RNA G-quadruplex folding is a multi-pathway process driven by conformational entropy

Marijana Ugrina, Ines Burkhart, Diana Müller, Harald Schwalbe, Nadine Schwierz

Angaben zur Veröffentlichung / Publication details:

Ugrina, Marijana, Ines Burkhart, Diana Müller, Harald Schwalbe, and Nadine Schwierz. 2024. "RNA G-quadruplex folding is a multi-pathway process driven by conformational entropy." *Nucleic Acids Research* 52 (1): 87–100. https://doi.org/10.1093/nar/gkad1065.

Nucleic Acids Research, 2024, **52**, 87–100
https://doi.org/10.1093/nar/gkad1065
Advance access publication date: 20 November 2023
Computational Biology

RNA G-quadruplex folding is a multi-pathway process driven by conformational entropy

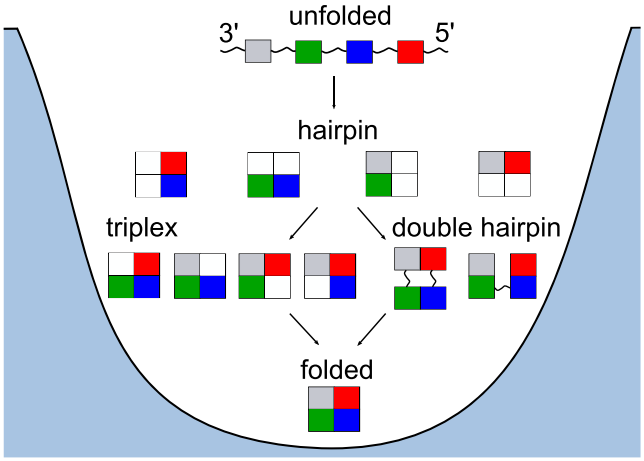
Marijana Ugrina [1,3], Ines Burkhart[2], Diana Müller[2], Harald Schwalbe [2] and Nadine Schwierz [1,*]

[1]Institute of Physics, University of Augsburg, Universitätsstraße 1, 86159 Augsburg, Germany
[2]Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt am Main, Max-von-Laue-Straße 7, 60438 Frankfurt am Main, Germany
[3]Department of Theoretical Biophysics, Max-Planck-Institute of Biophysics, Max-von-Laue-Straße 3, 60438 Frankfurt am Main, Germany
[*]To whom correspondence should be addressed. Tel: +49 821 598 3714; Email: nadine.schwierz@physik.uni-augsburg.de

Abstract

The kinetics of folding is crucial for the function of many regulatory RNAs including RNA G-quadruplexes (rG4s). Here, we characterize the folding pathways of a G-quadruplex from the telomeric repeat-containing RNA by combining all-atom molecular dynamics and coarse-grained simulations with circular dichroism experiments. The quadruplex fold is stabilized by cations and thus, the ion atmosphere forming a double layer surrounding the highly charged quadruplex guides the folding process. To capture the ionic double layer in implicit solvent coarse-grained simulations correctly, we develop a matching procedure based on all-atom simulations in explicit water. The procedure yields quantitative agreement between simulations and experiments as judged by the populations of folded and unfolded states at different salt concentrations and temperatures. Subsequently, we show that coarse-grained simulations with a resolution of three interaction sites per nucleotide are well suited to resolve the folding pathways and their intermediate states. The results reveal that the folding progresses from unpaired chain via hairpin, triplex and double-hairpin constellations to the final folded structure. The two- and three-strand intermediates are stabilized by transient Hoogsteen interactions. Each pathway passes through two on-pathway intermediates. We hypothesize that conformational entropy is a hallmark of rG4 folding. Conformational entropy leads to the observed branched multi-pathway folding process for TERRA25. We corroborate this hypothesis by presenting the free energy landscapes and folding pathways of four rG4 systems with varying loop length.

Graphical abstract



Introduction

RNA G-quadruplexes (rG4s) play a crucial role in a variety of physiological processes including regulation of transcription and translation (1,2) or pre-mRNA processing (3,4). G4-forming sequences are abundant in the human genome and non-canonical structures of guanosine-rich sequences can be formed by DNA and RNA. While long debated (5), advances in the development of highly sensitive approaches provided further evidence for the existence of rG4s *in vivo* and revealed their dynamic nature in cells (6). However, to resolve how rG4 folding and unfolding affects its function, a detailed understanding at the molecular level is required.

The overall architecture of G4s is striking: Four non-neighboring guanosine nucleobases assemble in quartets and several quartets stack onto one another to form a central channel. For rG4s, this channel is typically flanked by propeller loops, while a large variety of different loop arrangements have been reported for DNA G-quadruplexes (dG4s) (7). The

Received: February 10, 2023. **Revised:** September 25, 2023. **Editorial Decision:** October 23, 2023. **Accepted:** October 25, 2023

three-dimensional structures are stabilized by cations. In particular, K^+ and Na^+ are frequently observed in the channel and lead to the highest stabilization efficiency against denaturation among alkali and alkaline earth cations (8).

The telomeric repeat-containing RNA (TERRA) is an important rG4 example. Telomeres are specialized nucleoprotein structures that protect chromosomal DNA from progressive degradation. Transcription of the telomeric C-rich strand in the chromosomes produces TERRA, which has a canonical G-rich motif with sequence r(UUAGGG) (9,10).

TERRA is essential in maintaining genomic integrity by regulating telomerase activity and protecting the chromosome ends against degradation (11,12). Telomere dysfunction is connected to cell aging and cancer (13,14) and the design of small drug molecules targeting TERRA attracts therefore increasing scientific attention (15).

Despite the importance of rG4s, most scientific work has focused on dG4s so far (16). The scientific focus on dG4s can be rationalized by the fact that their existence in vivo has been demonstrated early on by different experimental techniques (17,18). By contrast, the existence of rG4s *in vivo* and their biological relevance are just now starting to be established (2,6,19,20). In addition, the investigation of rG4s requires much more sensitive experimental techniques since rG4s are more dynamic compared to their DNA counterparts (6).

Moreover, the folding intermediates of rG4s are transient, making their detection and resolution a major challenge (21). Consequently, the molecular details of folding have not been resolved so far and the intermediate states remain elusive.

Insights into the atomistic structure of folded G4s are obtained from nuclear magnetic resonance (NMR) spectroscopy (22–25) and X-ray crystallography (26). In addition, time-resolved NMR spectroscopy yields information on possible folding intermediates (27–29). Such experiments recently revealed that the folding kinetics of TERRA is fundamentally different compared to a homologous DNA sequence (30,31). The K^+-induced folding kinetics of TERRA is an order of magnitude faster ($k_1 = 1.45$/min) compared to the dG4 of identical nucleotide sequence ($k_1 = 0.41$/min) (21).

While some dG4 intermediates have been resolved at atomistic resolution by NMR (23,32–34), the intermediate states of rG4s remain elusive due to their short lifetimes (27). Further information on folding intermediates can be obtained from single-molecule techniques such as single-molecule Förster resonance energy transfer (smFRET). Based on the measured distance between strategically placed fluorophores, hairpin and triplex structures have been suggested as intermediates of two and three quartet rG4s and dG4s (35,36).

To complement experiments, biomolecular simulations are powerful tools which provide molecular insights into the folding pathways and intermediate states of rG4s and dG4s (16,37–41). At first sight, all-atom molecular dynamics (MD) simulations in explicit water seem to be the ideal computational method. For instance, atomistic simulations allow to resolve the conformational rearrangements in the folded state of G4s, their interactions with different cations, (42–44) or the stability of putative intermediates along the folding pathway (37,38). Moreover, combining atomistic simulations and enhanced sampling techniques such as replica exchange can improve the sampling of rare folding events (37–40). Still, G4 folding is on the timescale of minutes. Simulating a single folding pathway, let alone the large variety of different pathways, is therefore out of reach for atomistic simulations.

The limitations of atomistic simulations can be overcome by coarse-grained (CG) simulations. CG models simplify the atomistic representation by reducing the number of particles to three to seven beads per nucleotide and employing implicit solvent thereby reducing the computational costs further. For instance, the HiRE-DNA model revealed the potential of CG model to investigate dG4 folding by resolving numerous intermediates of human telomeric dG4s (41). By now a variety of CG models for RNA are available such as SimRNA (45), HiRE-RNA (46), seRNA (47) and TIS model (48–53). TIS, the three-interaction site model developed by Thirumalai and coworkers, is a simple low-resolution model. Its efficiency and accuracy to describe folding has been demonstrated for a variety of RNA molecules ranging in size from larger molecules such as ribozyme (52) to smaller RNAs like riboswitch (54), pseudoknot (50), and hairpin (49).

By combining CG simulations using the TIS model and circular dichroism (CD) experiments, we resolve the folding pathways and intermediate states of the TERRA rG4 sequence consisting of a monomeric 25-mer human telomeric RNA repeat. In a first step, we validate the CG model by CD experiments and obtain close agreement as judged by the population of folded states as a function of temperature and salt concentration. Quantitative agreement is obtained by a newly developed concentration matching procedure, which takes the ionic double layer into account via additional all-atom MD simulations. Subsequently, we resolve the folding pathways from the CG simulations and identify hairpins, triplexes and double-hairpins as intermediate states. Finally, we investigate the folding of three synthetic rG4 sequences with varying loop length and a plant rG4 sequence and confirm that branched multi-pathway folding is a characteristic feature of all rG4 systems.

Materials and methods

Coarse-grained and atomistic simulations

Simulation models

We modeled the monomeric 25-mer G4 from human telomeric RNA (TERRA25). As starting point, we used the NMR resolved structure of a dimeric 24-mer TERRA G4 in K^+ solution (PDB-ID: 2KBP from (55)) with sequence $(UAGGGUUAGGGU)_2$. In a first step, we used the ModeRNA homology modeling server (56) to connect the two strands of the RNA into a single strand through U12 and U13. We created an additional propeller loop and added a U25 to the 3‘-end of the strand. In a second step, we optimized the structure using the SimRNA server (45) and selected 10 minimum energy structures. We preformed MD simulations of the 10 structures for 100 ns and selected the one with lowest root mean square deviation (RMSD) with respect to the experimental structure (2KBP).

