Coarse-Grained Simulations Complemented by Atomistic Molecular Dynamics Provide New Insights into Folding and Unfolding of Human Telomeric G-Quadruplexes

Petr Stadlbauer, Liuba Mazzanti, Tristan Cragnolini, David J. Wales, Philippe Derreumaux, Samuela Pasquali, Jirí Sponer

JCTC Journal of Chemical Theory and Computation

Article

Coarse-Grained Simulations Complemented by Atomistic Molecular Dynamics Provide New Insights into Folding of Human Telomeric G-Quadruplexes

Petr Stadlbauer, Liuba Mazzanti, Tristan Cragnolini, David J. Wales, Philippe Derreumaux, Samuela Pasquali, and Jiri Sponer

J. Chem. Theory Comput., **Just Accepted Manuscript** • DOI: 10.1021/acs.jctc.6b00667 • Publication Date (Web): 21 Oct 2016

ACS Publications

Journal of Chemical Theory and Computation is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Coarse-Grained Simulations Complemented by Atomistic Molecular Dynamics Provide New Insights into Folding and Unfolding of Human Telomeric G-Quadruplexes

Petr Stadlbauer[1,2], Liuba Mazzanti[3], Tristan Cragnolini[4], David J. Wales[4], Philippe Derreumaux[3], Samuela Pasquali[3,] and Jiří Šponer[1,*]*

[1] Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

[2] Regional Centre of Advanced Technologies and Materials, Departments of Physical Chemistry, Faculty of Science, Palacký University, 17. listopadu 1192/12, 771 46 Olomouc, Czech Republic

[3] Laboratoire de Biochimie Théorique, IBPC, CNRS UPR9080, Université Sorbonne Paris Cite, Paris Diderot, 13 rue Pierre et Marie Curie, 75005 Paris, France

[4] Department of Chemistry, Cambridge University, Lensfield Road, Cambridge, CB2, 1EW, UK

ABSTRACT: G-quadruplexes are the most important non-canonical DNA architectures. Many quadruplex-forming sequences, including the human telomeric sequence d(GGGTTA)$_n$, have been investigated due to their implications in cancer and other diseases, and because of their potential in DNA-based nanotechnology. Despite availability of atomistic structural studies of folded G-quadruplexes, their folding pathways remain mysterious, and mutually contradicting models of folding coexist in the literature. Recent experiments convincingly demonstrated that G-quadruplex folding often takes days to reach the thermodynamics equilibrium. Based on atomistic simulations of diverse classes of intermediates of the G-quadruplex folding, we have suggested that the folding is an extremely multi-pathway process combining kinetic partitioning mechanism with conformational diffusion. However, complete G-quadruplex folding is far beyond the time scale of atomistic simulations. Here we for the first time use high-resolution coarse-grained simulations to investigate potential unfolding intermediates, whose structural dynamics are then further explored with all-atom simulations. This multiscale approach indicates how various pathways are interconnected in a complex network. For the first time, events of spontaneous conversions between different folds are observed. We demonstrate the inability of simple order parameters, such as radius of gyration or the number of native H-bonds, to describe the folding landscape of the G-quadruplexes. The study also provides information relevant to further development of the coarse-grained force field.

INTRODUCTION

A guanine quadruplex (GQ) is a non-canonical DNA or RNA structure, which can be formed by nucleotide sequences rich in guanine. Awareness of the importance of such structures has been increasing, mainly because GQ-forming sequences appear to be widespread in biological systems, and they are attractive for nanotechnology. In cells, potential GQ-forming sequences are found in gene expression regulatory regions, such as gene promoters or mRNA, and at chromosome ends in telomeres.[1-9] Human telomeres are formed from the repeat sequence d[GGGTTA].d[CCCTAA]$_n$, and terminate with up to 250 nts in single-stranded G-rich overhangs,[10-11] which can fold back on themselves and form GQs.[12-13] The length of telomeres is shortened during every DNA replication until a critical length is reached and apoptosis is

triggered.[14-16] Telomerase is an enzyme known to prolong telomeres, thus maintaining their length,[17] with overexpression in the majority of immortal cancer cells.[18] Since GQs may downregulate telomerase activity,[19] their stabilisation *in vivo* might help in cancer treatment.[18, 20] Understanding thermodynamics and kinetics of GQ folding is a cornerstone not only for drug targeting, but also for designing artificial GQs with predefined properties in other fields of science.

GQs are composed of stacked quartets of Gs (Figure 1). The quartet is a roughly planar assembly of four cyclically *cis*-Watson-Crick-Hoogsteen (*c*WH)[21] paired Gs. Four oxygens point towards the centre, creating a pocket of negative electrostatic potential. Upon stacking of the quartets, their centres form a GQ channel, which binds cations. The presence of cations inside the channel is essential for GQ stability. Mutual orientation of G-stretches in GQs can be either parallel or antiparallel and connecting loops can be lateral (edgewise), diagonal or propeller (double-chain reversal) (Figure 1). The mutual antiparallel orientation of two G-stretches requires, in any specific GG base-pair, one G with the glycosidic torsion angle χ in *anti* orientation and the other in *syn* orientation. Parallel strand orientation needs the paired Gs in the same orientation, either both *syn*, or both *anti*. If two consecutive Gs of a G-strand involved in two consecutive quartets adopt the same χ orientation, then the quartets are of the same directionality (either clockwise or counter-clockwise). If one G is *anti* and the other is *syn* oriented, the quartets are of opposite directionality.[22-23] These rules are an important aspect of GQ folding principles.

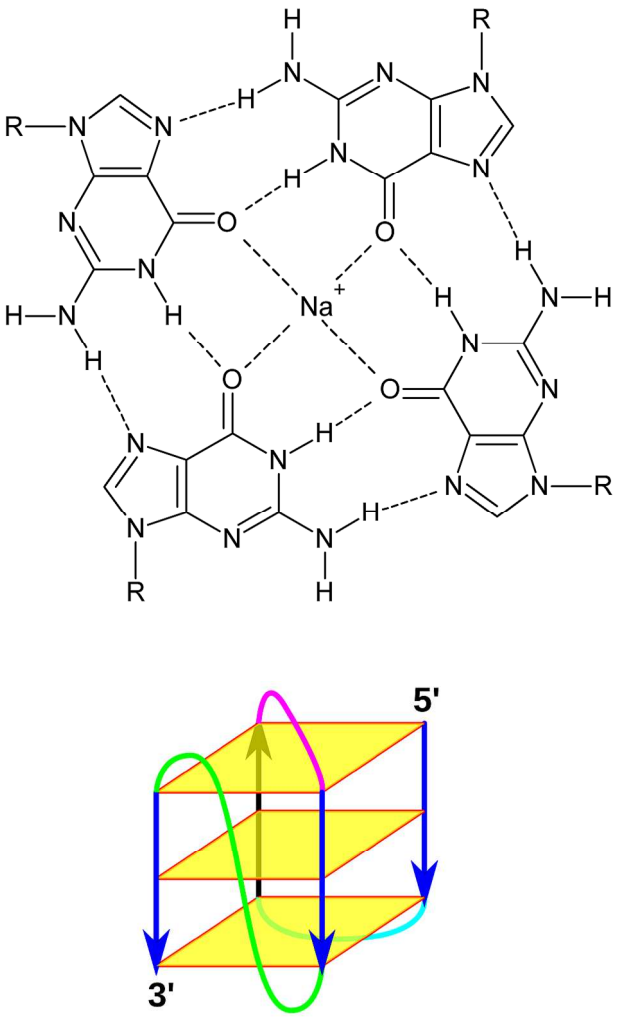


Figure 1. GQ structure. Top: A G-quartet with a sodium cation coordinated in its centre (it can either be coplanar with the quartet or reside between two consecutive quartets; potassium cation can only reside between two quartets). R stands for the sugar-phosphate moiety. Bottom: Schematic intramolecular three-quartet GQ. Quartets are simplified as yellow squares. Blue G-strands are mutually parallel, while the black one is antiparallel to them. The cyan loop is lateral, the mauve loop is diagonal. They connect antiparallel G-stretches. The green loop is propeller and it connects parallel G-stretches. Ions are not shown in the scheme. Although not shown in the Figure, a given GQ topology is interconnected with characteristic patterns of *anti* and *syn* orientations of the G-nucleotides.[23]

Human telomeric GQs are known for their structural polymorphism. Six different topologies have been elucidated in thermodynamic equilibrium so far.[24-32] The dominant conformation depends on many factors, among them the sequence of flanking residues at the 5′ and 3′ ends of the oligonucleotide, salt concentration and cation type, DNA strand concentration, or presence of cosolvents.[24-40] Coexistence of multiple species in equilibrium is common.[25, 28, 40-44]

There have been numerous studies of the GQ folding and unfolding processes. In contrast to earlier experiments suggesting fast millisecond GQ folding rates, many recent investigations have revealed that the folding takes hours to days, depending on the exact conditions of the experiment, and thus the dominant equilibrium structure.[45-52] Nevertheless, the principles of GQ

folding emerging from the available experimental literature are in many aspects ambiguous. Different studies suggest different time scales of the process as well as different types of intermediates. The ambiguity may reflect real variability in the GQ folding processes depending on the exact sequence and experimental conditions. In addition, since many experimental methods have limited structural and temporal resolution, the interpretation of the primary experimental data might be sometimes over-simplified.[39, 47, 53-54] Measurement times in some GQ folding experiments (such as many smFRET studies) are orders of magnitude shorter than the folding kinetics suggested by some other studied. It had been often anticipated that GQ folding happens from the unfolded state via a small number of distinct intermediates directly to the folded state in a simple funnel-like manner.[46, 55] Triplex and hairpin intermediates have been suggested, with some experimental support.[46, 55-60] In contrast, other works have argued that the folding likely follows a kinetic partition mechanism,[45, 48, 53] with competing and kinetically unrelated long-living intermediates. This is consistent with the long folding time scales and common coexistence of multiple GQ species in solution.[40, 43, 45, 48, 50-51, 61] Although involvement of G-hairpins and G-triplexes is likely, the most prominent competing substates primarily causing the extremely slow kinetics could rather be alternative four-stranded structures, i.e., alternatively folded GQs.[45, 48, 53] They may include four-stranded structures known from the atomistic experiments as well as other alternative folds that have not yet been experimentally detected.[51, 62] Four-stranded intermediates should have long lifetimes and are best-suited to serve as the dominant kinetic traps, explaining the long time scale of many folding experiments.[48, 51, 53] Assuming a simple funnel-like mechanism is contradictory to long folding timescale, as a funnel assumes fast unobstructed folding through a smooth folding landscape directly to one native state ensemble. An important factor, which might lead to apparently contradictory results for different experimental techniques, is that they work with different unfolded ensembles.[63] Chemically denatured, thermally denatured or force-denatured ensembles may explore very different regions of the free energy folding landscapes.[63] We have argued recently that an idealised (entirely unstructured single strand) unfolded ensemble of the human telomeric GQ sequence would populate all 4096 different combinations of *syn–anti* G-patterns, which can then initiate folding into a countless number of intermediate structures.[53] A fast folding process to one native fold would likely require that this broad *syn–anti* distribution is somehow narrowed already before the onset of the measurement in favour of the native *syn–anti* pattern.

To complement and interpret the experimental data, computational effort has been applied to study various aspects of GQ folding. Fully-atomistic molecular dynamics (MD) simulations are often a useful tool since they offer a good compromise between computational speed and accuracy. However, they are limited to the microsecond or perhaps millisecond time scale,[64] which is not sufficient when compared to the folding times of GQs. Thus, atomistic MD could only address faster stages of folding rather than the whole process. Full GQs,[65-69] hypothetical four-stranded intermediates with strand slippage,[62, 70] G-triplexes[46, 55, 71-72] and G-hairpins[53, 55, 73] have been investigated by MD simulations. The simulations revealed a kinetic obstacle posed by necessity to achieve a correct mixture of *syn* and *anti* Gs,[62] and hitherto unresolved difficulties in the formation of the propeller loops.[53] Enhanced sampling atomistic MD methods have also been used,[53, 71, 74-78] because they can achieve greater sampling of the conformational space. However, these methods have their own drawbacks, which are connected to the particular choice of sampling acceleration and which are not always sufficiently discussed. Thus the results should be interpreted with caution.[79-80] All methods using collective variables (CVs) such as metadynamics accelerate sampling of movements associated with the CVs.[80] The CVs represent low-

dimensional projections of the full coordinate space and if the real folding processes include slow motions not included in the CV space, the results may become inadequate. Thus, CV-based methods cannot sample transitions between diverse folds of GQs contributing to the kinetic partitioning and also adequate sampling of the unfolded state is challenging. Insufficient sampling of the unfolded state can cause spurious stabilisation of the initial folded state in the simulation. Simulations tearing GQs apart by pulling force inherently sample low-entropy ensembles; in addition the pulling is much faster than in the experiments. Thermal-acceleration methods (e.g. T-REMD) do not require CV-based description of the process but work with an ensemble denatured at high temperatures, which might differ from the experimental conditions and, in addition, the kinetic information is lost. Further, to depict folding, such simulation would have to be strictly initiated from the unfolded state and would be extremely time-consuming. This has so far been achieved only for G-hairpins and still not with a quantitative convergence.[53]

At first sight, atomistic simulations represent an ideal computational method to study the folding. However, closer inspection reveals that this approach has considerable limitations. First, the cost of atomistic simulations (even with enhanced sampling) is enormous, preventing investigations of the full biomolecular folding except for fast-folding biopolymers with evolutionary minimised frustration of the free energy folding landscape with simple funnel-like properties (e.g. some small proteins).[81] This is definitely not the case of GQs,[53] where formation of G-quartets, i.e. non-local contacts in the short sequences, partitions the folding landscape. GQs have extremely slow kinetics of folding compared to the number of residues. Although the atomistic force fields aim to mimic physical description of the studied molecules, they are based on fundamental and unphysical approximations such as lack of polarisation, atom-centred point charges, etc.[82] Despite continuous efforts to improve the atomistic force fields for nucleic acids it appears that their improvements have been reaching some plateau in the past years.[79, 83] Many important physical terms, significantly contributing to the total energy, are principally non-includable.[84-85] This may question our capability to accurately describe atomistic GQ folding even with unlimited computer resources. In a foreseeable future, the atomistic simulations will always be too slow and the atomistic force fields always insufficiently accurate. This justifies searching for alternative options.

The issue of slow sampling, limiting efficient exploration of GQ folding landscape, can be overcome by the means of coarse-graining. Coarse-grained (CG) MD simulations simplify the atomistic description of DNA by reducing the number of particles of the solute to a few beads per nucleotide.[86] In addition, CG models are usually parametrised for implicit solvent environment, which dramatically speeds up the computations (note that implicit solvent is unsuitable for atomistic simulations of GQs with standard force fields[87]). At first sight, the CG description is even less physical than the atomistic force field. However, the CG methods can be specifically tuned to include the key aspects to the folding process while disregarding less important details. Moreover, if a CG approach correctly captures the balance between important competing free energy basins (i.e. various GQ topologies and possibly other long-living intermediates), the description of fine details of transient regions may be less important. Indeed, striking insights and concepts have emerged from previous CG studies of protein and RNA folding.[81, 88]

Several CG nucleic acid models have been proposed in the past few years for various purposes, from studying the large-scale properties of DNA nanostructures to addressing RNA folding.[88-95] Some models have been used specifically to simulate whole GQs and their assemblies[96] and even

full GQ folding.[97] Nevertheless, the folding was achieved by driving the molecule from the unfolded state to the folded state, not through a spontaneous process.

In CG models, the significant speed-up is balanced by a more or less severe loss of accuracy of molecular interactions depending on the details preserved in the CG model. Among the possible CG models, the adaptation of HiRE-RNA version 3 to DNA (hereafter designated simply as HiRE-DNA)[98] employed in the present work is particularly suited for the study of GQs because of the relatively high detail retained in the description of the bases, allowing us to properly describe the non-canonical base-pairs formed in quartets and therefore capture the fundamental interactions in GQs.[99]

Here, we combine the advantages of CG simulations and standard atomistic simulations. A set of fixed-temperature and simulated tempering (ST) CG simulations are employed to unfold various human telomeric GQs to obtain physically plausible partially unfolded configurations. These structures are then subjected to standard fully-atomistic simulations for refinement and further exploration of the free energy basins associated with them. Similar strategy has been used earlier to investigate partially unfolded GQ structures obtained either by applying no-salt denaturing conditions,[62, 100] or using specific biases constructed via bias-exchange metadynamics.[75] We show that the CG approach can be used to efficiently initiate unbiased unfolding of the GQs from the folded state and also to achieve transitions of the DNA strand into another GQ topology with different strand orientation and *anti-syn* pattern. Such transitions have never been observed in the preceding computational studies on the human telomeric GQ.[75, 100] Albeit the present model remains far from being able to describe the full converged and unbiased GQ folding from the unfolded state, our work proves that CG methods can in the near future become a powerful tool in the portfolio of methods to search through folding landscapes of the GQs.

METHODS

G-quadruplex systems. Four human telomeric GQs and one mutated human telomeric GQ were considered (Figure 2): 1) three-quartet parallel-stranded d[A(GGGTTA)$_3$GGG] (PDB ID: 1KF1),[27] 2) three-quartet antiparallel basket-type d[A(GGGTTA)$_3$GGG] (PDB ID: 143D, first frame),[26] 3) three-quartet hybrid (3+1) type-1 d[TA(GGGTTA)$_3$GGG] (PDB ID: 2JSM, first frame),[32] 4) two-quartet antiparallel basket-type d[(GGGTTA)$_3$GGGT] (PDB ID: 2KF8, first frame),[24] and 5) antiparallel chair-type with mutated sequence d[A(GGGCTA)$_3$GGG] (PDB ID: 2KM3, first frame).[101]

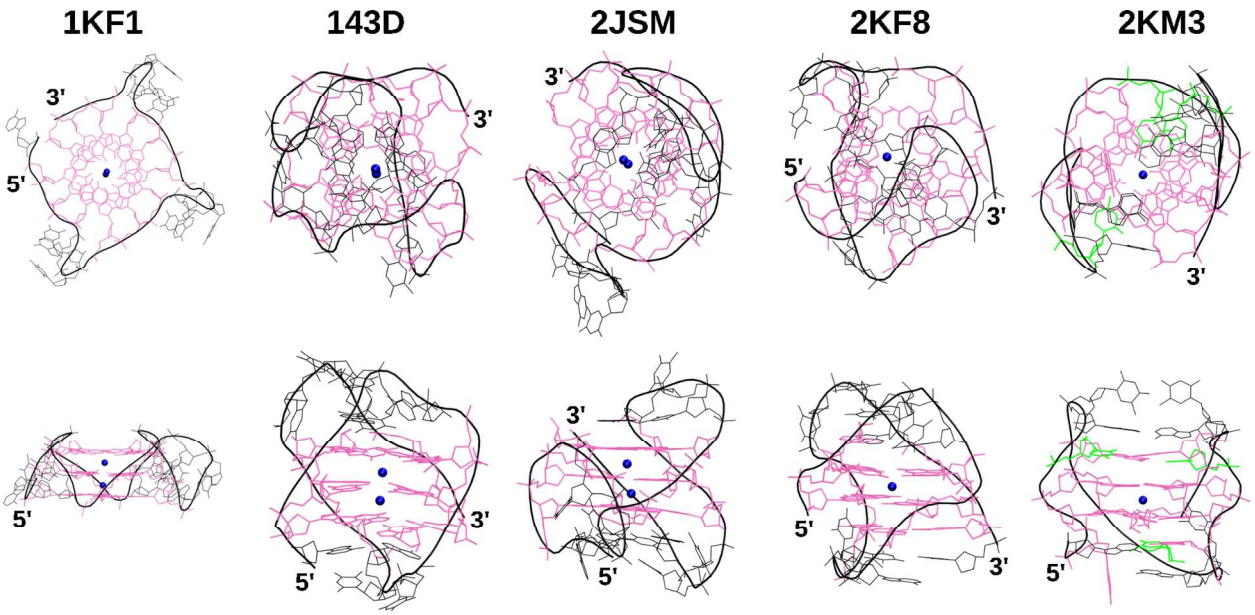


Figure 2. Top view (top) and side view (bottom) of the studied GQs. Gs are in mauve, loop residues and the backbone in black, and channel ions in blue. Cs in the case of 2KM3 are depicted in green.

Coarse-grained model. GQs are very stable structures at physiological conditions, when structural ions are present, and would not spontaneously unfold on the time scales accessible to atomistic simulations.[40, 62, 65, 100, 102] To generate partially unfolded states we used a CG representation of the system coupled to enhanced sampling techniques. This framework allows us to explore the conformational space more fully, including partially unfolded and completely unfolded states, while respecting a physical behaviour of the molecule.

The simulations were performed using the HiRE-DNA model and force field, having adapted version 3 of HiRE-RNA model for the study of DNA molecules.[98-99] HiRE-RNA has been presented in all detail in ref.[99] and in this section we are going to give a summary of its main features and the modifications that made it suitable for the study of single stranded DNA. Each nucleotide is represented by 6 or 7 beads, corresponding to the backbone heavy atoms P, O5′, C5′, C4′, and C1′, and to the centre of mass of each of the aromatic rings of the bases (G1, G2, A1, A2, C1, U1 or T1) (Figure 3).

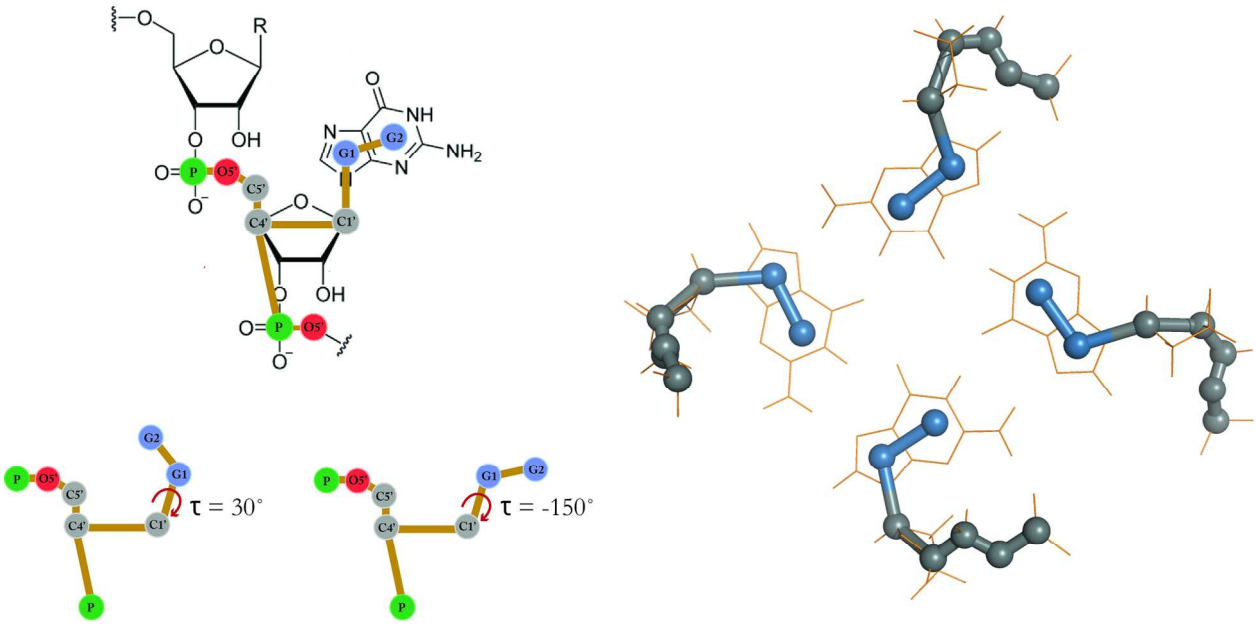


Figure 3. CG representation of a guanine nucleotide and a G-quartet. Top left: CG guanine nucleotide superposed to the atomistic structure. Bottom left: CG representation of *syn* (τ = 30°) and *anti* (τ = -150°) G-nucleotide conformations with τ equilibrium values in the dihedral potential (τ corresponds to the χ glycosidic torsion angle). Right: Atomistic and CG view of a G-quartet.

