Nucleic Acids Research, 2025, **53**, gkae1208
https://doi.org/10.1093/nar/gkae1208
Advance access publication date: 19 December 2024
Structural Biology

Interaction of N-methylmesoporphyrin IX with a hybrid left-/right-handed G-quadruplex motif from the promoter of the SLC2A1 gene

Paul Seth[1,†], Eric Xing[1,†], Andrew D. Hendrickson[1], Kevin Li[1], Robert Monsen[2], Jonathan B. Chaires[2], Stephen Neidle[3] and Liliya A. Yatsunyk[1,*]

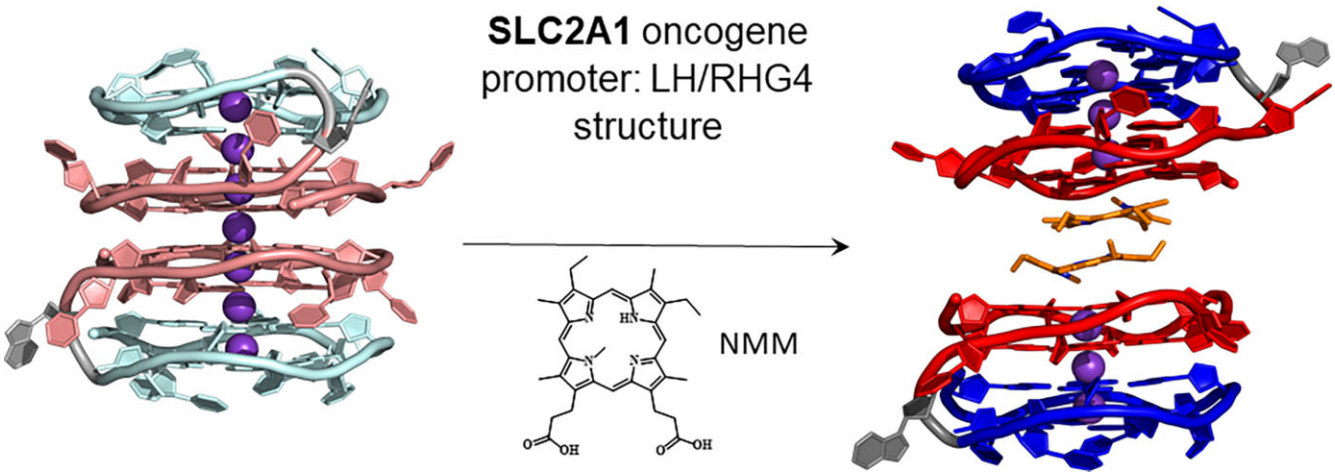
[1]Department of Chemistry and Biochemistry, Swarthmore College, 500 College Ave, Swarthmore, PA, 19081USA
[2]UofL Health, Brown Cancer Center, University of Louisville, 529 S. Jackson Street Louisville, KY, 40202USA
[3]School of Pharmacy, University College London, 29-39 Brunswick Square, London, WC1N 1AX, UK

[*]To whom correspondence should be addressed. Tel: +1 610 328 8558; Email: lyatsun1@swarthmore.edu
[†]The first two authors should be regarded as Joint First Authors.

Abstract

Left-handed G-quadruplexes (LHG4s) belong to a class of recently discovered noncanonical DNA structures under the larger umbrella of G-quadruplex DNAs (G4s). The biological relevance of these structures and their ability to be targeted with classical G4 ligands is underexplored. Here, we explore whether the putative LHG4 DNA sequence from the SLC2A1 oncogene promoter maintains its left-handed characteristics upon addition of nucleotides in the 5′- and 3′-direction from its genomic context. We also investigate whether this sequence interacts with a well-established G4 binder, N-methylmesoporphyrin IX (NMM). We employed biophysical and X-ray structural studies to address these questions. Our results indicate that the sequence d[G(TGG)$_3$TGA(TGG)$_4$] (termed here as SLC) adopts a two-subunit, four-tetrad hybrid left-/right-handed G4 (LH/RHG4) topology. Addition of 5′-G or 5′-GG abolishes the left-handed fold in one subunit, while the addition of 3′-C or 3′-CA maintains the original fold. X-ray crystal structure analyses show that SLC maintains the same hybrid LH/RHG4 fold in the solid state and that NMM stacks onto the right-handed subunit of SLC. NMM binds to SLC with a 1:1 stoichiometry and a moderate-to-tight binding constant of 15 μM^{-1}. This work deepens our understanding of LHG4 structures and their binding with traditional G4 ligands.

Graphical abstract



Introduction

G-quadruplex nucleic acids (G4s) are formed from the association of repeated short G-tracts of either DNA or RNA and are held together by G-quartets. An individual quartet comprises four in-plane guanine bases, held together by Hoogsteen hydrogen bonds. Several quartets stack on one another within a G4 and are held together by intervening loops and ion coordination (1–5). Putative G4-forming se-

Received: July 5, 2024. **Revised:** November 18, 2024. **Editorial Decision:** November 19, 2024. **Accepted:** November 21, 2024

quences are widespread in the human and other genomes, with occurrences in telomeric regions, promoters, 5′ and 3′-untranslated regions and translocations (5–8). The total number of G4 sequences in the human genome depends on assumptions about the length of loop regions (sequences linking individual G-tracts): bioinformatics estimates range from ca 350 000 to over 700 000 (9–11). The distribution of G4 sequences in the human genome is nonrandom, with significant over-representation in the promoter regions of many genes involved in proliferative processes and in human cancers (12). The number of G4s in active chromatin in cancer cells has been found experimentally to be ca 10 000 compared with ca 1500 in normal cells (13). These findings have led to the fruitful concept of targeting cancer gene promoters with selective small molecules to inhibit transcription of these genes (14–18).