The equilibrated TERRA25 structure with sequence 5′-UAGGGUUAGGGUUAGGGUUAGGGUU-3′ is shown in Figure 1A. It consists of four guanosine repeats. Each repeat is formed by three consecutive guanosines in anti-conformation (i.e. with a glycosidic torsion angle between 180° and 240°). The repeats are connected by propeller loops. The guanosines from the four repeats form three quartets that give the G4 its characteristic non-canonical structure.

To test if the branched multi-pathway folding process observed for TERRA25 also occurs in different systems, we investigated four additional rG4 systems (Table 1).

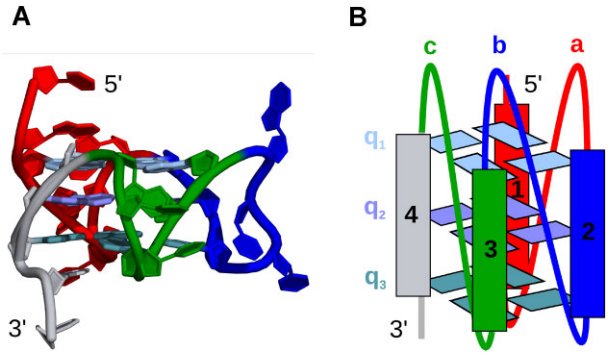


Figure 1. Monomeric 25-mer TERRA G4 with sequence 5′-UAGGGUUAGGGUUAGGGUUAGGGUU-3′ in the simulations. Simulation snapshot from all-atom MD simulations (**A**) and schematic representation (**B**). Each of the four GGG-repeats is represented in a different color (from 5′ to 3′: red (1), blue (2), green (3), gray (4). The three G-quartets consist of four guanosines from different repeats and are labeled q_1, q_2, q_3 in (B). The three propeller loops are labeled a, b, c. The connecting loops all have the sequence UUA. The sequence has a UU-overhang at the 3′ end and a UA-overhang at the 5′ end.

Table 1. Sequences of rG4 structures simulated in this work

Abbreviation	Sequence
TERRA25	UAGGGUUAGGGUUAGGGUUAGGGUU
U	UAGGGUGGGUGGGUGGGUU
UU	UAGGGUUGGGUUGGGUUGGGUU
UUU	UAGGGUUUGGGUUUGGGUUUGGGUU
Plant	GGGAGGGAAGGGGAAGGGG

Loop regions are marked in bold.

The first three systems are synthetic sequences of varying loop lengths. They contained only uridine nucleosites in the loops 5′-UAG$_3$U$_n$G$_3$U$_n$G$_3$U$_n$G$_3$UU-3′, where $n = 1, 2, 3$. The fourth system is a plant rG4 sequence that was previously studied by smFRET (36) and has the sequence 5′-GGGAGGGAAGGGGAAGGGG-3′. We employed the ModeRNA server (56) to create the initial folded structure and used the structure of TERRA25 as a reference. Since the last structure has more than three guanosines in the last two repeats, we used RNAFold web server from the ViennaRNA Package 2 (57) to model the secondary structure and generated the native structure accordingly.

All-atom MD simulations

The MD simulations were performed with Gromacs version 2018.1 (58). Periodic boundary conditions were used with the particle mesh Ewald method (59) to calculate long range electrostatics, with cubic interpolation and a Fourier spacing of 0.12 nm. The cutoff for Coulomb and Lennard–Jones interactions was 1.2 nm and long-range dispersion corrections for energy and pressure were applied. Bond to hydrogens were constrained with the LINCS algorithm (60) with order of four for the matrix inversion correction and one iterative correction. Simulations were performed for 100 ns with a 2-fs time step. TERRA25 was simulated using parameters from Amber99sb – ildn* force field (61) with parambsc0 (62) and χ_{OL3} (63) corrections and TIP3P water model (64). Optimized parameters were used for K^+ and Cl^- ions, which correctly reproduce the thermodynamic and kinetic properties of the ions in TIP3P water (65). We added the same number of anions and cations to obtain different bulk concentrations and 24 additional cations were used to neutralize the systems.

In the first set of simulations, we simulated the 10 conformations obtained from simRNA for which the interatomic potential energy is lowest. 1 M KCl was used in each of the 10 systems. The structures were initially placed in a rectangular box with dimensions $5.4 \times 5.0 \times 5.6$ nm^3. From these simulations, the equilibrated structure with the lowest RMSD with respect to the experimental structure was selected for all further simulations. In the RMSD calculations only heavy atoms from the central pore were used.

In a second set of simulations, we calculated the ion distribution profiles for TERRA25 to establish the relation between bulk salt concentration and the local concentration in the double layer.

We performed simulations of TERRA25 at eight concentrations of KCl, namely 0.48 mM, 16 mM, 53 mM, 82 mM, 0.14 M, 0.22 M, 1.03 M and 1.89 M. For the 0.48 mM KCl simulation, the box size with edge length 30 nm was used. It contained 878 660 water molecules. For the 16 mM KCl simulation, a box size with edge length 15 nm was used. The box contained 108 299 water molecules. For the KCl concentrations 0.22 and 1.89 M a box size with edge length of 10.0 nm was used. These systems contained 32 108 or 30 116 water molecules. For the 53 mM, 82 mM, 0.14 M and 1.03 M KCl concentrations, we used a smaller box with edge length 9.0 nm. The boxes contained between 22 744 and 23 551 water molecules. To investigate if the matching procedure for KCl is also valid for NaCl, we performed one additional 100 ns simulation of TERRA25 in 1 M NaCl using the Mamatkulov-Schwierz parameters (65). The simulation box had an edge length of 9 nm, and it contained 22.744 water molecules. RMSD, simulation snapshots, ion distribution profiles are shown in the Supporting Information (Supplementary Figure S1 and S2).

TERRA25 was placed in a cubic box and each system was equilibrated in NVT and NPT simulations. During equilibration harmonic restraints with force constant 1000 kJ/(mol nm^2) were applied on the heavy atoms of TERRA25 to prevent large conformational changes. The NVT equilibration was done in two parts. During the first 1 ns restraints of 2000 kJ/(mol nm^2) were applied to ions to prevent ion pairing before a complete hydration shell is formed. In the subsequent 1 ns, the restraints on the ions were released. As thermostat the velocity rescaling algorithm with a stochastic term (66) and with a coupling constant of 0.1 ps was used. During the 2 ns NPT equilibration, pressure coupling was performed using the Berendsen barostat (67) with a 1 ps coupling constant. In the production runs, the velocity rescaling thermostat (66) and the Parrinello-Rahman barostat (68) with a 2 ps coupling constant were used. Position restraints on TERRA25 were released during the production run. Each production run was 100 ns long. The radial concentration profile was calculated as follows: First, a 2D number density profile of K^+ around the central pore of the G4 was calculated using Gromacs and an in-house code (69). The radial concentration profile was obtained by integration and normalization.

CG simulations using TIS

The CG simulations were performed using the Three Interaction Sites (TIS) model developed by Thirumalai and coworkers (48–53). In the TIS model, each nucleotide is represented by three beads, located at the center of mass of the phosphate

group, sugar and nucleobase. The parameters for the canonical interactions between the beads were taken from previous work (48). The intramolecular attractive interactions were defined based on the residues that appear in the native structure of TERRA25. This ensures that the native structure is the minimum energy structure (70). Native hydrogen bonds and tertiary stacks were defined based on the modelled monomeric TERRA25 structure described above.

In order to capture folding intermediates that are not stabilized by native interactions, non-native secondary structure interactions were included via the base-stacking interactions of consecutive nucleotides and hydrogen-bond interactions between all nucleobases as in previous work (53). The parameters for these interactions were calculated by coarse-graining the standard A-form RNA helix (48).

A non-interacting uridine was added capping the structure at the 3′ end. The RNA was placed in a cubic box with an edge length of 70 nm.

The solvent was modeled implicitly, and we employ the Debye-Hückel approximation in combination with the concept of counterion condensation to take the ions into account (48). Specifically, the contribution to the energy function of TIS for the interaction between two phosphate groups is based on the linearized Debye-Hückel equation and is given by (48,71)

$$V(r_{ij}) = \frac{Q^2 e^2}{4\pi\epsilon_0\epsilon_r}\frac{e^{-r_{ij}/\lambda}}{r_{ij}}. \quad (1)$$

Here, r_{ij} is the distance between the two phosphate groups i and j, Q is the charge of a phosphate group, e is an elementary charge, ϵ_0 the dielectric constant of vacuum, ϵ_r is the dielectric constant of water and λ is the Debye length. Hence, the ionic concentration is included in the CG model via the Debye length

$$\lambda = \sqrt{\frac{\epsilon_0\epsilon_r k_B T}{2e^2 N_A c_{CG}}} \quad (2)$$

where N_A is Avogadro's constant and c_{CG} is the uniform ion concentration of the CG model.

Numerical integration of the equations of motion was performed using the leap-frog algorithm with time step $h = 0.05\tau$, where $\tau = 50$ fs as in previous work with the TIS model (49). The simulations were performed in the low friction regime to increase the sampling efficiency by reducing the viscosity of water by a factor of 100 (72). The cut-off for electrostatic interactions was set to 3 nm.

To investigate the temperature dependence of the population of folded rG4 states, we performed 120 simulations of independent copies of the system at each of temperatures in the temperature range from 1-120°C, leading to a temperature spacing between the copies of 1°C.

Each temperature was simulated at a concentration of c_{CG} = 150 mM (corresponding to c_{bulk} = 12 mM) for about 100 μs.

To investigate the concentration dependence of folding, we performed independent simulations of copies of the system at 16 different concentrations in a range of from 10 μM to 1 M (corresponding to $c_{bulk} = 0.06$ μM to $c_{bulk} = 430$ mM) at two different temperatures, 25°C and 60°C. Each setup was simulated for about 100 μs. The simulations were initiated from the folded state and 5 μs were neglected for equilibration in the analysis. To provide sufficient statistics at $T = 25$°C, additional simulations were performed at 0.1, 0.4 and 1 mM using 50 additional independent simulations of 20 μs.