The force field of the model is composed of local interactions accounting for the local stereochemistry, an excluded volume interaction giving a physical size to the beads, and non-local interactions accounting for base pairings, base stacking and electrostatics. Each interaction is given by an energy potential (and related forces) weighted by a constant ε, allowing to gauge the relative contribution of each term. Equilibrium values for the variables considered by the potentials were extracted by a statistical analysis of 200 PDB structures, including a wide range of different folds. Optimisation was carried out for the energetic constants in order to train the model to recognise correctly folded structures over unfolded, incorrectly folded or partially unfolded structures, using a training set of 16 RNA structures and 20 decoys each.[99]

Our representation corresponds to a reduction of 70% in number of particles compared to an atomistic description without hydrogens. The molecule's volume is preserved by the choice of the size of the beads implemented through a repulsive potential ensuring a minimal distance of approach of the beads corresponding to atomistic distances. The excluded volume potential is given by Eq. (1)

$$E_{ex}(d) = e^{-\kappa(d-d_0)} \quad (1)$$

with $d = |\vec{r}_{ij}| = |\vec{x}_i - \vec{x}_j|$ and d_0 a reference distance that also controls the coupling strength. Parameters are chosen such that the potential increases rapidly when the two particles approach an inter-penetration distance.

As in most force fields, local interactions are described as the sum of harmonic potentials for distances of two successive particles E_b, for angles of three connected particles E_a, and sinusoidal potentials for torsions of four connected particles E_d. Geometric parameters controlling bond distances, equilibrium angles and dihedrals, are specific for each type of nucleotide (A, G, C, U) given by Eq. (2)

$$E_b(d) = \epsilon_b k_b (d - d_0)^2 \quad (2)$$

with d the instantaneous beads distance, d_0 the equilibrium bond length, k_b the coupling constant, specific for the pair in question, and ϵ_b the optimised weight coefficient. Similarly, the potential for the bond angle θ is expressed by Eq. (3)

$$E_a(\theta) = \epsilon_a k_a (\theta - \theta_0)^2 \quad (3)$$

with θ_0 the equilibrium angle, k_a the coupling constant, and ϵ_a the weight coefficient.

The torsion potential is given by Eq. (4) in terms of the dihedral angle φ as

$$E_d(\varphi) = \epsilon_d k_d [1 + \cos(\varphi - \varphi_0)] \quad (4)$$

with k_d the coupling constant, ϵ_d the weight coefficient, φ_0 the equilibrium angle.

In physiological conditions, the phosphate groups of nucleic acids are negatively charged. We therefore associate a fix -1 charge at the bead representing the phosphate group. In the absence of explicit ions and water molecules, phosphate beads interact with one another through a Debye-Hückel potential expressed by Eq. (5)

$$E_{DH}(d) = \epsilon_{el} \frac{1}{4\pi\varepsilon_0\varepsilon_r} \frac{e^{-d/\lambda_D}}{d} \quad (5)$$

with $d = |\vec{r}_{ij}| = |\vec{r}_i - \vec{r}_j|$ the distance between the charged particles. This form of the electrostatic potential accounts only for ionic screening, and not for tightly-bound ions occupying specific positions in the molecule. Even though this constitutes a severe limitation for molecules such as G-quadruplexes, where quartets are stabilised by cations, it is worth mentioning that a proper description for ions in tight contact with the molecule is challenging for all models not giving a quantum description of the systems, including commonly used atomistic force fields.

A specific feature of HiRE-RNA is to give a detailed description of base-pairing and stacking by defining planes for the bases. Energies and forces depend on the relative orientations and positions of the planes. Given the analytical complexity of these two potentials and the fact that for the work we present here no modifications were introduced to their previous form, we refer the reader to the original publication in which they are explained in depth.[99] Here we present only a schematic overview focusing on the behaviour they induce on the bases relevant for the discussion of G-quadruplexes.

In our description stacking occurs if two nearby bases are vertically aligned and their planes have a parallel orientation. The stacking potential is the product of three terms enforcing the relative positions of two stacked bases (short-ranged Gaussian function), orientation of the plane of the bases (quadratic sinusoidal function), and bases vertical alignment (higher-order sinusoidal function). The effect of the three terms is illustrated in Supporting Figure S1. Equilibrium values for each term were extracted from the analysis of PDB structures averaging over all stacked bases, independently of their species. At this stage, the stacking potential has no sequence specificity.

The base-pairing potential is the product of two terms, one ensuring the correct geometry adopted by the bases in terms of distance of the interacting beads (short-ranged Gaussian function) and angles formed with the rest of the base i.e. α: $G1_i$-$G2_i$-$G2_j$ (higher-order sinusoidal function), and the other ensuring that the bases forming the pair are on the same plane (short-ranged Gaussian function) (Supporting Figure S2). HiRE-RNA currently accounts for 22 different possible base-pairs minima, taken as the most frequent base-pairs found in the NDB for each possible base-type combination (Supporting Figure S3). In agreement with the Leontis base-pairing classification,[21] in HiRE-RNA bases can form pairs on one of their three sides (Watson-Crick, Hoogsteen and Sugar edge), making it possible to form multiple base-pairs such

as triplets and quartets. This property of HiRE-RNA makes it suitable for the study of G-quartets where guanine bases interact simultaneously with two other G bases, forming hydrogen bonds on their Watson-Crick and Hoogsteen sides.

The coarse-grained HiRE-DNA model. The CG model described so far was developed to study folding of RNA molecules, but a version for DNA was already introduced in ref.[98], where we studied the assembly of DNA and RNA double helices. The backbone equilibrium parameters for DNA were chosen to be those of a B-form helix, while RNA molecules are modelled with equilibrium parameters for A-form helices. Replica exchange MD simulations were able to predict the duplex corresponding to the experimental structure as the most stable configuration and correctly reproduced specific heat curves and melting temperatures (T_m). Moreover, we showed that atomistic simulations started from conformations extracted as low energy states of the CG trajectory behaved indistinguishably from simulations originated from experimental structures.

The other significant difference introduced for the DNA is the presence of both *anti* and *syn* states for purines, which are modelled only as *anti* for RNA. In the case of G-quartet it is particularly important to make the distinction between *anti* and *syn* conformations of purines, leading to distinct organisation of the overall structure. In HiRE-DNA, dihedral parameters were modified from previous versions, where one equilibrium value was attributed to each torsional angle, to be able to make the distinction between *anti* and *syn* conformations of bases A and G. *Anti* and *syn* conformations are distinguished by the equilibrium value of τ for the dihedral C4-C1-G1-G2 and C4-C1-A1-A2, with a value $\tau_{anti} = 150°$ for the *anti* conformation and $\tau_{syn} = -30°$ for *syn* (Figure 3). The sinusoidal form of the potential attributes the same energy to the two possible configurations, separated by an energy barrier of ϵ_d.

Validity and limitations of the CG model. The high-resolution CG HiRE-RNA model was developed with the specific aim of studying nucleic acids folding and assembly, with emphasis on non-canonical (three-dimensional) structures. This is a different goal from many other CG approaches often aiming at prediction of secondary structures. Given that HiRE-RNA was trained to recognise folded structures and distinguish them energetically from partially unfolded structures, it is a good starting point to study also partially folded and misfolded states. It has been extensively tested on several benchmark molecules including hairpins, pseudoknots of various kinds, and double helices. For molecules up to 50 nucleotides long, folding was simulated without any addition of external sources of information, and for larger molecules, simulations were combined with other sources of information such as partial secondary structures.[98-99, 103] Using enhanced sampling methods such as parallel or simulated tempering, the model was able to fold the benchmark molecules to their native structures and to detect partially folded intermediates in agreement with experimental evidence.

The transferability of HiRE-RNA to high temperatures and to unfolded states remains still to be proven at this stage and it will be the focus of future work. In principle, because of the reduction in degrees of freedom resulting in an under-estimation of the entropy, thermodynamically optimised model needs free energy potentials with explicit temperature dependence. The introduction of such a term was done successfully by OxDNA[104] and Three-Interaction-Site (TIS)[105] CG models, where stacking depends explicitly on T in order to match T_m values of canonical duplexes over a wide range of temperatures (with a linear variation in strength of 6% over a range of 100 degrees for OxDNA). Such models thus work for canonical (Watson-Crick) duplexes. This procedure cannot be easily applied to systems dominated by non-canonical and multiple bases interactions, among other reasons because of a lack of experimental

data for a proper parameterisation. Indeed, the exhaustive thermodynamic parameterisation of OxDNA was based on SantaLucia's free energies for canonical base pairing and stacking,[106] but equivalent data do not exist for other kind of pairs, nor for triplets or quartets. Similarly, parameterisation of TIS was done based on experimental T_m values of dimers under the assumption of canonical base-pairings. To our knowledge, at the present state, because of the lack of suitable theoretical and experimental data on base pairs beyond canonical and on systems strongly departing from the double helix, no CG model exists that has both proper thermodynamic parameterisation and a full description of non-canonical and multiple base pairings.

Despite these limitations, in the case of HiRE-RNA, though no thermodynamic optimisation was performed, the melting curves extracted from our simulations showed a qualitative behaviour and T_m in good agreement with other sources, both theoretical and experimental. Simulations with our « twin » protein model OPEP[107] and with HiRE-RNA/DNA showed that we can predict heat capacity profiles similar to experiments with calculated T_m values within 20K to the experimental ones, indicating that the overall shape of the fluctuations in the potential energy of the conformations as a function of T and the thermodynamic up to T_m+10 K are very reasonable although they are not perfect. Calculated heat capacities of monomeric proteins above Tm are a bit smaller than the experimental values because the change in heat capacity upon unfolding results almost mostly from the increase in the hydration term and this cannot be accurately estimated using simplified beads for DNA bases and implicit solvent. This is a general problem intrinsic to the simplification introduced with coarse-graining and it is why we are very cautious about making any statement on thermodynamic behaviour of the system. Even for models with an explicit thermodynamic parametrisation such as OxDNA and TIS, T_m values are correctly predicted, but whether the behavior of the molecule at high temperatures is realistic or not remains to be shown.

For small pseudoknots HiRE-RNA was able to detect the two-state behaviour between the native conformation and the partially folded state where only one of the stems is formed, in agreement with OxRNA predictions.[108] For the more complex 50 nt triple helix pseudoknot of the human telomere, in addition to predicting the native triple helix as the most stable conformation, the model was able to propose two misfolded states and one partially folded intermediate in agreement with both experimental evidence and with prediction of the TIS model. [92, 109-110] Finally, our all-atom MD simulations starting from our list of CG partially unfolded states showed that 45 % of the atomistic simulations remained within the starting state for 1000 ns, with 30 % more stabilising on a state close to the starting state, providing evidence of the pertinence of the unfolded states predicted by our CG approach.

Regarding the present work, it is important to keep in mind that the CG model is not used to extract any thermodynamic or kinetic information, but i) to generate plausible partially folded conformations to be further investigated using atomistic MD and ii) to visualise larger-scale transitions through the GQ folding landscape that are far beyond those observable by atomistic simulations. HiRE-DNA seems adequate for the study of GQs thanks to its ability to account for non-canonical and multiple base-pairing and to study the interconversions between native, partially folded and unfolded structures in times within reach of current simulation methods. The transitions visualised by the CG approach (see below) are indeed much larger than those which have been achieved by atomistic simulations.

Simulated tempering coarse-grained simulations. CG MD simulations were performed for each of the GQs (1KF1, 143D, 2JSM, 2KF8 and 2KM3) using the HiRE-DNA force field. To

explore each conformational space around the native structure, simulated tempering (ST) was used to improve sampling with temperature ranging from 300 to 450K.[107] ST trajectories cannot be interpreted as real dynamics because of the changing temperature, however, they can easily generate new conformations as less stable portions of the molecule lose their structure faster upon temperature perturbations than energetically more stable parts.

Fixed temperature coarse-grained simulations. For each system we also performed a set of unfolding CG simulations at fixed temperatures, ranging from 300 to 400 K in steps of 10 K, each for 3 μs to follow the unfolding dynamics. The range was identified from the ST CG simulations of duplexes as appropriate for melting. In the absence of explicit solvent, a Langevin thermostat was used with a friction coefficient of 0.005 ns^{-1}.[107] Simulations were run using spherical reflective boundary conditions (non-periodic) and were started from the experimental configuration. The integration time step was set to 4 fs.

Due to the implicit solvent and bead representation the dynamics cannot reproduce exactly the dynamics in explicit solvent. The relationship between the time measured in the CG model and the time of an all-atom explicit solvent simulation is not straightforward and depends on the properties in question (diffusion coefficients, transition times between diverse conformations, etc.). Our experience suggests a 10- to 100-fold speed up compared to all atom MD in explicit solvent for fast-folding systems. For example, typical CG folding times of a 10 base pairs hairpin are of a few hundred of ns, while predictions of hairpins folding times as a function of base pairs extrapolated from experiments, range in the few microseconds.[111] This speed up is obviously still not sufficient to study full folding of GQs (with real times from hours to weeks) but represents a substantial improvement over the atomistic MD, as confirmed by the actual data presented below.

Analysis of coarse-grained simulations. For the analysis of CG trajectories, we monitor the presence of base-pairs, triplets and quartets based on energy criteria, together with *anti* and *syn* conformations of Gs. The base-pairing analysis is performed on reduced trajectories, which are built by extracting one frame every 4 ns from the original trajectories. In order to detect the existence of base-pairs in the CG structures we compute the base-pairing energy as described above. However, because of the short-ranged, narrowly peaked nature of the HB potential, base positions fluctuate in such a way that the instantaneous energy of a base-pair would often indicate loss of a base-pair even when the pair is still present. Thus, to detect base-pairs more coherently, we perform averaging of the base coordinates over a shifting time window (of 10 frames) to smear out fluctuations and we evaluate the base-pairing energy on the averaged coordinates. Indeed, a comparison between the instantaneous energy plots and the corresponding base-pairing plots shows that this procedure results in a much more reliable detection of formed pairs. We identify a pair as formed if the corresponding energy is in absolute value at least 10% of the maximum value that the specific pair can have. This cut-off is applied on the energy computed from the averaged coordinates. We have verified that small changes in the cut-off and in the number of frames over which we average do not affect the results. Base triplets and quartets are identified from base-pairs.

Refolding fully-atomistic MD simulations. With standard atomistic MD we attempted refolding of some unfolded structures selected from the CG simulations after putting them back in their natural environment with explicit water and ions. The starting structures were carefully selected from trajectories of the ST CG unfolding simulations, based on our expectations of which structure could be stabilised or even refolded. Atomistic structures were obtained through a reconstruction algorithm based on database searches of possible nucleotide conformers.[98] With

a full atomistic structure at our disposal, either one or two Na^+ cations (depending on the degree of perturbation of the selected structure) were manually placed into sites that were expected to form the GQ channel. The molecules were then solvated by a truncated octahedral box of water with minimal distance between the solute and the box border of 10 Å, using the TIP3P water model.[112] The simulation box was neutralised by Na^+ cations and finally 0.15 M excess NaCl was added (total concentration of Na^+ was thus about 0.35 M). The Joung and Cheatham TIP3P-adapted parameters for ions were used.[113] Solvation and addition of ions were performed in the xLEaP module of AMBER.[114]

Although Na^+ is not the native cation of most of the simulated GQs in thermodynamic equilibrium, the difference between Na^+ and K^+ can be neglected in our MD simulations. This can be justified by several reasons. First, the ions are merely described by their radius and potential well, thus no polarisability is included and the difference between Na^+ and K^+ is captured imperfectly. In the force-field approximation, Na^+ ions inside the GQ channel have properties (such as the effective ion size) that are roughly in-between those expected for real Na^+ and K^+ ions.[115] On the other hand, force-field K^+ ions may be somewhat over-sized. Second, both Na^+ and K^+ promote folding of GQs in real systems and readily stabilise GQs. Although ion-replacement sometimes may shuffle relative stabilities (populations) of different GQ folds (different conformational basins) in the thermodynamic equilibrium, it does not lead to any immediate disruption of the folded structures and does not obviate their basins from the landscape. The kinetics of ion replacement in the GQ channel is much faster than the GQ unfolding kinetics and eventual shifts of the GQ equilibrium populations after the salt change are thus slow. Therefore, both Na^+ and K^+ are equally suitable to attempt re-folding of partially unfolded GQ structures which still keep significant structural features of the target structures. One of the reasons why we have chosen Na^+ is that in some of our preceding studies we have noticed occasional expulsions of K^+ from the GQ channel,[115] though it remains to be seen if such problems may occur also with Joung and Cheatham TIP3P-adapted parameters. Thorough discussion of the inclusion of monovalent ions in simulations of GQs can be found elsewhere.[82, 100]

The bsc0[116] + χ_{OL4}[117] + $\varepsilon\zeta_{OL1}$[118] version (bsc0$\chi_{OL4}\varepsilon\zeta_{OL1}$) of the Cornell et al. force field[119-120] was used. The basic bsc0 modification suppresses detrimental non-canonical γ-trans substates,[116] χ_{OL4} improves the shape of the χ torsion profile so that it rebalances *syn*/*anti* conformations and $\varepsilon\zeta_{OL1}$ balances the BI/BII substates of DNA backbone; the last two modifications have proved useful both in B-DNA and GQ simulations.[117-118] For parameters, see AmberTools starting from May 2015 update or http://fch.upol.cz/ff_ol/. For overviews of DNA force fields, see refs.[64, 79]

The simulations were performed in the NPT ensemble, with the pressure held at 1 atm and temperature at 300 K using a Berendsen weak-coupling barostat and thermostat.[121] A 2 fs integration time step was used and all bonds involving hydrogen were constrained using SHAKE.[122] The particle-mesh Ewald method was used to calculate electrostatic interactions,[123-124] and the non-bonded cut-off was set to 9 Å. The equilibration protocol is given in the Supporting Information. The production phase was performed using the GPU CUDA version[125] of the pmemd module of AMBER 12.[114] The initial simulations were run for 200 ns and then the stable or improving trajectories (based on visual inspection) were prolonged to 1000 ns. The trajectories were post-processed in the ptraj module of AMBER[126] and visualised in VMD.[127]

Description of the trajectories. At first sight, it might seem desirable to visualise the simulation behaviour by simplified order parameters or some reaction coordinates. However, this common approach is not applicable to the presently-studied GQ folding landscape due to its

extreme degree of kinetic partitioning. There are no simple order parameters that could describe the GQ systems appropriately. Even a combination of several parameters, sometimes useful, cannot be interpreted intuitively, because the folding landscape of GQ is so complex that its description is principally not reducible to a few order parameters. Projection of such complex folding landscapes into simple order parameters may lead not only to physically incorrect models of folding mechanisms, but also to erroneous free-energy estimations, typically in cases, when the order parameter cannot distinguish between different kinetically unrelated free-energy minima. Note that cases where order parameters, transition states and reaction pathway become unsuitable for description of folding landscapes have been discussed in the literature.[128-135] The GQ folding landscape is an extreme example of such cases.

For the sake of demonstration, for each simulation we performed, we have calculated the evolution of order parameters often used as collective variables (CVs) in enhanced-sampling simulations, namely, RMSd (shown in the form of 2D RMSd), distance RMSd, total number of native hydrogen bonds, number of O6 – Na^+ contacts and radius of gyration. However, given the amount of the data, its complexity, and unsuitability of the order parameters, throughout the paper we have opted for a qualitative and schematic description of the simulation behaviour, representing the main features emerging from the detailed analyses. To present the data we mostly use idealised schemes that are standardly utilised in the GQ literature. Note that our simplified schemes do not show the helical twist of the GQ stems. Numerous quantitative graphs documenting the development (and unsuitability) of the order parameters are shown in the Supporting information appendix.

RESULTS

Simulated tempering coarse-grained simulations. All five systems unfolded during the 1.4 to 1.7 μs of ST CG simulations (corresponding to 48 hours real-time per trajectory on an Intel i5-4670 PC with 8GB of RAM). Next, we describe data for each system, focusing on the first steps leading to loss of the GQ organisation. The native structures and early unfolding stages of the GQs are shown in Figure 4.

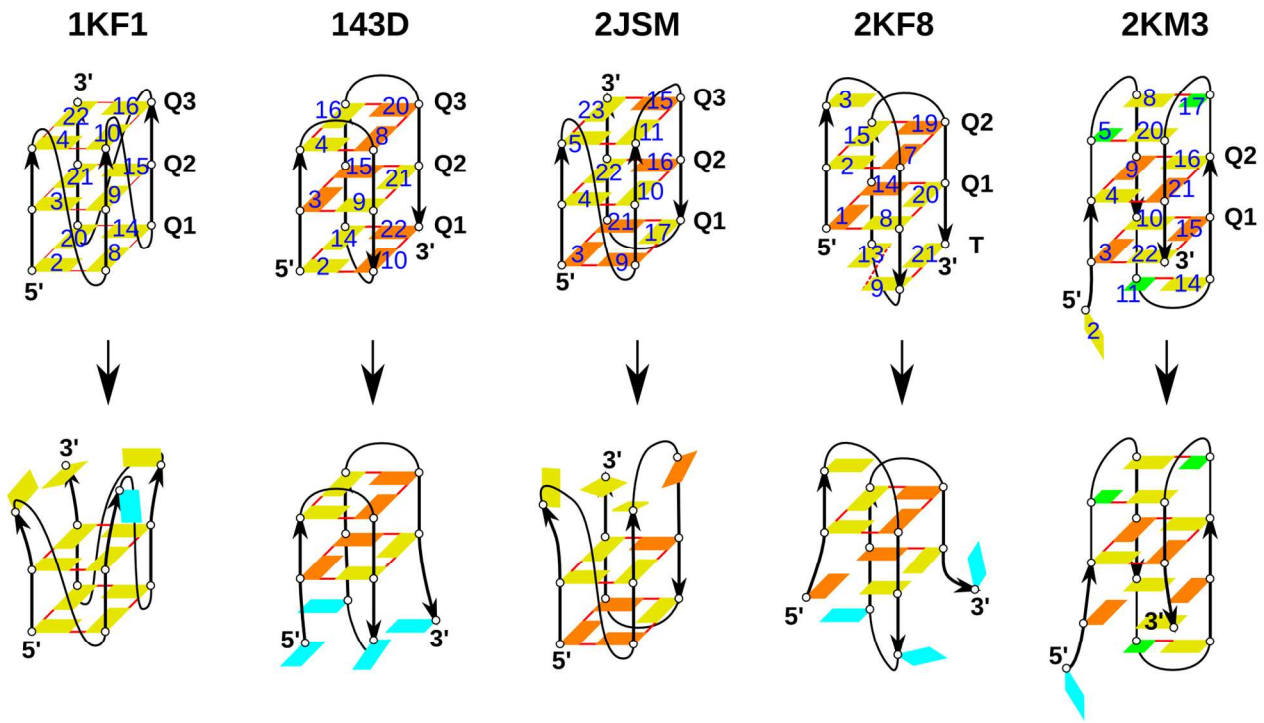


Figure 4. Schemes of native structures (top) and initial stage of unfolding (bottom) of GQs in ST CG simulations. First step of unfolding is disruption (loss of pairing) of one terminal quartet. G bases represented in yellow are in *anti* conformation while bases in orange are in *syn* conformation. Cyan bases fluctuate between *anti* and *syn*. Ions and loop residues are not shown. Cytosines in 2KM3 are shown as green squares. G-quartets (Q) are numbered from the 5′-end. 2KF8 contains a triplet (T). Numbering of Gs is shown by blue numbers.