The overwhelming majority of genomic intramolecular G4s characterized to date conforms to the general sequence type $G_{2\text{-}5}$ NL1 $G_{2\text{-}5}$ NL2 $G_{2\text{-}5}$ NL3 $G_{2\text{-}5}$, where the short G-tracts can vary in length between two and five Gs, and the NL loops, of general sequence, can have up to ca 26 nucleotides in each loop (9–11). G4 topology can vary, depending on loop size and sequence as well as G-tract size and the nature of the stabilizing cation, as has been shown in the variety of crystal and Nuclear Magnetic Resonance (NMR) G4 structures (4,19–27). All, however, have a right-handed (RH) fold. The sole exception is a small family of left-handed G4s (LHG4s) based on GGT repeats, which have been characterized by X-ray crystallography, NMR and biophysical methods (28–32). Specifically, two motifs, named *Motif1* and *Motif2*, with the 12-nt sequences d[G$(TGG)_3$TG] and d[$(GGT)_3$GTG], respectively, have been identified to fold into LHG4 structures (31). Both motifs contain three GG stretches and one split GG stretch. The presence of this split GG stretch enables *Motif1* and *Motif2* to fold into a LHG4. It appears from the data presented to date that small sequence changes can change the fold of such structures from left-handed (LH) to RH. The presence of a LHG4 sequence in a small number of genes has been previously noted (28), although not the occurrence presented here.

In the absence of any more generalized LHG4 sequences, we have searched the human genome for occurrences of potential LHG4s using *Motif1* or *Motif2*. It is apparent that such sequences are rare, especially compared with conventional RH ones. We report here on a LHG4 sequence within the SLC2A1 gene (Figure 1A). This gene encodes for the major glucose transporter protein GLUT-1, facilitating transport across the blood–brain barrier (33). It is overexpressed in many human cancers (34). Here we explore two questions, whether this putative LHG4 sequence adopts a stable LH fold within its genomic environment, and whether its LH character is stabilized by a conventional right-handed G4 (RHG4) ligand. To address these questions, we performed biophysical and structural characterizations of nine variants from the SLC2A1 gene either alone or in complex with a highly selective RHG4-binding ligand, N-methylmesoporphyrin IX (NMM; Figure 1B). NMM was chosen because of its preference for parallel G4 topology, as all reported LHG4s adopt a parallel G4 conformation (28–32). We determined the topology and stability of all variants, as well as the stoichiometry and binding constants for NMM binding. In addition, we solved crystal structures for one SLC2A1 variant alone and in complex with NMM.

Materials and methods

DNA, ligand and buffers

Lyophilized oligonucleotides were purchased from Integrated DNA Technologies (IDT; Coralville, IA) with standard desalting purification unless otherwise specified. DNA was hydrated in doubly distilled water to 1–2 mM and stored at −80°C. Extinction coefficients for all sequences were obtained using IDT's OligoAnalyzer 3.1 and DNA concentration was determined from UV-vis spectra collected at 95°C. DNA sequences, their molecular weights and extinction coefficients are listed in Supplementary Table S1. NMM (Thermo Fisher Scientific, former Frontier Scientific) stock was prepared in doubly distilled water and its concentration was determined via UV-vis scans using an extinction coefficient of 1.45×10^5 M^{-1} cm^{-1} at 379 nm (35). All biophysical experiments and crystallization trials were performed in a 100K buffer consisting of 10 mM lithium cacodylate pH 7.2 and 100 mM KCl. To induce G4 formation, DNA was diluted to a desired concentration with 100K buffer and heated at 90°C for 5 min, cooled slowly to room temperature over 4 h and equilibrated at 4°C overnight.

Circular dichroism scans, single point and full wavelength melting experiments

All circular dichroism (CD) experiments were conducted on a Jasco 1500 circular dichroism spectrophotometer equipped with a Koolance thermocontroller (±0.3°C error) in 1 cm quartz cuvettes. CD scans were collected at 20°C from 220 to 330 nm with 1 s averaging time, 2 nm bandwidth and 1 nm step. Three-to-ten scans were collected first for the buffer and then for each sample in the corresponding cuvettes. CD scans were averaged and the buffer contribution was removed. Data were zeroed using an average value for CD signal from 320 to 330 nm and were converted to the molar ellipticity $\Delta\varepsilon$ (deg cm^{-1} M^{-1}) using equation (1):

$$\Delta\varepsilon = \theta / (0.03298 \times c \times \ell), \quad (1)$$

where θ is the measured CD signal (mdeg), c is sample concentration (μM) and ℓ is cuvette thickness (cm).

CD melting experiments were conducted from 4 to 95°C with 1°C step, 1°C per min temperature rate, 8 s averaging time and 30 s equilibration time. The CD signal at 262 and 280 nm, the wavelengths corresponding to the RH and LH subunits, respectively, was monitored as a function of temperature. We also monitored 330 nm wavelength as a reference for instrument performance. The data at 262 and 280 nm were corrected using the CD signal at 330 nm. Melting temperatures, T_m, were determined either via derivative method or using a two-state approximation (36); the latter was applied to reversible systems with low hysteresis (<3°C). In the derivative method, melting data were typically smoothed, the first derivative was taken, and the value of T_m was determined visually. It is associated with 0.5°C error. Hysteresis was determined as the difference between T_m from the melting and cooling curves.