The folding and unfolding pathways were simulated at room temperature ($T = 25$°C) and at $c_{CG} = 50$ mM (corresponding to $c_{bulk} = 0.3$ mM). The salt concentration was chosen such that the probabilities of folded and unfolded structures are similar. The plant rG4 was simulated at c_{CG} = 400 mM (corresponding to $c_{bulk} = 80$ mM) and 25°C. In order to provide sufficient statistics, 100 copies of system were simulated. For 23 copies of the system, we performed 180 μs long simulations and for 77 copies we performed 52.5 μs long simulations. In total, 8182.5 μs of simulation time was used.

The simulations were initiated from the folded state and 5 μs were neglected for equilibration in the analysis. We observed about two folding/unfolding events per trajectory and the number of forward and backward transitions during 180 μs simulations were the same.

Intermediate states in TIS

A common simplification used in Gō-like models is that intramolecular attractive interactions are defined only between the residues that appear to be in contact in the native structure. This definition ensures that the native structure is the minimum energy structure. The drawback of such basic Gō-like models is that they cannot capture partially folded intermediate states stabilized by non-native interactions. The TIS model, used in this current work, is more advanced than standard Gō-like models. TIS explicitly includes non-native secondary structure interactions (48). In particular, all base-stacking interactions between consecutive nucleotides and hydrogen-bond interactions between any bases G and C, A and U, or G and U are included (48). Therefore, the model captures non-native secondary structure interactions as shown in previous work (54,73) and for TERRA25 (see Supplementary Figure S3). Moreover, it has been shown that TIS reproduces experimental thermodynamic and structural data for several different RNA molecules including ribozyme formation or riboswitch folding under a variety of solvent conditions (52,54,73). With these prerequisites, TIS is ideal to resolve intermediates states and their frequency of occurrence in the folding of rG4 systems.

Timescale, rates, and lifetimes in TIS

The dynamics of the system is based on the Langevin equation. The effect of the solvent is modeled implicitly via a Gaussian random force and a friction force which is assumed to be proportional to the viscosity η of the medium. The Langevin equation is solved numerically using a finite timestep $\Delta t = 2.5$ fs. The dynamics obtained from the solution of the Langevin equation sets the timescale of the coarse-grained simulations. However, relating the timescale of CG models to the timescale observed in experiments is challenging for two reasons: (i) The viscosity of the water in the CG simulations is reduced by a factor of 100 to enhance the conformational sampling as in previous work (48). (ii) The free energy folding landscape is smooth due to the coarsening of atomistic interactions and the implicit solvent.

One possibility to relate the scales is to introduce a time rescaling factor based on experimental folding rate (21)

$$\alpha = k_{FU}^{CG}/k_{FU}^{exp}. \quad (3)$$

With $k_{FU}^{exp} = 2.4 \cdot 10^{-2}$ s^{-1} and $k_{FU}^{CG} = 3.75 \cdot 10^{5}$ s$_{CG}^{-1}$, we obtain $\alpha = 1.56 \cdot 10^{7}$ for TERRA25. The corresponding rates and lifetimes are listed in Supplementary Tables S2 and S3 in the Supporting Information.

However, a single numerical factor is unlikely to capture the coarsening of complex atomistic interactions. We therefore provide the results in CG time units s_{CG} and introduce reduced rates $\tilde{k}_{ij} = k_{ij}^{CG}/k_{FU}^{CG}$ and lifetimes $\tilde{\tau}_i = \tau_i^{CG}/\tau_F^{CG}$ in addition to the rescaled results (Supplementary Tables S2 and S3). The kinetic rate coefficients k_{ij} for transitions $j \rightarrow i$ are calculated from

$$k_{ij} = \frac{N}{2t_j}, \quad k_{ji} = \frac{N}{2t_i} \tag{4}$$

where N is the total number of transitions between the two stable states and $t_{j/i}$ is the total time in the respective state. Note that for long equilibrium runs, the number of forward and reverse transitions are equal. Further information and the rate equations for the time evolution of the individual states can be found in the Supporting Information (section Rate equations for rG4 folding).

For first-order kinetics, the lifetime distribution is determined by the rates. Specifically, the mean lifetime of state j is calculated from

$$\langle \tau_j \rangle = 1/\sum_{i \neq j} k_{ij}. \tag{5}$$

Gibbs free energy

To determine the stability of the different rG4 systems the Gibbs free energy was calculated via

$$\Delta G = -k_B T \ln(K_{eq}), \quad \text{with} \quad K_{eq} = k_{FU}/k_{UF} \tag{6}$$

where K_{eq} is the equilibrium constant. The results are compared to the results from UV melting experiments (74). Since our simulations were performed at 298.15 K, we calculate the Gibbs free energy 298.15 K from the experimental entropy change and experimental enthalpy change (see Supporting Information, section Gibbs free energy from CG simulations and thermal melding curves for details). Moreover, the experimental results are the average over different loop sequences and the concentration of RNA and KCl ions were different in experiments and simulations. In the following, we therefore calculate the Gibbs free energy change with respect to a reference system. For the simulations, we chose the UUU system and the L333 system (74) for the experiments.

Number of native contacts and radius of gyration

To identify the transition between the folded and unfolded state, we calculated the radius of gyration r_g and the number of native contacts n_C between the guanosines in each quartet. $n_C = 12$ corresponds to the fully folded, native state. n_C is calculated using the coordination function of PLUMED (75) with cutoff distance between guanosine base beads $r_0 = 0.75$ nm, offset parameter $d_0 = 0.08$ nm and parameters $n = 35$ and $m = 55$.

Order parameters

To characterize the folding intermediates, we introduce two order parameters. The repeat order parameter counts the number of native contacts between repeats i and j (with $i \neq j$ and $i, j = 1 - 4$). It is calculated from the indicator function

$$h = \begin{cases} 1 & \text{for} \quad n_C > 2c \\ 1/(1 + e^{\kappa(n_C - \gamma c)}) & \text{otherwise} \end{cases} \tag{7}$$

where n_C is the number of native contacts between repeats i and j and the parameters $\kappa = -5, \gamma = 1.3, c = 1.4$. This choice provides a smooth function for the order parameter without fluctuations. The repeat order parameter $r_{ij} = h$ is one if all three native contacts have formed between the neighboring repeats i and j.

To resolve the twelve different intermediate states, we introduce two additional order parameters, s_1 and s_2, based on linear combinations of r_{ij}

$$s_1 = 3r_{12} + 6r_{23}, \quad s_2 = 3r_{34} + 6r_{14}. \tag{8}$$

The values for the repeat order parameters and s_1 and s_2 are illustrated Supplementary Figure S4. The quartet order parameter q_k counts the number of native contacts in each of the three quartets ($k = 1 - 3$). q_k is calculated from equation 7 with $\kappa = -4, \gamma = 1.9, c = 1.93$, and n_C being the number of native contacts in quartet k. The quartet order parameter is one if all four native contacts have formed in a quartet.

Fraction of folded structures

For the fraction of folded structures n_f, we calculated the total number of contacts between the guanosines and use equation (7) with $\kappa = -5, \gamma = 2$ and $c = 4.5$. n_f is one in the folded state. The fraction of folded structures was calculated from the time the molecule spends in the folded state relative to the total time. The error is calculated by block averaging over 2.5 μs blocks.

Concentration matching

In the CG simulations, the solvent is modeled implicitly, and the concentration of ions is homogeneous in the simulation box (Figure 2B). Due to the highly charged backbone of the rG4, cations are attracted to the phosphodiester backbone and form a pronounced electric double layer (Figure 2A). Since screening of ions in the vicinity of phosphodiester groups of RNA is a driving factor in RNA folding, the concentration in the CG model needs to be matched to the local concentration of ions in the electric double layer.

To achieve this, we developed the following matching procedure: Firstly, we derive the correlation between the bulk concentration (as measured in the experiments) and the local concentration in the double layer by performing all-atom MD simulations. The simulations were done at eight different KCl concentrations. Figure 2A shows the radial concentration profile at $c_{bulk} = 1.03$ M. The formation of an electric double layer is clearly visible. The highest peak ($r = 0.2$ nm) corresponds to K^+ in the central pore and the second peak ($r = 1.23$ nm) corresponds to adsorption of K^+ at the phosphate oxygens of the backbone. From the simulated profiles, we calculated the local concentration by averaging from the center of the channel up to the distance R_C. The distance R_C is a free parameter representing the distance of the edge of the bilayer from the geometrical center of the RNA. Since the results are not sensitive to the exact choice of R_C, we chose $R_C = 1.5$ nm, which best reproduced the experimental results from Figure 3. We calculated the bulk concentration by averaging the concentration profiles over the last 0.5 nm from the edge of the simulation box. Due to the high binding affinity of K^+, the bulk concentration vanishes in finite simulations

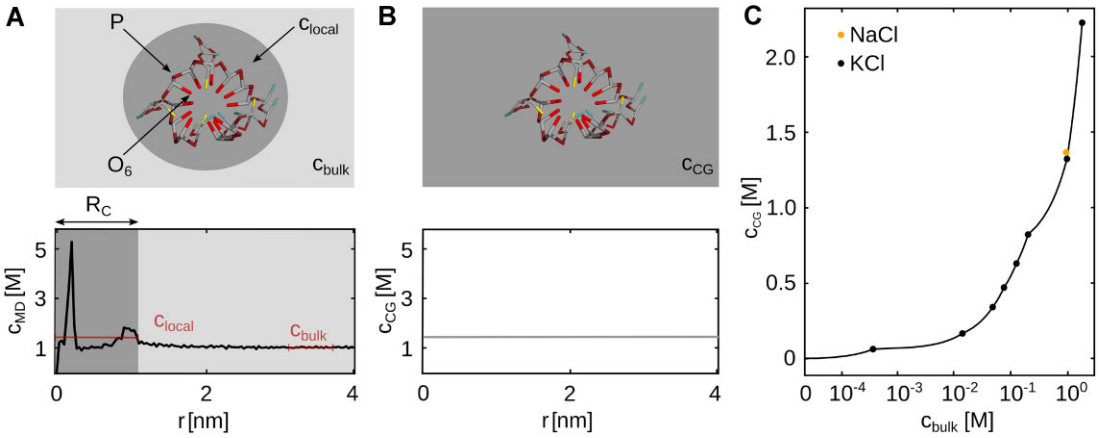


Figure 2. Concentration matching to mimic the effect of an ionic double layer in implicit solvent CG simulations. (**A**) Schematic representation of the ion atmosphere around the rG4 (top) and concentration profile $c_{MD}(r)$ of K^+ ions as function of the distance from the central channel from all-atom MD simulations (bottom). c_{bulk} is the bulk salt concentration and c_{local} the averaged concentration in the double layer within R_C. (**B**) Uniform concentration profile of the CG model $c_{CG} = c_{local}$ after matching to the local concentration in the double layer obtained from the MD simulations. (**C**) Matching relation of the bulk concentration c_{bulk} and the CG concentration c_{CG}. The matching is done for KCl. Note the logarithmic scale of the x-axis. One additional all-atom MD simulation was performed for $c_{bulk} = 1$ M NaCl.