1KF1 is composed of three all-*anti* quartets Q1, Q2, and Q3. Unfolding occurred almost simultaneously for all three Qs (Figure 5). In the experimental crystal conformation the T and A bases in each loop are tightly packed by stacking. Unfolding initiated with a destabilisation of the loop connecting G10 and G14 with the higher temperatures causing a loss of stacking and the unpaired bases orienting towards solvent. Upon unfolding G bases are free to fluctuate between *anti* and *syn* conformation (Figure 5). A transiently stable partially folded structure composed of two stacked triplets (G2-G9-G15 and G3-G8-G16) is also observed.

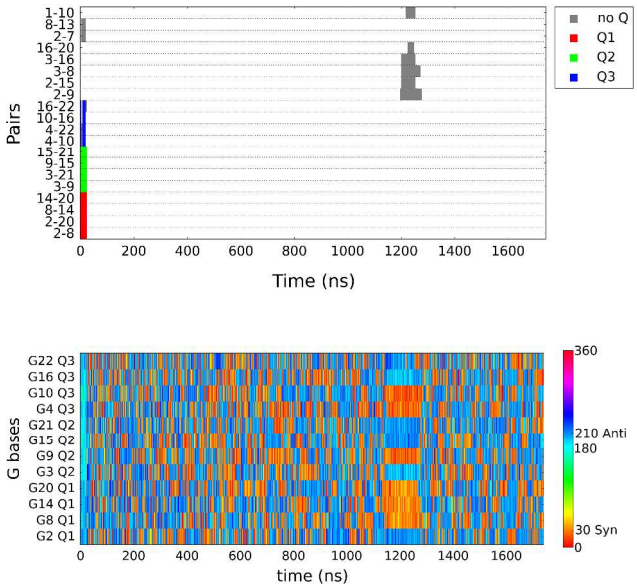


Figure 5. ST CG unfolding of 1KF1. Top: All base pairs detected during the simulation. Bottom: *Anti-syn* dynamics of the G bases. The native GQ unfolds in the first few tens of ns. After 1.1 μs, a transient triplex is formed for about 100 ns. See Supporting Information Figures S4-S7 for data for the remaining systems.

143D is composed of three stacked quartets, each one with two G bases in *anti* conformation and two bases in *syn* conformation. The main unfolding process was a destabilisation of the Q1 quartet (Figure 4) that partially opened and closed several times before complete unfolding (see Supporting Figure S4 for details). The quartet first opened briefly twice and both times came back to its native conformation to then open irreversibly. In the first two opening events all bases preserved their native *anti* and *syn* conformations, while after the third event, the bases fluctuated between *anti* and *syn* for the remainder of the trajectory (Figure S4). Opening of Q2 and Q3 occurred simultaneously. Once all the quartets unfolded, all Gs fluctuated freely between *anti* and *syn* conformations.

2JSM is composed of three stacked quartets, with mixed *syn* and *anti* conformations. Before complete unfolding (Supporting Figure S5) Q3 exhibited brief openings and closings (Figure 4). Q1 and Q2 were stable until complete unfolding occurred. The *anti* and *syn* conformations of all bases of the quartets were stable until unfolding occurred, and they fluctuated freely between the two states once the molecule adopted an open conformation (Figure S5). A short-lived structure containing a triplet (G2-G11-G16) and pairs of loop bases was also observed.

2KF8 comprises two stacked quartets, each with two *syn* and two *anti* Gs and a triplet T stacked with Q1. Unfolding occurred with the two quartets opening almost simultaneously (Supporting Figure S6). Prior to unfolding, the triplet fluctuated rapidly between stacked and unstacked states (Figure 4), and broke completely at unfolding. While the bases in Q1 and Q2 preserved their *anti* or *syn* conformation until unfolding occurred, the bases in the triplet changed from one state to the other several times (Figure S6).

2KM3 is composed of two stacked quartets with mixed *anti* and *syn* conformations. In addition, C5-G20 and G8-C17 canonical base-pairs are stacked with Q2, and C11-G14 pair is stacked with Q1. Unfolding occurred with Q1 (Figure 4) and Q2 opening almost simultaneously (Supporting Figure S7). Prior to unfolding, Q1 was unstable with G3 and G10 fluctuating between paired and unpaired positions. G3 briefly switched conformation between *syn* and *anti*

two times before unfolding, while all other G bases constituting the quartet remained in their native τ conformation until unfolding took place (Figure S7). The GC base pairs remained formed for some time after the loss of the GQ structure, with C5-G20 breaking, reforming, and breaking. G8-C17 broke to form G9-C17, followed by G10-G17 at 800 ns. Other non-native pairs were also formed (Figure S7).

Fixed temperature coarse-grained simulations. In this section we analyse the overall behaviour of the molecules at different temperatures in 3μs-long fixed-temperature CG MD simulations. For each system, we run eleven simulations: at 300, 310,… , 390 and 400 K.

The 1KF1 system was the least stable of the studied GQ topologies, as it completely unfolded at 330 K and at all higher temperatures. At the same time, this system provided the most interesting dynamics and we have observed two rare cases of its folding to alternative GQ topologies.

Already at 300 K, short openings of Q1 and Q3 were observed, with one base detaching and returning. Simulations at 310 and 320 K showed the rare unfolding-alternative folding events (Figure 6).

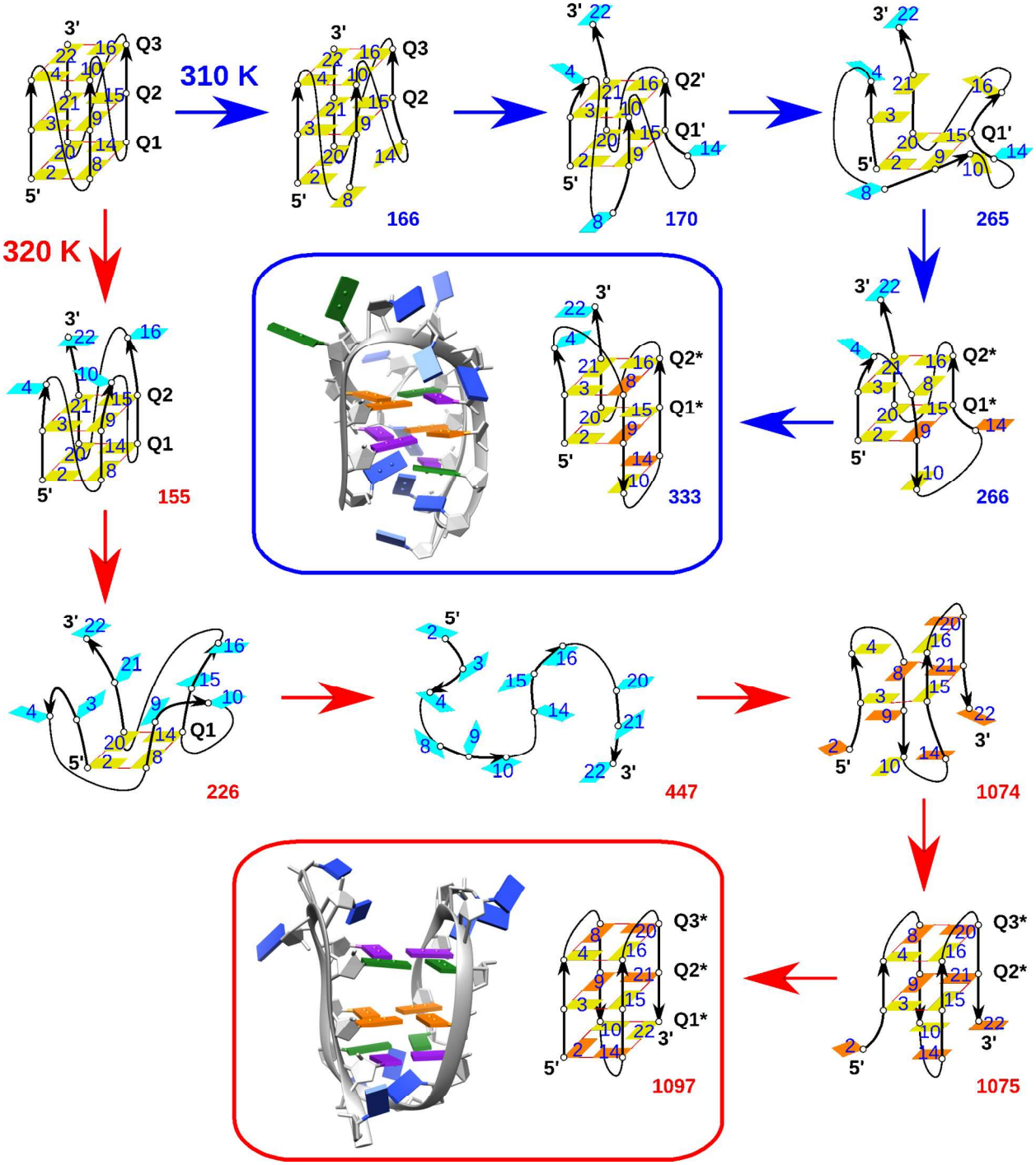


Figure 6. Unfolding-folding events in fixed temperature CG simulations of the parallel stranded GQ 1KF1 at 310 K and 320 K. At 310 K, a new two-quartet hybrid (3+1) type GQ with double strand-slippage is formed (in the blue box). At 320, a new three-quartet antiparallel alternative basket-type GQ is formed (in the red box). Folding mechanism of the two GQs is different. The structural schemes (yellow, orange and cyan bases) are visualised as in the Figure 4; numbering of Gs is shown by blue numbers inside the bases. The number next to each scheme denotes time (ns), when the structure appears in the simulation. In the structural snapshots of the two newly folded GQs, bases in mauve belong to the first quartet (Q1) in the original 1KF1 structure, those

in orange to Q2 and those in green to Q3. The remaining bases (flanking and loop Ts and As) are in blue.

At 310 K, opening of Q3 was immediately followed by simultaneous strand slippage of two strands, so that two quartets Q1' (G2-G9-G15-G20) and Q2' (G3-G10-G16-G21) were formed. Unfolding continued by disruption of Q2', the second G-strand turned upside down and two new quartets were formed: Q1* (G2-G9-G15-G20) and Q2* (G3-G8-G16-G21). G9 gained its *syn* conformation by the strand reorientation and G8 became *syn* after a while. Together with the formation of Q2*, a base-pair G10-G14 was formed and remained stable for the rest of the simulation (Figures 6 and 7). The resulting GQ was similar to the hybrid (3+1) type-2 GQ, however, with only two quartets. It formally corresponds to a hybrid (3+1) structure with double-strand slippage. Note that structures with slipped strands have been suggested to be commonly populated intermediates during the GQ folding processes.[62]

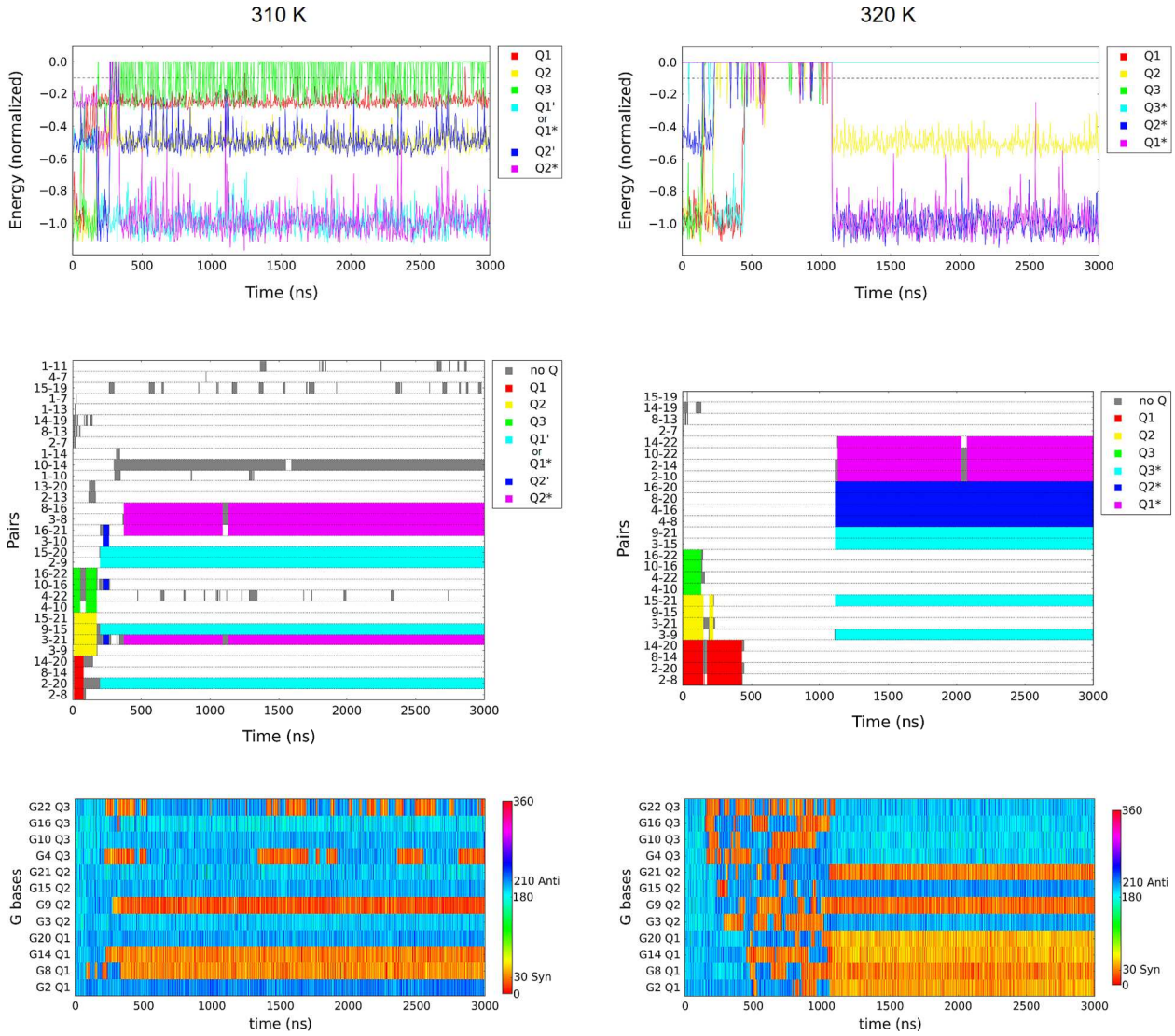


Figure 7. Fixed-temperature CG simulation of 1KF1 at 310K and 320 K. Top row: Overall energy of base-pairing of the quartets as a function of time computed as the sum of the energies

of the single pairs, where each pair's energy has been normalised to its maximum possible absolute value (the 10% cutoff on the pairs' energies is shown as a dashed line). Middle row: All base pairs observed in the simulations coloured as a function of the quartet they belong to. Bottom row: *Anti-syn* conformation of all G bases. Structural development in the two simulations is depicted in the Figure 6.

At 320 K, the 1KF1 molecule unfolded completely by the quartet-by-quartet disruption. In the unfolded state, all Gs fluctuated between *syn* and *anti*. After undergoing several unsuccessful folding attempts, a sudden reorganisation led to formation of a nucleation centre, which finally evolved to an antiparallel three-quartet GQ topology, stable for the rest of the simulation. Upon reaching the new structure, the bases adopted a mixed *anti-syn* organisation, different from the all-*anti* conformation of the starting structure of 1KF1 (Figures 6 and 7). The new conformation had an alternative basket-type topology, with strand progression opposite to 143D. During its folding, transient base-pairs formed between As and Ts, and between Gs and Ts. These base-pairs helped to navigate the system through the folding landscape towards further structuring. Two consecutive pairs seemed to be enough to stabilise the structure for some time, and allowed other parts to come close and to start forming some transient GG pairs, and eventually a first transient quartet G3-G8-G15-G21. This served as a nucleation centre for the other Gs to approach through stacking. Once the overall structure was compact again and all Gs were close to one another in the middle of the folded strand, the second G-stretch slipped by one level and two new quartets (Q2* and Q3*) formed. Q1* formed after a while, because G22 first needed to switch from *syn* to *anti*.

In the light of these results, we ran ten new CG simulations of this system, hoping to evidence additional unfolding-folding events. Unfortunately, no other unfolding-folding events were observed, illustrating the rarity of successful folding attempts.

No folding was achieved for the other systems after unfolding. The 143D molecule unfolded completely at temperatures above 360 K. Starting at 310 K transient openings of Q1 and Q2 were observed, becoming longer-lived as the temperature increased. There seemed to be a loss of stability related to a loss of stacking between Q1 and Q2. The 2JSM structure unfolded at 350 K. At lower temperatures the two external quartets, Q1 and Q3, exhibited some opening, becoming longer-lived as the temperature increased. The 2KF8 unfolded at 380 K. Both Q1 and Q2 remained closed until 350 K where they started exhibiting some openings and closings. The triplet was disrupted already at 300 K. The 2KM3 system was either stable as a GQ or converted to a hairpin at higher temperatures. At the lowest temperatures the GQ was mainly fully folded, with transient openings of Q1. At temperatures above 350 K the stable configuration was the hairpin formed by G10-G15, G9-G16, G8-C17, A7-T18, T6-A19, C5-G20, and G4-G21 base-pairs (Figure 8). GG pairs did not form quartets at these temperatures. The hairpin structure was irregular, but appeared to be rather stable.

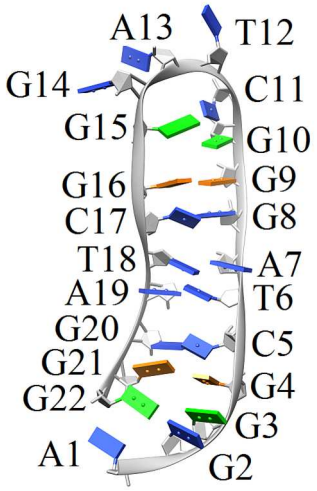


Figure 8. Hairpin intermediate in unfolding of 2KM3. G bases forming the first quartet (Q1) in 2KM3 are represented in light green and of Q2 in orange.

Stabilising fully-atomistic MD simulations. We have carried out a series of explicit-solvent atomistic simulations (Table 1) using starting structures selected along the CG unfolding trajectories.

Table 1. The fully-atomistic MD simulations and their outcome (see Supporting Information appendix for time development of selected order parameters).

Simulation name	Time when the starting structure was taken from the ST CG simulation (ns)	Simulation length (ns)	Main simulation outcome
1KF1 – see Figure 9, Supporting Information appendix Figures SA1-SA6			
A	19.21	1000	Stabilisation of a slipped GQ
B	19.32	200	Unfolding to coil
C	20.42	1000	Stabilisation of the starting structure
143D – see Supporting Figure S9, Supporting Information appendix Figures SA7-SA20			
A	103.74	1000	Stabilisation of the starting structure
B	103.76	1000	Stabilisation of the starting structure
C	105.82	1000	Stabilisation of the starting structure
D	108.02	1000	Stabilisation of the starting structure
E	140.12	1000	Stabilisation of the starting structure
F	140.14	1000	Stabilisation of the starting structure
G	266.38	200	Unfolding to coil
2JSM – see Supporting Figure S10, Supporting Information appendix Figures SA21-SA32			
A	85.02	1000	Modest improvement
B	85.04	1000	Stabilisation of the starting structure
C	96.80	1000	Completely refolded GQ
D	89.98	1000	Stabilisation of the starting structure
E	122.84	1000	Unfolding to coil
F	138.26	200	Unfolding, a parallel G-hairpin survived

2KF8 – see Figure 10, Supporting Information appendix Figures SA33-SA40			
A	492.90	1000	Almost refolded GQ
B	492.92	1000	Unfolding to coil
C	492.94	200	Unfolding to various G-triplexes
D	493.00	200	Unfolding to coil
2KM3 – see Supporting Figure S11, Supporting Information appendix Figures SA41-SA50			
A	326.72	200	Unfolding to an antiparallel G-triplex
B	326.96	200	Unfolding to coil
C	328.48	200	Stabilisation of the starting structure
D	328.76	1000	Unfolding to an antiparallel hairpin
E	332.48	200	Stabilisation of the starting structure

1KF1. Three different snapshots of the ST CG simulation for the parallel stranded GQ were selected. All three structures were modestly unfolded and similar in shape, with two quartets preserved (Q1 and Q2) and one quartet disrupted (Q3). The molecule did not regain the native folded state in any of the three simulations (Figure 9). Stabilisation of the Q1+Q2 stem was observed in one run (1KF1-**C**) (Supporting Figure S8). A different structure with two quartets was stabilised in another simulation (1KF1-**A**): the Q1+Q2 stem was stable till 125 ns, then the first G-stretch slipped by one level so that G2 was expelled from the stem, G3 became part of Q1 and G4 was incorporated into Q2. This slipped arrangement was stable to the end of the run. The simulation 1KF1-**B** led to loss of the stem by the strand slippage mechanism[62] within 20 ns.

