How to cite: *Angew. Chem. Int. Ed.* **2024**, *63*, e202313226
doi.org/10.1002/anie.202313226

VIP DNA Quadruplex Very Important Paper

DNA Quadruplex Structure with a Unique Cation Dependency

Martin Gajarsky, Petr Stadlbauer, Jiri Sponer, Anne Cucchiarini, Michaela Dobrovolna, Vaclav Brazda, Jean-Louis Mergny, Lukas Trantirek, and Martina Lenarcic Zivkovic**

Abstract: DNA quadruplex structures provide an additional layer of regulatory control in genome maintenance and gene expression and are widely used in nanotechnology. We report the discovery of an unprecedented tetrastranded structure formed from a native G-rich DNA sequence originating from the telomeric region of *Caenorhabditis elegans.* The structure is defined by multiple properties that distinguish it from all other known DNA quadruplexes. Most notably, the formation of a stable so-called KNa-quadruplex (KNaQ) requires concurrent coordination of K^+ and Na^+ ions at two distinct binding sites. This structure provides novel insight into G-rich DNA folding under ionic conditions relevant to eukaryotic cell physiology and the structural evolution of telomeric DNA. It highlights the differences between the structural organization of human and nematode telomeric DNA, which should be considered when using *C. elegans* as a model in telomere biology, particularly in drug screening applications. Additionally, the absence/presence of KNaQ motifs in the host/parasite introduces an intriguing possibility of exploiting the KNaQ fold as a plausible antiparasitic drug target. The structure's unique shape and ion dependency and the possibility of controlling its folding by using low-molecular-weight ligands can be used for the design or discovery of novel recognition DNA elements and sensors.

Introduction

G-quadruplexes (G4s) constitute the most studied class of noncanonical (non-B) DNA structures that form within G-rich, evolutionarily conserved regulatory regions of the prokaryotic and eukaryotic genomes (reviewed in[1]). The formation of G4s results from the stacking of two or more planar (G:G:G:G) quartets stabilized by Hoogsteen-type guanine-guanine base pairing and the coordination of metal ions in the central channel. In mammals, the formation of G4s in gene promoters contributes to regulating gene expression.[1,2] The formation of G4s in human telomeric DNA interferes with critical telomeric functions, allowing their exploitation as plausible anticancer targets.[3] Notably, several different G4 topologies have been shown to form within human telomeric DNA: The formation of different topologies was shown to depend on the choice of the studied oligonucleotide sequence (*i.e.*, constructs of tandem repeats varying in primary sequence length and the nature of 5′ and 3′ flanking residues) and the environmental (ionic) conditions.[4] The observed topological variability of human telomeric G4s and their sensitivity to experimental conditions expose the complexity of the conformational landscape of G-rich telomeric DNA.

Numerous biophysical studies have confirmed the formation of G4s in telomeric DNA fragments from many distinct species (from ciliates and fungi to plants and humans), establishing G4-forming potential as its evolutionarily conserved structural hallmark.[5] In spite of that, exceptions to this notion were recently reported.[5d,6] Arguably, the most significant exception has been the unsettled case of telomeric DNA from *Caenorhabditis elegans.* A recent NMR study demonstrated that a 20 nucleotide-long construct, derived from the telomeric sequence of *C. elegans*, d[(GGCTTA)$_3$GG] (referred to as Ce20), folds into a two-

[*] Dr. M. Gajarsky, A. Cucchiarini, Prof. L. Trantirek, Dr. M. Lenarcic Zivkovic
Central European Institute of Technology (CEITEC), Masaryk University
Kamenice 753/5, 62500 Brno (Czech Republic)
E-mail: lukas.trantirek@ceitec.muni.cz

Dr. P. Stadlbauer, Prof. J. Sponer, M. Dobrovolna, Prof. V. Brazda, Dr. J.-L. Mergny
Institute of Biophysics, Czech Academy of Sciences
Kralovopolska 135, 61265 Brno (Czech Republic)

A. Cucchiarini, Dr. J.-L. Mergny
Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS, Inserm, Institut Polytechnique de Paris
91120 Palaiseau (France)

M. Dobrovolna, Prof. V. Brazda
Faculty of Chemistry, Brno University of Technology
Purkynova 464, 61200 Brno (Czech Republic)

Dr. M. Lenarcic Zivkovic
Slovenian NMR Centre, National Institute of Chemistry
Hajdrihova 19, 1000 Ljubljana (Slovenia)
E-mail: martina.lenarcic@ki.si

Dr. M. Gajarsky
Current address: Center for Molecular Medicine Cologne, University of Cologne
50931 Cologne (Germany)

quartet antiparallel G4 in the presence of K^+ ions as well as in a buffer emulating the complex ion composition of the intracellular space (Figure 1A).[5d,7] Interestingly, spectroscopic data of its single-nucleotide extended variant, d-[(GGCTTA)$_3$GGC] (hereafter designated as Ce21), in a complex ionic environment provided evidence for the formation of a non-G4 structure (Figure 1B).[5d] These conflicting results have reopened the debate on the role of G4 structures in telomere biology and the suitability of simplistic model buffer systems, primarily based on a single type of counterion, in biomolecular structural studies. As *C. elegans* is a widely used model system to study the mechanism of telomere maintenance,[8] we deemed it essential to gain insights into the structural details of Ce21 and explore its potential sensitivity to environmental conditions.

Here, we show that in a buffer containing the two most natively abundant counterions in the intracellular space (K^+ and Na^+), Ce21 folds into an unprecedented tetrastranded structure, referred to as KNa-quadruplex (KNaQ), which is stabilized by the concurrent binding of K^+ and Na^+ ions in two distinct cation binding sites in the structure's core. In addition to the unique ion dependency, the KNaQ structure differs from closely related G4 structures by the absence of stacked G-quartets, a different groove width, and susceptibility toward G4 binding ligands. While Ce21-related sequences are extremely rare in the human genome, they are abundant in the genomes of (non)parasitic helminths, indicating the potential use of the KNaQ fold as an antiparasitic drug target. Moreover, the structure's unique shape and ion dependency and the possibility of controlling its folding by using low-molecular-weight ligands, as demonstrated here, can be exploited for the design or discovery of novel recognition DNA elements and sensors.