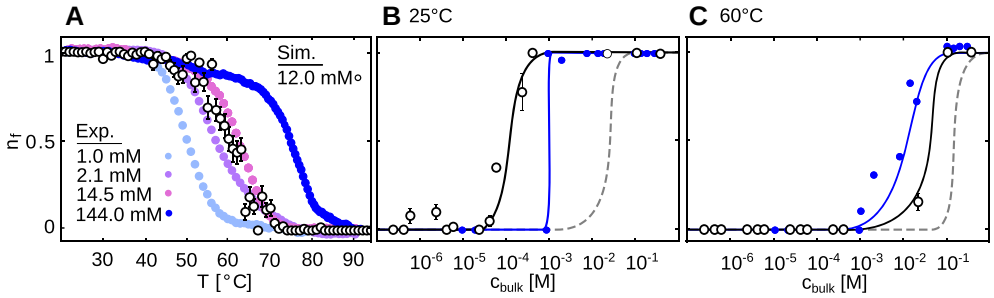


Figure 3. Comparison of the population of the folded state n_f from simulations and experiments. (**A**) n_f as function of the temperature from experimental melting curves and CG simulations. (**B, C**) n_f as function of the K^+ concentrations at 25°C and 60°C. With increasing salt concentration, the equilibrium is shifted to the folded state with $n_f = 1$. The lines are a guide to the eye and indicate the folded/unfolded transition region. The dashed lines are the uncorrected results neglecting the effect of the ionic double layer.

boxes at low concentrations. To provide consistent results, we therefore used the anion profiles to determine the bulk concentration (Supplementary Figure S2).

We performed a linear interpolation between the datapoints, and the resulting matching relation is shown in Figure 2C and Supplementary Table S1. The results from the matching procedure are validated by CD experiments as will be discussed further below.

Experiments

Sample preparation

TERRA25 (5′-UAGGGUUAGGGUUAGGGUUAGGGUU-3′) was purchased from Dharmacon (Cambridge, UK). The oligomer was HPLC purified using a tetrabutylammonium acetate buffer and precipitated with five volumes of $LiClO_4$ (2% in acetone (w/v)) at –20°C over night. Subsequently, desalting with an ultracentrifuge filtration device (Vivacon 2 kDa cut-off, VWR) was performed to remove residual Li^+ cations and acetone.

Circular dichroism (CD) experiments

CD measurements were performed using a JASCO J-810 spectropolarimeter with a Peltier temperature control system. 10 μM RNA were used per measurement and dissolved in ddH_2O containing 1, 2.1, 14.5 or 144 mM KCl. A quartz cuvette with 2 mm pathlength was used. The CD spectra were recorded with a scanning speed of 50 nm/min and three accumulations between 200 and 320 nm. CD melting curves were recorded at the maximum of the CD signal (264 nm) with a heating rate of 0.5°C/min. Results from melting and annealing curves at 144 mM KCl confirm that folding and unfolding of TERRA25 proceed along the same pathway (Supplementary Figure S6).

Native polyacrylamide gel electrophoreses (PAGE)

Correct folding of TERRA25 was checked via native PAGE (15% acrylamide gel) (Supplementary Figure S5). 400 pmol samples were loaded in 50% glycerol and 5 mM KCl. Gels were prepared in 50 mM Tris–borate buffer (pH 8.4) supplemented with 5 mM KCl. Bands were separated in the same buffer at 0.8 W for 90 min with water cooling (13°C water temperature). The bands have been visualized by Stains-All.

Fraction of folded structures from CD melting curves

Conversion of CD absorbance into population of the folded state was calculated according to the method highlighted by Mergny and Lacroix (76). Firstly, upper and lower baselines were corrected for temperature correction to determine the CD signal for the unfolded and folded states, respectively. The

baselines were then used to normalize the absorbance values and yield fraction of folded states.

The melting temperatures were calculated by fitting the normalized melting curves to a sigmoidal function

$$n_f = 1/\left(1 + e^{(T_m - T)/s}\right) \quad (9)$$

where T_m is the melting temperature and s the slope of the curve. The fitting was done according to the nonlinear least-squares Marquardt–Levenberg algorithm.

Results and discussion

Quantitative comparison of the fraction of folded structures from simulations and experiments

Initially, we validated the CG model by CD experiments and provided a direct comparison of the fraction of folded structures in dependence of the temperature or the K^+ concentration (Figure 3). To obtain quantitative agreement between experiments and CG simulations, the ion atmosphere surrounding the highly charged G4 has to be taken into account. Since the cations are attracted to the negatively charged phosphodiester backbone while the anions are repelled, a pronounced double layer is formed. Given the high negative charge, the local ion concentration in the double layer is significantly larger compared to bulk (Figure 2C). It is therefore crucial to take the local concentration in the CG simulations into account to correctly capture the population of the folded state at a given bulk salt concentration. To achieve this, we derived a matching relation from all-atom MD simulations which relates the bulk salt concentration to the local concentration in the double layer (Figure 2C). Based on the relation, the CG concentration was chosen such that it reproduces the average concentration in the double layer.

To validate the procedure, we first investigated the temperature-induced unfolding of the TERRA25 and obtained melting curves from experiments and simulations (Figure 3A). At low temperatures, rG4 is folded in presence of K^+ ($n_f = 1$). With increasing temperature, the equilibrium shifts toward the unfolded state and n_f decreases. At high temperatures, rG4 is unfolded ($n_f = 0$). The melting point T_m depends on the salt concentration (Table 2) and the rG4 becomes more stable with increasing K^+ concentration. The direct comparison of experiments and simulations reveals a similar dependence of n_f on the temperature (Figure 3A). Moreover, the melting temperature of 60°C from the CG simulations at 12 mM (derived from the matching relation in Figure 2C) is in between the experimental results of 59°C and 62°C for 2.1 and 14.5 mM (see Table 2). Note that without the matching relation, a melting temperature of 60°C corresponds to 150 mM salt concentration. Hence, without the details of the ionic double layer, the CG model clearly overrates the population of the unfolded state.

The importance of including the local salt concentration via the matching relation becomes even more evident from the dependence of n_f on the concentration at 25°C or 60°C (Figure 3B, C). Without matching, the CG model overrates the destabilization of the folded state, and the folded/unfolded transition is predicted at too high salt concentrations. On the other hand, quantitative agreement of CD experiments and CG simulations is obtained by taking the local concentration in the double layer into account via the matching relation from additional MD simulations.

Table 2. Melting temperatures of TERRA25 at different KCl concentratio ns from CD experiments and CG simulations using concentration matching from all-atom MD simulations

	c_{bulk} (mM)	c_{CG} (mM)	T_m (°C)
Exp.	1.0		50.67 ± 0.04
	2.1		59.21 ± 0.09
	14.5		62.6 ± 0.2
	144.0		76.9 ± 0.1
Sim.	12	150	60.1 ± 0.1

Folding and unfolding of TERRA25 is a sequential but multi-pathway process

After successful validation, we analyzed the CG simulations to extract information regarding the folding and unfolding of TERRA25.

Figure 4B shows the number of native contacts n_c and the radius of gyration r_g for an 80 μs simulation at 25°C and c_{bulk} = 0.3 mM. Multiple transitions between the folded ($n_c = 12$ and $r_g \approx 15$ Å) and unfolded ($n_c = 0$ and $r_g \approx 24$) states are observed. Energy and entropy contributions to the folding are obtained by calculating the free energy profile $F(r_g)$ (Figure 4C). The folded state is characterized by the sharp and deep minimum in the free energy profile while the unfolded state is broader with r_g ranging from 15 to 25 Å. This reflects an energy-entropy compensation mechanism in which the folded state is energetically favored while the unfolded state competes with a larger entropy due to a variety of unfolded conformations. The folded and unfolded structures are separated by a barrier of 5 k_BT. The two-dimensional free energy landscape $F(n_c, r_g)$ in the region of this barrier reveals several distinct intermediate states (Figure 4C, inset). The intermediate states are broad indicating that multiple different conformations contribute to them (as will be discussed in detail further below).

Figure 5 shows the computation of one representative folding pathway of TERRA25. The number of native contacts n_C increases stepwise while the radius of gyration r_g decreases (Figure 5A). The repeat order parameter r_{ij} and the quartet order parameter q_i reveal that rG4 is formed repeat after repeat and not quartet after quartet (Figure 5B, C), suggesting that stacking interactions drive the initial formation of rG4 rather than Hoogsten interactions. In this example, repeat 1 and 4 assemble first and form the parallel hairpin H4 without propeller loop (Figure 5D, E at $t = 47$ ns_{CG}). Subsequently, repeat 3 assembles and forms the triplex structure T3 with three planar units formed by three h-bonded guanosines, called triads. At this point, the propeller loop c is formed at the 3′ end (Figure 5D, E at $t = 125$ ns_{CG}). Folding is completed once the last repeat assembles, creating the loops a and b (Figure 5D,E at $t = 145$ ns_{CG}) and all three quartets are formed simultaneously as visible from the quartet order parameter in Figure 5C.

In summary, during rG4 folding the repeats assemble and the corresponding propeller loops are formed. During the folding, two intermediate structures are observed, a hairpin and a triplex, supporting previous observations and predictions of hairpin and triplex structures as intermediates in the folding of rG4s and dG4s in experimental (27,29,35,77) and computational (39,78) work. In addition, we observe intermediates with non-native interactions (Supplementary Figure S3). These off-pathway intermediates are transient and unfold quickly and were therefore not analyzed in more detail.