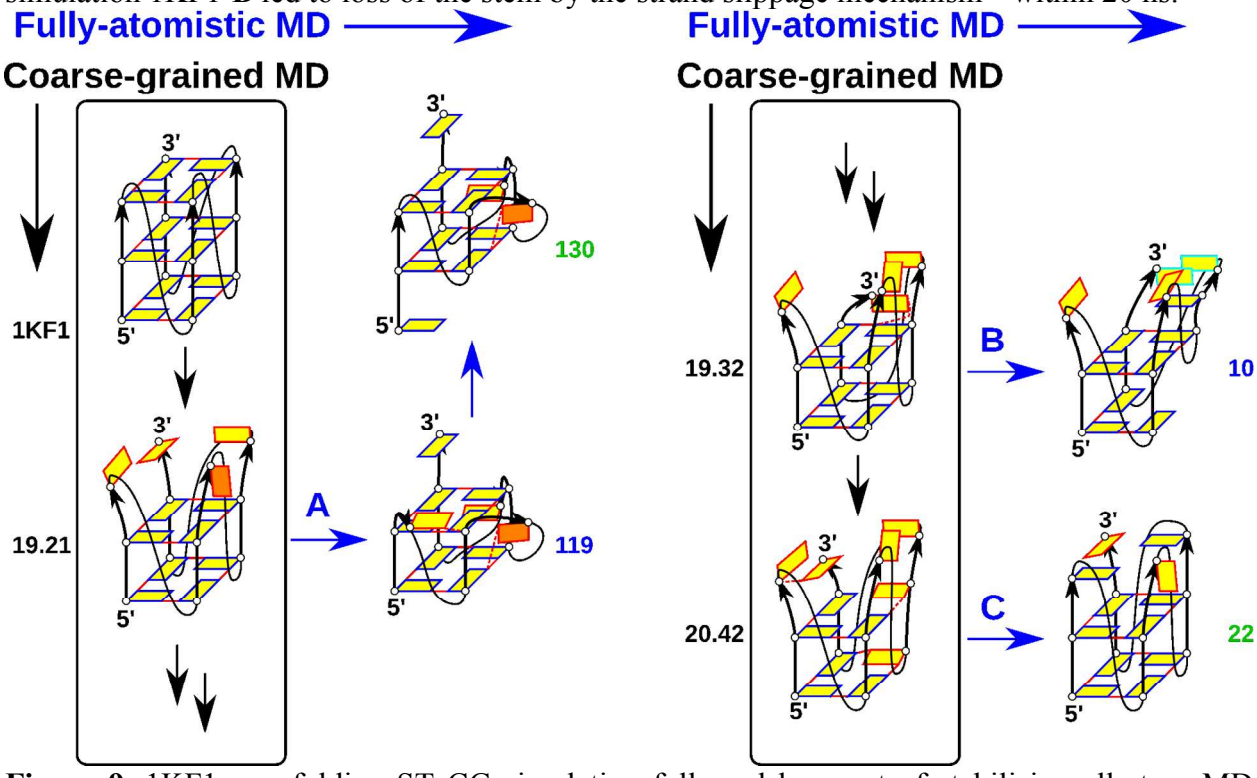


Figure 9. 1KF1 – unfolding ST CG simulation followed by a set of stabilising all-atom MD simulations. The black box contains intermediates occurring in the unfolding CG simulation. Development of the CG simulation goes vertically down from the experimental structure and continues in the right part of the Figure. Black numbers next to the intermediates depict the time when the structure was observed in the CG simulation. These intermediates served as starting

structures in subsequent all-atom MD simulations. Their progression is shown by blue arrows and the numbers show the time in ns at which the given structure occurred. The green numbers mean that the particular structure remained stable till the simulation end and blue numbers mean that the depicted structure was subsequently lost. Simulation B leads to complete structure loss after 10 ns. The structural schemes are visualised as follows: Gs are depicted as rectangles; yellow and orange mean *anti* and *syn* conformation, respectively. Solid red lines denote two GQ-like H-bonds, while dashed lines any other H-bonding. The colours of the rectangle edges are used to visualise the way how the Gs mutually interact. Gs that have the same colour of their edge are either stacked together or coplanar and H-bonded. The remaining Gs, either lacking any direct interactions, or H-bonded, but not coplanar, are indicated by a red edge. Flanking residues, loop bases and ions are not shown. See the Supporting Information appendix Figures SA1-SA6 for development of selected order parameters.

143D. Seven snapshots taken from the ST CG simulation were used for atomistic MD. The starting structures for 143D-**A** to **F** contained a stem with two quartets (Q2 and Q3), and the quartet Q1 was disrupted. In all six simulations, the Q2+Q3 stem was stable throughout, with a channel cation between Q2 and Q3. Four runs exhibited an additional channel cation above Q2. This ion binding site was created with the aid of Hoogsteen edges of some Gs of the disrupted Q1 and exhibited, albeit infrequent, ion exchanges with the bulk on the simulation time scale. Nevertheless, Q1 never reformed (Supporting Figure S9). The starting structure of the simulation 143D-**G** possessed no full quartets, Q1 was completely disrupted and Q2 and Q3 were in a form of imperfect triplet close to the fourth G, keeping the GQ-like shape. The arrangement was lost immediately and the structure adopted a coil-like conformation.

2JSM. Six snapshots from the ST CG simulation were chosen for atomistic MD simulations. The GQ was successfully refolded in one run, the intermediates (the initial arrangements) were stabilised in three cases, and further degradation was observed in two runs (Supporting Figure S10). The starting structures in the simulations 2JSM-**A** to **D** contained a two-quartet stem Q1+Q2. In two simulations (2JSM-**B** and 2JSM-**D**), no improvement was achieved. In one simulation (2JSM-**A**), a G-triplet was formed instead of Q3 and was sampling several local G-triplet substates described in the literature.[72] The fourth G, whose incorporation into the Q3 would mean full refolding of the GQ, remained stacked on the triplet. In the run 2JSM-**C**, a G-triplet was formed at Q3 immediately after the simulation started, and at 200 ns, the fourth G was incorporated so that the GQ was fully refolded. In the simulation 2JSM-**E**, the structure possessing Q1 and misfolded Q2 further unfolded. Right after the start, it split into two hairpins, one with a propeller loop and one with a lateral loop. During the course of the run a transient triplex-like structure was observed. The simulation 2JSM-**F** started from an arrangement with only Q1 preserved. The structure quickly split into two hairpins, one with the propeller loop and one with the lateral loop. The one with the lateral loop had imperfect base-pairing and unfolded. The one with the propeller loop remained stable, probably due to adjacent bases stacked on both sides of the hairpin.

2KF8. Four structures generated in the ST CG simulation were selected for atomistic simulations. The GQ structure was nearly restored in one run, the intermediate stabilised in another, and in two runs the intermediates were lost (Figure 10). The simulation 2KF8-**A** led to a quite spectacular restoration of Q1 and Q2 (Figure 11). In the starting structure, no quartets were present, but the respective bases were close in space and in favourable orientation for *c*WH base-pairing. Indeed, immediately after the start, the quartets refolded and remained stable. On the

other hand, the G-triplet near the 5′-end, which is present in the native 2KF8 structure, was not reformed. The simulation 2KF8-**B** started with Q2 formed and Q1 lacking one G. After 340 ns, Q2 was broken and the structure unfolded. In the simulation 2KF8-**C**, Q2 was slightly perturbed in the starting structure. Various G-triplets (including misfolded ones) were formed in the course of the run, but no quartet formation occurred. The simulation 2KF8-**D** led to entire unfolding of the starting structure.

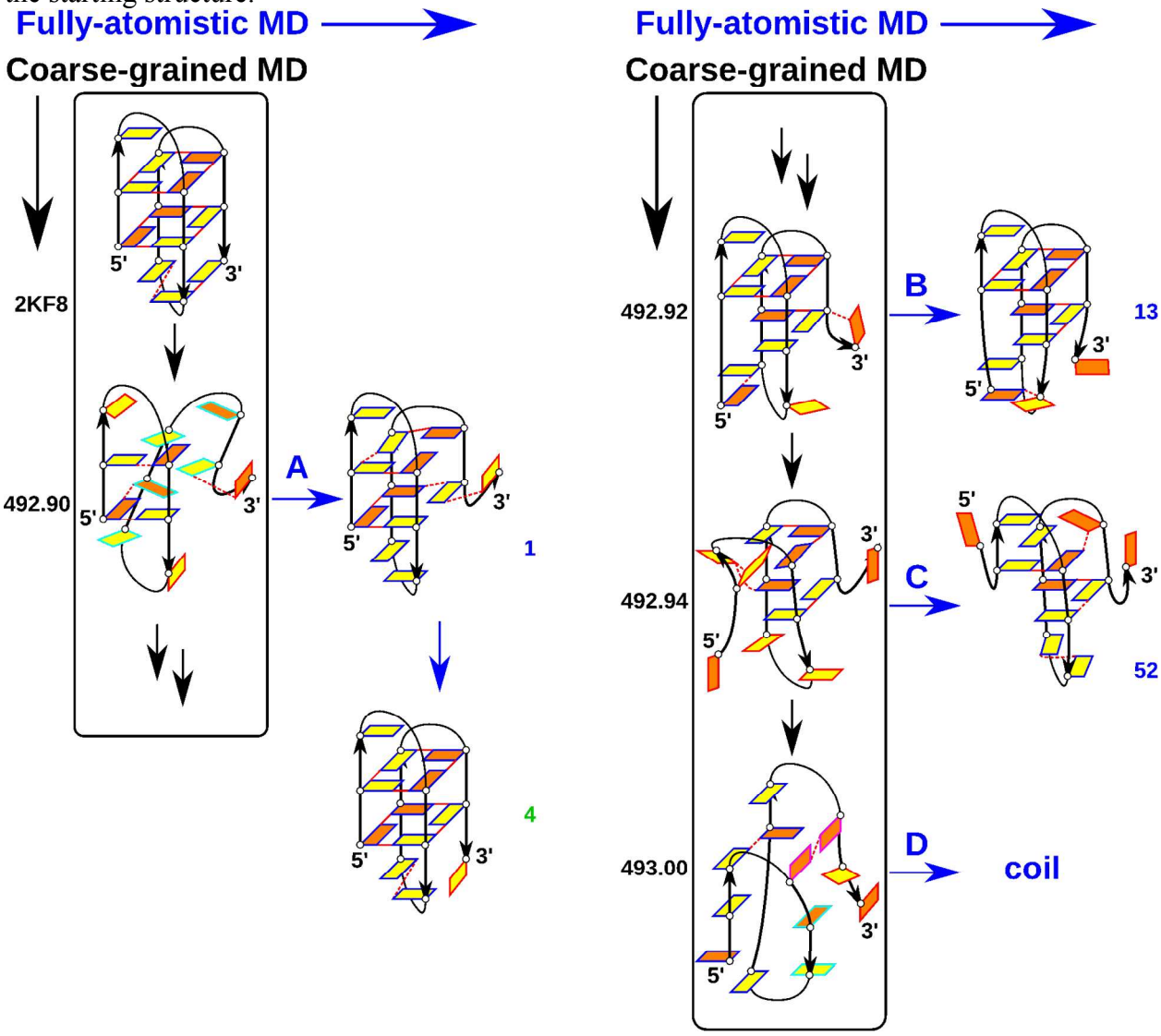


Figure 10. 2KF8 – unfolding ST CG simulation followed by a set of all-atom MD simulations. The simulation **A** leads to almost refolded native GQ. The structural schemes are visualised as in the Figure 9. See the Supporting Information appendix Figures SA33-SA40 for development of selected order parameters.

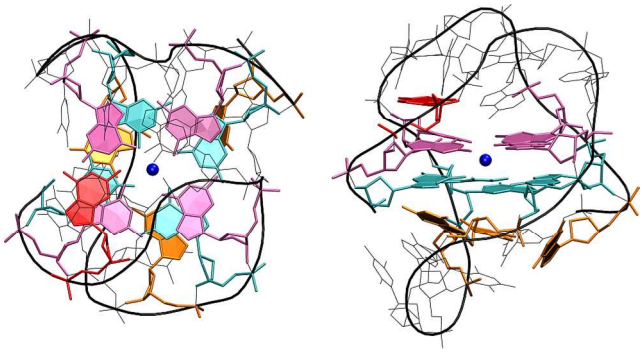


Figure 11. Top and side view of the final structure in the simulation 2KF8-**A**. Bases of Q1 are shown in cyan, Q2 in mauve, G-triplet bases in orange, G3 is red, cations are blue, and loop nucleotides and the backbone are black.

2KM3. Five snapshots from the ST CG simulation were chosen as starting points for refolding attempts. All the initial structures were either triplexes or triplex-like structures, but none of the runs led to formation of quartets (Supporting Figure S11). In the simulations 2KM3-**A**, **C** and **E**, the starting structures slightly deteriorated. The run 2KM3-**B** led to full unfolding of the triplex into a coil-like structure. At 150 ns of the run 2KM3-**D**, the triplex-like arrangement transformed into a long hairpin structure that was stable till the end of the simulation (having the same base-pairing as the hairpin shown in Figure 8). Pairing occurred between the G-stretches 1 and 4 and between the G-stretches 2 and 3. Such a structure could be a potential intermediate for a quick folding event. Nevertheless, if some GQ folded from this hairpin, it would be severely misfolded compared to the 2KM3 fold, because some Gs changed χ conformation during the transition to the hairpin.

DISCUSSION

Coarse-grained (CG) simulations do not aspire to give detailed quantitative information. Rather, they provide valid qualitative insights allowing exploration of molecular rearrangements to an extent currently unreachable for atomistic MD. New configurations are obtained following a physical description of the relevant interactions to which the molecule is subject, and are therefore worth exploring further with atomistic MD as belonging to possible free-energy basins along the folding pathways.

Several recent studies attempted to unfold three-quartet GQs in atomistic simulations and then investigate properties of the potential intermediates along the unfolding pathways.[62, 75, 100] One approach utilised the no-salt environment, i.e., ion depletion. This method resembles (reversed) stop-flow experiments and is justified by the assumption that real GQ rearrangements are likely to take place when the GQs are temporarily destabilised during the genuine ion-exchange processes.[62, 100] Another study unfolded the human telomeric GQ using bias-exchanges metadynamics, which required *a priori* selection of collective variables alongside which the unfolding then progressed.[75]

In the present work, we have achieved the unfolding using unbiased CG simulations followed by *a posteriori* reconstruction of atomistic structures for explicit solvent simulations. Despite the simplified CG model, the CG method is a powerful complement to the other techniques. It reaches much further into the unfolded ensemble than the atomistic approaches since it allows extensive sampling of the *anti* ↔ *syn* transitions, which is absolutely essential to describe the GQ folding landscape.[53, 62] We have evidenced, albeit as very rare single-molecule events, spontaneous unfolding-folding events from one GQ topology to another (Figure 6).

Initial stages of unfolding: quartet loss or strand slippage/departure? The ST CG unfolding simulations reveal a preferred unfolding mechanism with consecutive disruption of quartets, starting from the outer quartets. Such rearrangements may be relevant to the real molecules as they may mimic one-by-one loss of channel cations correlated with quartet-by-quartet disruption of the GQ core (although the ions are not explicitly present in the CG model, the ion effects can be included implicitly). The process could resemble melting of the GQ, mimicked by the temperature jumps in ST CG simulations or elevated temperature in fixed-temperature CG simulations. The CG simulations performed here preserve, in their simplified description, the chemical environment.

The CG simulations complement previous atomistic MD studies, where unfolding was induced by complete ion depletion,[62, 100] i.e., by a drastic change of the molecular chemical environment at the beginning of the computer experiments. The atomistic no-salt simulations initiated unfolding via strand-slippage for all-parallel all-*anti* GQs. For the other types of GQ topologies, the no-salt simulations initiated unfolding either by splitting the GQ stems into two two-stranded halves or by single strand unbinding.

The initial parts of unfolding pathways suggested by atomistic no-salt[62, 100] and elevated temperature CG (present work) simulations differ and may lead to different unfolded ensembles. In the case of the CG simulations, the driving force is elevated temperature, while it is chemical denaturation in no-salt simulations. Nevertheless, we suggest that all observed mechanisms of unfolding initiation, i.e., outer quartet loss, G-strand departures and strand slippages may co-exist for real GQ systems. For example, the outer quartet loss may occur during periods of reduction of the number of the ions in the channel (one ion for three quartets), which is a much more likely event than the complete ion depletion. Let us reiterate that folded GQ molecules often have single-molecule lifetimes of minutes to hours (depending on the fold and environment).[39, 47] This indicates that in real molecules only a tiny fraction of initial unfolding attempts (major structural fluctuations) crosses the energy barrier for unfolding and progresses to full unfolding.

Previous MD studies have shown that strand slippage may be the dominant mechanism of (un)folding for all-*anti* parallel-stranded GQs, because all their quartets are of the same directionality.[62, 70, 100] Vertical strand-slippage movements were not observed in the other folds as they would lead to *syn* – *anti* conflicts.[62, 72] Here we show that strand-slippage movements can be relevant also for the other folds, when one of the G-quartets melts and the remaining two have the same directionality (Supporting Figure S12). Strand slippage can thus facilitate restructuring of partially folded intermediates, as demonstrated by the CG simulation of 1KF1 at 320 K (Figure 6, the transition 1074 → 1075). Here, the slippage occurred in the partially folded antiparallel intermediate with one quartet and one triplet. Although all four G-strands contained both *syn* and *anti* Gs, those Gs that would form a quartet of opposite directionality (G2, G10, G14 and G22) were not yet part of the transient GQ core. This allowed the strand slippage resulting ultimately in formation of the Q1* quartet.

Sampling of *syn*/*anti* distributions. The CG model was able to deal with the challenge of sampling *syn* and *anti* Gs. Note that the twelve Gs in the nucleotide sequence can adopt 2^{12} = 4096 *syn*/*anti* combinations, which are compatible with numerous diverse GQ folds.[53]

New mechanism for strand re-orientations. The CG model not only showed that *syn*-*anti* sampling is allowed when G is not H-bonded with other nucleotides, but also that χ (τ in the CG model) torsion can be flipped by rotation of a whole G-strand when one quartet is present (cf. Figure 6, the transition 265 → 266 at 310 K). It remains to be seen if such strand rotations would

be possible at the atomistic level of description, but we do not see any obvious steric hindrance that would prohibit them. Such rearrangements would allow strand reorientations while keeping one of the quartets fully paired. This may allow substantial structural transitions between different GQ folding basins without the need of a full unfolding. At the atomistic description, such transitions could involve a series of consecutive intermediates stabilised by various sets of non-native H-bonds, through which the molecule would diffuse from one fold to another.

GQ folding cannot be monitored by simple order parameters and collective variables. Due to complexity of the folding landscape of GQs, their folding cannot be monitored or visualised by simple order parameters and collective variables. We have explicitly documented this for a number of common order parameters such as the RMSd criteria, number of native H-bonds and several others. An example is shown in the Figure 12. All the results are in details shown in Supporting Information appendix. Specifically, the number of native H-bonds is an unsuitable parameter as different GQ folds that are competing on the folding landscape have very different sets of native H-bonds (see Figures 4 and 7). For example, while the 1KF1 and 143D structures both have 24 native H-bonds, they share only 2 of them. Thus, a common order parameter based on the number of native H-bonds is absolutely unsuitable to describe competition between these two structures. Note that the GQ folding landscape (see Introduction) may contain enormous numbers of major competing folds. Six of them actually have already been visualised as thermodynamic minima in structural experiments under different conditions, while detectable coexistence of multiple folds is common.

We suggest that human telomeric DNA GQ folding is principally non-describable by a limited set of order parameters and any such attempt would result in erroneous model of folding. The exceptionally complex kinetic partitioning folding landscape of GQs prevents any meaningful low-dimensional projections.[129] These would misrepresent or even remove barriers. Integrating over all but one or two degrees of freedom may produce population distributions where the connectivity information that determines transition rates is lost. If we choose an order parameter that averages over states on different sides of a high barrier, then kinetically isolated configurations can appear to be connected. In other words, GQ folding, with kinetics occurring on a time scale of days, is based on entirely different principles than funnel-like folding of fast-folding proteins or simple RNA and DNA hairpins. There is ~7-8 orders of magnitude difference in the folding rate between GQs and the fast-folding systems. Concepts of transitions states ensembles, transition paths and reaction coordinates are to our opinion inappropriate to describe GQ folding as a whole. GQ folding landscape could, in principle, be in future described by some comprehensive Markov State Model[128, 136] or disconnectivity graphs.[129]

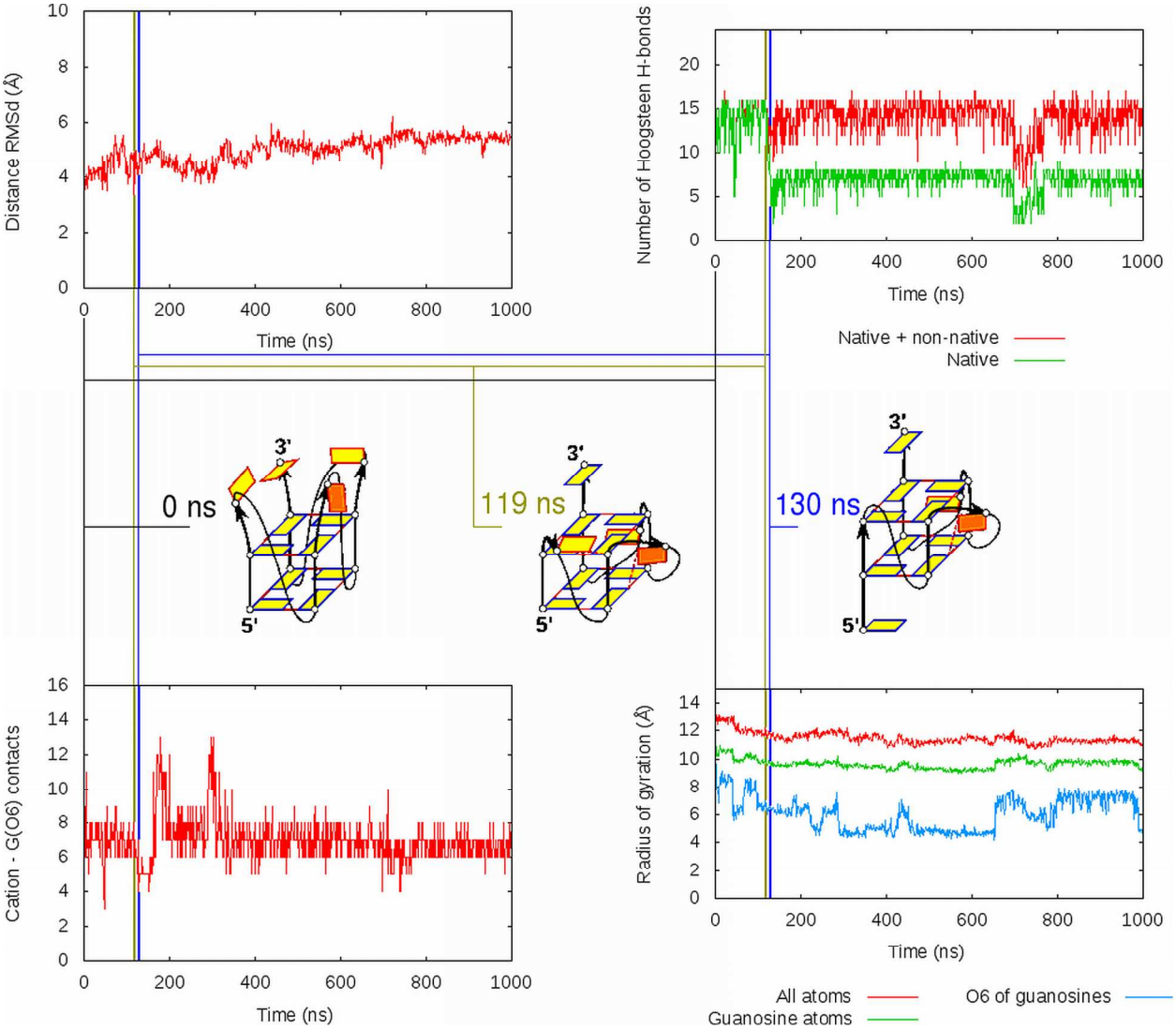


Figure 12. Development of various order parameters in the simulation 1KF1-**A**. The graphs are shown along with representative conformations found in the simulation, the rightmost being dominant till the end of the simulation, except of a temporary unfolding between 700 and 800 ns. The development of various quantities does not reflect the structural changes. For example, the partial unfolding is evident only in the graph monitoring the number of hydrogen bonds. On the other hand, the CV monitoring the number of native H-bonds has its own severe limitations, for example the decrease in the number of native bonds at ~120 ns cannot distinguish between quartet melting and strand slippage. It is absolute unsuitable to monitor transitions from one topology into another, as it does not differentiate most of the other folded structures from the unfolded state.