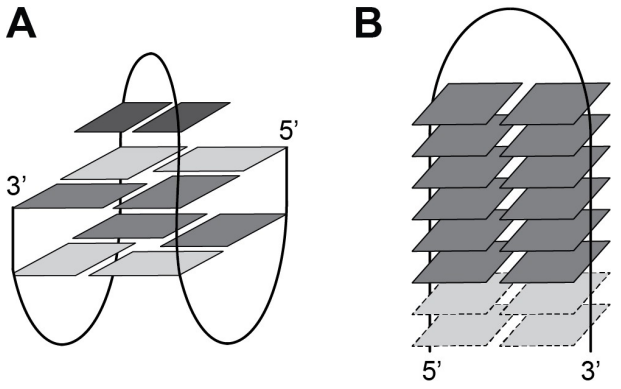

Figure 1. Schematic representations of (A) a G4 formed by Ce20 of *C. elegans* telomeric sequence, determined in K^+-based solution,[7] and (B) a foldback structure proposed to form by Ce21 in complex ionic conditions.[5d]

Results and Discussion

Insights into the cation-dependent folding of the C. elegans G-rich telomeric DNA

We separately characterized Ce21, a 21-nucleotide repeat sequence d[(GGCTTA)$_3$GGC], defining the *C. elegans* (nematode) telomeric DNA in buffers containing Na^+, K^+, Li^+, and NH_4^+ ions and in a buffer with complex ionic composition [25 mM NaPOi, 110 mM KCl, 10.5 mM NaCl, 1 mM $MgCl_2$, 130 nM $CaCl_2$, pH = 7.5], referred to as the intracellular buffer (ICB), using NMR and circular dichroism (CD) spectroscopies. The CD spectral shapes and the absence of the imino signals in the 1D ^{1}H NMRspectra of Ce21 provided evidence that it did not form any higher-order DNA structure in the presence of Li^+ and NH_4^+ ions (Figure 2A, B). Conversely, the characteristic pattern of the 1D ^{1}H NMR spectrum, together with the CD profile, showed the formation of an antiparallel G4 in the presence of K^+ ions (Figures 1A and 2A,B), in line with Marquevielle *et al.*[7] Notably, the folding of Ce21 in ICB led to a structure that was fundamentally different from the G4 formed in the K^+-based solution, as substantiated by the altered CD profile and signal pattern in the imino region of the corresponding NMR spectrum (Figure 2A,B). The positions of imino signals in the NMR spectrum of Ce21 acquired in the presence of Na^+ ions resembled those observed in the NMR spectrum recorded in ICB (Figure 2A). However, based on the comparison of signal intensities of the C9 H6 proton (*i.e.*, the non-overlapped aromatic signal representing the structure in ICB), we deduced that the folded species in NaCl represents a population of approximately 5% in comparison to that of the ICB, while the rest of the oligonucleotide remained unfolded (Figure S1). Additionally, differences in chemical shifts of signals between δ 11.0 and 11.5 ppm observed in ICB and Na^+-based solution indicate that these two structures may not be identical (Figure 2A). The differences may arise from different numbers and/or geometries of coordinated cations. The almost completely unfolded state of Ce21 in the presence of Na^+ ions was corroborated by the shape of the corresponding CD spectrum, which was similar to those acquired under conditions that prohibited Ce21 folding (in the presence of Li^+ and NH_4^+) (Figure 2B). Additionally, Mg^{2+} and Ca^{2+} ions, also present in the ICB buffer, were shown not to influence the folding of Ce21 (Figure S2). Overall, these data indicate that the formation and stabilization of Ce21 structure in ICB critically depend on the concurrent presence of K^+ and Na^+ ions.

To substantiate the effect of both ions on the folding of Ce21, we separately annealed the oligonucleotide in the presence of 100 mM NaCl or KCl. We then titrated Na^+- and K^+-based solutions with increasing concentrations of KCl or NaCl, respectively. The addition of KCl to the Na^+-based solution of Ce21 resulted in sharpening and an increase of the imino signals in NMR spectra in a [K^+]-dependent manner, eventually leading to an NMR spectral fingerprint identical to the one acquired in ICB (Figures 2C and S3A). In contrast, adding NaCl to a K^+-based solution

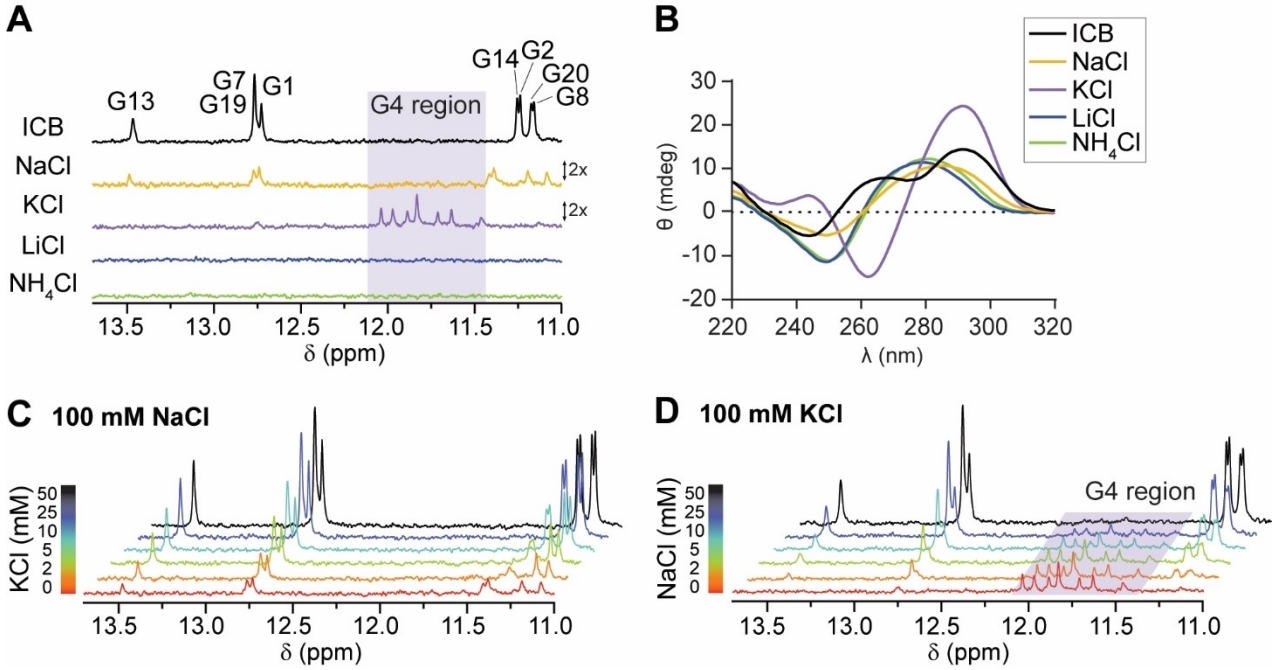

Figure 2. Influence of the cation type on Ce21 folding. (A) Imino regions of 1D ^{1}H NMR spectra and (B) CD spectra of Ce21 in ICB and different single cation-supplemented Tris-HCl buffers. (C and D) Imino regions of 1D ^{1}H NMR spectra of Ce21 in 100 mM NaCl and 100 mM KCl titrated with KCl and NaCl, respectively.

completely changed the signal pattern in the NMR spectrum of Ce21 (Figure 2D). With increasing NaCl concentration, the G4 imino signals gradually decreased and finally disappeared in the noise at 50 mM NaCl. Most importantly, with decreasing signals of G4, intensities of the imino, amino, and aromatic signals specific to the unprecedented structure simultaneously increased (Figures 2D and S3B). The titration experiments (Figure 2C, D) corroborated that the formation of a stable Ce21 structure observed in ICB depends on the simultaneous presence of K^+ and Na^+ ions.