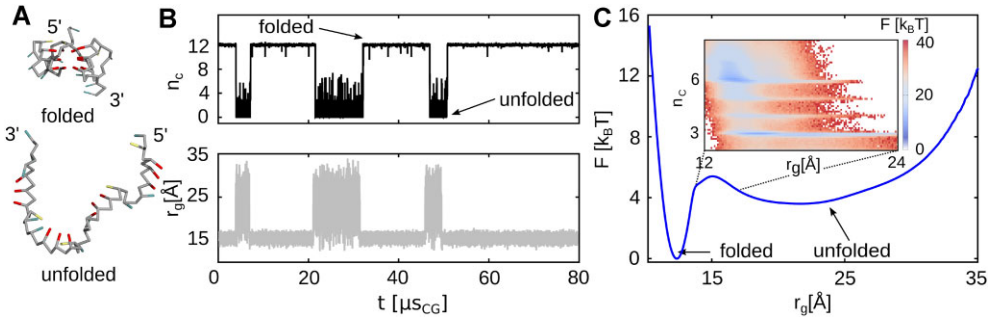


Figure 4. Folding pathway and folding free energy landscape of TERRA25 from CG simulations. (**A**) Snapshots in the folded and unfolded state from the CG simulations. Guanosine is represented in red, uridine in teal and adenosine in yellow. (**B**) Number of native contacts n_C and radius of gyration r_g as function of simulation time. (**C**) One-dimensional free energy profile F as a function of r_g. The inset shows the two-dimensional free energy landscape as function of r_g and n_C in the region of the barrier.

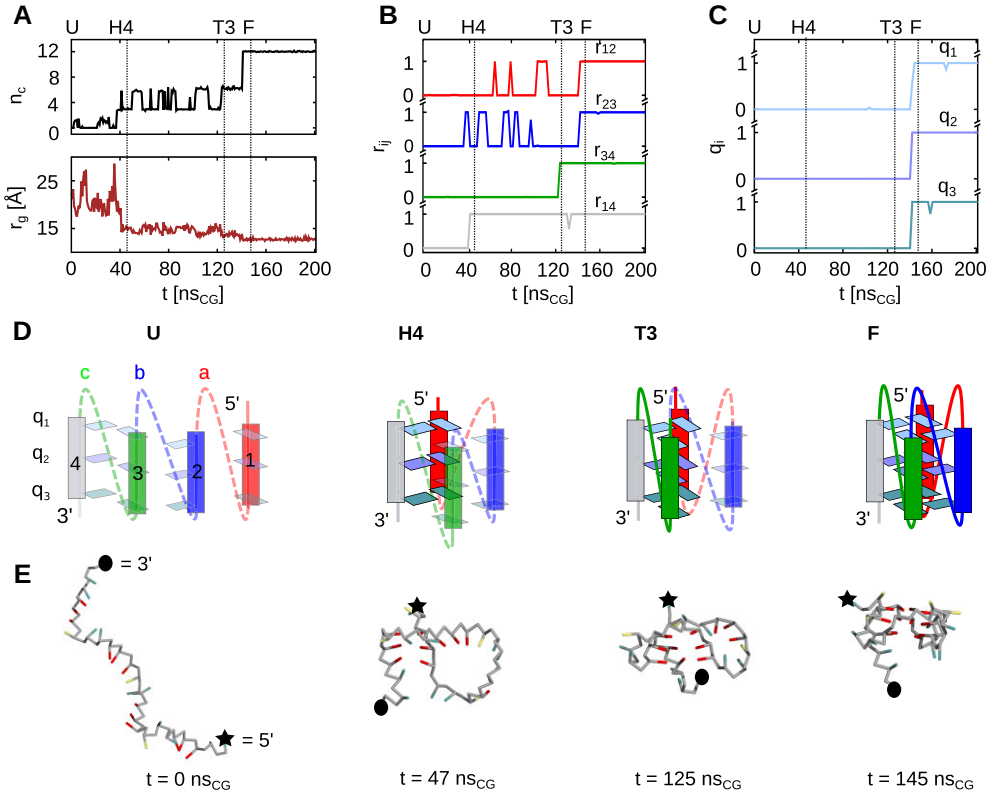


Figure 5. Representative folding pathway and intermediates of TERRA25 from CG simulations. (**A**) Number of native contacts n_C and radius of gyration r_g during the folding transition. (**B**) Repeat order parameter r_{ij} describing the assembly of repeat i and j (with $i \neq j$ and $i, j = 1 - 4$). r_{ij} is one if three native contacts have formed neighboring repeats. (**C**) Quartet order parameter q_i describing the formation of the guanosine quartets ($i = 1 - 3$). q_i is one if all four native contacts have formed in a quartet. (**D**) Schematic representation of the folding pathway. Assembled parts of the structure (repeats, quartets, propeller loops) are shown in solid colors; non-assembled parts are transparent or dashed. (**E**) Simulation snapshots along the folding pathway at times indicated in (A)–(C).

From simple combinatorics, there are four possibilities to form a hairpin from a nucleic acid structure with four repeats since there are only parallel orientations in rG4s (16,21). Similarly, there are four possibilities to form a triplex. For the folding of G4s the question arises whether all these intermediates can be observed or whether their formation is inhibited kinetically or energetically.

To gain further insights, we determined the folding pathways from CG simulations and classified the intermediates from about 200 folding and unfolding events (Figures 6 and 7). The results reveal that all four possible hairpin conformations are accessible in the simulations. However, their lifetimes $\tilde{\tau} = \tau/\tau_F$ relative to the folded state are slightly different. The hairpin structure formed by the two terminal guanosine repeats (labeled H4 in Figure 6) is most stable ($\tilde{\tau} = 3.106 \times 10^{-4}$; $\tau_{resc} = 0.2975$ s).

Hairpin structures with propeller loops have shorter lifetimes. The shortest lifetime is observed for the hairpin structure formed by the first two repeats at the 3′ end (labeled H3, $\tilde{\tau} = 1.381 \times 10^{-4}$; $\tau_{resc} = 0.1323$ s).

Moreover, a hairpin intermediate at the 5′ end, similar to H1 in Figures 6 and 7, was recently observed in smFRET ex-

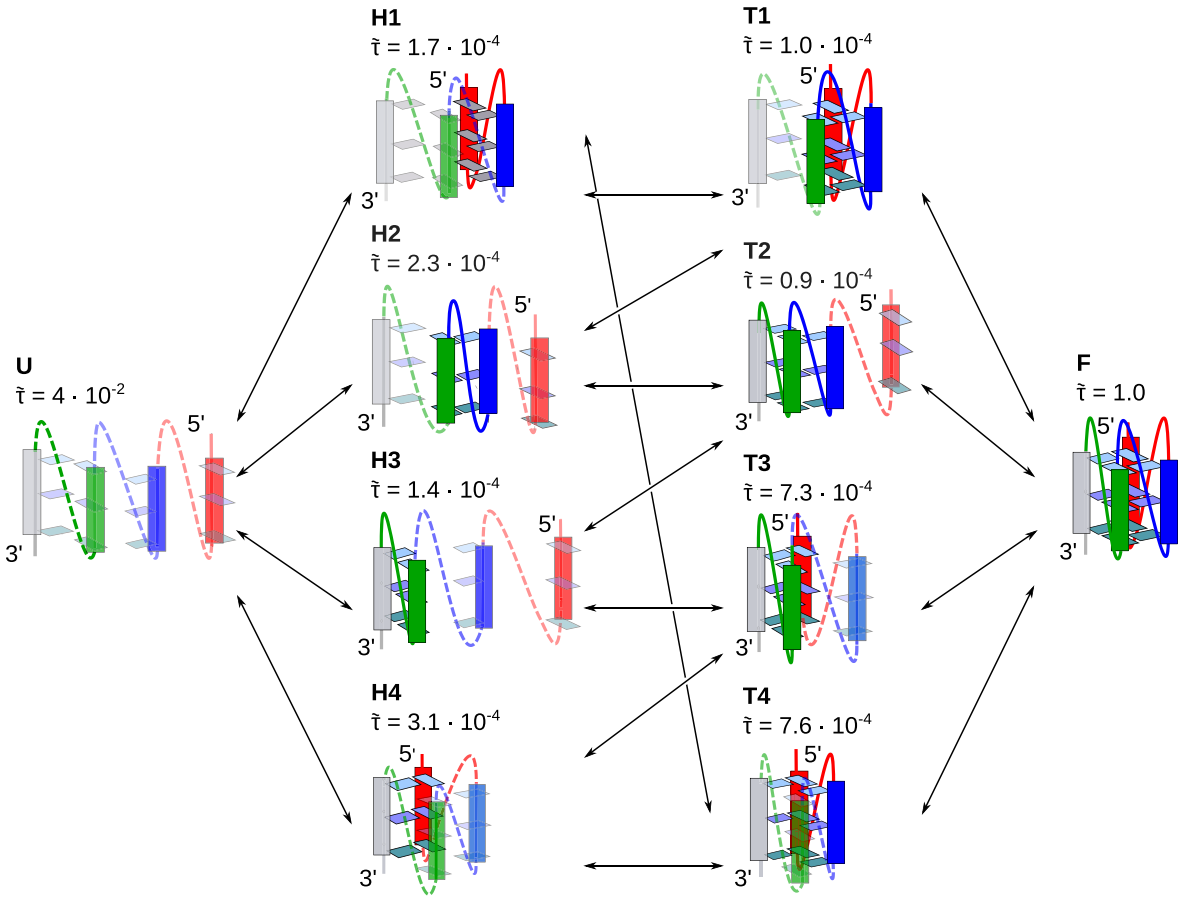


Figure 6. Folding and unfolding pathways of TERRA25. The arrows correspond to the transitions between the conformations. The pathways pass through two intermediate states: hairpin and triplex state. In total four different hairpin (H1, H2, H3, H4) and four different triplex (T1, T2, T3, T4) conformations are observed. The rescaled lifetimes $\tilde{\tau} = \tau/\tau_F$ are shown above the structures with $\tau_F = 6 \times 10^{-5}$ s_{CG}. The rescaled lifetimes are related to the experimental timescale via $\tau_{resc} = \alpha\tilde{\tau}\tau_F$ with $\alpha = 1.56 \times 10^7$. The reduced, CG and rescaled lifetimes are listed in the Supporting Information (Supplementary Table S3).

periments on the folding pathway of a plant rG4 (36). The situation is similar for the triplex intermediates. Again, all four possible triplex conformations are observed in the simulations. Triplex intermediates with $\tilde{\tau} = 7.293 \times 10^{-4}$; $\tau_{resc} = 0.6986$ s (T3) and $\tilde{\tau} = 7.632 \times 10^{-4}$; $\tau_{resc} = 0.7310$ s (T4) are formed by the two terminal repeats and one central repeat (T3 and T4 in Figure 6). The other triplexes (T1, T2) have slightly shorter lifetimes.