CG model allows changes of the GQ topology. The fixed temperature CG simulation was able to turn partially unfolded parallel stranded all-*anti* GQ 1KF1 into a two-quartet hybrid (3+1) GQ with mixed *syn* and *anti* Gs. It was also able to fully unfold the parallel stranded structure and fold it into an antiparallel alternative basket-type GQ (Figure 6). This is the first time that GQ structures were spontaneously formed in a simulation. Obviously, both new GQs have a combination of *syn*/*anti* nucleotides that do not correspond to any known human telomeric GQ

structure. However, it is not surprising because there could be many possible GQ architectures yet to be discovered experimentally, as pointed out in previous studies.[23, 46, 51, 53, 62] Many of them can be present only transiently in the process of folding and evade direct experimental observation with the available techniques.[48] Thus, the CG model is capable of investigating transitions between different GQ topologies, though the number of possible structures would require enormous computer resources to get a converged picture of the transitions. Also, fine-tuning of the CG potential will be necessary to appropriately balance relative stability of the different folds. The hairpin structure (Figure 8) may also potentially represent a new type of intermediates in the folding, though its formation might have been facilitated by the three T → C substitutions in the sequence, enhancing possibility of formation of canonical base pairs.

The results obviously might indicate that the newly formed GQ topologies are more stable in the CG description than the native 1KF1 structure. However, because we have seen only one transition event in each case, no thermodynamic conclusions can be made at this point. In addition, the 1KF1 structure is known to be preferred by crystal packing and in presence of certain cosolvents,[36] so the result may even be correct. Nevertheless, it may be necessary to specifically refine the CG DNA model based on known experimental data about GQ topology rules and other sources of data.[22-23, 69]

Atomistic simulations. Atomistic MD simulations aiming to stabilise intermediates obtained during unfolding ST CG simulations provided three basic scenarios: stabilisation of the intermediates with no further changes, partial or even complete GQ refolding, and progressive unfolding. In the parallel-stranded GQ simulations, strand slippage was the characteristic mechanism for unfolding and refolding (Figure 9). If further unfolding occurred in the antiparallel and hybrid GQs simulations, triplexes, hairpins and coil arrangements were observed (Figures S9 – S11). These data are quite consistent with previously published atomistic MD studies.[62, 70] In the cases of refolding of the antiparallel and hybrid species, slow restoration of quartets was observed (Figures 10, S10). The number of observed refolding trajectories was much lower than the number of steady or unfolding evolutions, confirming that spontaneous quartet formation is a rather rare event at the time scale of contemporary unbiased atomistic simulations.

Diversity of GQ rearrangements evidenced so far in computational studies. The events of unfolding for the parallel stranded GQ, its change into the antiparallel basket-type GQ via quartet nucleation, reversal of direction of one strand and formation of the hybrid (3+1) type GQ (Figure 6), gradual melting of the GQs (Figure 4), and scarce quartet reformation in the atomistic simulations (Figures 10, S10) point to high diversity of the (un)folding pathways (Figure 13). The folded GQs may be structured from (partially) unfolded, kinetically trapped, conformations by rare transition events.[63] The other (un)folding pathways observed in this work involve triplex and hairpin intermediates, whose existence has been suggested from experimental and previous computational data.[46, 53, 55, 58-60, 72-73] These intermediates could in principle participate in fast folding tracks with only a few intermediates as well as in the slow diffusive movements through the rugged free energy landscapes.[53]

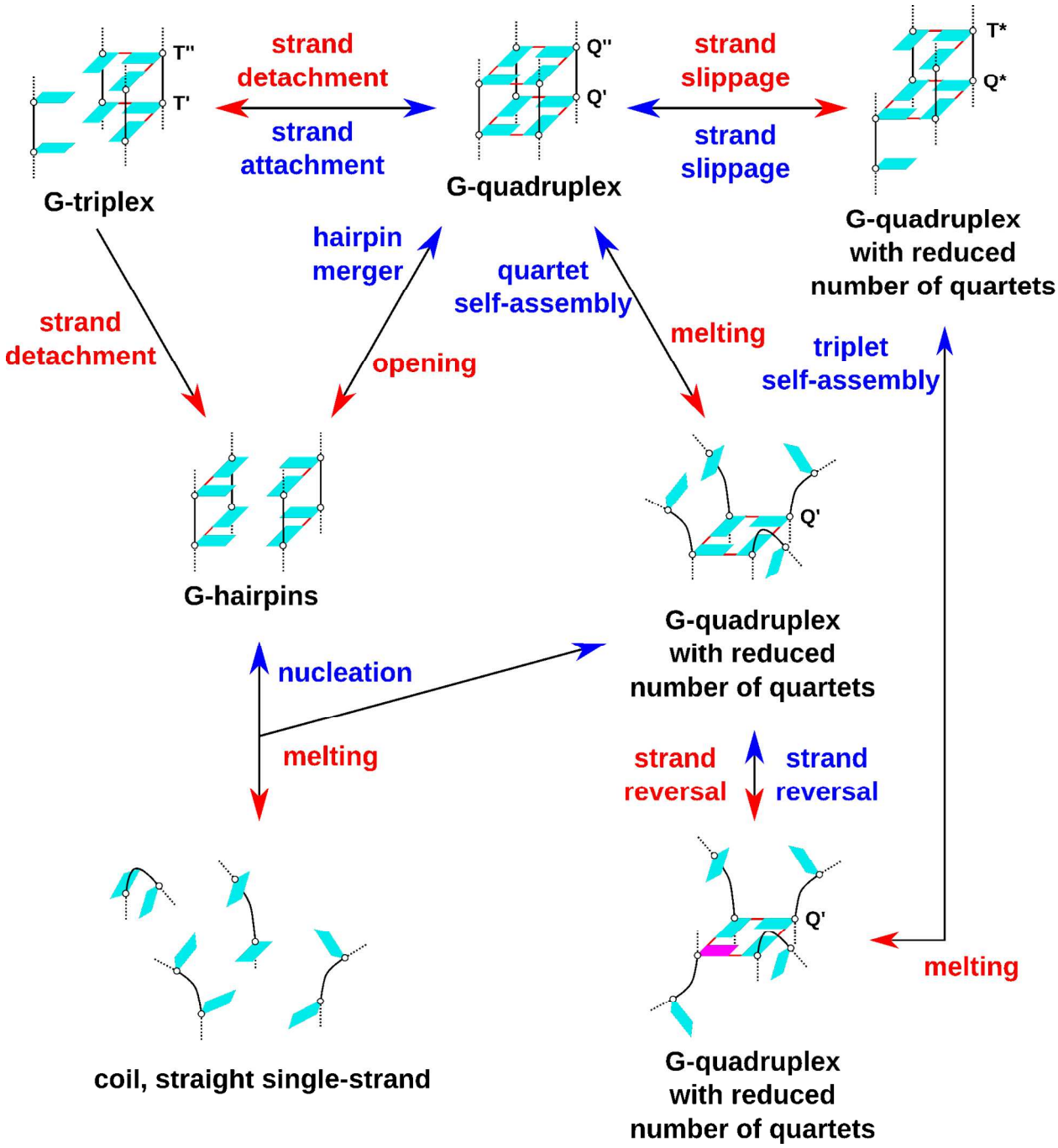


Figure 13. Sketch of various main types of structures and folding/unfolding interconversions between various intermediates for human telomeric GQs observed in simulation studies that have been carried here and elsewhere.[53, 62, 72, 75] For the sake of simplicity, the G-stem is marked only by two quartets, irrespective of whether it consists of two or three consecutive quartets. Gs are in cyan, Hoogsteen H-bonds are shown as red lines. Vertical strand slippage (top right) is possible only when all involved quartets have the same directionality (see also Figure S12). Mauve guanine in the bottom right scheme means that its glycosidic χ torsion angle is inverted upon the strand reversal (rotation) transition, i.e. it changes either from *syn* to *anti* or vice versa, while continuously keeping unperturbed G-quartet base-pairing. *Syn* to *anti* dynamics of unpaired Gs

(not shown) has been observed in the earlier T-REMD simulations[53] and in the present CG simulations.

Figure 13 summarises types of movements and interconversions between various potential intermediates in (un)folding of human telomeric GQs directly observed in this work and in previous atomistic MD studies.[62, 72, 75] Although the simulation description remains far from providing a full description of folding, the Figure suggests an enormous complexity of the folding landscape. We assume that all kinds of movements (and many others) coexist in structural dynamics of real GQ-folding sequences, leading to very complex folding landscape, which has been so far obscured by structural and temporal resolutions of the available techniques.

Advantages and limitations of the CG approach. Evolution of the CG trajectories is typically much faster compared to the atomistic simulations, due to the decreased number of particles and reduced spectrum of possible interactions in our CG model. Certainly, lack of explicit solvent plays a role. Taken together, these effects significantly reduce available conformational space, and thus avoid trapping the molecules in many deep free-energy basins and speed up sampling. Nevertheless, this is a double-edged sword and one should always be aware of what the model is able to describe and what is beyond its resolution. Every approximation may compromise the accuracy and excessive simplification can lead to physically meaningless simulation outcomes. Finding an optimal balance between computational speed and desired accuracy is a cornerstone of a successful CG model. As mentioned in the Methods section, one of the limitations of the current CG model is that only selected base-pairs are included in the potential function (Figure S3) and base-phosphate interactions are not defined at all. In addition, explicit ions are not considered. Hence, G-hairpin intermediates with reversed Watson-Crick GG base-pairs could not be attained, although they are suggested to form during folding of various GQ forming sequences.[48, 137 53] As the inclusion of other types of base-pairs not yet accounted for in HiRE-DNA can be easily achieved, modelling of base-phosphate interactions, and inclusion of explicit ions in the model, are our future most challenging development goals. On the other hand, the model in its current form showed qualitative match of GQ disintegration temperatures with melting temperatures found experimentally, though experimental GQ melting temperatures should in general be treated with caution.[138-142]

Our CG representation can certainly be improved. A common issue observed in our simulations was a loss of helical twist and non-systematic swelling of GQ cores, where the distance between one terminal quartet and the middle quartet was approximately 1.5 times greater than the distance between the other terminal quartet and the middle quartet (Figure 14). Preliminary results suggest that swelling is related to a combination of backbone dihedral interactions and lack of base rigidity. When bases are made more rigid by potentials favouring positioning of all nucleotide particles in the same plane, swelling is realistically reduced. Untwisting is most likely due to our representation of the stacking potential, where the energetic minimum is given by the bases piled precisely on top of each other, not including the shift typically observed in nucleic acids helices. Work is in progress to represent this term in the stacking energy. Nevertheless, already with the present model we have been able to observe many exciting structural developments and successfully reconstruct atomistic models of the CG trajectories for further analysis in atomistic simulations. The CG model can certainly be specifically trained for GQs, taking into account for example different free energies of different *syn-anti* patterns in the GQ stems[69] and other data.

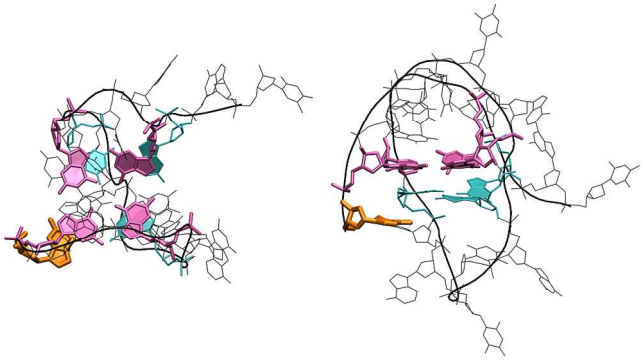


Figure 14. Underestimation of helical twist and non-systematic swelling in the CG model (converted to atomistic structures). Top and side view of the input structure for the simulation 2KF8-**C** shows underestimated helical twist of the stem and unphysical prolongation of the stacking distance between G1 (orange) and the bases of Q2 (mauve). The other Q1 bases (cyan) are in correct distance to Q2. Loop residues and the backbone are black.

CONCLUSIONS

We have carried out a first coarse-grained (CG) simulation study of unfolding of several topologies of human telomeric GQs, using the DNA version of the high-resolution CG model HiRE-DNA, supplemented by a series of explicit-solvent atomistic simulations. Our work confirms the clear potential of the CG methodology to be used in studies of GQs.

This simulation study has revealed additional complexity in the free-energy surface of the human telomeric sequence. New transition mechanisms between GQ topologies were observed, such as gradual melting of GQs initiated by loss of one of the outer quartets and strand slippage of partially unfolded GQs with antiparallel strands. The most interesting observations are probably two cases of the full conversion of the parallel-stranded all-*anti* GQ into alternative topologies, namely a two-quartet hybrid (3+1) GQ with a double-strand slippage and an antiparallel three-quartet basket-type GQ, albeit with opposite strand progression compared to the experimentally known basket topology. Each transition followed a significantly different folding pathway, nucleation from fully unfolded molecule vs. strand reversal in a partially folded GQ (Figure 6). The strand reversal is a large scale movement changing the direction of the strand while preserving one fully paired quartet via an *anti* - *syn* guanine flip (Figure 13). The results further support the unprecedented complexity of the folding landscape of the GQ-forming DNA sequences anticipated in preceding computational studies, which apparently is beyond the resolution of available experimental techniques. The very rugged folding landscape likely contains numerous individual free energy basins with diverse stabilities, lifetimes and routes to the unfolded ensemble. Some of them may act as long-living kinetic traps, in a highly complex network of folding and rearrangement pathways. The CG simulations thus further support the model of kinetic partitioning and diverse conformational diffusion processes taking place during GQs folding.

Combining CG sampling with fully-atomistic MD simulations appears to be a promising approach in studying principles of folding of GQs in future. Time scales affordable by CG will help us study events that are beyond the reach of classical atomistic MD. All-atom MD instead keeps its central role in the refinement of structures obtained by the CG model and in their detailed local analysis.

ASSOCIATED CONTENT

Supporting Information. Equilibration protocol, ST CG methodology, Figures S1 – S12 and Figures SA1 – SA50 summarising development of selected order parameters are available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

S.P.: Phone: +33 1 58 41 51 69 E-mail: samuela.pasquali@ibpc.fr. J.S.: Phone: +420 541 517 133; E-mail: sponer@ncbr.muni.cz.

ACKNOWLEDGEMENT

P.S. and J.S. were supported by the Czech Science Foundation grant 16-13721S. Institutional funding was in addition obtained from the project LO1305 (NPU programme) of the Ministry of Education, Youth and Sports of the Czech Republic. J.S. acknowledges support by Praemium Academiae. S.P. and P.D. were supported in part by the "Initiative d'Excellence" programme from the French State ("DYNAMO", ANR-11-LABX-0011-01) and P.D. also by IUF.

REFERENCES

(1) Huppert, J. L.; Balasubramanian, S., G-Quadruplexes in Promoters Throughout the Human Genome. *Nucleic Acids Res.* **2007**, *35*, 406-413.

(2) Huppert, J. L., Four-Stranded Nucleic Acids: Structure, Function and Targeting of G-Quadruplexes. *Chem. Soc. Rev.* **2008**, *37*, 1375-1384.

(3) Qin, M.; Chen, Z.; Luo, Q.; Wen, Y.; Zhang, N.; Jiang, H.; Yang, H., Two-Quartet G-Quadruplexes Formed by DNA Sequences Containing Four Contiguous GG Runs. *J. Phys. Chem. B* **2015**, *119*, 3706-3713.

(4) Duan, X.-L.; Liu, N.-N.; Yang, Y.-T.; Li, H.-H.; Li, M.; Dou, S.-X.; Xi, X.-G., G-Quadruplexes Significantly Stimulate Pif1 Helicase-catalyzed Duplex DNA Unwinding. *J. Biol. Chem.* **2015**, *290*, 7722-7735.

(5) Cayrou, C.; Coulombe, P.; Puy, A.; Rialle, S.; Kaplan, N.; Segal, E.; Méchali, M., New Insights into Replication Origin Characteristics in Metazoans. *Cell Cycle* **2012**, *11*, 658-667.

(6) Paeschke, K.; Bochman, M. L.; Garcia, P. D.; Cejka, P.; Friedman, K. L.; Kowalczykowski, S. C.; Zakian, V. A., Pif1 Family Helicases Suppress Genome Instability at G-Quadruplex Motifs. *Nature* **2013**, *497*, 458-462.

(7) Xu, S.; Li, Q.; Xiang, J.; Yang, Q.; Sun, H.; Guan, A.; Wang, L.; Liu, Y.; Yu, L.; Shi, Y., et al., Directly Lighting up RNA G-quadruplexes from Test Tubes to Living Human Cells. *Nucleic Acids Res.* **2015**, *43*, 9575-9586.

(8) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S., High-throughput Sequencing of DNA G-quadruplex Structures in the Human Genome. *Nat. Biotech.* **2015**, *33*, 877-881.

(9) Kwok, C. K.; Marsico, G.; Sahakyan, A. B.; Chambers, V. S.; Balasubramanian, S., rG4-seq Reveals Widespread Formation of G-Quadruplex Structures in the Human Transcriptome. *Nat. Methods* **2016**, *13*, 841-844.

(10) Moyzis, R. K.; Buckingham, J. M.; Cram, L. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J. R., A Highly Conserved Repetitive DNA Sequence, (TTAGGG)n, Present at the Telomeres of Human Chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 6622-6626.

(11) Wright, W. E.; Tesmer, V. M.; Huffman, K. E.; Levene, S. D.; Shay, J. W., Normal Human Chromosomes Have Long G-rich Telomeric Overhangs at One End. *Genes Dev.* **1997**, *11*, 2801-2809.
(12) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S., Quantitative Visualization of DNA G-quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, *5*, 182-186.
(13) Lam, E. Y. N.; Beraldi, D.; Tannahill, D.; Balasubramanian, S., G-quadruplex Structures Are Stable and Detectable in Human Genomic DNA. *Nat. Commun.* **2013**, *4*, 1796.
(14) Allsopp, R. C.; Vaziri, H.; Patterson, C.; Goldstein, S.; Younglai, E. V.; Futcher, A. B.; Greider, C. W.; Harley, C. B., Telomere Length Predicts Replicative Capacity of Human Fibroblasts. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 10114-10118.
(15) Engelhardt, M.; Martens, U. M., The Implication of Telomerase Activity and Telomere Stability for Replicative Aging and Cellular Immortality (Review). *Oncol. Rep.* **1998**, *5*, 1043-1052.
(16) Aubert, G.; Lansdorp, P. M., Telomeres and Aging. *Physiol. Rev.* **2008**, *88*, 557-579.
(17) Greider, C. W.; Blackburn, E. H., Identification of a Specific Telomere Terminal Transferase Activity in Tetrahymena Extracts. *Cell* **1985**, *43*, 405-413.
(18) Neidle, S., Human Telomeric G-quadruplex: The Current Status of Telomeric G-quadruplexes as Therapeutic Targets in Human Cancer. *Febs J.* **2010**, *277*, 1118-1125.
(19) Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M., Inhibition of Telomerase by G-quartet DNA Structures. *Nature* **1991**, *350*, 718-720.
(20) Balasubramanian, S.; Neidle, S., G-quadruplex Nucleic Acids as Therapeutic Targets. *Curr. Opin. Chem. Biol.* **2009**, *13*, 345-353.
(21) Leontis, N. B.; Westhof, E., Geometric Nomenclature and Classification of RNA Base Pairs. *RNA* **2001**, *7*, 499-512.
(22) da Silva, M. W., Geometric Formalism for DNA Quadruplex Folding. *Chem. Eur. J.* **2007**, *13*, 9738-9745.
(23) Karsisiotis, A. I.; O'Kane, C.; da Silva, M. W., DNA Quadruplex Folding Formalism - A Tutorial on Quadruplex Topologies. *Methods* **2013**, *64*, 28-35.
(24) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T., Structure of the Human Telomere in K+ Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers. *J. Am. Chem. Soc.* **2009**, *131*, 4301-4309.
(25) Lim, K. W.; Ng, V. C. M.; Martin-Pintado, N.; Heddi, B.; Phan, A. T., Structure of the Human Telomere in Na+ Solution: An Antiparallel (2+2) G-quadruplex Scaffold Reveals Additional Diversity. *Nucleic Acids Res.* **2013**, *41*, 10556-10562.
(26) Wang, Y.; Patel, D. J., Solution Structure of the Human Telomeric Repeat d[AG(3)(T(2)AG(3))3] G-tetraplex. *Structure* **1993**, *1*, 263-282.
(27) Parkinson, G. N.; Lee, M. P. H.; Neidle, S., Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA. *Nature* **2002**, *417*, 876-880.
(28) Ambrus, A.; Chen, D.; Dai, J. X.; Bialis, T.; Jones, R. A.; Yang, D. Z., Human Telomeric Sequence Forms a Hybrid-Type Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel Strands in Potassium Solution. *Nucleic Acids Res.* **2006**, *34*, 2723-2735.
(29) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D., Structure of the Hybrid-2 Type Intramolecular Human Telomeric G-quadruplex in K+ Solution: Insights into Structure Polymorphism of the Human Telomeric Sequence. *Nucleic Acids Res.* **2007**, *35*, 4927-4940.

