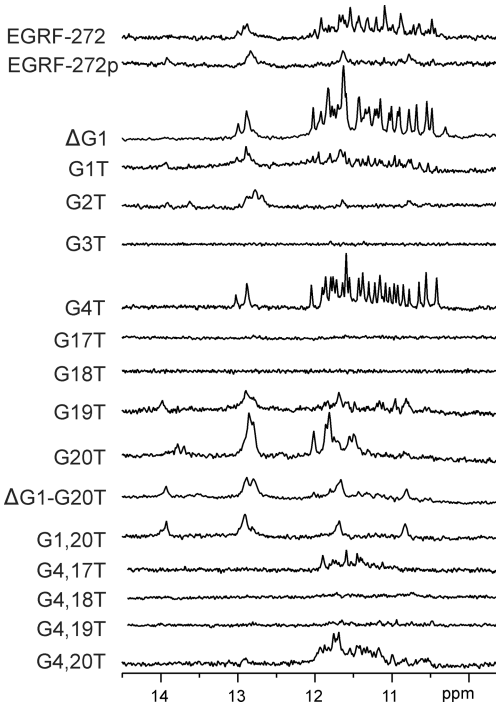

Figure 7. Imino region of ^{1}H NMR spectra of EGFR-272 and mutants. NMR spectra were recorded at 0.1 mM concentration of the oligonucleotide per strand, pH 7.0, 25°C on 800 MHz spectrometer. The concentration of K^{+} ions was 50 mM for G3T, G17T, G18T, ΔG1,G20T; 70 mM for ΔG1, G4T, G4,18T and 100 mM for all other samples. The sequences of mutants are presented in Table 1.

272p), the removal of G1 residue (ΔG1) and 14 oligonucleotides with one or two G→T mutations (Figure 7).

For ΔG1 and G4T mutants, we observed 24 narrow and resolved signals in ^{1}H NMR spectra between δ 10.30 and 12.05 ppm indicating formation of well-defined G4 forms (vide infra). In the case of G1T and G4,20T mutants, the high number of imino signals in ^{1}H NMR spectra demonstrated the presence of multiple G4 structures in the samples. These results suggested that the removal of G1 in ΔG1 was more favorable to stabilize two major forms in NMR samples in comparison to its mutation with T (G1T). Interestingly, the phosphate group added at the 5′-end in the EGFR-272p mutant greatly influenced the folding of the oligonucleotide as it can derive from the observed four broad signals in imino region between δ 10.70 and 14.00 ppm. A similar pattern of imino signals in ^{1}H NMR spectra was detected also for G2T, ΔG1-G20T and G1,20T mutants. This pattern of signals does not support formation of G4 structures. While changes in the first G-tracts resulted in resolved signals in NMR spectra of two different major G4 forms, modifications in the third tract did not lead to well-defined structures. In the case of G17T, G18T, G4,18T and G4,19T mutants, no or only minor signals were detected in the imino region of ^{1}H NMR spectra indicating that the modified guanine residues play a vital role in G-tetraplex folding and structural integrity. No imino signals were detected also for G3T mutant. For G19T and G20T we observed additional signal at around δ 13.8 ppm besides imino signals involved in Hoogsteen and Watson–Crick G-C base pairs. This signal could indicate an additional base paired thymine residue or the presence of the form observed also in the NMR samples of some other mutants (e. g. G1,20T).

Interestingly, by comparing double mutants (ΔG1-G20T and G1,20T) we observed that they share a very similar pattern of signals in the imino region of ^{1}H NMR spectra (Figure 7). On the other hand, for double mutants of residue G4 and all four guanine residues in the third G-tract we observed no signals in imino region for G4,18T and G4,19T. Only imino signals between δ 11.10 and 11.90 ppm were observed for G4,17T mutant. The imino region of the ^{1}H NMR spectrum of G4,20T mutant was more similar to EGFR-272 with signals between δ 10.43 and 12.00 ppm and a very broad signal at δ 12.90 ppm (Figure 7). It is noteworthy that similar spectral characteristics point to comparable structural elements, but not necessary to the same structure.