Interestingly, a triplex similar to T1 has been observed in smFRET experiments for a plant rG4 (36). Moreover, RNA triplexes with propeller loops have been proposed for rG4s and dG4s with two quartets (35).

In addition to hairpin and triplex, an additional class of intermediates is observed: Two hairpins are created sequentially and form a double-hairpin intermediate state (Figure 7). In pathways containing a double-hairpin intermediate, no triplex state is observed. The double-hairpin structure D1, in which each hairpin comprises one free end, is the most stable $\tilde{\tau} = 5.198 \times 10^{-4}$; $\tau_{resc} = 0.4980$ s .

This is further supported by a similar double-hairpin conformation from NMR spectroscopy on a hybrid G4 of human telomeric DNA (29).

In summary, the pathways between the unfolded and folded state of TERRA25 pass through two intermediate states. The first intermediate is a hairpin and the second one is either a triplex or a double-hairpin state. Each state comprises several different conformations leading to the high conformational entropy of the intermediate states. In total, four different hairpin, four different triplex and two different double-hairpin structures are observed along the folding pathways of TERRA25. Based on the rescaling factor (Eq. 3) the lifetimes of the intermediates are estimated to be on the order of a hundred milliseconds (Supplementary Table S3). Note however that the rescaling factor provides only a rough estimate due to the coarsening of complex molecular interactions into a single scalar. Still, additional all-atom simulations of the hairpin intermediates reveal that these structures remain stable for >500 ns (Supplementary Figure S7). By contrast, if the dangling ends are removed the unfolding kinetics is much faster in agreement with previous MD simulations (37,79).

Conformational entropy as hallmark of rG4 folding

The results for TERRA25 show that the intermediates are degenerate with two to four alternative conformations per state leading to a branched multi-pathway folding process driven by conformational entropy. Here, the immediate question arises whether entropy driven multi-pathway folding is a general feature of rG4 systems. To investigate this hypothesis, we investigate three synthetic rG4 sequences with vary-

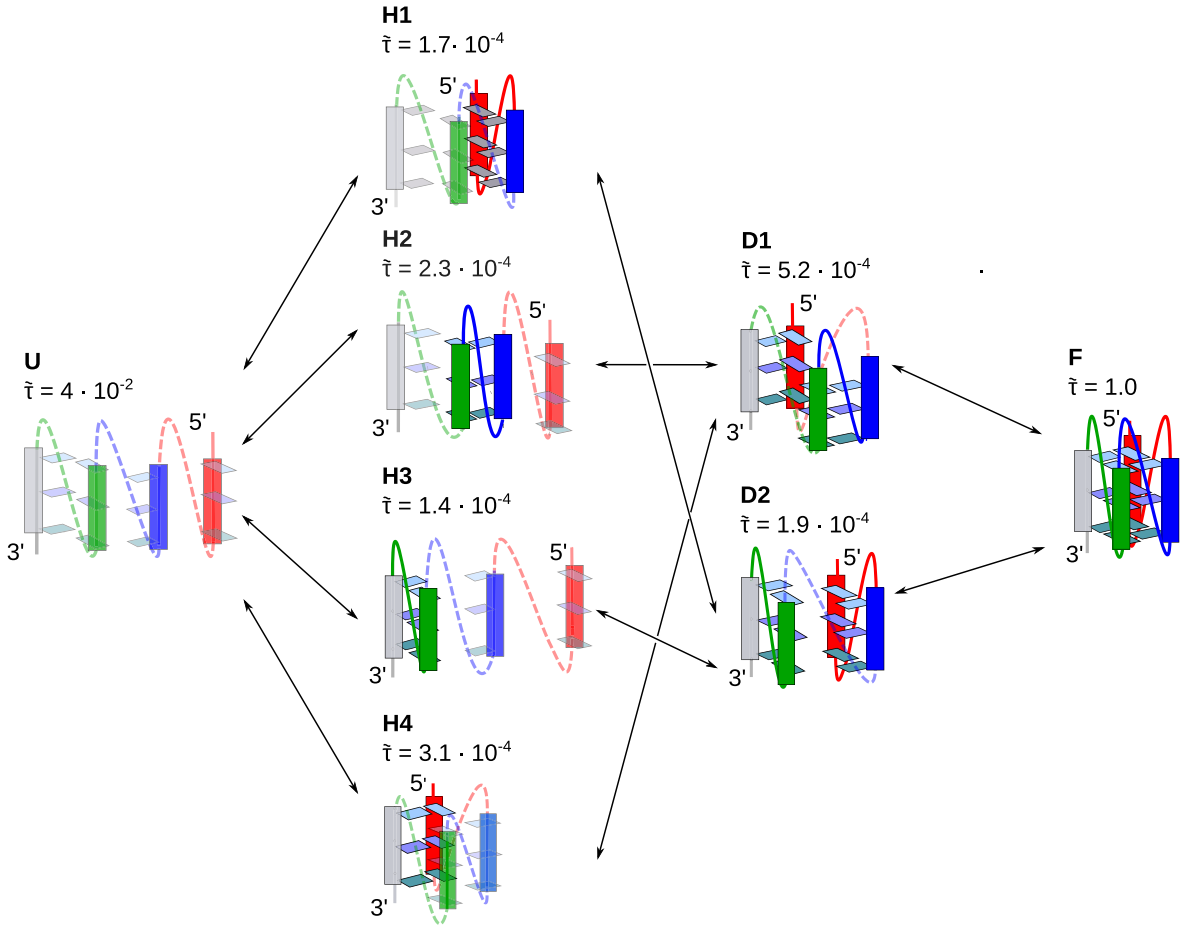


Figure 7. Folding and unfolding pathways of TERRA25. The arrows correspond to the transitions between the conformations. The pathways pass through two intermediate states: hairpin and double-hairpin state. As in Figure 5, four different hairpin (H1, H2, H3, H4) and two different double-hairpin (D1, D2) conformations are observed. The rescaled lifetimes $\tilde{\tau} = \tau/\tau_F$ are shown above the structures with $\tau_F = 6 \cdot 10^{-5}$ s_{CG}. The rescaled lifetimes are related to the experimental timescale via $\tau_{resc} = \alpha\tilde{\tau}\tau_F$ with $\alpha = 1.56 \times 10^7$. The reduced, CG and rescaled lifetimes are listed in the Supporting Information. (Supplementary Table S3)

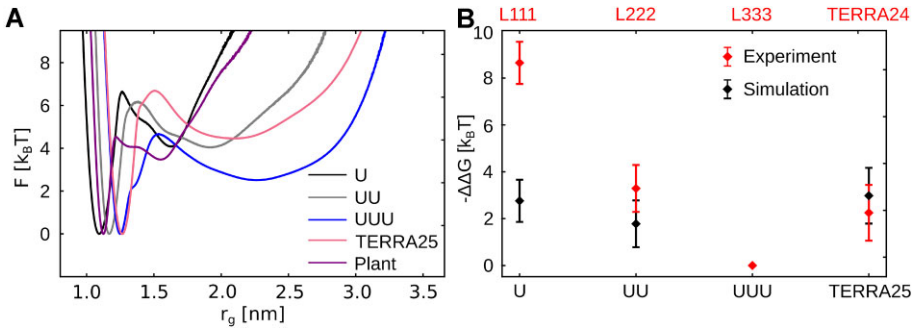


Figure 8. Free energy profiles and stability of different rG4 systems (Table 1). (**A**) Free energy profile as a function of the radius of gyration r_g. (**B**) Gibbs free energy and van't Hoff Gibbs free energy change for the four different systems. The experimental data was calculated from ΔH-TΔS (see Supporting Information) based on the thermodynamic parameters provided in (74). The Gibbs free energy difference $\Delta\Delta G = \Delta G - \Delta G_{ref}$ was calculated with respect to a reference system. For the simulations, we chose the UUU system and the L333 system (74) for the experiments. Errors of results from simulations were calculated from block averaging by dividing trajectories into 10 equal blocks. Errors for experimental values are taken from ref. (74). Values for $\Delta H - T\Delta S$ from (74) as well as Gibbs free energies calculated from simulations are presented in Supplementary Table S4.

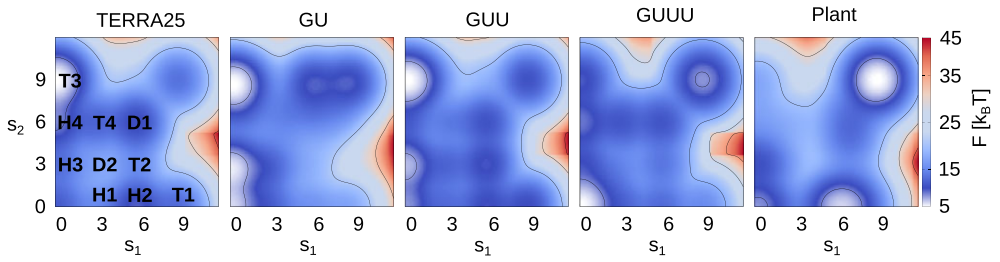


Figure 9. Free energy landscapes of rG4 folding as function of the order parameters s_1 and s_2 (Eq. 8). The positions of the intermediate states are indicated on the left. The free energies of folded and unfolded states are rescaled for visualization.

ing loop length and a plant rG4 sequence with mixed loops (see Table 1). Initially, we calculate the free energy profiles and determine the stability of the five systems. The free energy profiles as function of the radius of gyration r_g all have similar shapes (Figure 8A). The position of the first minimum increases with increasing loop length as expected. The height of the barrier, which separates the folded and unfolded state, increases with decreasing loop length for synthetic G4 sequences indicating that the loop length affects the folding kinetics. The most pronounced differences are the height and width of the second minimum. Clearly, short sequences have a narrower unfolded state due to the smaller number of possible unfolded conformations and hence a smaller conformational entropy. The entropic contribution leads to the highest free energy level for U, followed by UU and UUU. A general trend emerges for the stability namely that rG4 stability decreases with increasing loop length (Figure 8B). The results for UU, UUU and TERRA25 are in quantitative agreement with the experimental results. For the one nucleotide loop deviations are observed which likely result from the different loop sequence used in experiments and simulations.