(30) Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D., Structure of the Intramolecular Human Telomeric G-quadruplex in Potassium Solution: A Novel Adenine Triple Formation. *Nucleic Acids Res.* **2007**, *35*, 2440-2450.
(31) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J., Structure of the Human Telomere in K+ Solution: An Intramolecular (3+1) G-Quadruplex Scaffold. *J. Am. Chem. Soc.* **2006**, *128*, 9963-9970.
(32) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J., Structure of Two Intramolecular G-quadruplexes Formed by Natural Human Telomere Sequences in K+ Solution. *Nucleic Acids Res.* **2007**, *35*, 6517-6525.
(33) Abu-Ghazalah, R. M.; Rutledge, S.; Lau, L. W. Y.; Dubins, D. N.; Macgregor, R. B.; Helmy, A. S., Concentration-Dependent Structural Transitions of Human Telomeric DNA Sequences. *Biochemistry* **2012**, *51*, 7357-7366.
(34) Renciuk, D.; Kejnovska, I.; Skolakova, P.; Bednarova, K.; Motlova, J.; Vorlickova, M., Arrangements of Human Telomere DNA Quadruplex in Physiologically Relevant K+ Solutions. *Nucleic Acids Res.* **2009**, *37*, 6625-6634.
(35) Palacky, J.; Vorlickova, M.; Kejnovska, I.; Mojzes, P., Polymorphism of Human Telomeric Quadruplex Structure Controlled by DNA Concentration: A Raman Study. *Nucleic Acids Res.* **2013**, *41*, 1005-1016.
(36) Buscaglia, R.; Miller, M. C.; Dean, W. L.; Gray, R. D.; Lane, A. N.; Trent, J. O.; Chaires, J. B., Polyethylene Glycol Binding Alters Human Telomere G-quadruplex Structure by Conformational Selection. *Nucleic Acids Res.* **2013**, *41*, 7934-7946.
(37) Hansel, R.; Lohr, F.; Foldynova-Trantirkova, S.; Bamberg, E.; Trantirek, L.; Dotsch, V., The Parallel G-quadruplex Structure of Vertebrate Telomeric Repeat Sequences Is Not the Preferred Folding Topology under Physiological Conditions. *Nucleic Acids Res.* **2011**, *39*, 5768-5775.
(38) Hansel, R.; Lohr, F.; Trantirek, L.; Dotsch, V., High-Resolution Insight into G-Overhang Architecture. *J. Am. Chem. Soc.* **2013**, *135*, 2816-2824.
(39) Wang, Z.-F.; Li, M.-H.; Hsu, S.-T. D.; Chang, T.-C., Structural Basis of Sodium–Potassium Exchange of a Human Telomeric DNA Quadruplex Without Topological Conversion. *Nucleic Acids Res.* **2014**, *42*, 4723-4733.
(40) Zhang, X.; Xu, C.-X.; Di Felice, R.; Sponer, J.; Islam, B.; Stadlbauer, P.; Ding, Y.; Mao, L.; Mao, Z.-W.; Qin, P. Z., Conformations of Human Telomeric G-Quadruplex Studied Using a Nucleotide-Independent Nitroxide Label. *Biochemistry* **2016**, *55*, 360-372.
(41) Noer, S. L.; Preus, S.; Gudnason, D.; Aznauryan, M.; Mergny, J.-L.; Birkedal, V., Folding Dynamics and Conformational Heterogeneity of Human Telomeric G-quadruplex Structures in Na+ Solutions by Single Molecule FRET Microscopy. *Nucleic Acids Res.* **2016**, *44*, 464-471.
(42) Singh, V.; Azarkh, M.; Exner, T. E.; Hartig, J. S.; Drescher, M., Human Telomeric Quadruplex Conformations Studied by Pulsed EPR. *Angew. Chem., Int. Ed.* **2009**, *48*, 9728-9730.
(43) Azarkh, M.; Singh, V.; Okle, O.; Dietrich, D. R.; Hartig, J. S.; Drescher, M., Intracellular Conformations of Human Telomeric Quadruplexes Studied by Electron Paramagnetic Resonance Spectroscopy. *ChemPhysChem* **2012**, *13*, 1444-1447.
(44) Petraccone, L.; Malafronte, A.; Amato, J.; Giancola, C., G-quadruplexes from Human Telomeric DNA: How Many Conformations in PEG Containing Solutions? *J. Phys. Chem. B* **2012**, *116*, 2294-2305.

(45) Long, X.; Stone, M. D., Kinetic Partitioning Modulates Human Telomere DNA G-Quadruplex Structural Polymorphism. *PLoS One* **2013**, *8*, e83420.
(46) Gray, R. D.; Trent, J. O.; Chaires, J. B., Folding and Unfolding Pathways of the Human Telomeric G-Quadruplex. *J. Mol. Biol.* **2014**, *426*, 1629-1650.
(47) You, H.; Zeng, X.; Xu, Y.; Lim, C. J.; Efremov, A. K.; Phan, A. T.; Yan, J., Dynamics and Stability of Polymorphic Human Telomeric G-Quadruplex under Tension. *Nucleic Acids Res.* **2014**, *42*, 8789-8795.
(48) Bessi, I.; Jonker, H. R.; Richter, C.; Schwalbe, H., Involvement of Long-Lived Intermediate States in the Complex Folding Pathway of the Human Telomeric G-Quadruplex. *Angew. Chem., Int. Ed.* **2015**, *54*, 8444-8448.
(49) Armstrong, R. E.; Riskowski, R. A.; Strouse, G. F., Nanometal Surface Energy Transfer Optical Ruler for Measuring a Human Telomere Structure. *Photochem. Photobiol.* **2015**, *91*, 732-738.
(50) Xue, Y.; Liu, J.-Q.; Zheng, K.-W.; Kan, Z.-Y.; Hao, Y.-H.; Tan, Z., Kinetic and Thermodynamic Control of G-Quadruplex Folding. *Angew. Chem., Int. Ed.* **2011**, *50*, 8046-8050.
(51) Lannan, F. M.; Mamajanov, I.; Hud, N. V., Human Telomere Sequence DNA in Water-Free and High-Viscosity Solvents: G-Quadruplex Folding Governed by Kramers Rate Theory. *J. Am. Chem. Soc.* **2012**, *134*, 15324-15330.
(52) Gabelica, V., A Pilgrim's Guide to G-Quadruplex Nucleic Acid Folding. *Biochimie* **2014**, *105C*, 1-3.
(53) Stadlbauer, P.; Kuhrova, P.; Banas, P.; Koca, J.; Bussi, G.; Trantirek, L.; Otyepka, M.; Sponer, J., Hairpins Participating in Folding of Human Telomeric Sequence Quadruplexes Studied by Standard and T-REMD Simulations. *Nucleic Acids Res.* **2015**, *43*, 9626-9644.
(54) Yu, Z.; Mao, H., Non-B DNA Structures Show Diverse Conformations and Complex Transition Kinetics Comparable to RNA or Proteins—A Perspective from Mechanical Unfolding and Refolding Experiments. *Chem. Rec.* **2013**, *13*, 102-116.
(55) Mashimo, T.; Yagi, H.; Sannohe, Y.; Rajendran, A.; Sugiyama, H., Folding Pathways of Human Telomeric Type-1 and Type-2 G-quadruplex Structures. *J. Am. Chem. Soc.* **2010**, *132*, 14910-14918.
(56) Koirala, D.; Mashimo, T.; Sannohe, Y.; Yu, Z. B.; Mao, H. B.; Sugiyama, H., Intramolecular Folding in Three Tandem Guanine Repeats of Human Telomeric DNA. *Chem. Commun.* **2012**, *48*, 2006-2008.
(57) Zhang, A. Y. Q.; Balasubramanian, S., The Kinetics and Folding Pathways of Intramolecular G-Quadruplex Nucleic Acids. *J. Am. Chem. Soc.* **2012**, *134*, 19297-19308.
(58) Li, Y.; Liu, C.; Feng, X. J.; Xu, Y. Z.; Liu, B. F., Ultrafast Microfluidic Mixer for Tracking the Early Folding Kinetics of Human Telomere G-Quadruplex. *Anal. Chem.* **2014**, *86*, 4333-4339.
(59) Rajendran, A.; Endo, M.; Hidaka, K.; Sugiyama, H., Direct and Single-Molecule Visualization of the Solution-State Structures of G-Hairpin and G-Triplex Intermediates. *Angew. Chem.* **2014**, *126*, 4191-4196.
(60) Rajendran, A.; Endo, M.; Hidaka, K.; Teulade-Fichou, M.-P.; Mergny, J.-L.; Sugiyama, H., Small Molecule Binding to a G-hairpin and a G-triplex: A New Insight into Anticancer Drug Design Targeting G-rich Regions. *Chem. Commun.* **2015**, *51*, 9181-9184.
(61) Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T., Extreme Conformational Diversity in Human Telomeric DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 18938-18943.

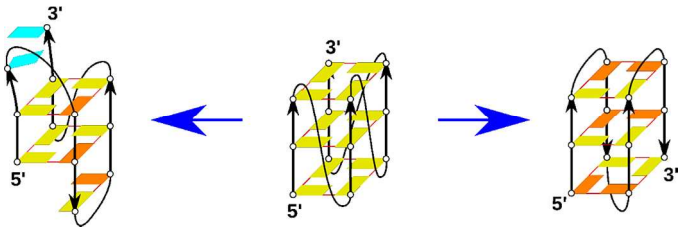
(62) Stadlbauer, P.; Krepl, M.; Cheatham, T. E., 3rd; Koca, J.; Sponer, J., Structural Dynamics of Possible Late-Stage Intermediates in Folding of Quadruplex DNA Studied by Molecular Simulations. *Nucleic Acids Res.* **2013**, *41*, 7128-7143.
(63) Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C., Theoretical Perspectives on Protein Folding. *Annu. Rev. Biophys.* **2010**, *39*, 159-183.
(64) Galindo-Murillo, R.; Robertson, J. C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham, T. E., Assessing the Current State of AMBER Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114-4127.
(65) Islam, B.; Sgobba, M.; Laughton, C.; Orozco, M.; Sponer, J.; Neidle, S.; Haider, S., Conformational Dynamics of the Human Propeller Telomeric DNA Quadruplex on a Microsecond Time Scale. *Nucleic Acids Res.* **2013**, *41*, 2723-2735.
(66) Wei, D.; Husby, J.; Neidle, S., Flexibility and Structural Conservation in a c-KIT G-quadruplex. *Nucleic Acids Res.* **2015**, *43*, 629-644.
(67) Ghosh, S.; Jana, J.; Kar, R. K.; Chatterjee, S.; Dasgupta, D., Plant Alkaloid Chelerythrine Induced Aggregation of Human Telomere Sequence-A Unique Mode of Association between a Small Molecule and a Quadruplex. *Biochemistry* **2015**, *54*, 974-986.
(68) Reshetnikov, R.; Golovin, A.; Spiridonova, V.; Kopylov, A.; Sponer, J., Structural Dynamics of Thrombin-Binding DNA Aptamer d(GGTTGGTGTGGTTGG) Quadruplex DNA Studied by Large-Scale Explicit Solvent Simulations. *J. Chem. Theory Comput.* **2010**, *6*, 3003-3014.
(69) Cang, X. H.; Sponer, J.; Cheatham, T. E., Explaining the Varied Glycosidic Conformational, G-Tract Length and Sequence Preferences for Anti-Parallel G-Quadruplexes. *Nucleic Acids Res.* **2011**, *39*, 4499-4512.
(70) Stefl, R.; Cheatham, T. E.; Spackova, N.; Fadrna, E.; Berger, I.; Koca, J.; Sponer, J., Formation Pathways of a Guanine-Quadruplex DNA Revealed by Molecular Dynamics and Thermodynamic Analysis of the Substates. *Biophys. J.* **2003**, *85*, 1787-1804.
(71) Limongelli, V.; De Tito, S.; Cerofolini, L.; Fragai, M.; Pagano, B.; Trotta, R.; Cosconati, S.; Marinelli, L.; Novellino, E.; Bertini, I., et al., The G-Triplex DNA. *Angew. Chem., Int. Ed.* **2013**, *52*, 2269-2273.
(72) Stadlbauer, P.; Trantirek, L.; Cheatham, T. E., 3rd; Koca, J.; Sponer, J., Triplex Intermediates in Folding of Human Telomeric Quadruplexes Probed by Microsecond-scale Molecular Dynamics Simulations. *Biochimie* **2014**, *105*, 22-35.
(73) Li, W.; Hou, X.-M.; Wang, P.-Y.; Xi, X.-G.; Li, M., Direct Measurement of Sequential Folding Pathway and Energy Landscape of Human Telomeric G-quadruplex Structures. *J. Am. Chem. Soc.* **2013**, *135*, 6423-6426.
(74) Li, H.; Cao, E. H.; Gisler, T., Force-induced Unfolding of Human Telomeric G-quadruplex: A Steered Molecular Dynamics Simulation Study. *Biochem. Biophys. Res. Commun.* **2009**, *379*, 70-75.
(75) Bian, Y.; Tan, C.; Wang, J.; Sheng, Y.; Zhang, J.; Wang, W., Atomistic Picture for the Folding Pathway of a Hybrid-1 Type Human Telomeric DNA G-quadruplex. *PLoS Comput. Biol.* **2014**, *10*, e1003562.
(76) Yang, C.; Jang, S.; Pak, Y., Multiple Stepwise Pattern for Potential of Mean Force in Unfolding the Thrombin Binding Aptamer in Complex with Sr2+. *J. Chem. Phys.* **2011**, *135*, 225104.
(77) Kim, E.; Yang, C.; Pak, Y., Free-energy Landscape of a Thrombin-binding DNA Aptamer in Aqueous Environment. *J. Chem. Theory Comput.* **2012**, *8*, 4845-4851.

(78) Bergues-Pupo, A. E.; Arias-Gonzalez, J. R.; Moron, M. C.; Fiasconaro, A.; Falo, F., Role of the Central Cations in the Mechanical Unfolding of DNA and RNA G-quadruplexes. *Nucleic Acids Res.* **2015**, *43*, 7638-7647.
(79) Sponer, J.; Banas, P.; Jurecka, P.; Zgarbova, M.; Kuhrova, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M., Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. *J. Phys. Chem. Lett.* **2014**, *5*, 1771-1782.
(80) Abrams, C.; Bussi, G., Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2013**, *16*, 163-199.
(81) Best, R. B., Atomistic Molecular Simulations of Protein Folding. *Curr. Opin. Struct. Biol.* **2012**, *22*, 52-61.
(82) Sponer, J.; Cang, X. H.; Cheatham, T. E., Molecular Dynamics Simulations of G-DNA and Perspectives on the Simulation of Nucleic Acid Structures. *Methods* **2012**, *57*, 25-39.
(83) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E., Highly Sampled Tetranucleotide and Tetraloop Motifs Enable Evaluation of Common RNA Force Fields. *RNA* **2015**, *21*, 1578-1590.
(84) Gkionis, K.; Kruse, H.; Platts, J. A.; Mladek, A.; Koca, J.; Sponer, J., Ion Binding to Quadruplex DNA Stems. Comparison of MM and QM Descriptions Reveals Sizable Polarization Effects Not Included in Contemporary Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 1326-1340.
(85) Kruse, H.; Mladek, A.; Gkionis, K.; Hansen, A.; Grimme, S.; Sponer, J., Quantum Chemical Benchmark Study on 46 RNA Backbone Families Using a Dinucleotide Unit. *J. Chem. Theory Comput.* **2015**, *11*, 4972-4991.
(86) Cragnolini, T.; Derreumaux, P.; Pasquali, S., Ab Initio RNA Folding. *J. Phys.: Condens. Matter* **2015**, *27*, 233102.
(87) Sponer, J.; Spackova, N., Molecular Dynamics Simulations and Their Application to Four-stranded DNA. *Methods* **2007**, *43*, 278-290.
(88) Hyeon, C.; Thirumalai, D., Mechanical Unfolding of RNA Hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6789-6794.
(89) Xia, Z.; Gardner, D. P.; Gutell, R. R.; Ren, P., Coarse-Grained Model for Simulation of RNA Three-Dimensional Structures. *J. Phys. Chem. B* **2010**, *114*, 13497-13506.
(90) Dans, P. D.; Walther, J.; Gomez, H.; Orozco, M., Multiscale Simulation of DNA. *Curr. Opin. Struct. Biol.* **2016**, *37*, 29-45.
(91) He, Y.; Maciejczyk, M.; Ołdziej, S.; Scheraga, H. A.; Liwo, A., Mean-Field Interactions between Nucleic-Acid-Base Dipoles can Drive the Formation of a Double Helix. *Phys. Rev. Lett.* **2013**, *110*, 098101.
(92) Cho, S. S.; Pincus, D. L.; Thirumalai, D., Assembly Mechanisms of RNA Pseudoknots Are Determined by the Stabilities of Constituent Secondary Structures. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 17349-17354.
(93) Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S., A Systematically Coarse-grained Model for DNA and Its Predictions for Persistence Length, Stacking, Twist, and Chirality. *J. Chem. Phys.* **2010**, *132*, 035105.
(94) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K., DNA Nanotweezers Studied with a Coarse-Grained Model of DNA. *Phys. Rev. Lett.* **2010**, *104*, 178101.
(95) Markegard, C. B.; Fu, I. W.; Reddy, K. A.; Nguyen, H. D., Coarse-Grained Simulation Study of Sequence Effects on DNA Hybridization in a Concentrated Environment. *J. Phys. Chem. B* **2015**, *119*, 1823-1834.

(96) Rebic, M.; Mocci, F.; Laaksonen, A.; Ulicny, J., Multiscale Simulations of Human Telomeric G-Quadruplex DNA. *J. Phys. Chem. B* **2015**, *119*, 105-113.
(97) Wu, X.; Xu, P. J.; Wang, J. G.; Xu, Y.; Fu, T.; Zhao, M. X.; Zhang, D. P.; Liu, J. H.; Shen, H. J.; Xiu, Z. L., et al., Theoretical Studies on the Folding Mechanisms for Different DNA G-quadruplexes. In *Advance in Structural Bioinformatics*, Wei, D.; Xu, Q.; Zhao, T.; Dai, H., Eds. Springer-Verlag Berlin: Berlin, 2015; Vol. 827, pp 123-141.
(98) Cragnolini, T.; Derreumaux, P.; Pasquali, S., Coarse-Grained Simulations of RNA and DNA Duplexes. *J. Phys. Chem. B* **2013**, *117*, 8047-8060.
(99) Cragnolini, T.; Laurin, Y.; Derreumaux, P.; Pasquali, S., Coarse-Grained HiRE-RNA Model for ab Initio RNA Folding beyond Simple Molecules, Including Noncanonical and Multiple Base Pairings. *J. Chem. Theory Comput.* **2015**, *11*, 3510-3522.
(100) Islam, B.; Stadlbauer, P.; Krepl, M.; Koca, J.; Neidle, S.; Haider, S.; Sponer, J., Extended Molecular Dynamics of a c-kit Promoter Quadruplex. *Nucleic Acids Res.* **2015**, *43*, 8673-8693.
(101) Lim, K. W.; Alberti, P.; Guedin, A.; Lacroix, L.; Riou, J.-F.; Royle, N. J.; Mergny, J.-L.; Phan, A. T., Sequence Variant (CTAGGG)n in the Human Telomere Favors a G-quadruplex Structure Containing a G·C·G·C Tetrad. *Nucleic Acids Res.* **2009**, *37*, 6239-6248.
(102) Zhu, H.; Xiao, S.; Liang, H., Structural Dynamics of Human Telomeric G-Quadruplex Loops Studied by Molecular Dynamics Simulations. *PLoS One* **2013**, *8*, e71380.
(103) Pasquali, S.; Derreumaux, P., HiRE-RNA: A High Resolution Coarse-Grained Energy Model for RNA. *J. Phys. Chem. B* **2010**, *114*, 11957-11966.
(104) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K., Structural, Mechanical, and Thermodynamic Properties of a Coarse-Grained DNA Model. *J. Chem. Phys.* **2011**, *134*, 085101.
(105) Denesyuk, N. A.; Thirumalai, D., Coarse-Grained Model for Predicting RNA Folding Thermodynamics. *J. Phys. Chem. B* **2013**, *117*, 4901-4911.
(106) SantaLucia, J., A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 1460-1465.
(107) Sterpone, F.; Melchionna, S.; Tuffery, P.; Pasquali, S.; Mousseau, N.; Cragnolini, T.; Chebaro, Y.; St-Pierre, J.-F.; Kalimeri, M.; Barducci, A., et al., The OPEP Protein Model: from Single Molecules, Amyloid Formation, Crowding and Hydrodynamics to DNA/RNA Systems. *Chem. Soc. Rev.* **2014**, *43*, 4871-4893.
(108) Sulc, P.; Romano, F.; Ouldridge, T. E.; Doye, J. P. K.; Louis, A. A., A Nucleotide-Level Coarse-Grained Model of RNA. *J. Chem. Phys.* **2014**, *140*, 235102.
(109) Theimer, C. A.; Blois, C. A.; Feigon, J., Structure of the Human Telomerase RNA Pseudoknot Reveals Conserved Tertiary Interactions Essential for Function. *Mol. Cell* **2005**, *17*, 671-682.
(110) Cho, H. M.; Chu, J.-W., Inversion of Radial Distribution Functions to Pair Forces by Solving the Yvon–Born–Green Equation Iteratively. *J. Chem. Phys.* **2009**, *131*, 134107.
(111) Neupane, K.; Ritchie, D. B.; Yu, H.; Foster, D. A. N.; Wang, F.; Woodside, M. T., Transition Path Times for Nucleic Acid Folding Determined from Energy-Landscape Analysis of Single-Molecule Trajectories. *Phys. Rev. Lett.* **2012**, *109*, 068102.
(112) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

(113) Joung, I. S.; Cheatham, T. E., Determination of Alkali and Halide Monovalent Ion Parameters for Use In Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020-9041.
(114) Case, D. A. D., T.; Cheatham, T. E., III; Simmerling, C. L.;; Wang, J. D., R. E.; Luo, R.; Walker, R., Zhang, W.;; Merz, K. M. R., S.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A.W.;; Kolossvai, I. W., K. F. ; Paesani, F.; Vanicek, J.; Wolf, R.M.; Liu, J.; Wu, X.;; Brozell, S. R. S., T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D.R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui,; C.; Babin, V. L., T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*, University of California: San Francisco, 2012.
(115) Fadrna, E.; Spackova, N.; Sarzynska, J.; Koca, J.; Orozco, M.; Cheatham, T. E.; Kulinski, T.; Sponer, J., Single Stranded Loops of Quadruplex DNA as Key Benchmark for Testing Nucleic Acids Force Fields. *J. Chem. Theory Comput.* **2009**, *5*, 2514-2530.
(116) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M., Refinenement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007**, *92*, 3817-3829.
(117) Krepl, M.; Zgarbova, M.; Stadlbauer, P.; Otyepka, M.; Banas, P.; Koca, J.; Cheatham, T. E.; Jurecka, P.; Sponer, J., Reference Simulations of Noncanonical Nucleic Acids with Different Chi Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8*, 2506-2520.
(118) Zgarbova, M.; Luque, F. J.; Sponer, J.; Cheatham, T. E.; Otyepka, M.; Jurecka, P., Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013**, *9*, 2339-2354.
(119) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A., Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers - Charge Derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* **1995**, *16*, 1357-1377.
(120) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
(121) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
(122) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of Cartesian Equations of Motion of a System with Constraints - Molecular Dynamics of N-alkans. *J. Comput. Phys.* **1977**, *23*, 327-341.
(123) Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald - An N.log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
(124) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
(125) Salomon-Ferrer, R.; Goetz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878-3888.
(126) Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.
(127) Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Mol. Graphics Modell.* **1996**, *14*, 33-38.