Multidimensional melting curves were obtained for the SLC variant using the strategy and analysis described by Gray and Chaires (36). Specifically, full wavelength CD scan melting experiments were conducted from 4 to 95°C with 1°C step, 1°C per min temperature rate, 30 s equilibration time and three scans for each temperature. The rest of the parameters were set as above. Each scan was processed as above, and the overall system was analyzed using the Singular Value Decompo-

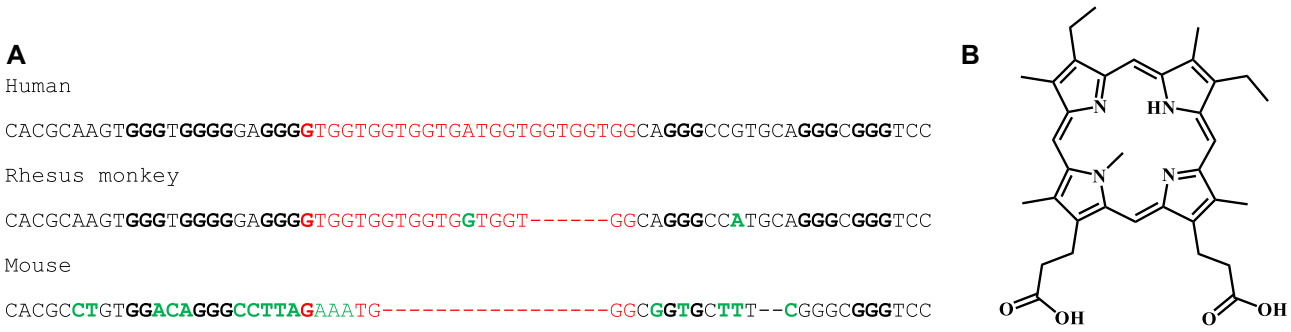


Figure 1. (**A**) Alignment of the G-rich sequence found in the SLC2A1 promoter with that in two other mammalian species. The LH/RHG4 sequence is shown in red, flanking G-tracts are in bold, and sequence changes from the human are colored green. The alignment was performed using the browser tools in the UCSC Genome Browser (https://genome.ucsc.edu/). Only the red section of the SLC2A1 gene was interrogated in this work. (**B**) Structure of NMM: 8,13-diethyl-3,7,12,17,23-pentamethyl-21H,23H-porphine-2,18-dipropanoic acid.

sition (SVD) method. Additional details are provided in the Supplementary Material section titled 'Singular Value Decomposition applied to the full wavelength CD melting'. The data were best fitted to a sequential melting model with two intermediates. Such analysis also yields thermodynamic parameters for the three melting transitions.

Native polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis (PAGE) samples contained 50 μM DNA in 100K buffer and were weighted down with 7% w/v sucrose prior to loading. Fifteen percent native polyacrylamide gels were made with 10 mM KCl and 1× Tris-Borate-EDTA. Gels were pre-migrated at 150 V for 30 min, loaded with 10 μl of each sample and allowed to run for 150 min at 150 V at room temperature. A tracking dye was used to monitor gel progress and an oligothymidylate ladder consisting of dT_{15}, dT_{24}, dT_{30} and dT_{57} or dT_{60} was used as a length marker. DNA bands were visualized using Stains-All and the resulting gel was captured using a conventional smartphone camera.

UV-vis spectroscopy

A Varian Cary 3500 UV-vis spectrophotometer equipped with a temperature controller was used to collect UV-vis data (±0.3°C error) from 220 to 330 nm for DNA and from 350 to 480 nm for NMM at 0.5–1 nm intervals with 0.1 s averaging time, 300 nm per min scan rate, and 2 nm spectral bandwidth.

UV-vis titration of SLC and 2R with NMM

NMM samples were prepared in 1 cm methyl methacrylate cuvettes to target an absorbance of ∼0.5 (3–4 μM NMM with a volume of 1000 μl). DNA stock solutions at ∼400 μM were prepared to target a final [DNA]/[NMM] ratio of at least 5. To maintain a constant NMM concentration in the cuvette throughout the titration, DNA was annealed at double the target concentration and diluted with a 2× NMM sample. This DNA sample was titrated into a cuvette with NMM, the mixture was equilibrated for 2 min after which the UV-vis spectrum was collected. The titration continued until no further changes were observed in the UV-vis spectrum after three additions of DNA. The volume of DNA added (50–70 μl total), λ_{max}, and absorbance at λ_{max} were monitored throughout the titration. Data were processed using SVD (37) followed by direct fit, as described in detail in the SI section titled 'Singular Value Decomposition and analysis of binding isotherms'. Reported K_a values represent the average of three trials with associated errors.

Thermal difference spectra

Thermal difference spectra (TDS) (38) data were obtained by subtracting UV-vis scans taken at 20°C after 5 min of equilibration from scans taken at 90°C after 10 min of equilibration. The low and high temperature limits are determined by the temperatures at which the DNA is (mostly) folded and unfolded.

X-ray crystallography

Ninety-six well crystallization trays were set by a SPT Labtech Mosquito robot (equipped with a humidity chamber) using the hanging drop vapor diffusion method. We utilized two commercial screens, Helix (Molecular Dimensions) (39) and Natrix (Hampton Research). Each drop contained 0.2 μl of SLC or SLC–NMM complex and 0.2 μl of the screen condition. The wells contained 100 μl of the screen conditions.

Crystallization of native SLC

The SLC sample for crystallization was prepared at 1 mM using High-Performance Liquid Chromatography (HPLC)-purified DNA in a 100K buffer. Initial were obtained in a condition containing 0.08 M NaCl, 0.4 M sodium cacodylate pH 6 and 45% 2-methyl-2,4-pentanediol (MPD). Optimization did not yield higher quality crystals. Therefore, the initial condition was used to obtain diffraction quality crystals. These crystals were flash frozen in liquid nitrogen without additional cryoprotection.