The free energy landscapes (Figure 9) for the five systems reveal that the branched multi-pathway folding is a general feature for all rG4 systems investigated here. The order parameters s_1 and s_2 allow us to resolve the different intermediate states and confirm that the 12 intermediate states occur in all rG4 systems. Interestingly, the distribution of the states is not uniform but depends on the exact sequence (see Supplementary Table S3 in the Supporting Information for the populations in the intermediate states). Again, the H1 intermediate state for the plant rG4 agrees with the predictions from smFRET (36). However, due to the positioning of the labels and the limited resolution in those experiments further work by smFRET is required to resolve all intermediates.

Conclusion

The folding kinetics of rG4s is essential for their function in cellular processes. In the current work, we resolve the folding pathways of a G4 derived from human telomeric repeat-containing RNA (TERRA25) by combining all-atom MD, CG simulations and CD experiments. Our results reveal a branched multi-pathway folding process for TERRA25 driven by conformational entropy. Moreover, four rG4 systems with varying loop length highlight the fundamental importance of conformational entropy in rG4 folding and reveal branched multi-pathway folding as a characteristic feature of rG4 systems.

Initially, we developed a matching procedure based on atomistic simulations in explicit water to correctly capture the ionic double layer in the implicit solvent CG simulations. The matching correctly reproduces the local salt concentration in the vicinity of the negatively charged phosphate groups and therefore yields close agreement of the fraction of folded structures obtained from CG simulations and experiments at a given salt concentration.

Folding of rG4s is on the timescale of minutes (21) and therefore out of reach for atomistic simulations. CG simulations using the TIS model, on the other hand, reach beyond this limit and allow us to resolve the folding pathways and intermediate states with sufficient statistics. The simulations reveal on-pathways intermediates corresponding to hairpin, triplex or double-hairpin states for all rG4 systems. The intermediates possess a high conformational entropy with two to four structures contributing to each state. However, properties that are easy to measure such as the end-to-end distance are insufficient to discriminate between the different structures. Here, the insights from the simulations can guide the experiments to detect all intermediate structures, for instance by strategically placed fluorophores in smFRET experiments (35,36).

In the future, more complex models could be employed to further explore the off-pathway intermediates including strand-shifts or syn-anti G-patterns. In any case, combining atomistic MD, CG simulations and experiments is a valuable starting point to capture the influence of the ionic environment and to resolve the multi-pathway folding landscape.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

Supplementary data

Supplementary Data are available at NAR Online.

Funding

DFG [CRC902]; Emmy Noether program [315221747]; The authors gratefully acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) under the NHR project b119ee. NHR funding is provided by federal and Bavarian state authorities. NHR@FAU hardware is partially funded by the German Research Foundation (DFG), number 440719683. GOETHE HLR is also acknowledged for supercomputing access. The open access publication of this article was supported by the DFG sponsored Open Access Fund of

the University of Augsburg. BMRZ is supported by the state of Hesse; N.S. and M.U. thank Dave Thirumalai, Naoto Hori, Hung T. Nguyen and Debayan Chakraborty for sharing their code and helpful discussions; M.U. thanks Jürgen Köfinger and Sergio Cruz-Leòn for many fruitful discussions.

Conflict of interest statement

None declared.

References

1. Wanrooij,P.H., Uhler,J.P., Simonsson,T., Falkenberg,M. and Gustafsson,C.M. (2010) G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 16072–16077.
2. Kharel,P., Becker,G., Tsvetkov,V. and Ivanov,P. (2020) Properties and biological impact of RNA G-quadruplexes: from order to turmoil and back. *Nucleic Acids Res.*, **48**, 12534–12555.
3. Dumas,L., Herviou,P., Dassi,E., Cammas,A. and Millevoi,S. (2021) G-Quadruplexes in RNA biology: recent advances and future directions. *Trends Biochem. Sci.*, **46**, 270–283.
4. Zhang,J., Harvey,S.E. and Cheng,C. (2019) A high-throughput screen identifies small molecule modulators of alternative splicing by targeting RNA G-quadruplexes. *Nucleic Acids Res.*, **47**, 3667–3679.
5. Guo,J.U. and Bartel,D.P. (2016) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, **353**, 5371-1–5371-8.
6. Chen,X.C., Chen,S.B., Dai,J., Yuan,J.H., Ou,T.M., Huang,Z.S. and Tan,J.H. (2018) Tracking the dynamic folding and unfolding of RNA G-quadruplexes in live cells. *Angew. Chem. Int. Ed.*, **57**, 4702–4706.
7. Karsisiotis,A.I., O'Kane,C. and Webba da Silva,M. (2013) DNA quadruplex folding formalism - a tutorial on quadruplex topologies. *Methods*, **64**, 28–35.
8. Bhattacharyya,D., Arachchilage,G.M. and Basu,S. (2016) Metal cations in G-quadruplex folding and stability. *Front. Chem.*, **4**, 1–14.
9. Azzalin,C.M., Reichenbach,P., Khoriauli,L., Giulotto,E. and Lingner,J. (2007) Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science (New York, NY)*, **318**, 798–801.
10. Azzalin,C.M. and Lingner,J. (2008) Telomeres: the silence is broken. *Cell Cycle*, **7**, 1161–1165.
11. Agarwala,P., Pandey,S. and Maiti,S. (2015) The tale of RNA G-quadruplex. *Org. Biomol. Chem.*, **13**, 5570–5585.
12. Biffi,G., Tannahill,D. and Balasubramanian,S. (2012) An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *J. Am. Chem. Soc.*, **134**, 11974–11976.
13. Gilley,D., Tanaka,H. and Herbert,B.S. (2005) Telomere dysfunction in aging and cancer. *Int. J. Biochem. Cell Biol.*, **37**, 1000–1013.
14. Marzano,F., Rapacciuolo,A., Ferrara,N., Rengo,G., Koch,W.J. and Cannavo,A. (2021) Targeting GRK5 for treating chronic degenerative diseases. *Int. J. Mol. Sci.*, **22**, 1–17.
15. Zhao,C., Qin,G., Niu,J., Wang,Z., Wang,C., Ren,J. and Qu,X. (2021) Targeting RNA G-quadruplex in SARS-CoV-2: a promising therapeutic target for COVID-19?. *Angew. Chem. Int. Ed.*, **60**, 432–438.
16. Šponer,J., Bussi,G., Stadlbauer,P., Kührová,P., Banáš,P., Islam,B., Haider,S., Neidle,S. and Otyepka,M. (2017) Folding of guanine quadruplex molecules–funnel-like mechanism or kinetic partitioning? An overview from MD simulation studies. *Biochim. Biophys. Acta - Gen. Subj.*, **1861**, 1246–1263.
17. Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Plückthun,A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 8572–8577.
18. Paeschke,K., Juranek,S., Simonsson,T., Hempel,A., Rhodes,D. and Lipps,H.J. (2008) Telomerase recruitment by the telomere end binding protein-β facilitates G-quadruplex DNA unfolding in ciliates. *Nat. Struct. Mol. Biol.*, **15**, 598–604.
19. Yang,X., Cheema,J., Zhang,Y., Deng,H., Duncan,S., Umar,M.I., Zhao,J., Liu,Q., Cao,X., Kwok,C.K., *et al.* (2020) RNA G-quadruplex structures exist and function in vivo in plants. *Genome Biol.*, **21**, 1–23.
20. Kharel,P., Fay,M., Manasova,E.V., Anderson,P.J., Kurkin,A.V., Guo,J.U. and Ivanov,P. (2023) Stress promotes RNA G-quadruplex folding in human cells. *Nat Commun.*, **14**, 1–10.
21. Müller,D., Bessi,I., Richter,C. and Schwalbe,H. (2021) The folding landscapes of human telomeric RNA and DNA G-quadruplexes are markedly different. *Angew Chem. Int. Ed.*, **60**, 10895–10901.
22. Adrian,M., Heddi,B. and Phan,A.T. (2012) NMR spectroscopy of G-quadruplexes. *Methods*, **57**, 11–24.
23. Wirmer-Bartoschek,J., Bendel,L.E., Jonker,H.R., Grün,J.T., Papi,F., Bazzicalupi,C., Messori,L., Gratteri,P. and Schwalbe,H. (2017) Solution NMR structure of a ligand/Hybrid-2-G-quadruplex complex reveals rearrangements that affect ligand binding. *Angew. Chem. Int. Ed.*, **56**, 7102–7106.
24. Binas,O., Bessi,I. and Schwalbe,H. (2020) Structure validation of G-rich RNAs in noncoding regions of the human genome. *Chembiochem*, **21**, 1656–1663.
25. Bessi,I., Bazzicalupi,C., Richter,C., Jonker,H.R., Saxena,K., Sissi,C., Chioccioli,M., Bianco,S., Bilia,A.R., Schwalbe,H., *et al.* (2012) Spectroscopic, molecular modeling, and NMR-spectroscopic investigation of the binding mode of the natural alkaloids berberine and sanguinarine to human telomeric G-quadruplex DNA. *ACS Chem. Biol.*, **7**, 1109–1119.
26. Campbell,N.H. and Parkinson,G.N. (2007) Crystallographic studies of quadruplex nucleic acids. *Methods*, **43**, 252–263.
27. Bessi,I., Jonker,H.R., Richter,C. and Schwalbe,H. (2015) Involvement of long-lived intermediate states in the complex folding pathway of the human telomeric G-quadruplex. *Angew. Chem. Int. Ed.*, **54**, 8444–8448.
28. Grün,J.T., Blümler,A., Burkhart,I., Wirmer-Bartoschek,J., Heckel,A. and Schwalbe,H. (2021) Unraveling the kinetics of spare-tire DNA G-quadruplex folding. *J. Am. Chem. Soc.*, **143**, 6185–6193.
29. Frelih,T., Wang,B., Plavec,J. and Šket,P. (2020) Pre-folded structures govern folding pathways of human telomeric G-quadruplexes. *Nucleic Acids Res.*, **48**, 2189–2197.
30. Largy,E., König,A., Ghosh,A., Ghosh,D., Benabou,S., Rosu,F. and Gabelica,V. (2022) Mass spectrometry of nucleic acid noncovalent complexes. *Chem. Rev.*, **122**, 7720–7839.
31. Marchand,A. and Gabelica,V. (2016) Folding and misfolding pathways of G-quadruplex DNA. *Nucleic Acids Res.*, **44**, 10999–11012.
32. Cerofolini,L., Amato,J., Giachetti,A., Limongelli,V., Novellino,E., Parrinello,M., Fragai,M., Randazzo,A. and Luchinat,C. (2014) G-triplex structure and formation propensity. *Nucleic Acids Res.*, **42**, 13393–13404.
33. Dai,J., Carver,M., Punchihewa,C., Jones,R.A. and Yang,D. (2007) 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.*, **35**, 4927–4940.
34. Dai,J., Punchihewa,C., Ambrus,A., Chen,D., Jones,R.A. and Yang,D. (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: A novel adenine triple formation. *Nucleic Acids Res.*, **35**, 2440–2450.
35. Zhang,A.Y. and Balasubramanian,S. (2012) The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids. *J. Am. Chem. Soc.*, **134**, 19297–19308.
36. Wang,L., Xu,Y.-P., Bai,D., Shan,S.-W., Xie,J., Li,Y. and Wu,W.-Q. (2022) Insights into the structural dynamics and helicase-catalyzed