(128) Noe, F.; Fischer, S., Transition Networks for Modeling the Kinetics of Conformational Change in Macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154-162.
(129) Wales, D. J., Energy Landscapes: Some New Horizons. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3-10.
(130) Krivov, S. V.; Karplus, M., Hidden Complexity of Free Energy Surfaces for Peptide (Protein) Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14766-14770.
(131) Krivov, S. V.; Karplus, M., One-Dimensional Free-Energy Profiles of Complex Systems: Progress Variables that Preserve the Barriers. *J. Phys. Chem. B* **2006**, *110*, 12689-12698.
(132) Muff, S.; Caflisch, A., Kinetic Analysis of Molecular Dynamics Simulations Reveals Changes in the Denatured State and Switch of Folding Pathways upon Single-Point Mutation of a β-sheet Miniprotein. *Proteins: Structure, Function and Genetics* **2008**, *70*, 1185-1195.
(133) Krivov, S. V.; Karplus, M., Diffusive Reaction Dynamics on Invariant Free Energy Profiles. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13841-13846.
(134) Dickson, B. M.; Makarov, D. E.; Henkelman, G., Pitfalls of Choosing an Order Parameter for Rare Event Calculations. *J. Chem. Phys.* **2009**, *131*, 074108.
(135) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L., Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291-318.
(136) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S., To Milliseconds and Beyond: Challenges in the Simulation of Protein Folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 58-65.
(137) Ceru, S.; Sket, P.; Prislan, I.; Lah, J.; Plavec, J., A New Pathway of DNA G-Quadruplex Formation. *Angew. Chem., Int. Ed.* **2014**, *53*, 4881-4884.
(138) Mergny, J. L.; Phan, A. T.; Lacroix, L., Following G-quartet Formation by UV-Spectroscopy. *Febs Lett.* **1998**, *435*, 74-78.
(139) Gray, R. D.; Chaires, J. B., Analysis of Multidimensional G-Quadruplex Melting Curves. In *Curr. Protoc. Nucleic Acid Chem.*, **2011**, doi: 10.1002/0471142700.nc1704s45
(140) Buscaglia, R.; Gray, R. D.; Chaires, J. B., Thermodynamic Characterization of Human Telomere Quadruplex Unfolding. *Biopolymers* **2013**, *99*, 1006-1018.
(141) Li, Y. Y.; Macgregor, R. B., A Thermodynamic Study of Adenine and Thymine Substitutions in the Loops of the Oligodeoxyribonucleotide HTel. *J. Phys. Chem. B* **2016**, *120*, 8830-8836.
(142) Gray, R. D.; Buscaglia, R.; Chaires, J. B., Populated Intermediates in the Thermal Unfolding of the Human Telomeric Quadruplex. *J. Am. Chem. Soc.* **2012**, *134*, 16834-16844.



TOC graphics

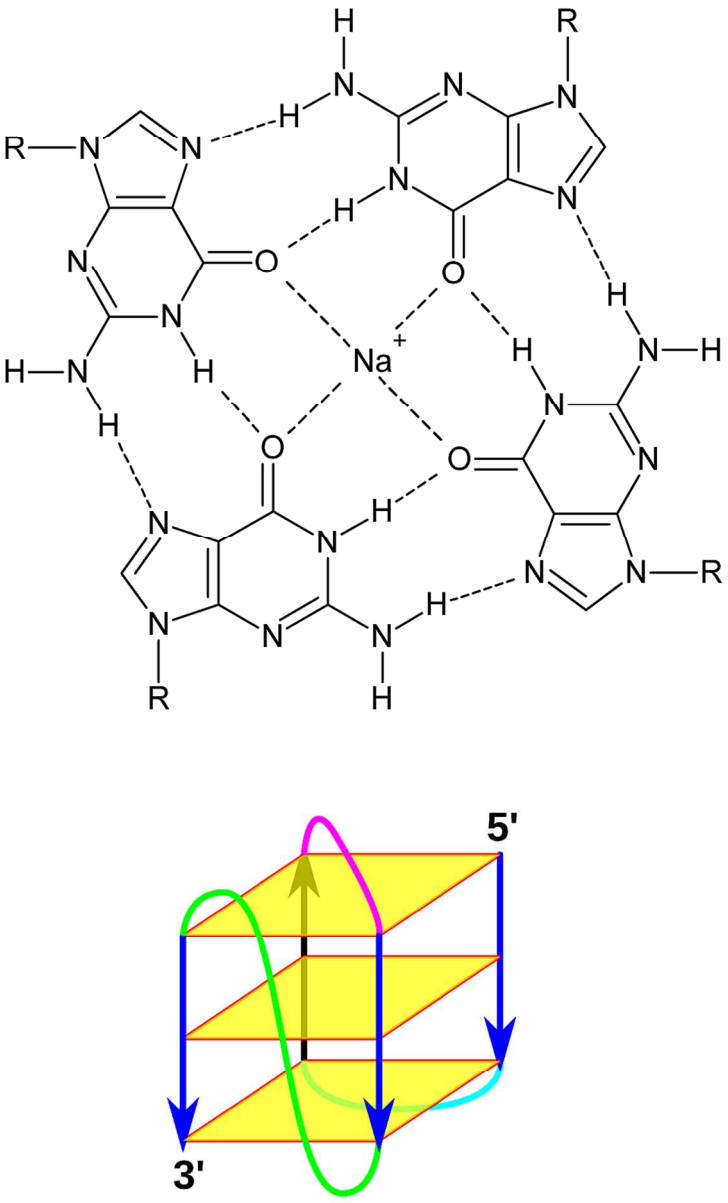


Figure 1. GQ structure. Top: A G-quartet with a sodium cation coordinated in its centre (it can either be coplanar with the quartet or reside between two consecutive quartets; potassium cation can only reside between two quartets). R stands for the sugar-phosphate moiety. Bottom: Schematic intramolecular three-quartet GQ. Quartets are simplified as yellow squares. Blue G-strands are mutually parallel, while the black one is antiparallel to them. The cyan loop is lateral, the mauve loop is diagonal. They connect antiparallel G-stretches. The green loop is propeller and it connects parallel G-stretches. Ions are not shown in the scheme. Although not shown in the Figure, a given GQ topology is interconnected with characteristic patterns of anti and syn orientations of the G-nucleotides.23

80x134mm (300 x 300 DPI)

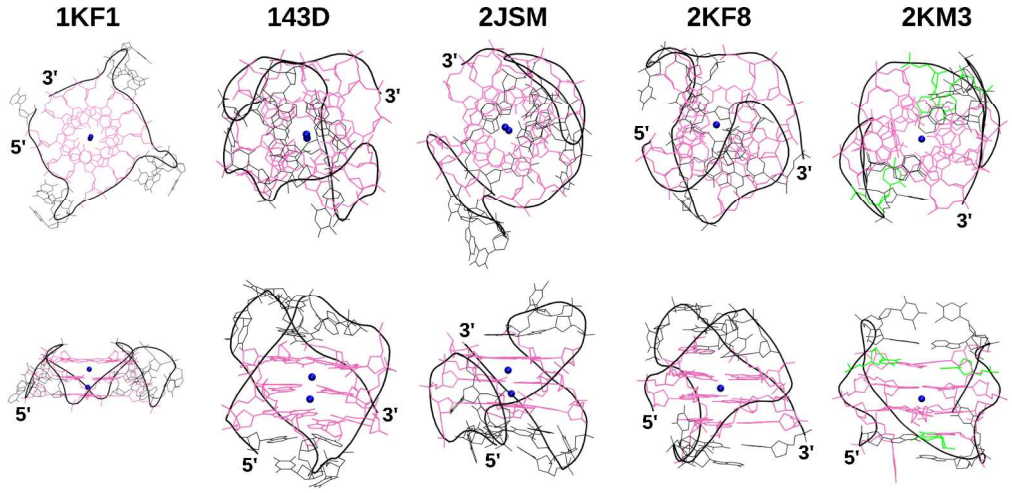


Figure 2. Top view (top) and side view (bottom) of the studied GQs. Gs are in mauve, loop residues and the backbone in black, and channel ions in blue. Cs in the case of 2KM3 are depicted in green.

171x82mm (300 x 300 DPI)

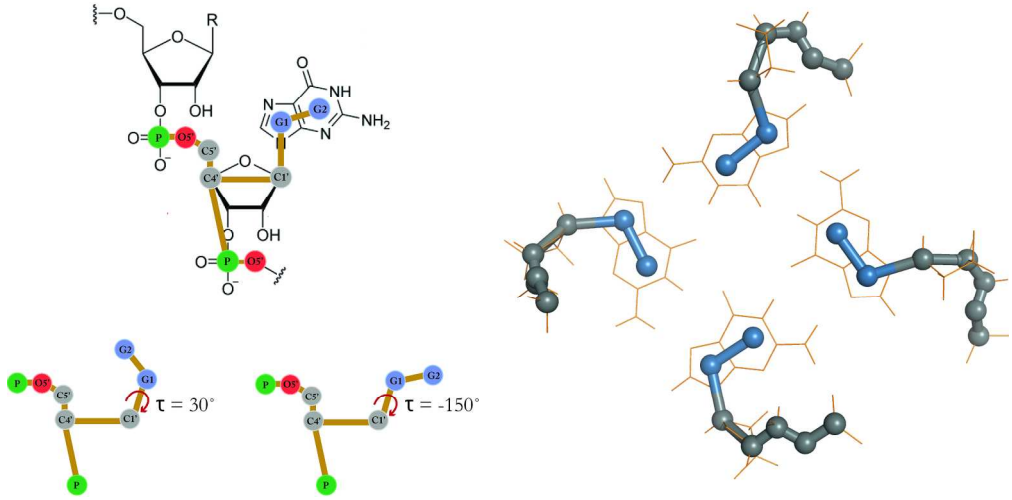


Figure 3. CG representation of a guanine nucleotide and a G-quartet. Top left: CG guanine nucleotide superposed to the atomistic structure. Bottom left: CG representation of syn (τ = 30°) and anti (τ = -150°) G-nucleotide conformations with τ equilibrium values in the dihedral potential (τ corresponds to the χ glycosidic torsion angle). Right: Atomistic and CG view of a G-quartet.

177x88mm (300 x 300 DPI)

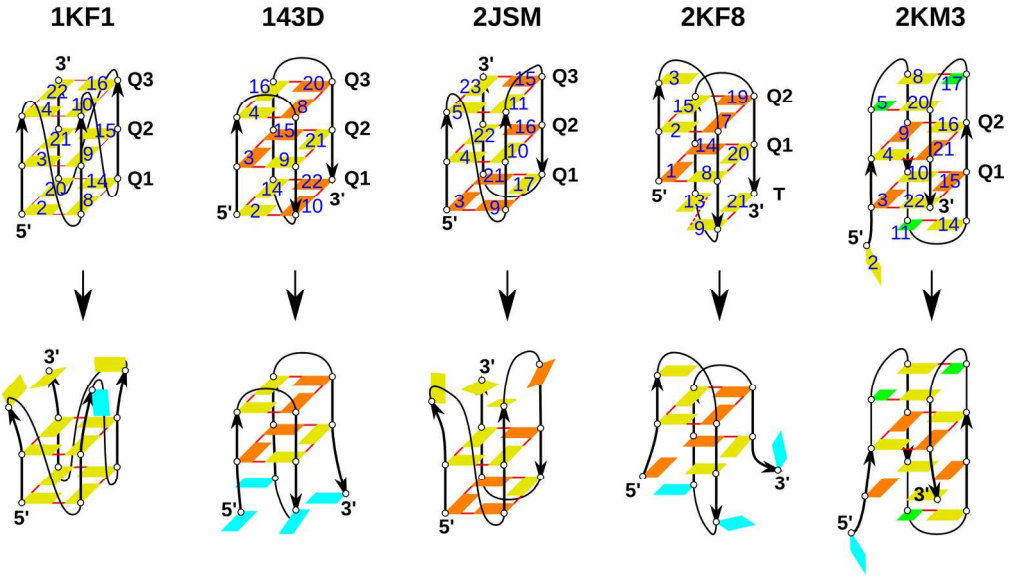


Figure 4. Schemes of native structures (top) and initial stage of unfolding (bottom) of GQs in ST CG simulations. First step of unfolding is disruption (loss of pairing) of one terminal quartet. G bases represented in yellow are in anti conformation while bases in orange are in syn conformation. Cyan bases fluctuate between anti and syn. Ions and loop residues are not shown. Cytosines in 2KM3 are shown as green squares. G-quartets (Q) are numbered from the 5′-end. 2KF8 contains a triplet (T). Numbering of Gs is shown by blue numbers.

171x97mm (300 x 300 DPI)

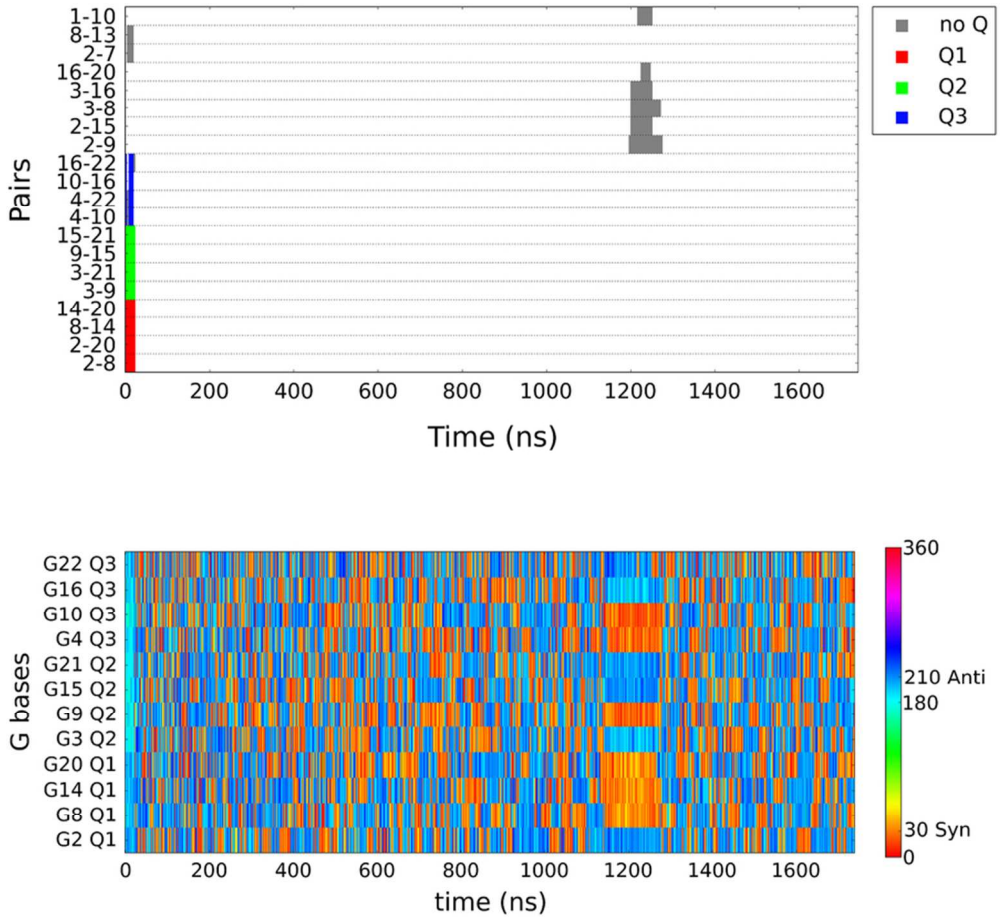


Figure 5. ST CG unfolding of 1KF1. Top: All base pairs detected during the simulation. Bottom: Anti-syn dynamics of the G bases. The native GQ unfolds in the first few tens of ns. After 1.1 μs, a transient triplex is formed for about 100 ns. See Supporting Information Figures S4-S7 for data for the remaining systems.

76x69mm (300 x 300 DPI)

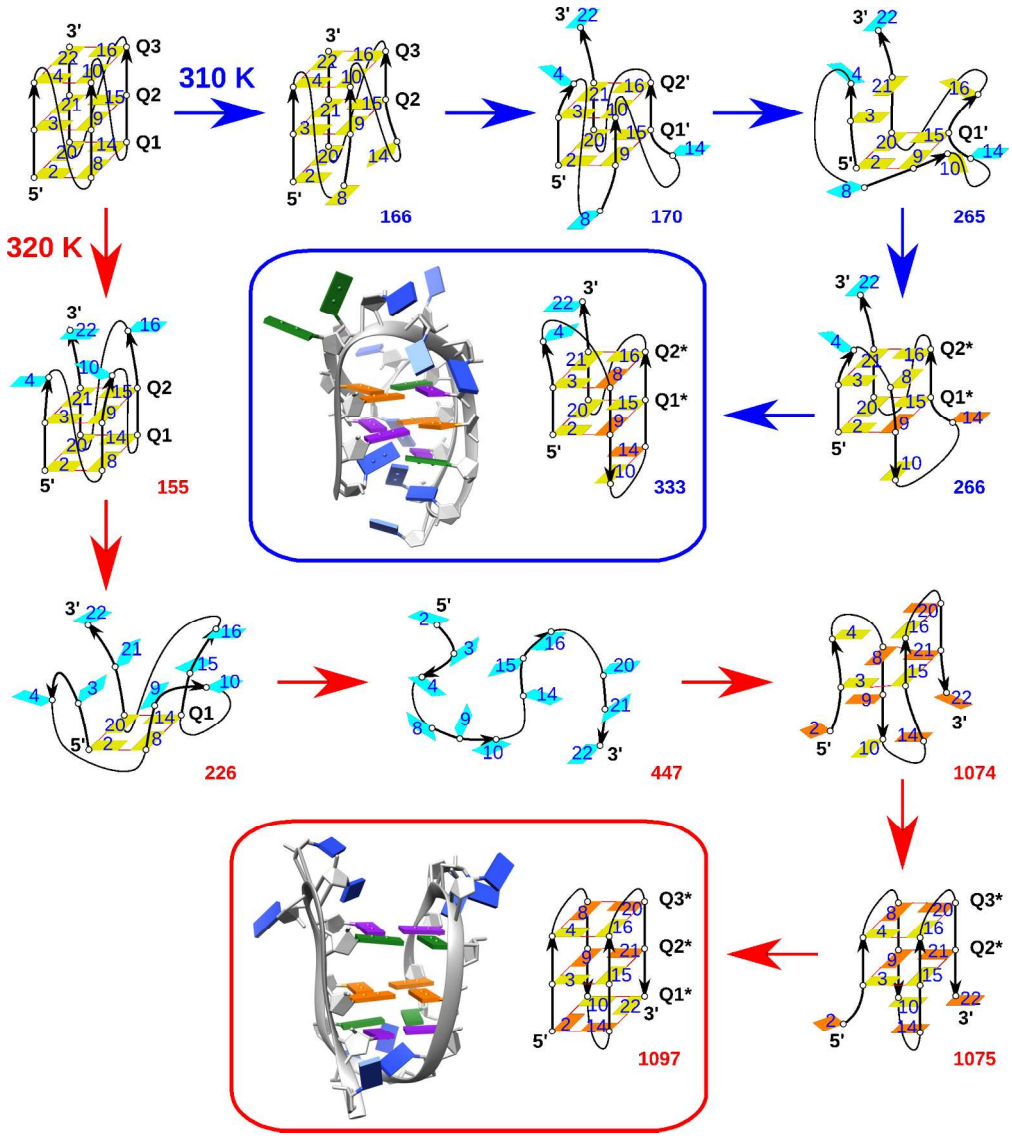


Figure 6. Unfolding-folding events in fixed temperature CG simulations of the parallel stranded GQ 1KF1 at 310 K and 320 K. At 310 K, a new two-quartet hybrid (3+1) type GQ with double strand-slippage is formed (in the blue box). At 320, a new three-quartet antiparallel alternative basket-type GQ is formed (in the red box). Folding mechanism of the two GQs is different. The structural schemes (yellow, orange and cyan bases) are visualised as in the Figure 4; numbering of Gs is shown by blue numbers inside the bases. The number next to each scheme denotes time (ns), when the structure appears in the simulation. In the structural snapshots of the two newly folded GQs, bases in mauve belong to the first quartet (Q1) in the original 1KF1 structure, those in orange to Q2 and those in green to Q3. The remaining bases (flanking and loop Ts and As) are in blue.

177x199mm (300 x 300 DPI)

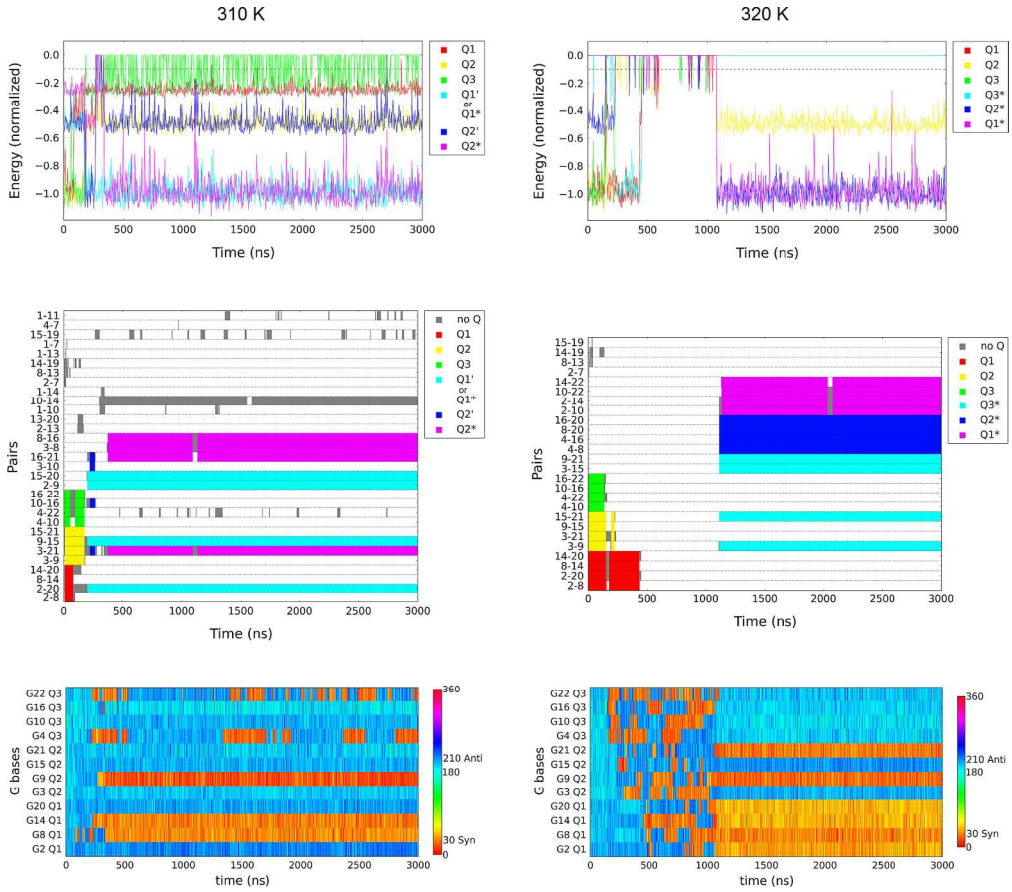


Figure 7. Fixed-temperature CG simulation of 1KF1 at 310K and 320 K. Top row: Overall energy of base-pairing of the quartets as a function of time computed as the sum of the energies of the single pairs, where each pair's energy has been normalised to its maximum possible absolute value (the 10% cutoff on the pairs' energies is shown as a dashed line). Middle row: All base pairs observed in the simulations coloured as a function of the quartet they belong to. Bottom row: Anti-syn conformation of all G bases. Structural development in the two simulations is depicted in the Figure 6.