Due to the better peaks resolution provided by ΔG1 and G4T mutants, they were studied in more detail with the use of 2D NOESY spectra. In the case of both mutants, we could distinguish between two sets of signals based on intensity in the imino region between δ 10.30 and 12.05 ppm of 1D ^{1}H and 2D NOESY spectra (Figure 7, Supplementary Figures S7 and 8). Each set was comprised of 12 signals. The number and intensity of signals suggested the formation of two monomeric G4s each consisting of three G-quartets. The ratio between the two forms was ≈65:35 and ≈60:40 for ΔG1 and G4T, respectively and was not dependent on K^{+} ion concentration as shown by titration experiments (Supplementary Figure S9). For both mutants we detected at least three broad signals in the region between δ 12.80 and 13.00 ppm that are characteristics for imino protons involved in Watson–Crick GC base pairs. The aromatic-anomeric region of 2D NOESY NMR spectra revealed four very intense intermolecular NOE cross-peaks between H8 and H1′ protons that indicated a *syn* orientation of at least four guanine residues in the case of ΔG1 as well as G4T (Supplementary Figures S10 and 11). This is consistent with one of the G4 forms occupying a (3+1) hybrid strand orientation for each mutant. However, further investigation beyond the scope of this study is needed for more detailed high-resolution structural data.

To asses if the predicted two folded forms identified for ΔG1 and G4T are adopted also by the wt oligonucleotide, we referred to the basic spectra obtained by deconvolution of CD profile of EGFR-272. We found that also the experimentally acquired CD spectra of ΔG1, G1T and G4T are well described by proper combination of these two datasets (Supplementary Figure S12). Interestingly, the relative weights of the two forms cluster our oligonucleotides into two groups: the first one comprising of EGFR-272 and G1T in which Form 2 counts for ≈70% of the entire population, the second one groups ΔG1 and G4T in which, in agreement with NMR data, the ratio is reversed to about 60% of Form 1 (Figure 8).

The third loop contributes in defining the relative distribution among the two main G4 forms

The so far collected data confirmed that when EGFR-272 and its related mutants fold into G4 two forms are the most relevant. However no explanation was derived for the criti-

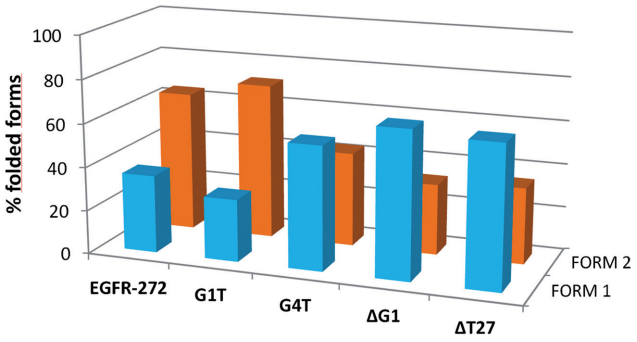

Figure 8. Prediction of the distribution of Form 1 and Form 2 in the folded population as derived from deconvolution of the experimental CD spectra of selected oligonucleotides based on the SVD-generated dataset reported in Figure 3A.

cal contribution of G20 to the folding efficiency. In particular, the massive reduction of G4 formation by G20T suggests that both forms might benefit from interactions with G20. It is worth to remind that according to NMR and footprinting data, G20 results inserted in a very long loop and any condition that freezes it in a fixed position is expected to be favorable for G4 stability. Actually, NMR of G4 folded systems always contains peaks attributable to GC pairs.

With this in mind, we observed that the second loop contains two cytosines that eventually can pair with G20 to form an extended cupping element and thus strengthening its contribution to G4 stability. Furthermore, also within the third loop two Watson–Crick GC pairs can occur between G20-C21 and G25-C26. This pairing is expected to support the formation of a stable hairpin within the loop thus decreasing its flexibility. To dissect among these two different models, we analyzed by CD spectroscopy an additional set of EGFR-272 mutants in which we selectively modified these two loops (Figure 9).

The mutation of the cytosines at position 12 and 13 (C12,13T) did not affect to a significant extent the G4 formation thus ruling out any relevant contribution from them to the G4 folding process.