High-resolution NMR structure of Ce21 in the concurrent presence of K^+ and Na^+ ions

We elucidated the structure of Ce21 in ICB using thorough assignment approaches based on site-specific isotopic labels and through-bond correlations using NMR spectroscopy following standard protocols (Figures S4 and S5).[9] Structural calculations were conducted based on 481 NOE-derived distance restraints along with 20 hydrogen bonds, 21 torsion angles, 12 planarity, and 2 repulsive distance restraints using restrained simulated annealing protocol (Table 1). The resulting structure shows that Ce21, in the presence of K^+ and Na^+ ions, folds into an unprecedented tetrastranded structure, referred to as KNa-quadruplex (KNaQ) (Figures 3A,B and S6). The KNaQ structure features a well-defined compact core composed of a single G-quartet sandwiched between GC-based structural elements consisting of two G–C Watson–Crick base pairs on each side of the G-quartet (Figure 3). The single G-quartet, comprised of residues G2·G20·G14·G8, is marked by an *anti:syn:anti:syn* arrangement of glycosidic conformations with an anticlockwise progression of hydrogen bonds (Figure 3B). *Syn* glycosidic conformations for residues G8 and G20 were substantiated by their intense intra-residual H8-H1′ NOE cross-peaks (Figure S5B) and corroborated by the downfield chemical shift of C8 carbon atoms at around δ_C

Table 1: NMR statistics of the Ce21 KNaQ structure.

NMR restraints	
NOE-derived distance restraints	
Total	481
Intra-residual	255
Inter-residual	226
Sequential	167
Long-range	59
Hydrogen bond restraints	20
Torsion angle restraints	21
Planarity restraints	12
Repulsive distance restraints	2
Structure Statistics	
Violations	
Mean NOE restraint violation (Å)	0.055 ± 0.008
Max. NOE restraint violation (Å)	0.074
Max. torsion angle restraint violation (°)	0
Deviations from idealized geometry	
Bonds (Å)	0.0131 ± 0.0003
Angles (°)	2.491 ± 0.044
Pairwise heavy atom RMSD (Å)	
Overall	1.22 ± 0.30
Without loop residues	0.82 ± 0.18

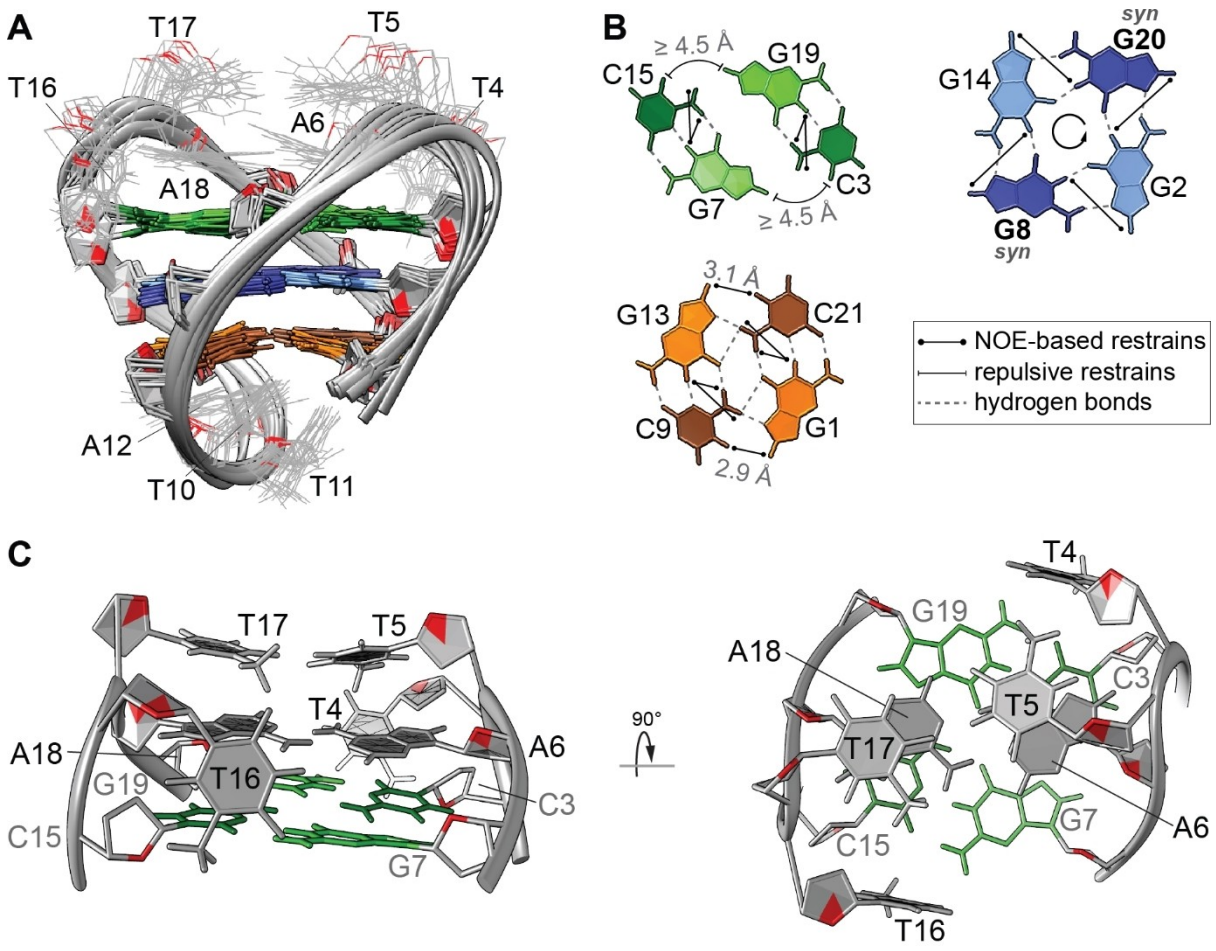

Figure 3. High-resolution structure of Ce21 KNaQ in ICB buffer. (A) Superposition of the ten lowest-energy solution structures of Ce21 (PDB ID: 8BZU, BMRB ID: 34780) in ICB. (B) Arrangements of the slipped G7·C15 and G19·C3 base pairs (green), G2·G20·G14·G8 G-quartet (blue), and direct G1·C21·G13·C9 quartet (orange). (C) Side view and bird's eye view of the T4-T5-A6 and T16-T17-A18 loops stacked around the slipped G7·C15 and G19·C3 base pairs.