unfolding of plant RNA G-quadruplexes. *J. Biol. Chem.*, **298**, 102165.
37. Stadlbauer,P., Kührová,P., Banáš,P., Koča,J., Bussi,G., Trantírek,L., Otyepka,M. and Šponer,J. (2015) Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations. *Nucleic Acids Res.*, **43**, 9626–9644.
38. Stadlbauer,P., Kührová,P., Vicherek,L., Banáš,P., Otyepka,M., Trantírek,L. and Šponer,J. (2019) Parallel G-triplexes and G-hairpins as potential transitory ensembles in the folding of parallel-stranded DNA G-Quadruplexes. *Nucleic Acids Res.*, **47**, 7276–7293.
39. Luo,D. and Mu,Y. (2016) Computational insights into the stability and folding pathways of human telomeric DNA G-quadruplexes. *J. Phys. Chem. B*, **120**, 4912–4926.
40. Limongelli,V., De Tito,S., Cerofolini,L., Fragai,M., Pagano,B., Trotta,R., Cosconati,S., Marinelli,L., Novellino,E., Bertini,I., *et al.* (2013) The G-triplex DNA. *Angew. Chem. Int. Ed.*, **52**, 2269–2273.
41. Stadlbauer,P., Mazzanti,L., Cragnolini,T., Wales,D.J., Derreumaux,P., Pasquali,S. and Šponer,J. (2016) Coarse-grained simulations complemented by atomistic molecular dynamics provide new insights into folding and unfolding of human telomeric G-quadruplexes. *J. Chem. Theory Comput.*, **12**, 6077–6097.
42. Lemkul,J.A. (2020) Same fold, different properties: polarizable molecular dynamics simulations of telomeric and TERRA G-quadruplexes. *Nucleic Acids Res.*, **48**, 561–575.
43. Siebenmorgen,T. and Zacharias,M. (2017) Origin of ion specificity of telomeric DNA G-quadruplexes investigated by free-energy simulations. *Biophys. J.*, **112**, 2280–2290.
44. Salsbury,A., Dean,T. and Lemkul,J.A. (2020) Polarizable molecular dynamics simulations of two c-kit oncogene promoter G-quadruplexes: effect of primary and secondary structure on loop and ion sampling. *J. Chem. Theory Comput.*, **16**, 3430–3444.
45. Boniecki,M.J., Lach,G., Dawson,W.K., Tomala,K., Lukasz,P., Soltysinski,T., Rother,K.M. and Bujnicki,J.M. (2015) SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.*, **44**, e63.
46. Pasquali,S. and Derreumaux,P. (2010) HiRE-RNA: a high resolution coarse-grained energy model for RNA. *J. Phys. Chem. B*, **114**, 11957–11966.
47. Cruz-León,S., Vázquez-Mayagoitia,A., Melchionna,S., Schwierz,N. and Fyta,M. (2018) Coarse-grained double-stranded RNA model from quantum-mechanical calculations. *J. Phys. Chem. B*, **122**, 7915–7928.
48. Denesyuk,N.A. and Thirumalai,D. (2013) Coarse-grained model for predicting RNA folding thermodynamics. *J. Phys. Chem. B*, **117**, 4901–4911.
49. Hyeon,C. and Thirumalai,D. (2005) Mechanical unfolding of RNA hairpins. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 6789–6794.
50. Denesyuk,N.A., Hori,N. and Thirumalai,D. (2018) Molecular simulations of ion effects on the thermodynamics of RNA folding. *The J. Phys. Chem. B*, **122**, 11860–11867.
51. Hori,N., Denesyuk,N.A. and Thirumalai,D. (2016) Salt effects on the thermodynamics of a frameshifting RNA pseudoknot under tension. *J. Mol. Biol.*, **428**, 2847–2859.
52. Denesyuk,N.A. and Thirumalai,D. (2015) How do metal ions direct ribozyme folding?. *Nat. Chem.*, **7**, 793–801.
53. Nguyen,H.T., Hori,N. and Thirumalai,D. (2019) Theory and simulations for RNA folding in mixtures of monovalent and divalent cations. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 21022–21030.
54. Fuks,C., Falkner,S., Schwierz,N. and Hengesbach,M. (2022) Combining coarse-grained simulations and single molecule analysis reveals a three-state folding model of the Guanidine-II riboswitch. *Front. Mol. Biosci.*, **9**, 1–16.
55. Martadinata,H. and Phan,A.T. (2009) Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric rna sequences in K+ solution. *J. Am. Chem. Soc.*, **131**, 2570–2578.
56. Rother,M., Rother,K., Puton,T. and Bujnicki,J.M. (2011) ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.*, **39**, 4007–4022.
57. Hjelm,B.E., Rollins,B., Morgan,L., Sequeira,A., Mamdani,F., Pereira,F., Damas,J., Webb,M.G., Weber,M.D., Schatzberg,A.F., *et al.* (2019) Splice-break: Exploiting an RNA-seq splice junction algorithm to discover mitochondrial DNA deletion breakpoints and analyses of psychiatric disorders. *Nucleic Acids Research*, **47**, 1–14.
58. Hess,B., Kutzner,C., van der Spoel,D. and Lindahl,E. (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theo. Comp.*, **4**, 435–447.
59. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
60. Hess,B., Bekker,H., Berendsen,H.J. and Fraaije,J.G. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463–1472.
61. Bayly,C.I., Merz,K.M., Ferguson,D.M., Cornell,W.D., Fox,T., Caldwell,J.W., Kollman,P.A., Cieplak,P., Gould,I.R. and Spellmeyer,D.C. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
62. Pérez,A., Marchán,I., Svozil,D., Sponer,J., Cheatham,T.E., Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.*, **92**, 3817–3829.
63. Zgarbová,M., Otyepka,M., Šponer,J., Mládek,A., Banáš,P., Cheatham,T.E. and Jurečka,P. (2011) Refinement of the Cornell etal. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.*, **7**, 2886–2902.
64. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
65. Mamatkulov,S. and Schwierz,N. (2018) Force fields for monovalent and divalent metal cations in TIP3P water based on thermodynamic and kinetic properties. *J. Chem. Phys.*, **148**, 074504-1–074504-11.
66. Bussi,G., Donadio,D. and Parrinello,M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**, 14101.
67. Berendsen,H.J., Postma,J.P., Van Gunsteren,W.F., Dinola,A. and Haak,J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
68. Parrinello,M. and Rahman,A. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.*, **52**, 7182–7190.
69. Cruz-León,S., Vanderlinden,W., Müller,P., Forster,T., Staudt,G., Lin,Y.Y., Lipfert,J. and Schwierz,N. (2022) Twisting DNA by salt. *Nucleic Acids Res.*, **50**, 5726–5738.
70. Taketomi,H., Ueda,Y. and Gō,N. (1975) Studies on protein folding, unfolding and fluctuations by computer simulation. *Int. J. Pept. Protein Res.*, **7**, 445–459.
71. Sharp,K.A. and Honig,B. (1990) Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J. Phys. Chem.*, **94**, 7684–7692.
72. Honeycutt,J.D. and Thirumalai,D. (1992) The nature of folded states of globular proteins. *Biopolymers*, **32**, 695–709.
73. Hori,N. and Thirumalai,D.(2023)Watching ion-driven kinetics of ribozyme folding and misfolding caused by energetic and topological frustration one molecule at a time. *Nucleic Acids Res.*, **51**, 10737–10751.
74. Zhang,A.Y., Bugaut,A. and Balasubramanian,S. (2011) A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry*, **50**, 7251–7258.
75. Abrams,J.B. and Tuckerman,M.E. (2008) Dynamics without coordinate transformations. *J. Phys. Chem. B*, **112**, 15742–15757.

76. Mergny,J.L. and Lacroix,L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.
77. Gray,R.D., Trent,J.O. and Chaires,J.B. (2014) Folding and unfolding pathways of the human telomeric G-quadruplex. *J. Mol. Biol.*, **426**, 1629–1650.
78. Bian,Y., Tan,C., Wang,J., Sheng,Y., Zhang,J. and Wang,W. (2014) Atomistic picture for the folding pathway of a Hybrid-1 type human telomeric DNA G-quadruplex. *PLoS Comput. Biol.*, **10**, e1003562.
79. Havrila,M., Stadlbauer,P., Kührová,P., Banáš,P., Mergny,J.L., Otyepka,M. and Šponer,J. (2018) Structural dynamics of propeller loop: towards folding of RNA G-quadruplex. *Nucleic Acids Res.*, **46**, 8754–8771.

Received: February 10, 2023. **Revised:** September 25, 2023. **Editorial Decision:** October 23, 2023. **Accepted:** October 25, 2023