Crystallization of SLC–NMM complex

The SLC–NMM complex for crystallization was prepared by annealing 0.5 mM HPLC-purified SLC with NMM at a 1:1 ratio in 100K buffer. The HPLC purification of the DNA was crucial as the SLC–NMM crystals obtained from desalted DNA were of significantly poorer quality (resolution of 2.2 versus 1.45 Å). Initial hits were obtained in Natrix 2–19 and 2–32. Both conditions contained 0.04 M sodium cacodylate pH 6, 0.08 M NaCl and 12 mM spermine but differed in MPD content (45 and 30%, respectively). Optimization of these conditions was performed manually in 24-well plates using 1 μl of SLC–NMM complex and 1 μl of the crystallization condition. The final optimized condition contained 0.04 M sodium cacodylate pH 6, 0.08 M NaCl, 2 mM spermine

Table 1. Crystallographic statistics

	Native SLC	SLC–NMM
Resolution range, Å	30.523–2.388	35.39–1.45
Highest resolution shell, Å	2.429–2.388	1.56–1.45
Space group	*C*2	$P4_12_12$
Unit cell dimensions		
a, b, c (Å)	64.1, 31.5, 31.5	30.90, 30.90, 141.54
α, β, γ (°)	90, 104.3, 90	90, 90, 90
Unique reflections	2469	13 141
Redundancy	3.3(3.4)	12.2(12.4)
Completeness (%)	98.7(100)	100(100)
I/sigma	8.4(1.1)	21(1.2)
R-merge	0.103(1.383)	0.049(1.789)
R_{work}/R_{free} (%)	23.34/25.43	17.83/22.68
Molecular entities in ASU	1	1
Number of atoms	544	653
DNA	530	530
NMM		36 (43)[a]
Solvent	1	62
Potassium ion	3.5	3
Spermine		14
Sodium ion	1	
MPD	8	8
Overall B-factor for ASU (Å^2)	68.4	42.3
RMS deviations		
Bond length (Å)	0.004	0.006
Bond angles (°)	0.668	1.009
PDB ID	**9C46**	**8TAA**

[a]Note that the NMM molecule comprises 43 nonhydrogen atoms, although only 36 are visible in the electron density.

and 45% MPD. Crystals were flash frozen in liquid nitrogen without additional cryoprotection.

Data collection and structure solution

Native SLC: Six data sets were collected on three crystals at the National Synchrotron Light Source II 17-ID-1 (AMX) synchrotron facility. Raw diffraction data was processed using autoPROC (40). The structure was solved in the *C*2 space group via molecular replacement (MR) using Phenix.Phaser (41) and the SLC–NMM structure (PDB id 8TAA) as a model. NMM and solvents were removed from the model before MR but potassium ions were kept. Extensive model building cycles were performed in Coot 0.8.8 (42) followed by Phenix.Refine. The asymmetric unit (ASU) contains a single DNA chain, one MPD and one water molecule.

SLC–NMM complex: Six data sets on two crystals were collected at the Advanced Photon Source 24 ID-C synchrotron facility. Raw diffraction data were processed using XDS (43). The structure was solved in the $P4_12_12$ space group via MR using Phenix.Phaser (41) and a G4 model containing left-/right-handed G4 (LH/RHG4; PDB id 6QJO). Extensive manual model building cycles were performed in Coot 0.8.8 (42) followed by Phenix.Refine. The ASU contains a single DNA chain, one NMM molecule, one spermine, one MPD and 62 water molecules. T20-G21 and T23-G24 were split to accommodate different backbone conformations. The major and minor conformations of the T20-G21 fragment were refined at 0.64 and 0.36 occupancy, respectively. The major and minor conformations of the T23-G24 fragment were refined at 0.72 and 0.28 occupancy, respectively. There was clear electron density for the porphyrin ring and N-methyl group of NMM. Peripheral substituents were modeled as five methyl groups and three ethyl groups due to the poor electron density; electron density for the propionate groups was not visible, presumably due to their high mobility. Data collection and refinement statistics for both structures are presented in Table 1.

Structural analysis

Structural analysis was performed using 3DNA and ASC-G4 (44) as described in our previous work (45). Water–water and water–DNA hydrogen-bond contacts for SLC–NMM complex were analyzed and visualized using Pymol (https://pymol.org/) with a cutoff for a maximum hydrogen-bond distance of 3.2 Å.

Results and discussion

We aimed to determine whether, for a specific gene within the human genome, *Motif1* with the sequence d[G(TGG)$_3$TG] and *Motif2* with the sequence d[(GGT)$_3$GTG] fold into putative LHG4s. We performed a BLAST search on the human genome (Ensembl v. GRCh37 at https://ensemblgenomes.org/) and found at least 10 different matches across eight chromosomes. One of those sequences is within the SLC2A1 gene, located in NG_008232.1 : 1329–1353 (reverse complement). This gene encodes for GLUT-1, the glucose carrier protein, which is highly overexpressed in many human cancers (33,34). The LHG4 sequence is 247–258 nucleotides upstream of the start of the SLC2A1 gene (5309), and is thus within the promoter region. We have examined the conservation of this and the surrounding sequence in various eukaryotic species (Figure 1A). The full sequence only occurs in humans; the LHG4 sequence also occurs in rhesus monkeys (we have not examined genomes of other primates). This sequence is missing in the mouse genome, as it is in dogs. The complete sequence is missing in 20 other lower mammals examined in this study (data not shown). It is notable that both 5′ and 3′ flanking

Table 2. SLC2A1 variants and their thermodynamic stability determined via CD melting. LHG4 *Motif1* is colored in blue, 5′- and 3′-overhangs relative to the first sequence, SLC, are in red. The adenine linker is underlined. The G4 conformation was determined based on CD signatures and PAGE