140 ppm (Figure S4B), respectively. G–C Watson–Crick base pairs that are found on both sides of the G-quartet (*i.e.*, G1·C21 and G13·C9 on the 5′-, and G7·C15 and G19·C3 on the 3′-side) decisively differ in their arrangements. G1·C21 and G13·C9 are arranged into a direct G·C·G·C quartet *via* their major groove edges (Figures 3B and S7). This alignment is supported by the observation of strong NOE connections (interproton distances ≈3 Å) between the guanine H8 proton from one G–C base pair and the cytosine H5 proton from neighboring G–C base pair (Figures 3B and S5B). In contrast, G7·C15 and G19·C3 are separated and displaced in opposite directions to the helical axis, as shown by the absence of H8-H5 NOE cross-peaks and the observation of very weak H8-amino NOEs between two neighboring base pairs (Figures 3B, S5B, and S7B).

Antiparallel strands of KNaQ are linked by three lateral loops and form two wide and two narrow grooves (Figure 3A). The positions of residues from both loops that reside above the slipped G7·C15 and G19·C3 base pairs and span narrow grooves (*i.e.*, T4-T5-A6 and T16-T17-A18) are defined by a high number (almost 100) of sequential and long-range NOEs (Figures S5B and S8). Observing such a high number of structure-defining NOE connections between these residues corroborates the formation of a very compact and well-defined part of the structure (Figure 3A, C). Tightly packed loop residues were, in the first round of our structural calculations, forcing G7·C15 and G19·C3 base pairs spatially too close and into the formation of a direct G·C·G·C quartet, which was inconsistent with NMR data (see above). To solve this artifact, we introduced two repulsive distance restraints between C3 H42 - G7 N7 and C15 H42 - G19 N7 atoms. These restraints keep G7·C15 and G19·C3 base pairs apart in space consistently with NMR observations. Importantly, they do not influence the overall heart-shaped structure of Ce21 nor result in significant differences in NOE-derived distance restraint violations (Figure S9, Table S1). Noteworthy, symmetrical positions of narrow grooves-spanning loops agree with very similar chemical shifts and patterns of NOEs observed for residues on equivalent positions of C3-G7 and C15-G19 segments (for a more detailed description of NOEs, see Figure S8). On the other side, loop residues T10-T11-A12 represent the most dynamic part of the structure (Figure 3A); thus, only sequential NOE cross-peaks between loop-forming residues

and residues C9 and G13 were observed, while no NOE connectivities to G1 and C21 were detected (Figure S8C).

While the widths of the wide grooves of KNaQ (20.7 ± 0.4 Å) are comparable with values in the antiparallel G4s (20.0 ± 1.0 Å), narrow grooves are wider by approximately 2 Å (10.1 ± 0.3 Å in KNaQ compared to 8.1 ± 1.2 Å in G4s) (Figure S10).[4k,10] Despite the presence of wider narrow grooves, A6 and A18 from the first and third lateral loops of Ce21 stack on top of both slipped G–C base pairs and, together with tightly packed T4, T5, T16, and T17, form a well-defined capping structure. Stacking of residues from the lateral loops bridging narrow grooves is often observed in other tetrastranded structures such as G4s and is believed to significantly contribute to the stability of the structure.[4c,11] The well-defined core of the KNaQ structure, together with tightly packed loop residues that form a capping structure, is also in accord with the melting temperature of Ce21 (T_m 51 °C), which is higher by approximately 11 °C than that of a two-quartet G4 formed by a nucleotide shorter Ce20 (T_m 40 °C)[7] (Figure S11).

MD simulations reveal distinct coordination sites for K^+ and Na^+ ions

Different arrangements of GC-based structural elements above and below the G-quartet indicate the existence of two distinct binding sites (BS1 and BS2) suitable to coordinate different types of cations (Figure 4A). To determine the interaction sites of K^+ and Na^+ ions, we used unrestrained molecular dynamic (MD) simulations (starting with the experimentally determined NMR structure) in an explicit solvent involving the simultaneous presence of both ions. MD simulations were initiated from two distinct ionic configurations: KNa and NaK. In the KNa configuration, K^+ and Na^+ ions were placed into putative binding sites BS1 and BS2, respectively, while in the NaK configuration, the positions of both ions were swapped (Figure 4A,B). In both KNa and NaK MD simulations (two 10 μs replicas for each starting configuration), the KNaQ fold remained stable. Most notably, despite starting from two distinct ionic configurations, all MD simulations reproducibly ended in an identical layout corresponding to the KNa state, suggesting that the coordination of K^+ and Na^+ ions in BS1 and BS2, respectively, was responsible for the formation of a stable KNaQ structure (Figure 4B,C). As observed by MD simulations, the coordination site for Na^+ ion resides between the G-quartet and direct G1·C21·G13·C9 quartet, in which Na^+ is stabilized by six guanine O6 atoms from both quartets (Figure S12A). Formation of G1·C21·G13·C9 quartet brings G1 and G13 spatially close enough to Na^+ to ensure proper coordination. The coordination site for the K^+ ion in Ce21 is located between the G-quartet and two G–C base pairs, where K^+ is coordinated by four O6 atoms from the G-quartet and two pairs of O6 and N7 atoms from G7 and G19, respectively (Figure S12B). G7·C15 and G19·C3 are separated and displaced in opposite directions to the helical axis to provide enough space for the K^+ ion and form a slipped major groove K^+-mediated G·C·G·C quartet.[12] These observations align with previous reports demonstrating that forming a direct G·C·G·C quartet favors the coordination of Na^+ ion,[13] while a larger atomic radius of K^+ ion demands a more open conformation in the form of a slipped G·C·G·C quartet.[14] The coordination of K^+ ion in BS1 is thus responsible for the slipped arrangement of the G7·C15·G19·C3 quartet, leading to a more open structure above the G-quartet. This results in the characteristic heart-like shape of KNaQ and noticeably influences the width of the narrow grooves in the Ce21 structure (see above).