On the opposite, the G25T-C26T double mutant showed a remarkably reduced tendency to fold. The CD spectrum showed that this mutation mostly influenced the parallel form, while the hybrid one was more preserved in comparison to EGFR-272. Unfavorable effects of G25 and C26 modifications on folding and structural integrity were evident also from ^{1}H NMR spectrum (Supplementary Figure S13). The hump together with many superimposed and overlapped imino signals between δ 10.70 and 12.12 ppm indicated formation of aggregates and several multimeric structures. Very broad imino signals characteristic for GC base pairs were observed at δ 12.83 ppm. They probably do not belong to the same base pairs as in EGFR-272 due to their very different fingerprint of imino signals in comparison to G25T-C26T mutant (Supplementary Figure S13). These results suggest that the third loop is organized in a hairpin-like structure stabilized by two GC base pairs. We believe that hairpin-like structures drive the G4 formation, since the decreased folding of the G25T-C26T mutant is most likely due to disruption of GC base pairs. It is noteworthy that pre-folded hairpin structures were observed in the absence of KCl as well (Supplementary Figure S1).

According to this model, it derives that in the wild-type EGFR-272, an unpaired T (T27) connects the G-tetrad and the steam of the loop (Figure 10). This thymine appears to be required to allow the orthogonal positioning of the loop double helical domain across a medium wide groove as the one present in an all parallel structure as the one corresponding to Form 2 of EGFR-272 (41).

Conversely, from a wide groove, as the one occurring between two antiparallel-oriented strands as expected in the hybrid Form 1, the hairpin can directly exit in a coaxial orientation with reference to the G4 core without loosing any base stacking and thus not impairing the overall stability of the system. Here, deletion of the connecting T27 on EGFR-272 (ΔT27) provided an oligonucleotide that preserves the ability to fold into G4. Additionally, it did not showed time-dependent structural rearrangements in agreement with the indication that the slow forming parallel Form 2 is not expected to favorably accommodate the hairpin steam when the oligonucleotide contains this constrain. Consistently, Form 1 dominates Form 2 in the folded fraction (Figure 8).

DISCUSSION

EGFR represents a highly valuable target for anticancer treatment but with high demanding need for novel strategies to suppress its activity. In this connection, the presence of G-rich domains within its gene promoter is an attractive starting point which prompted us to explore its potential conformational rearrangements. Our data confirmed that the sequence located at position −272 from TSS can fold into G4 in the presence of physiological concentration of K^+ ions. Interestingly, this sequence is not fully folded at lower metal ion concentrations thus suggesting a reliable structural plasticity within the cell which can be proficiently exploited to regulate gene expression by small ligands targeting.

Although it was not possible to solve a unique G4 structure, interesting peculiar structural features emerged. First of all, in our experimental conditions, two main folded forms were identified in solution and we succeeded to attribute them to a kinetically favored hybrid G4 corresponding to a (3+1) arrangement and to a slower forming parallel one, both comprising three stacked G-quartets. The easiest model would correspond to the full conversion of the first structure into the second one. However, this was not the case since we always found the two of them simultaneously present in solution. The resolution of the two forms was further hampered by their unexpected comparable affinities for the metal ion as well as by their similar thermal stabilities. Since EGFR-272 is quite rich in guanines, this led us to hypothesize the formation of distinct G4 structures deriving from the recruitment of different residues in the G-tetrads. In particular, our evidences indicated that among the four guanines at 5′-end, it exists the possibility of a shift of the residues involved in G4 pairing. Consistently, the recruitment of residues 1–3 and 2–4 to produce the two forms could be envisaged. Nevertheless, mutation studies elicited this hypothesis since both G1T and G4T containing only three guanines at 5′-end still contains a balanced contribu-