Name	Sequence 5' → 3'	G4 fold	T_m (°C)	Hysteresis (°C)
SLC	GTGGTGGTGGTGATGGTGGTGGTGG	LH/RH	52 ± 11 60.6 ± 0.9 75.8 ± 0.3[a]	N/A
1R	GTGGTGGTGGTGATGGTGGTGGTGGC	LH/RH	57.8 ± 0.9 71.0 ± 0.8[b]	N/A
2R	GTGGTGGTGGTGATGGTGGTGGTGGCA	LH/RH	56.1 ± 0.6[c]	3.6 ± 0.9[c]
1L	GGTGGTGGTGGTGATGGTGGTGGTGG	RH	62.7 ± 0.3	1.4 ± 0.7
2L	GGGTGGTGGTGGTGATGGTGGTGGTGG	RH	60.6 ± 0.4	0.4 ± 0.6
1LR	GGTGGTGGTGGTGATGGTGGTGGTGGC	RH	59.5 ± 0.3	2.4 ± 0.6
2LR	GGGTGGTGGTGGTGATGGTGGTGGTGGCA	RH	57.3 ± 0.3	0.6 ± 0.5
1L2R	GGTGGTGGTGGTGATGGTGGTGGTGGCA	RH	59.6 ± 0.3	1.5 ± 0.9
2L1R	GGGTGGTGGTGGTGATGGTGGTGGTGGC	RH	57.5 ± 0.3	0 ± 0.3

a) The raw melting data displayed a non-sigmoidal melting curve with multiple transitions; T_m values are listed for each transition determined via SVD analysis of full wavelength CD melts
b) The raw melting data displayed a non-sigmoidal melting curve with multiple transitions; T_m values were obtained using derivative method
c) Value determined using 280 nm wavelength; the T_m value determined from the 262 nm wavelength is 58.6 ± 0.7 °C with 3.0 ± 1.0 °C hysteresis

sequences in the human and primate sequences contain three additional G-tracts each. There are no further G-tracts or any other obvious higher order structure forming sequences in their vicinity.

Design of the SLC2A1 variants

The identified SLC2A1 sequence consists of the LHG4 *Motif1* connected to a RHG4 motif with the sequence d[$(TGG)_4$] by an adenine linker (Table 2). Thus, we expect this DNA to adopt a LH/RHG4 fold. We designed nine native variants by extending the sequence in the 3′ direction by one or two nucleotides (using 1R or 2R in the name for 'Right') and in the 5′ directions (using 1L or 2L in the name for 'Left') according to its genomic context. These variants allow us to investigate how the 5′ and 3′ extensions affect G4 folding and, specifically, whether they preserve the LHG4 topology in one subunit.

Biophysical characterization of SLC2A1 variants

We used TDS to verify G4 formation, CD to determine G4 topology and stability, and PAGE to assess homogeneity of SLC variants (Figure 2). The TDS signatures display a characteristic trough at 297 nm, suggesting that all constructs fold into G4s (Figure 2A). The CD signatures of all variants fall into two groups, marked by red and green lines on Figure 2B. SLC, 1R and 2R (two variants with 3′-C or 3′-CA overhangs, respectively) display a CD signature with a peak at 255–260 nm and a prominent trough at 279 nm, characteristic of a hybrid LH/RHG4 topology (28). This topology was confirmed in our X-ray study (see below). PAGE for SLC, 1R and 2R display one prominent band with some smear or minor bands (Figure 2C). The second group of DNA structures displays a CD signal with a peak at 263 nm and a trough at about 240 nm, characteristic of a parallel RHG4 (46). This group contains every variant with one or two guanines at the 5′ end (1L, 2L, 1LR, 2LR, 1L2R and 2L1R) regardless of the 3′ additions. On PAGE, members of this group run as a smear without well-defined bands suggesting that they may possibly adopt several similar G4 conformations. Despite this heterogeneity, the six RHG4 variants, as well as 2R, display relatively simple single-wavelength thermal denaturation transitions (Supplementary Figure S1) with apparent melting temperatures from 56.9 to 62.5°C (Figure 2D and Table 2). Thus, addition of flanking nucleotides did not affect SLC's stability to a great extent. We also extracted the enthalpies (Supplementary Table S1), but, in view of the observed heterogeneity of the samples (Figure 2C), those lack true thermodynamic significance. We report them here because such data provide a phenomenological characterization of the steepness and widths of the melting transitions.

The observed thermal stabilities of our four tetrad, 'two-subunit' structures of the RHG4 SLC2A1 variants differ from those of four tetrad, 'one-subunit' RHG4s formed by telomeric DNA from *Tetrahymena thermophila* (PDB id 6W9P, 7JKU, 7LL0 and PDB 6XT7). These four tetrad, 'one-subunit' variants show a half transition at a higher temperature of 74–80°C in 10 mM K^+ buffer (47). It is well known that the stability of G4 structures increases with increasing K^+ concentration (48), thus we would expect yet higher stability of these variants in the 100K buffer used in this work. AT11-B0, a four tetrad, 'two-subunit' RHG4 structure (PDB id 2N3M for a variant with a TT bulge), displayed a T_m of 59°C in buffer with 90 mM K^+ (49), similar to the T_m values for RHG4 SLC2A1 variants. Thus, RHG4 SLC2A1 variants may possibly adopt a similar four-tetrad 'two-subunit' structure.