Trajectories of the NaK MD simulation showed mechanistic details of the ion exchange process underlying the transition from NaK to the KNa state. The transition was characterized by two main subsequent events (Figure 4B, Movie S1): I) concurrent rearrangement of a slipped G7·C15·G19·C3 quartet above the G-quartet in the presence of Na^+ at BS1 into a direct G·C·G·C quartet and transient opening of the G1·C21·G13·C9 quartet below the

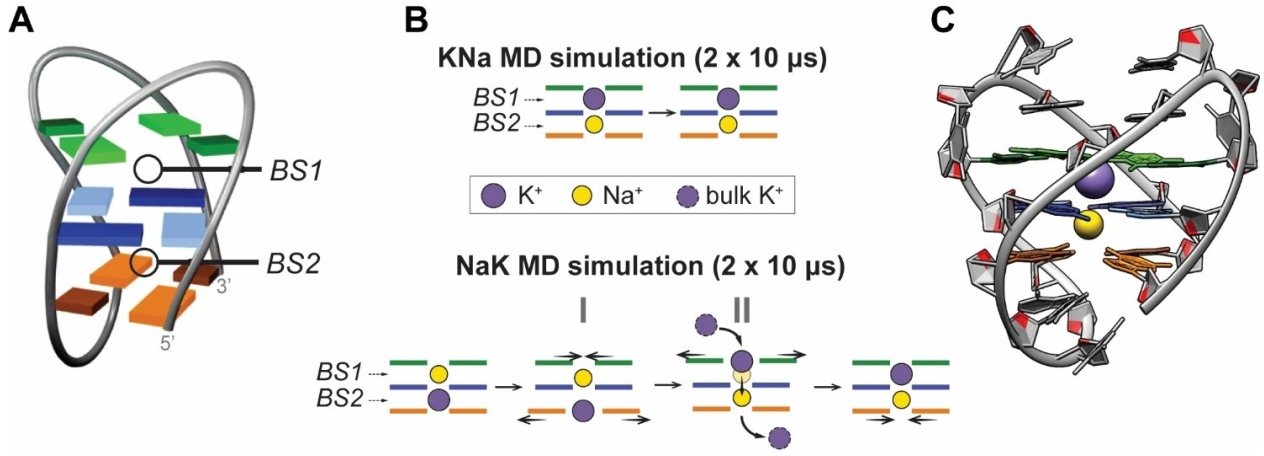

Figure 4. (A) Schematic representation of Ce21 with highlighted putative cation binding sites, BS1 and BS2. (B) Localizations of K^+ and Na^+ ions within the Ce21 core during MD simulations. (C) Representative structure of Ce21 with coordinated K^+ and Na^+ ions from MD simulations.

G-quartet, which enabled the relocalization of K^+ from BS2 toward a bulk solution; this step occurred shortly after the start of the simulation (at 5 ns in one simulation and 17 ns in the other); and II) a complex ion exchange step initiated by the approach of K^+ from the bulk solution above the newly formed direct G·C·G·C quartet near BS1 and its gradual rearrangement into slipped conformation (at 190 ns and 1829 ns, respectively); this was followed by movement of K^+ into BS1 and concurrent progression of Na^+ toward BS2 coupled by the immediate displacement of K^+ from BS2 to bulk solvent (at 205 ns and 1855 ns, respectively). In contrast to NaK simulations, we observed no ion exchange between either BS1 or BS2 and the bulk solution along KNa MD trajectories. Overall, the MD simulations identified Ce21 in the KNa configuration as a stable arrangement under complex ionic conditions in the simultaneous presence of K^+ and Na^+ (Figure 4C).

Positions of ions, substantiated with MD simulations, align with the results of NMR titration experiments (Figure 2C,D). Changes in spectral fingerprint and intensities of imino signals of guanines that delineate both binding sites support the coordination of K^+ and Na^+ ions into BS1 and BS2, respectively (for a detailed description, see Figure S13).

Small molecular weights ligands can discriminate KNaQ from G4 fold

In addition to differences in groove widths, the major structural features that discriminate KNaQ from closely related telomeric G4 structures are the following: I) The KNaQ core contains only a single G-quartet, while G4s contain, by definition, at least two stacked G-quartets, and II) simultaneous coordination of K^+ and Na^+ is required to form a stable KNaQ structure, while a single cation type stabilizes the known intramolecular telomeric G4 structures. Although mixed di-ionic K^+-NH_4^+ and K^+-Na^+ G4 species were previously observed, they corresponded to short-lived intermediate states in the exchange process from pure Na^+/NH_4^+ to K^+ form.[15] To the best of our knowledge, KNaQ is the first native tetrastranded intramolecular DNA structure that requires two different monovalent cations for its formation and stabilization. At the same time, it also represents the first instance where direct and slipped major groove G·C·G·C quartets coexist within the same structure, which is a lineal consequence of the coordination of two different types of ions within distinct binding sites.

On the other hand, similar to G4s, the KNaQ is tetrastranded, and its features expose the aromatic surface(s), which, as we hypothesize, are an underlying factor for its reactivity toward G4-specific ligands.[16] Indeed, the titrations with prototypic G4 ligands, namely, Braco19 and $PhenDC_3$ (Figures S14 and S15), confirmed their interaction with KNaQ, yet *via* distinct mechanisms. Adding Braco19 to Ce21 induced changes in both NMR and CD spectra (Figures 5A,B, and S15A). While the Braco19-induced negative ellipticity at 260 nm in CD spectra could indicate a ligand-induced refolding into an antiparallel G4, the minimal variations in the CD signals band intensities at 240 and 280 nm were suggestive of direct interaction between Ce21 and the ligand (Figure 5B). The preservation of the KNaQ fingerprint in the NMR spectra upon the addition of Braco19 corroborated the direct interaction of this ligand with Ce21 (Figures 5A and S15A).

On the contrary, the changes in the CD/NMR spectra upon the addition of $PhenDC_3$ showed ligand-induced refolding to an antiparallel G4 upon the formation of a ligand-G4 complex (Figures 5A, B, and S15B). The absence of a KNaQ fingerprint supported the refolding of Ce21 KNaQ into G4. The signals of imino protons of guanines constituting both G·C·G·C quartets in Ce21 (namely, G1, G7, G9, and G13) moved upfield upon the addition of $PhenDC_3$ to a region between δ 10.5 and 12.0 ppm, which is typical for Hoogsteen-base paired guanines and signature of G-quartet formation (Figures 5A and S15B).[9] These observations were consistent with a previous study demonstrating that $PhenDC_3$ could induce refolding of a hybrid G4 occupied by human telomeric DNA to an antiparallel G4 *via* a conformational selection mechanism.[17] Comparison of the imino regions of the 1D ^1H NMR spectra of Ce21:$PhenDC_3$ 1:2 complexes in ICB and 100 mM KCl showed matching signal fingerprints, which indicates that the binding of $PhenDC_3$ to KNaQ in ICB and the G4 in KCl leads to the formation of the same complex (Figure S15C). Next to the unique counterion dependency of KNaQ formation, $PhenDC_3$-induced refolding of the Ce21 structure to antiparallel G4 can be exploited for the design of molecular switches/sensors.

Conversely, the G4-selective porphyrin derivative NMM showed no interaction with Ce21 (Figures 5A, B, and S14). While the NMM autofluorescence significantly increases upon binding to validated G4-based targets regardless of their topology, it remains at the basal level in the presence of non-G4 structures, including KNaQ (Figure 5C). Similar behavior was observed for another broadly used G4-selective turn-on autofluorescence probe, thioflavin T (Figure S16).[16,18] These observations lay the basis for using NMM and thioflavin T as turn-on fluorescent probes to discriminate between KNaQ and G4 structures.