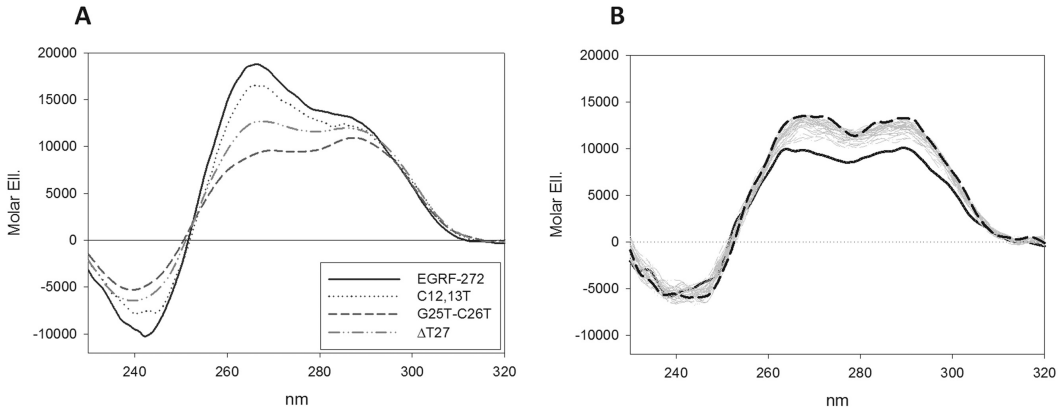

Figure 9. CD spectra of EGFR-272-related sequences containing mutations in the loops acquired in 10 mM Tris, 200 mM KCl, pH 7.5, 25°C (**A**). Variation of the CD spectra acquired in 10 mM Tris, pH 7.5, 25°C of 4 μM ΔT27 upon addition of 200 mM KCl as a function of incubation time (1 acquisition/120 s); dashed black line correspond to the end of the titration (**B**).

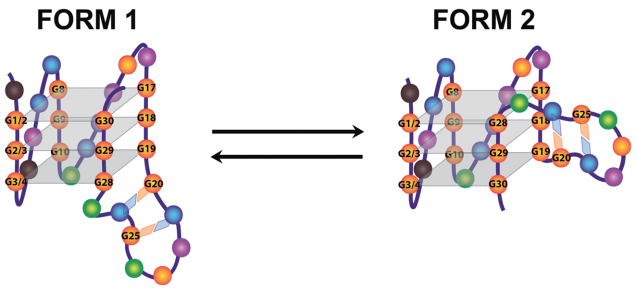

Figure 10. Schematic drawing of the main structural equilibrium occurring within EGFR-272. Guanines are shown in orange, adenines in magenta, cytosines in blue and thymines in green.

tion of the hybrid and parallel G4 folded forms. Only the absence of an unpaired nucleotide at 5′-modulates this ratio toward a preference for the hybrid form, as shown by the shared structures distribution between G4T and ΔG1.

By merging these data with the observation that G20 was not directly involved in G-tetrad formation, the emerging picture indicates that both G4 forms derive from pairing of the same guanines. This produces a general model in which three G-quartet stack one over the other and where the combination of 3/4, 6 and 8 nt long loops fit with the quite relevant concentration of KCl required to fully fold the sequence.

Unexpectedly, highly detrimental for G4 formation resulted the mutation of G20 located in the longest terminal loop which affect both the parallel and the hybrid form (G20T). This foresees a clear role of G20 as a stabilizing element. Insertion of proper mutations in the loops excluded the G20 pairing with a cytosine of the second loop. Conversely, it indicated that such a function is related to the formation of a hairpin within the third loop. It has been reported that G4 containing a long loop arranged in a hairpin can be joined to the duplex domain with different arrangement and energetic contribution according to the G4 geometry at the loop insertion points (41–43). By applying the proposed models to our sequence, we can propose that in the all parallel structure, the steam can be orthogonally inserted in the G4 groove. In our case, the energetic stabilization can be partly reduced by the presence of just one unpaired nucleotide at the junction (T27). Upon removal of this thymine (ΔT27), the stability of the overall system is not hampered. This suggests that the direct exit of the steam from a wide groove as the one occurring between the two antiparallel-oriented strands in a (3+1) structure. This corresponds also to the kinetically favored form which maximally benefits the preformed arrangement of the sequence.

Thus, in conclusion the unique base composition of the long-third loop of EGFR-272 is suitable to provide the energetic contributions that are required to describe the fascinating structural equilibria of EGFR-272.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We acknowledge M. Folini for support with bioinformatics analysis.

FUNDING

University of Padova [CPDA147272 to C.S.]; University of Padova PhD Fellowships (to C.C., RR); Cariparo PhD Fellowships (to M.L.G.); Slovenian Research Agency [P1–242, J1–6733]. Funding for open access charge: University of Padova [CPDA147272].
Conflict of interest statement. None declared.

REFERENCES

1. Normanno,N., De Luca,A., Bianco,C., Strizzi,L., Mancino,M., Maiello,M.R., Carotenuto,A., De Feo,G., Caponigro,F. and Salomon,D.S. (2006) Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene*, **366**, 2–16.
2. Lemmon,M.A., Schlessinger,J. and Ferguson,K.M. (2014) The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harb. Perspect. Biol.*, **6**, a020768.
3. Holowka,D. and Baird,B. (2017) Mechanisms of epidermal growth factor receptor signaling as characterized by patterned ligand activation and mutational analysis. *Biochim. Biophysica Acta*, **1859**, 1430–1435.