Stability of the SLC variant determined using full wavelength CD melting

The SLC and 1R variants displayed nonsigmoidal melting curves with multiple transitions, indicating the presence of multiple species during the unfolding process (Supplementary Figure S1). We further investigated the complexity of the SLC melting profile through multidimensional full-wavelength CD melting experiments (Figure 3, and Supplementary Figure S2–S3) (36). The data were analyzed via SVD and four significant species were identified (Supplementary Figure S2A–C). Five models were explored (Supplementary Figure S2D) and the three transitions model (i.e. Folded → Intermediate1 (I1) → Intermediate2 (I2) → Unfolded) yielded the best fits based on Chi-squared and the Akaike information criterion val-

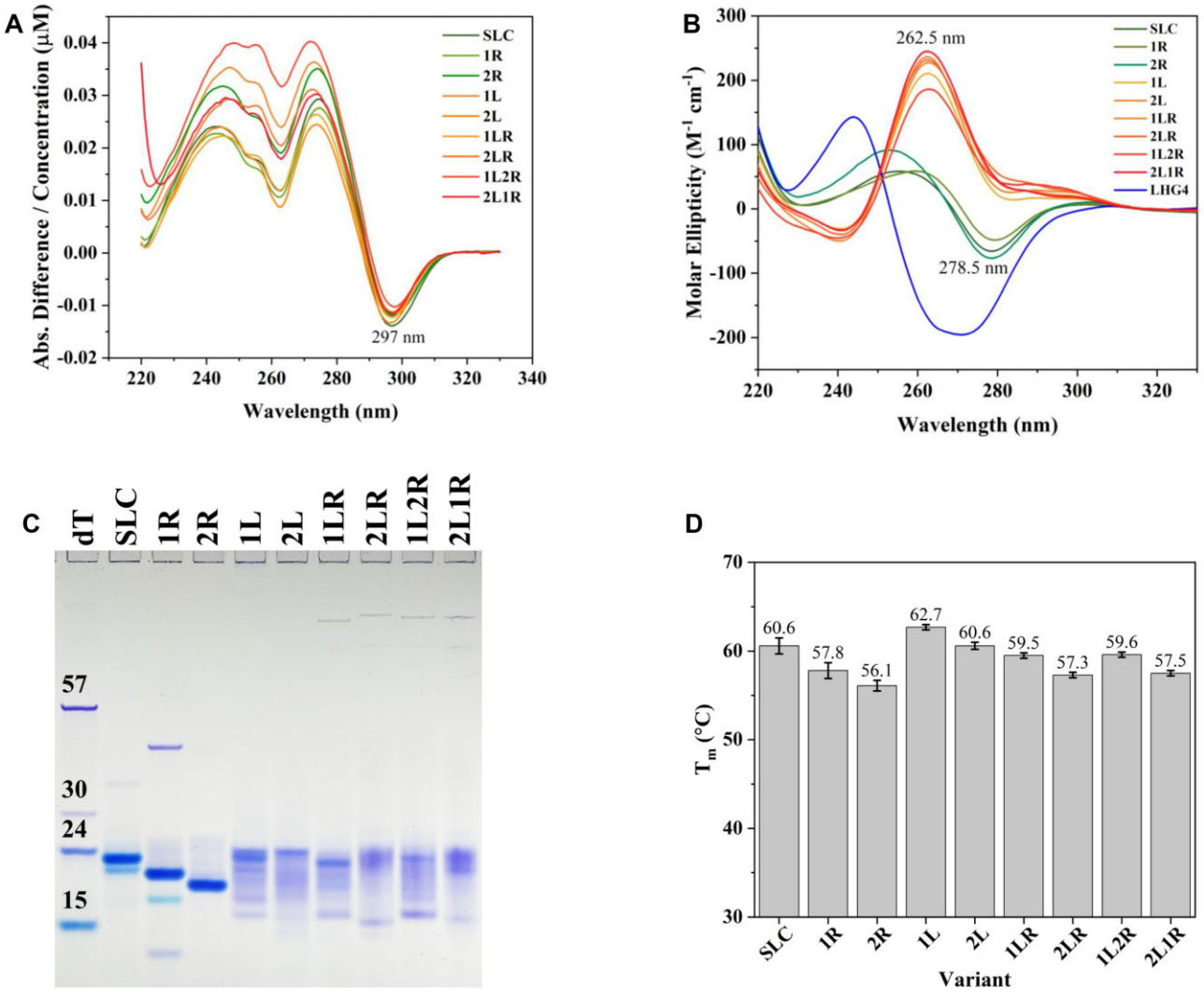


Figure 2. Biophysical characterization of SLC variants. (**A**) TDS. (**B**) CD spectra. LH/RHG4s are shown in shades of green, and RHG4s are shown in shades of red-orange. A representative spectrum of a previously characterized LHG4 (PDB id 4U5M) is shown in blue as a reference. (**C**) Fifteen percent native PAGE gel. (**D**) Melting temperatures of SLC2A1 variants were determined from CD melting data using a two-state model. For SLC, T_m was obtained via analysis of the full wavelength CD melting. For 1R, the melting temperature was obtained using the derivative method. Only the values for the dominant transition are shown for SLC and 1R. Error bars represent standard deviation across 2–3 trials. DNA was annealed at ∼50 μM for PAGE and ∼4 μM for other experiments in 100K buffer.