There are several potential, non-exclusive reasons for the inability of NMM to bind KNaQ: (*i*) restricted accessibility of the terminal G·C·G·C quartets, (*ii*) non-planarity of these quartets, and (more broadly) (*iii*) suboptimal geometry to mediate effective stacking interaction with this heteroaromatic planar ligand. Upon examining the 3D KNaQ structure, it becomes evident that accessibility of the ligand is substantially restricted by the loop residues on the side of the slipped G7·C15·G19·C3 quartet (Figure 3C). Additionally, the inability of NMM to interact with this quartet may be explained by its distinctive slipped arrangement, which provides a very different geometry compared to a classical G-quartet. Although the direct G1·C21·G13·C9 quartet on the other side of the molecule is more accessible, in order to provide proper coordination of Na^+ ion, G1·C21, and G13·C9 base pairs must tilt (on average for approximately 14°) with respect to the ideal planarity (Figures 3A and S17). The effective stacking of NMM to the

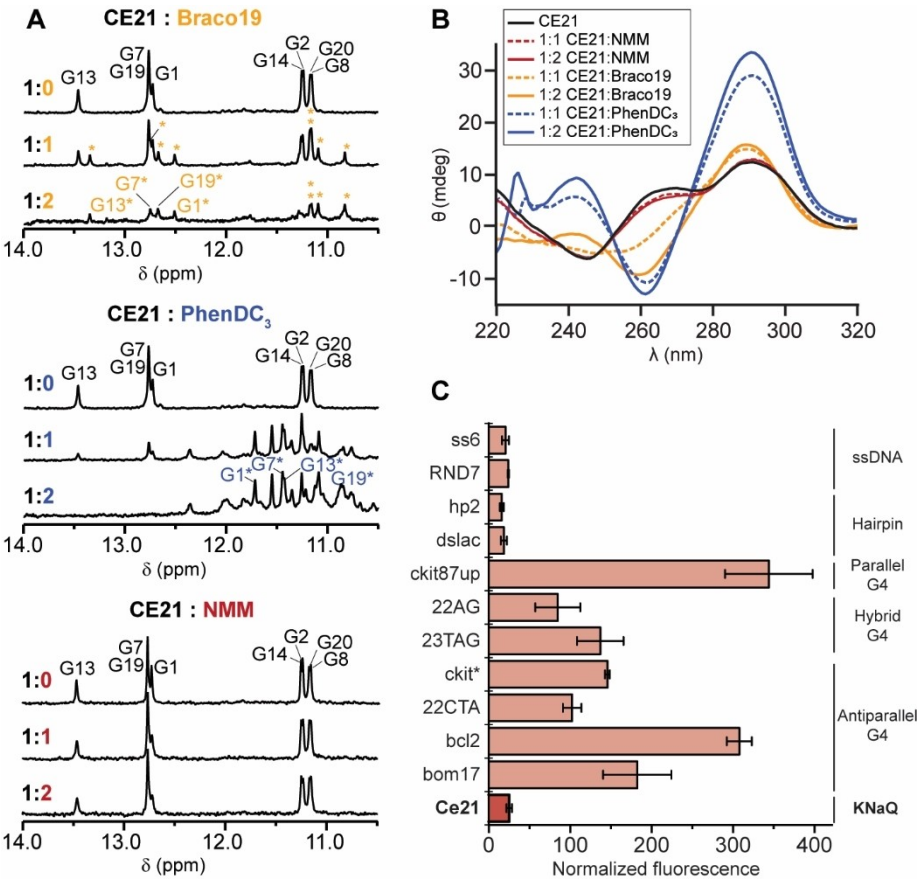

Figure 5. Interaction of Ce21 KNaQ with ligands. (A) 1D ^{1}H NMR and (B) CD titration experiments of Ce21 with Braco19 (top), PhenDC$_3$ (middle), and NMM (bottom) indicated direct interaction, ligand-induced refolding, and no interaction, respectively. Asterisks in (A) mark new signals appearing after ligand addition (for more details, see Figure S15). (C) Comparison of NMM autofluorescence in the presence of Ce21 in ICB, different G4 topologies, and non-G4 controls (*i.e.*, hairpins and single-stranded (ss) DNA), presented as mean ± s.d. of two technical replicates. Sequences of oligonucleotides in (C) are reported in Table S2.

direct G1·C21·G13·C9 quartet is thus most probably impeded by its non-planarity.

On the other hand, the lack of fluorescence enhancement of ThT in the presence of KNaQ cannot definitively rule out their interaction. The observed absence of fluorescence enhancement suggests either that ThT does not bind to KNaQ or that its binding mode to KNaQ does not efficiently restrict the rotation of the ThT benzothiazole ring relative to the aminobenzene ring in the excited state. This restriction is a fundamental prerequisite for inducing fluorescence enhancement through binding.[19]

Internal and terminal modifications of Ce21 and potential use of KNaQ as an antiparasitic drug target

KNaQ structure formation is not limited to the parent Ce21 sequence. As indicated by the characteristic pattern of NMR spectra, modifications at internal sites (*e.g.*, loop extensions, site-specific substitutions in loops, and fluorophore attachment at position T11), as well as modifications at both ends (*e.g.*, 5′- and 3′-end sequence extensions and 3′-fluorophore attachment) of the Ce21 sequence were well tolerated (Figure S18). Spectral fingerprint typical for KNaQ fold was observed in ten different loop variations (*i.e.*, substitutions of individual bases, elongations of loops, or both). The introduction of shorter two-nucleotide TT loops led to the formation of G4 without observation of the KNaQ fingerprint.

Notably, regardless of the numerous sequence variants supporting KNaQ formation, bioinformatics analysis showed that KNaQ-prone sequences in the human genome are rare, with only five occurrences found (Table S3). In contrast, such sequences were found to be abundant at internal and telomeric sites of nonparasitic (*C. elegans*) and parasitic helminths, such as soil-transmitted *Ascaris lumbricoides* and *Trichuris trichiura*, which together are responsible for infecting billions of people and livestock worldwide (Table S3).[20] The absence/presence of KNaQ motifs in the host/parasite opens an intriguing possibility of exploiting the KNaQ fold as a plausible antiparasitic drug target.