4. Goffin,J.R. and Zbuk,K. (2013) Epidermal growth factor receptor: pathway, therapies, and pipeline. *Clin. Therapeutic.*, **35**, 1282–1303.
5. Hata,A.N., Niederst,M.J., Archibald,H.L., Gomez-Caraballo,M., Siddiqui,F.M., Mulvey,H.E., Maruvka,Y.E., Ji,F., Bhang,H.E., Krishnamurthy Radhakrishna,V. *et al.* (2016) Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat. Med.*, **22**, 262–269.
6. Morgillo,F., Della Corte,C.M., Fasano,M. and Ciardiello,F. (2016) Mechanisms of resistance to EGFR-targeted drugs: lung cancer. *ESMO Open*, **1**, e000060.
7. Collie,G.W. and Parkinson,G.N. (2011) The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.*, **40**, 5867–5892.
8. Wong,H.M., Stegle,O., Rodgers,S. and Huppert,J.L. (2010) A toolbox for predicting g-quadruplex formation and stability. *J. Nucleic Acids*, **2010**, 564946–564951.
9. Hansel-Hertsch,R., Di Antonio,M. and Balasubramanian,S. (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.*, **18**, 279–284.
10. Čeru,S., Šket,P., Prislan,I., Lah,J. and Plavec,J. (2014) A new pathway of DNA G-quadruplex formation. *Angew. Chem. Int. Ed.*, **53**, 4881–4884.
11. Marušič,M. and Plavec,J. (2015) The effect of DNA sequence directionality on G-quadruplex folding. *Angew. Chem. Int. Ed.*, **54**, 11716–11719.
12. Dolinnaya,N.G., Ogloblina,A.M. and Yakubovskaya,M.G. (2016) Structure, properties, and biological relevance of the DNA and RNA G-quadruplexes: overview 50 years after their discovery. *Biochemistry (Mosc.)*, **81**, 1602–1649.
13. Brčić,J. and Plavec,J. (2017) ALS and FTD linked GGGGCC-repeat containing DNA oligonucleotide folds into two distinct G-quadruplexes. *Biochim. Biophys. Acta*, **1861**, 1237–1245.
14. Marušič,M., Hošnjak,L., Krafčikova,P., Poljak,M., Viglasky,V. and Plavec,J. (2017) The effect of single nucleotide polymorphisms in G-rich regions of high-risk human papillomaviruses on structural diversity of DNA. *Biochim. Biophys. Acta*, **1861**, 1229–1236.
15. Trajkovski,M., Webba da Silva,M. and Plavec,J. (2012) Unique structural features of interconverting monomeric and dimeric G-quadruplexes adopted by a sequence from the intron of the N-myc gene. *J. Am. Chem. Soc.*, **134**, 4132–4141.
16. Moye,A.L., Porter,K.C., Cohen,S.B., Phan,T., Zyner,K.G., Sasaki,N.i., Lovrecz,G.O., Beck,J.L. and Bryan,T.M. (2015) Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.*, **6**, 7643–7655.
17. Ray,S., Bandaria,J.N., Qureshi,M.H., Yildiz,A. and Balci,H. (2014) G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 2990–2995.
18. Fujii,T., Podbevšek,P., Plavec,J. and Sugimoto,N. (2017) Effects of metal ions and cosolutes on G-quadruplex topology. *J. Inorg. Biochem.*, **166**, 190–198.
19. Miyoshi,D., Nakao,A. and Sugimoto,N. (2003) Structural transition from antiparallel to parallel G-quadruplex of d(G4T4G4) induced by Ca2+. *Nucleic Acids Res.*, **31**, 1156–1163.
20. Marušič,M., Šket,P., Bauer,L., Viglasky,V. and Plavec,J. (2012) Solution-state structure of an intramolecular G-quadruplex with propeller, diagonal and edgewise loops. *Nucleic Acids Res.*, **40**, 6946–6956.
21. Gajarský,M., Živković,M.L., Stadlbauer,P., Pagano,B., Fiala,R., Amato,J., Tomáška,L., Šponer,J., Plavec,J. and Trantírek,L. (2017) Structure of a Stable G-Hairpin. *J. Am. Chem. Soc.*, **139**, 3591–3594.