ues (50). Analysis of the data yielded melting parameters, species distributions, and basis CD spectra for the four significant components (Figure 3C and D for one data set and Supplementary Figure S3 for another) (51). Note that, due to the nature of SVD analysis, some of the basis spectra may not have an exact physical meaning or reflect the actual spectra of intermediates. The basis CD spectra of the folded and unfolded species matched the initial (4°C) and final (95°C) spectra in the raw data, corresponding to the LH/RHG4 and unstructured DNA, respectively (Figure 3D). I1 had an unidentifiable CD spectrum, with a peak at ∼240 nm and a broad signal. Interestingly, the I2 intermediate has a CD spectrum similar in shape to that of a LHG4, with a peak at ∼245 nm and a broad trough at ∼270 nm. Thus, I2 would be consistent with a species with an unfolded RH subunit but intact LH subunit. Notably, this theoretical spectrum is identifiable in the raw data at ∼70°C. The melting temperatures for each unfolding step are 52 ± 11°C, 60.6 ± 0.9°C, and 75.8 ± 0.5°C, respectively (Supplementary Figure S4C). The first transition from the folded species to I1 is poorly defined (hence large error), and I1 disappears together with the folded species as the temperature increases (Figure 3C). The next transition from I1 to I2 is dominant and closely matches the melting of RH subunits in other SLC2A1 variants (Figure 2D). The final transition is only present in SLC and 1R structures and corresponds to the melting of the LH subunit at a significantly higher temperature >71°C and with higher enthalpy (397 versus 290 kJ mol^{-1} for LH versus RH subunits of SLC, respectively). When the melting temperatures of the SLC2A1 variants are compared with the second transition in SLC (I1 → I2), all variants but 1L are less stable than SLC, although destabilization is not greater than 4°C.

From the species distribution diagram, the folded species predominates up to ∼60°C, with a gradual partial unfolding into the unidentified intermediate I1 (Figure 3C). From ∼60°C to ∼75°C, I2, the LHG4 species, dominates and fully unfolds above ∼80°C. Although we did not investigate the 1R variant by a full CD scan method, the similarity of its melting curve

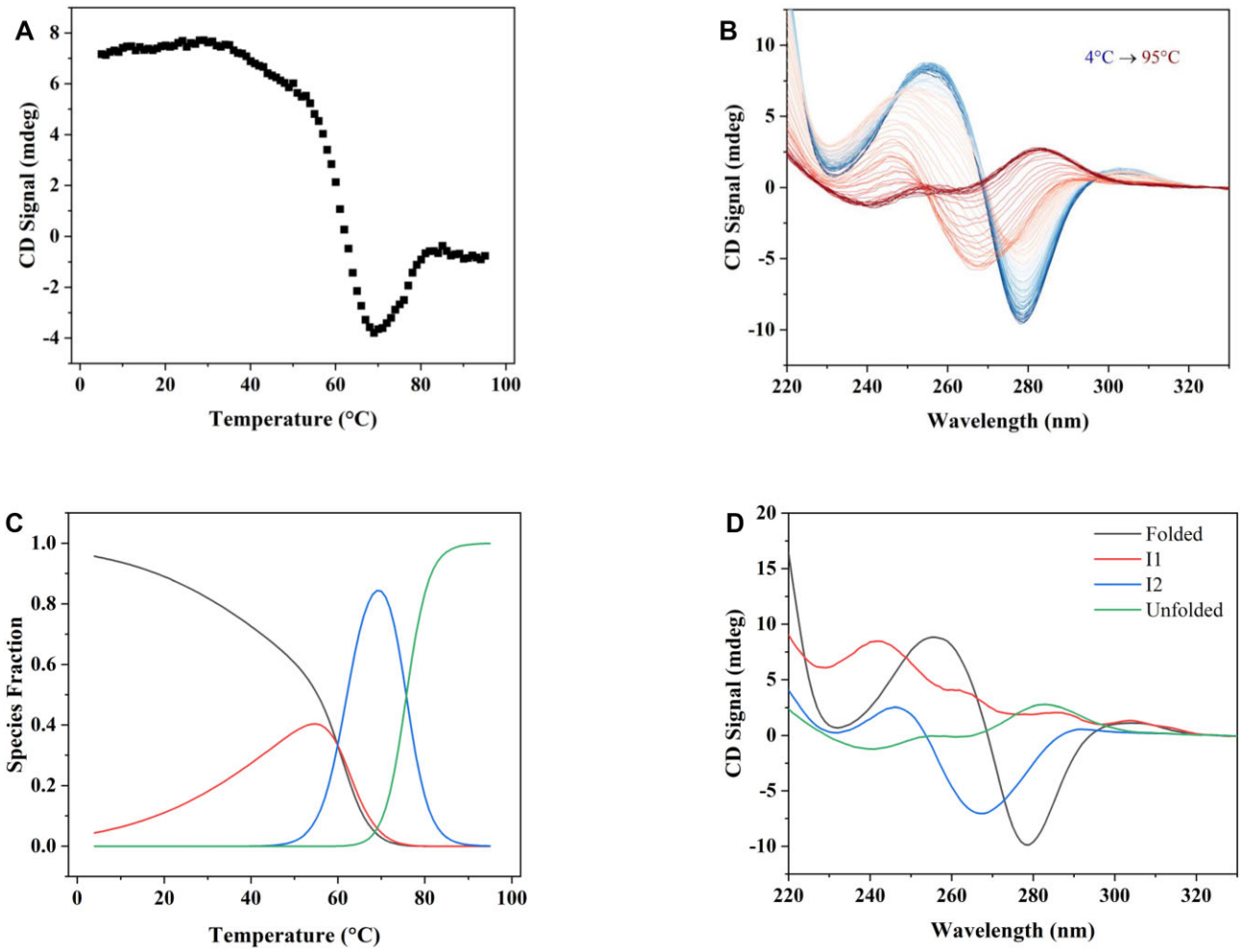


Figure 3. Determination of SLC intermediates in the full wavelength CD melting experiment. (**A**) CD melting curve of SLC monitored at 262 nm. (**B**) Representative full wavelength CD scan collected over the 4–95°C range. (**C**) Distribution diagram produced via SVD analysis with the assumptions of a three transitions model (F →I1→I2→U). (**D**) The basis CD spectra of four species. Samples were prepared at ∼4 μM in 100K buffer.

to that for SLC (Supplementary Figure S1) suggests that this structure might undergo a similar thermal unfolding process.