Conclusion

The G4-forming potential has been considered an evolutionarily conserved structural hallmark of telomeric DNA. In this work, we structurally characterized one of the postulated exceptions to the *G4 rule* and found that Ce21, a G-rich sequence derived from telomeric repeats of *C. elegans*, folds into unprecedented tetrastranded structure in the presence of both K^+ and Na^+ ions. The structure is characterized by a single G-quartet sandwiched between a slipped G·C·G·C quartet on one side and a direct G·C·G·C quartet on the other side. The presence of different arrangements of G·C·G·C quartets on each side of the G-quartet allows the binding of K^+ and Na^+ in two structurally distinct sites. Importantly, the NMR spectral signatures characteristic of KNaQ, unlike those of G4, were observed in extended *C. elegans* telomeric constructs suggesting the physiological relevance of the KNaQ fold.[5d]

The KNaQ fold is not correctly recognized by the current generation of bioinformatics tools for the prediction of DNA/RNA secondary structures, including those optimized for G-rich sequences.[21] Along with recently reported non-G4 structures folding from sequences with predicted strong G4-forming potential or, conversely, G-rich sequences with very low G4-forming potential forming unusual G4 structures,[6,22] KNaQ exposes a deficiency in current prediction models and shows that the structural landscape of G-rich DNA sequences is potentially much more complex than previously understood.

The KNaQ structure offers a fresh perspective on G-rich DNA folding under ionic conditions relevant to eukaryotic cell physiology. Its existence draws attention to common malpractice used in nucleic acid structural biology, using simplistic buffers based on a single counterion type (in most cases, K^+-based buffers) for structural studies. Although this practice is applied to limit conformational polymorphism, it can cause inaccuracies in identifying physiologically relevant structures by disregarding the fact that biologically active DNA/RNA motifs have evolved and function in a complex ionic environment.

Additionally, this structure provides insight into telomeric DNA structural evolution. From an interspecies perspective, it supports the hypothesis that the different telomeric DNA secondary structures, KNaQ in *C. elegans* and G4 in *Homo sapiens*, might mediate the same biological function. Analogously, at the intraspecies level, the observed coexistence of different telomeric G4 folding topologies in telomeric DNA, *e.g.*, hybrid-2 and two-quartet antiparallel in humans[3a,23] or parallel and antiparallel in *S. cerevisiae*,[24] might be explained by neutral evolution (functional equivalency of distinct folding topologies).[25]

Finally, our data indicate that the differences between the structural organization of human and nematode telomeric DNA need to be considered when using *C. elegans* as a model in telomere biology, particularly in drug screening applications.

Supporting Information

The authors have cited additional references within the Supporting Information.[26–42]

Acknowledgements

This work was supported by the Czech Science Foundation [19–26041X to L.T. and M.G., 21–23718S to J.S. and P.S.] and Czech Ministry of Education, Youth and Sports [MSCAfellow2@MUNI, grant no. CZ.02.2.69/0.0/0.0/18.070/0009846 to M.L.Z., SYMBIT, grant no. CZ.02.1.01/0.0/0.0/15 003/0000477 to J.L.M., J.S. and V.B.]. The authors acknowledge projects enabling access to research infrastructure: We acknowledge Josef Dadok NMR Centre of CIISB, Instruct-CZ Centre, supported by MEYS CR (LM2023042) and European Regional Development Fund-Project "UP CIISB" (No. CZ.02.1.01/0.0/0.0/18 046/0015974), and the CERIC-ERIC Consortium [project number 20192131].

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Atomic coordinates of the double-ion dependent DNA (KNaQ) structure of Ce21 in ICB have been deposited in the PDB under accession code 8BZU. The list of chemical shifts has been deposited in the BMRB under accession code 34780. MD trajectories can be obtained from the corresponding authors upon a reasonable request due to their large size. The bioinformatics analysis was conducted using in-house written Python code (Supporting Information, Script S1) and publicly available software.

Keywords: DNA · quadruplex · unique cation dependency · NMR spectroscopy · telomere

[1] R. J. Wellinger, D. Sen, *Eur. J. Cancer* **1997**, *33*, 735–749.
[2] V. Brázda, M. Bartas, R. P. Bowater, *Trends Genet.* **2021**, *37*, 730–744.
[3] a) T. M. Bryan, *Molecules* **2020**, *25*, 3686; b) S. Asamitsu, N. Shioda, H. Sugiyama, in *Annual Reports in Medicinal Chemistry*, Elsevier, Amsterdam, **2020**, pp. 77–99; c) N. Kosiol, S. Juranek, P. Brossart, A. Heine, K. Paeschke, *Mol. Cancer* **2021**, *20*, 40.
[4] a) Y. Wang, D. J. Patel, *Structure* **1993**, *1*, 263–282; b) G. N. Parkinson, M. P. H. Lee, S. Neidle, *Nature* **2002**, *417*, 876–880; c) J. Dai, C. Punchihewa, A. Ambrus, D. Chen, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2007**, *35*, 2440–2450; d) J. Dai, M. Carver, C. Punchihewa, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2007**, *35*, 4927–4940; e) A. T. Phan, V. Kuryavyi, K. N. Luu, D. J. Patel, *Nucleic Acids Res.* **2007**, *35*, 6517–6525; f) K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J.

Patel, K. N. Luu, A. T. Phan, *J. Am. Chem. Soc.* **2009**, *131*, 4301–4309; g) B. Heddi, A. T. Phan, *J. Am. Chem. Soc.* **2011**, *133*, 9824–9833; h) K. W. Lim, V. C. M. Ng, N. Martín-Pintado, B. Heddi, A. T. Phan, *Nucleic Acids Res.* **2013**, *41*, 10556–10562; i) P. Galer, B. Wang, P. Šket, J. Plavec, *Angew. Chem. Int. Ed.* **2016**, *55*, 1993–1997; j) C. Lin, D. Yang, in *Telomeres and Telomerase* (Ed.: Z. Songyang), Springer New York, New York, **2017**, pp. 171–196; k) C. Liu, B. Zhou, Y. Geng, D. Y. Tam, R. Feng, H. Miao, N. Xu, X. Shi, Y. You, Y. Hong, B. Z. Tang, P. K. Lo, V. Kuryavyi, G. Zhu, *Chem. Sci.* **2019**, *10*, 218–226.

[5] a) E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco, E. H. Blackburn, *Cell* **1987**, *51*, 899–908; b) A. T. Phan, *FEBS J.* **2010**, *277*, 1107–1117; c) P. L. T. Tran, J.-L. Mergny, P. Alberti, *Nucleic Acids Res.* **2011**, *39*, 3282–3294; d) P. Školáková, S. Foldynová-Trantírková, K. Bednářová, R. Fiala, M. Vorlíčková, L. Trantírek, *Nucleic Acids Res.* **2015**, *43*, 4733–4745; e) W.-Q. Wu, M.-L. Zhang, C.-P. Song, *J. Biol. Chem.* **2020**, *295*, 5461–5469.

[6] M. Gajarský, M. L. Živković, P. Stadlbauer, B. Pagano, R. Fiala, J. Amato, L. Tomáška, J. Šponer, J. Plavec, L. Trantírek, *J. Am. Chem. Soc.* **2017**, *139*, 3591–3594.