22. Kocman,V. and Plavec,J. (2014) A tetrahelical DNA fold adopted by tandem repeats of alternating GGG and GCG tracts. *Nat. Commun.*, **5**, 5831–5842.
23. Kocman,V. and Plavec,J. (2017) Tetrahelical structural family adopted by AGCGA-rich regulatory DNA regions. *Nat. Commun.*, **8**, 15355–15370.
24. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
25. Rhodes,D. and Giraldo,R. (1995) Telomere structure and function. *Curr. Opin. Struct. Biol.*, **5**, 311–322.
26. Laguerre,A., Hukezalie,K., Winckler,P., Katranji,F., Chanteloup,G., Pirrotta,M., Perrier-Cornet,J.M., Wong,J.M. and Monchaud,D. (2015) Visualization of RNA-quadruplexes in live cells. *J. Am. Chem. Soc.*, **137**, 8521–8525.
27. Henderson,A., Wu,Y., Huang,Y.C., Chavez,E.A., Platt,J., Johnson,F.B., Brosh,R.M. Jr, Sen,D. and Lansdorp,P.M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.*, **42**, 860–869.
28. Shivalingam,A., Izquierdo,M.A., Marois,A.L., Vysniauskas,A., Suhling,K., Kuimova,M.K. and Vilar,R. (2015) The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat. Commun.*, **6**, 9178–9188.
29. Kotar,A., Wang,B., Shivalingam,A., Gonzalez-Garcia,J., Vilar,R. and Plavec,J. (2016) NMR structure of a triangulenium-based long-lived fluorescence probe bound to a G-quadruplex. *Angew. Chem. Int. Ed. Engl.*, **55**, 12508–12511.
30. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
31. Balasubramanian,S., Hurley,L.H. and Neidle,S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.*, **10**, 261–275.
32. Grand,C.L., Han,H., Munoz,R.M., Weitman,S., Von Hoff,D.D., Hurley,L.H. and Bearss,D.J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol. Cancer Therapeut.*, **1**, 565–573.
33. Howell,R.M., Woodford,K.J., Weitzmann,M.N. and Usdin,K. (1996) The chicken beta-globin gene promoter forms a novel 'cinched' tetrahelical structure. *J. Biol. Chem.*, **271**, 5208–5214.
34. Rigo,R., Palumbo,M. and Sissi,C. (2017) G-quadruplexes in human promoters: a challenge for therapeutic applications. *Biochim. Biophys. Acta*, **1861**, 1399–1413.
35. Mergny,J.L., Li,J., Lacroix,L., Amrane,S. and Chaires,J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.
36. DeSa,R.J. and Matheson,I.B. (2004) A practical approach to interpretation of singular value decomposition results, *Methods Enzymol.*, **384**, 1–8.
37. Hendler,R.W. and Shrager,R.I. (1994) Deconvolutions based on singular value decomposition and the pseudoinverse: a guide for beginners, *J. Biochem. Biophys. Methods*, **28**, 1–33.
38. Webba da Silva,M. (2007) NMR methods for studying quadruplex nucleic acids. *Methods*, **43**, 264–277.
39. Adrian,M., Heddi,B. and Phan,A.T. (2012) NMR spectroscopy of G-quadruplexes. *Methods*, **57**, 11–24.
40. Sun,D. and Hurley,L.H. (2010) Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. *Methods Mol. Biol.*, **608**, 65–79.
41. Lim,K.W. and Phan,A.T. (2013) Structural basis of DNA quadruplex–duplex junction formation. *Angew. Chem. Int. Ed.*, **52**, 8566–8569.
42. Lim,K.W., Jenjaroenpun,P., Low,Z.J., Khong,Z.J., Ng,Y.S., Kuznetsov,V.A. and Phan,A.T. (2015) Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study. *Nucleic Acids Res.*, **43**, 5630–5646.
43. Onel,B., Carver,M., Wu,G., Timonina,D., Kalarn,S., Larriva,M. and Yang,D. (2016) A new G-quadruplex with hairpin loop immediately upstream of the human BCL2 P1 promoter modulates transcription. *J. Am. Chem. Soc.*, **138**, 2563–2570.