173x152mm (299 x 299 DPI)

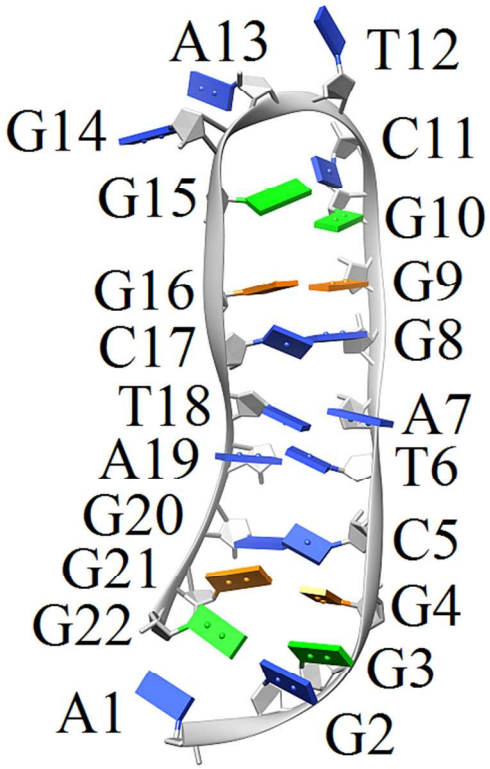


Figure 8. Hairpin intermediate in unfolding of 2KM3. G bases forming the first quartet (Q1) in 2KM3 are represented in light green and of Q2 in orange.

83x65mm (300 x 300 DPI)

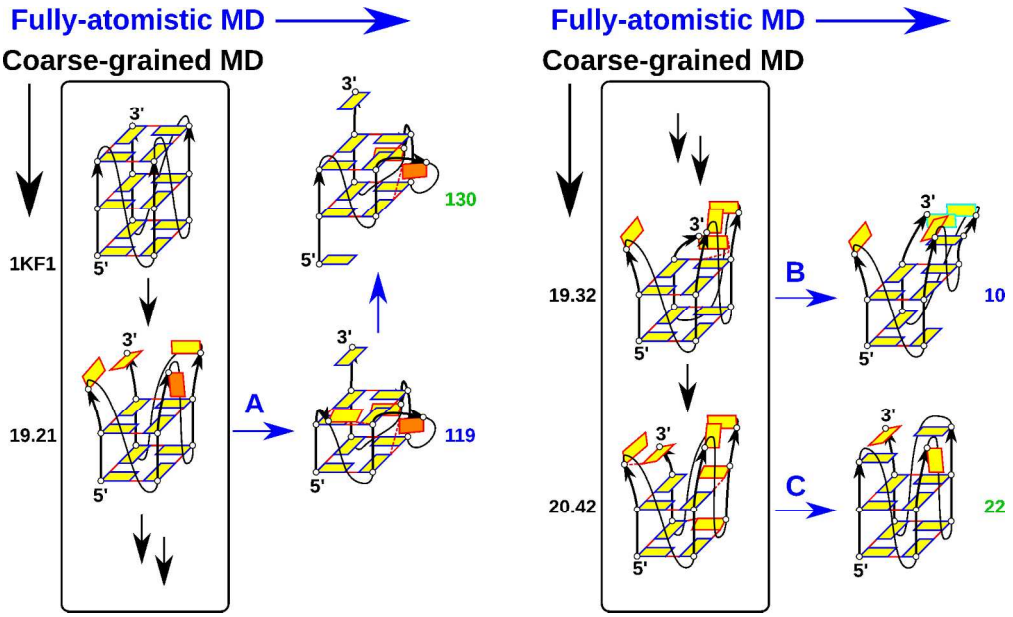


Figure 9. 1KF1 – unfolding ST CG simulation followed by a set of stabilising all-atom MD simulations. The black box contains intermediates occurring in the unfolding CG simulation. Development of the CG simulation goes vertically down from the experimental structure and continues in the right part of the Figure. Black numbers next to the intermediates depict the time when the structure was observed in the CG simulation. These intermediates served as starting structures in subsequent all-atom MD simulations. Their progression is shown by blue arrows and the numbers show the time in ns at which the given structure occurred. The green numbers mean that the particular structure remained stable till the simulation end and blue numbers mean that the depicted structure was subsequently lost. Simulation B leads to complete structure loss after 10 ns. The structural schemes are visualised as follows: Gs are depicted as rectangles; yellow and orange mean anti and syn conformation, respectively. Solid red lines denote two GQ-like H-bonds, while dashed lines any other H-bonding. The colours of the rectangle edges are used to visualise the way how the Gs mutually interact. Gs that have the same colour of their edge are either stacked together or coplanar and H-bonded. The remaining Gs, either lacking any direct interactions, or H-bonded, but not coplanar, are indicated by a red edge. Flanking residues, loop bases and ions are not shown. See the Supporting Information appendix Figures SA1-SA6 for development of selected order parameters.

177x107mm (300 x 300 DPI)

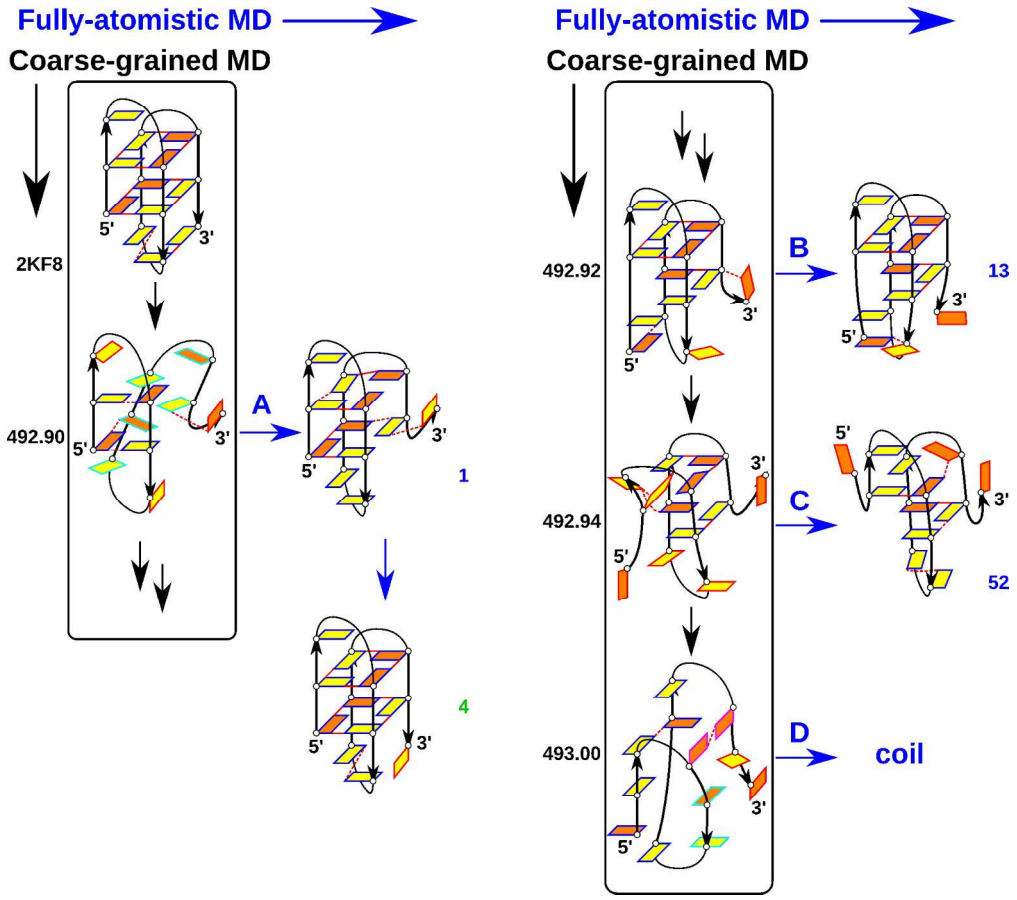


Figure 10. 2KF8 – unfolding ST CG simulation followed by a set of all-atom MD simulations. The simulation A leads to almost refolded native GQ. The structural schemes are visualised as in the Figure 9. See the Supporting Information appendix Figures SA33-SA40 for development of selected order parameters.

177x157mm (300 x 300 DPI)

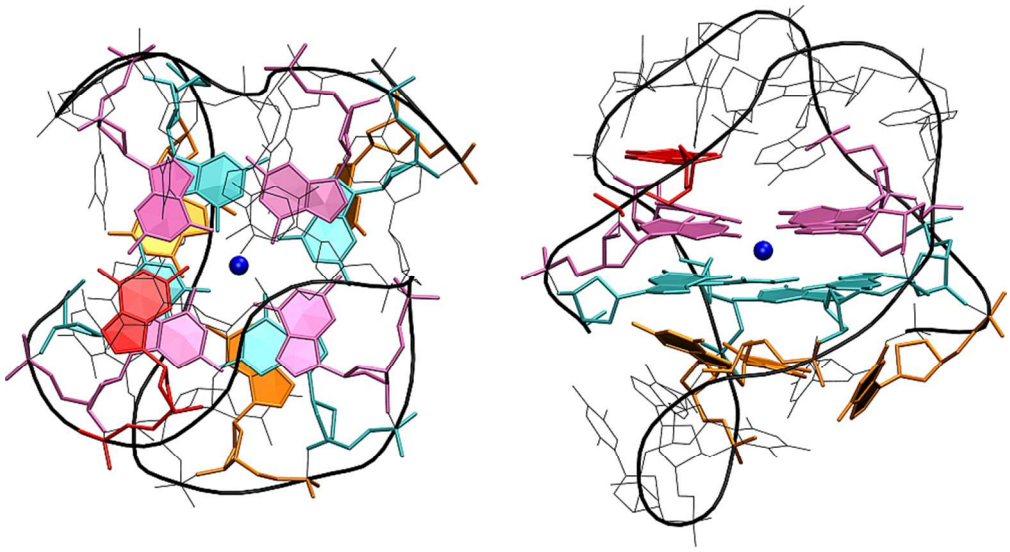


Figure 11. Top and side view of the final structure in the simulation 2KF8-A. Bases of Q1 are shown in cyan, Q2 in mauve, G-triplet bases in orange, G3 is red, cations are blue, and loop nucleotides and the backbone are black.

83x45mm (300 x 300 DPI)

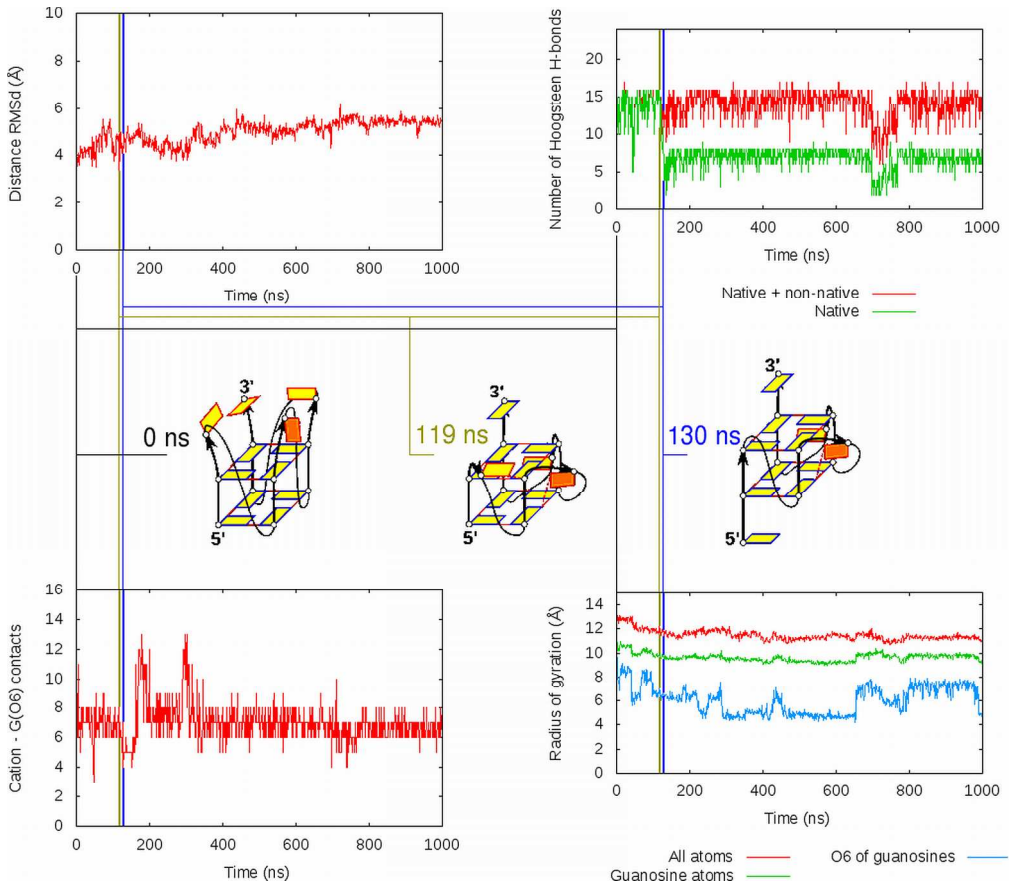


Figure 12. Development of various order parameters in the simulation 1KF1-A. The graphs are shown along with representative conformations found in the simulation, the rightmost being dominant till the end of the simulation, except of a temporary unfolding between 700 and 800 ns. The development of various quantities does not reflect the structural changes. For example, the partial unfolding is evident only in the graph monitoring the number of hydrogen bonds. On the other hand, the CV monitoring the number of native H-bonds has its own severe limitations, for example the decrease in the number of native bonds at ~120 ns cannot distinguish between quartet melting and strand slippage. It is absolute unsuitable to monitor transitions from one topology into another, as it does not differentiate most of the other folded structures from the unfolded state.

171x149mm (300 x 300 DPI)

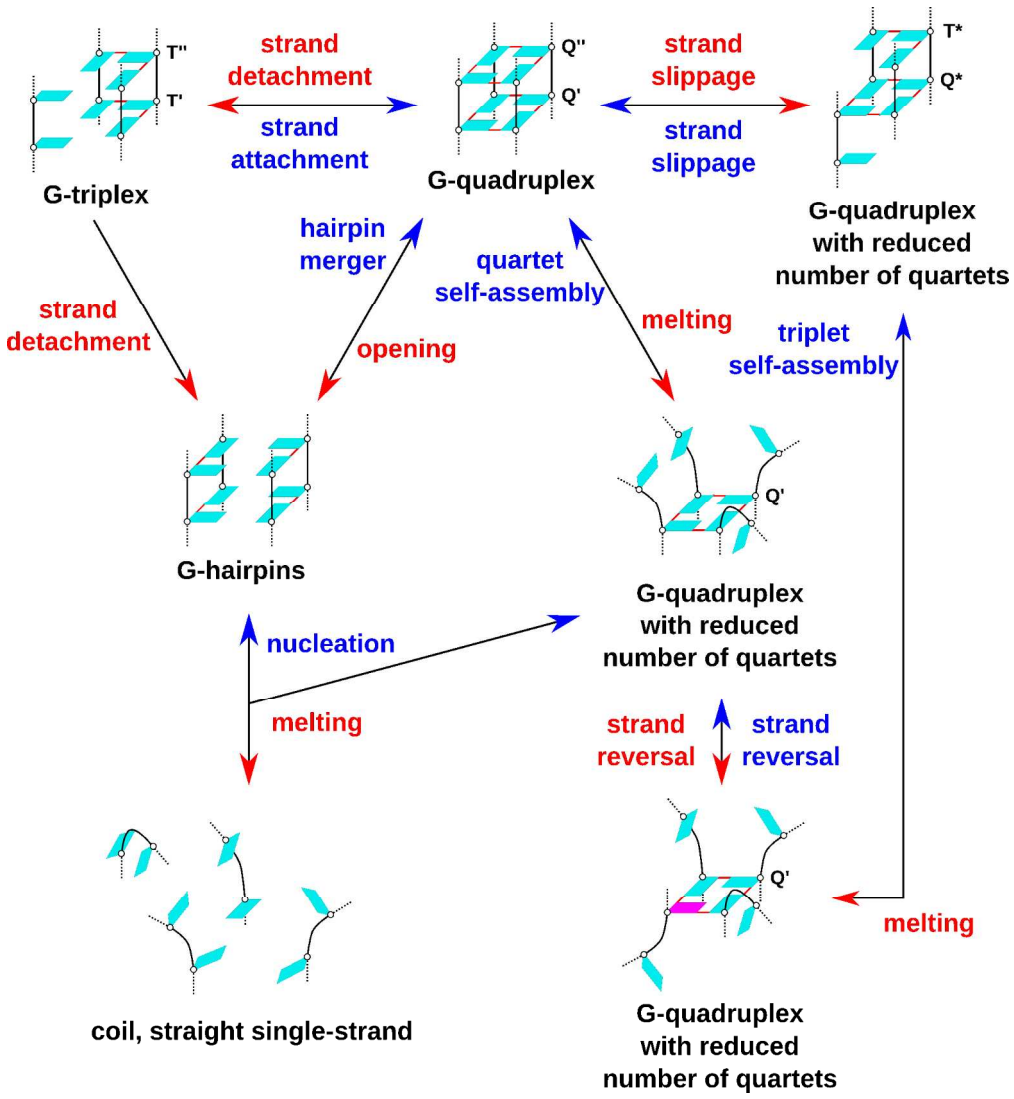


Figure 13. Sketch of various main types of structures and folding/unfolding interconversions between various intermediates for human telomeric GQs observed in simulation studies that have been carried here and elsewhere.53, 62, 72, 75 For the sake of simplicity, the G-stem is marked only by two quartets, irrespective of whether it consists of two or three consecutive quartets. Gs are in cyan, Hoogsteen H-bonds are shown as red lines. Vertical strand slippage (top right) is possible only when all involved quartets have the same directionality (see also Figure S12). Mauve guanine in the bottom right scheme means that its glycosidic χ torsion angle is inverted upon the strand reversal (rotation) transition, i.e. it changes either from syn to anti or vice versa, while continuously keeping unperturbed G-quartet base-pairing. Syn to anti dynamics of unpaired Gs (not shown) has been observed in the earlier T-REMD simulations53 and in the present CG simulations.

175x190mm (300 x 300 DPI)

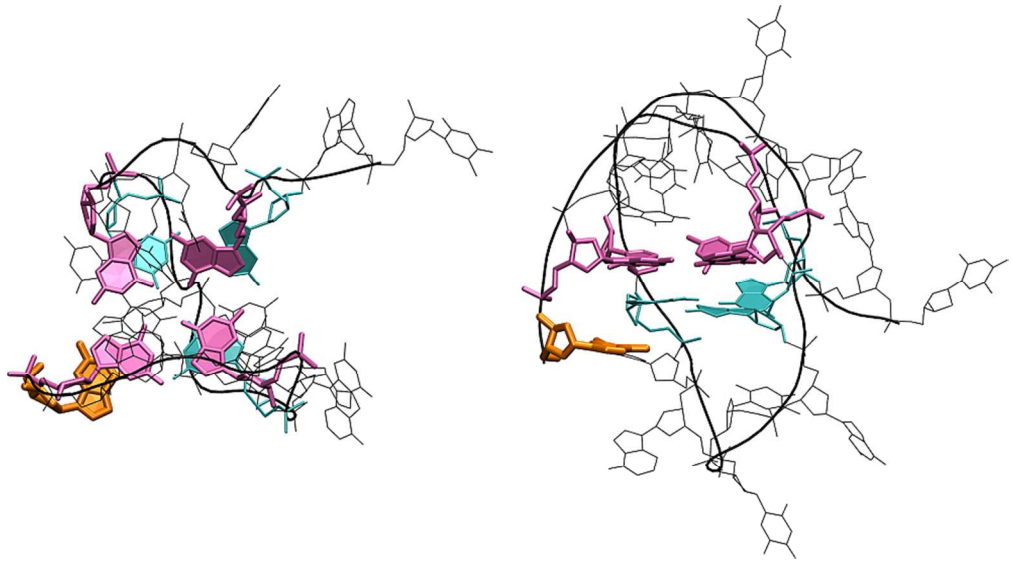
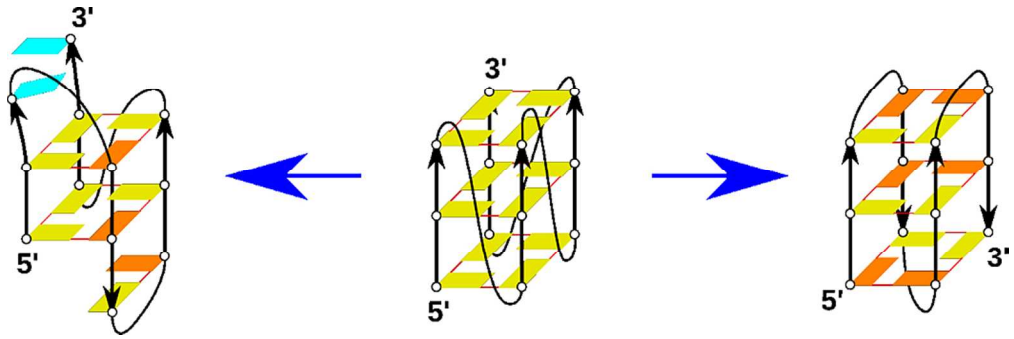


Figure 14. Underestimation of helical twist and non-systematic swelling in the CG model (converted to atomistic structures). Top and side view of the input structure for the simulation 2KF8-C shows underestimated helical twist of the stem and unphysical prolongation of the stacking distance between G1 (orange) and the bases of Q2 (mauve). The other Q1 bases (cyan) are in correct distance to Q2. Loop residues and the backbone are black.

83x46mm (300 x 300 DPI)



TOC graphics

88x29mm (300 x 300 DPI)