[7] J. Marquevielle, A. De Rache, B. Vialet, E. Morvan, J.-L. Mergny, S. Amrane, *Nucleic Acids Res.* **2022**, *50*, 7134–7146.

[8] D. H. Lackner, J. Karlseder, *Worm* **2013**, *2*, e21073.

[9] M. Adrian, B. Heddi, A. T. Phan, *Methods* **2012**, *57*, 11–24.

[10] a) P. Schultze, R. F. Macaya, J. Feigon, *J. Mol. Biol.* **1994**, *235*, 1532–1547; b) K. W. Lim, P. Alberti, A. Guédin, L. Lacroix, J.-F. Riou, N. J. Royle, J.-L. Mergny, A. T. Phan, *Nucleic Acids Res.* **2009**, *37*, 6239–6248.

[11] a) R. Tippana, W. Xiao, S. Myong, *Nucleic Acids Res.* **2014**, *42*, 8106–8114; b) J. Jana, Y. M. Vianney, N. Schröder, K. Weisz, *Nucleic Acids Res.* **2022**, *50*, 7161–7175.

[12] a) N. Zhang, A. Gorin, A. Majumdar, A. Kettani, N. Chernichenko, E. Skripkin, D. J. Patel, *J. Mol. Biol.* **2001**, *312*, 1073–1088; b) N. Escaja, B. Mir, M. Garavís, C. González, *Molecules* **2022**, *27*, 5287.

[13] A. Kettani, S. Bouaziz, A. Gorin, H. Zhao, R. A. Jones, D. J. Patel, *J. Mol. Biol.* **1998**, *282*, 619–636.

[14] S. Bouaziz, A. Kettani, D. J. Patel, *J. Mol. Biol.* **1998**, *282*, 637–652.

[15] a) P. Šket, M. Črnugelj, J. Plavec, *Nucleic Acids Res.* **2005**, *33*, 3691–3697; b) Z.-F. Wang, M.-H. Li, S.-T. D. Hsu, T.-C. Chang, *Nucleic Acids Res.* **2014**, *42*, 4723–4733.

[16] V. Pirota, M. Stasi, A. Benassi, F. Doria, in *Annual Reports in Medicinal Chemistry* (Ed.: S. Neidle), Academic Press, New York, **2020**, pp. 163–196.

[17] A. Ghosh, M. Trajkovski, M.-P. Teulade-Fichou, V. Gabelica, J. Plavec, *Angew. Chem. Int. Ed.* **2022**, *61*, e202207384.

[18] A. Renaud de la Faverie, A. Guédin, A. Bedrat, L. A. Yatsunyk, J.-L. Mergny, *Nucleic Acids Res.* **2014**, *42*, e65.

[19] V. I. Stsiapura, A. A. Maskevich, V. A. Kuzmitsky, V. N. Uversky, I. M. Kuznetsova, K. K. Turoverov, *J. Phys. Chem. B* **2008**, *112*, 15893–15902.

[20] J. Bethony, S. Brooker, M. Albonico, S. M. Geiger, A. Loukas, D. Diemert, P. J. Hotez, *Lancet* **2006**, *367*, 1521–1532.

[21] V. Brázda, J. Kolomazník, J. Lýsek, M. Bartas, M. Fojta, J. Šťastný, J.-L. Mergny, *Bioinformatics* **2019**, *35*, 3493–3495.

[22] a) V. Kocman, J. Plavec, *Nat. Commun.* **2014**, *5*, 5831; b) V. Kocman, J. Plavec, *Nat. Commun.* **2017**, *8*, 15355; c) M. L. Živković, M. Gajarský, K. Beková, P. Stadlbauer, L. Vicherek, M. Petrová, R. Fiala, I. Rosenberg, J. Šponer, J. Plavec, L. Trantírek, *Nucleic Acids Res.* **2021**, *49*, 2317–2332; d) S. Roschdi, J. Yan, Y. Nomura, C. A. Escobar, R. J. Petersen, C. A. Bingman, M. Tonelli, R. Vivek, E. J. Montemayor, M. Wickens, S. G. Kennedy, S. E. Butcher, *Nat. Struct. Mol. Biol.* **2022**, *29*, 1113–1121.

[23] R. Hänsel, F. Löhr, L. Trantirek, V. Dötsch, *J. Am. Chem. Soc.* **2013**, *135*, 2816–2824.

[24] K. Jurikova, M. Gajarsky, M. Hajikazemi, J. Nosek, K. Prochazkova, K. Paeschke, L. Trantirek, L. Tomaska, *J. Biol. Chem.* **2020**, *295*, 8958–8971.

[25] M. W. Gray, J. Lukes, J. M. Archibald, P. J. Keeling, W. F. Doolittle, *Science* **2010**, *330*, 920–921.

[26] W. Lee, M. Tonelli, J. L. Markley, *Bioinformatics* **2015**, *31*, 1325–1327.

[27] R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka, T. E. I. Cheatham, *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127.

[28] D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, L. Wilson, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York, P. A. Kollman, AMBER 20, University of California, San Francisco, **2020**.

[29] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.

[30] M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo, P. Jurečka, *J. Chem. Theory Comput.* **2015**, *11*, 5723–5736.

[31] a) M. Zgarbová, F. J. Luque, J. Šponer, T. E. Cheatham, M. Otyepka, P. Jurečka, *J. Chem. Theory Comput.* **2013**, *9*, 2339–2354; b) A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, M. Orozco, *Biophys. J.* **2007**, *92*, 3817–3829.

[32] M. Krepl, M. Zgarbová, P. Stadlbauer, M. Otyepka, P. Banáš, J. Koča, T. E. Cheatham, P. Jurečka, J. Sponer, *J. Chem. Theory Comput.* **2012**, *8*, 2506–2520.

[33] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[34] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, *J. Phys. Chem.* **1987**, *91*, 6269–6271.

[35] I. S. Joung, T. E. Cheatham, *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

[36] D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, AMBER 12, University of California, San Francisco, **2012**.

[37] P. Stadlbauer, B. Islam, M. Otyepka, J. Chen, D. Monchaud, J. Zhou, J.-L. Mergny, J. Šponer, *J. Chem. Theory Comput.* **2021**, *17*, 1883–1899.

[38] a) R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, R. C. Walker, *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888; b) D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y.

Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York, P. A. Kollman, AMBER 18, University of California, San Francisco, **2018**.

[39] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.

[40] C. W. Hopkins, S. Le Grand, R. C. Walker, A. E. Roitberg, *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874.

[41] K. L. Howe, B. J. Bolt, M. Shafie, P. Kersey, M. Berriman, *Mol. Biochem. Parasitol.* **2017**, *215*, 2–10.

[42] W. J. Kent, *Genome Res.* **2002**, *12*, 656–664.

Manuscript received: September 6, 2023
Accepted manuscript online: December 24, 2023
Version of record online: January 15, 2024