Biophysical characterization of interactions between LH/RHG4 SLC variants and NMM

Here, we focus on the three variants that maintained the LH/RHG4 fold—SLC, 1R, and 2R. We used CD scan, thermal melt, and PAGE to determine the conformational selectivity and stabilizing effect of NMM on these variants. NMM is known to selectively bind and stabilize RHG4s, favoring the parallel topology (52). We aimed to determine whether NMM would bind the RH subunit, coerce the LH subunit to the RH conformation, or both. We examined the effect of NMM at 0, 1, 2, 5, and 10 equivalents (eq.) Data for SLC are shown in Figure 4 and Supplementary Figure S4, while data for 1R and 2R are shown in Supplementary Figures S5 and S6, respectively.

The CD scan on SLC in the presence of increasing amount of NMM revealed a noticeable increase in the 262-nm peak indicative of RHG4. This change is accompanied by a modest decrease in the 280-nm trough associated with the LHG4 fold but only at ≥5 eq. of NMM (Figure 4A). We interpret the CD data to suggest that NMM enhances the fold of the RHG4 subunit without affecting the LHG4 subunit (at least at lower equivalency of NMM). CD melting data indicate that NMM both stabilizes SLC (Figure 4B) and simplifies its transition to a 'two-state like' melt (Supplementary Figure S4). Without NMM, three transitions are observed for SLC with T_m of 52°C, 60.6°C, and 75.8°C. The addition of NMM leads to a single transition with T_m of 69.7 ± 1°C at 1 eq. of NMM. Relative to the dominant SLC transition at 60.6°C, NMM stabilizes SLC by 9.1°C. This degree of stabilization is at the lower end of the typical NMM stabilization range for parallel RHG4s, of 11–23°C (52). It is important to point out that the bulk of NMM's stabilizing effect is observed at 1 eq. with only a small gradual increase in the T_m with the subsequent equivalents of NMM—by up to 3°C with 2–10 equivalents (Figure 4B). PAGE data confirm interactions of NMM with SLC, where even at one equivalent, only a single fast-moving band predominates and persists at a higher equivalence of NMM (Figure 4C). Because the SLC–NMM complex moves faster than SLC alone, it is possible that SLC adopts a more compact conformation and/or that NMM's negative charge contributes to the observed faster mobility.

In summary, our biophysical characterization of SLC–NMM interactions indicates that NMM increases the RH character of SLC without affecting its LH character. It stabilizes SLC by ∼10°C even at 1 eq. and leads to increased homogeneity of the SLC structure. Our data for the 1R

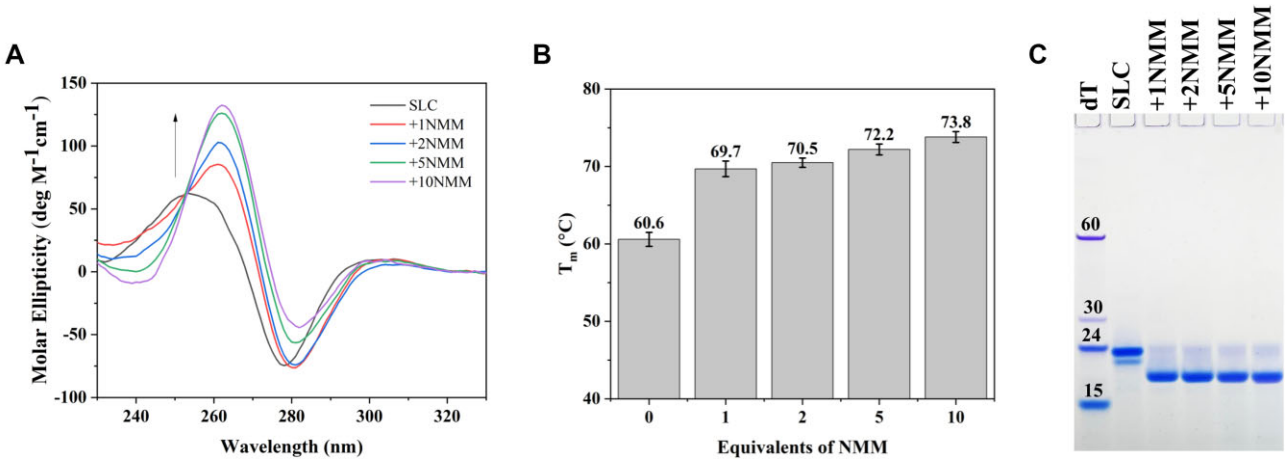


Figure 4. Titration of SLC with 1–10 eq. of NMM. (**A**) CD spectra. (**B**) Stabilization temperatures determined via CD melting studies. Error bars represent standard deviation across three trials. (**C**) Fifteen percent native PAGE. DNA was annealed at ∼50 μM for PAGE and ∼4 μM for all other experiments in the 100K buffer. Note, that G4 formation was also verified via TDS.

and 2R variants in the presence of 1–10 eq. of NMM (Supplementary Figure S5–S6) are consistent with the data presented for SLC, suggesting similar binding modes. The effect of another highly selective G4 ligand, PhenDC3, on LHG4 structure has been reported earlier (53). This study showed that at ≥3 eq. of PhenDC3, this ligand converts LHG4 into RHG4 but does not otherwise interact with LHG4.

UV-vis titration of SLC and 2R with NMM

We determined the binding affinity and stoichiometry of the SLC–NMM and 2R–NMM complexes using UV-vis titration (Figure 5 for SLC and Supplementary Figure S7 for 2R). Binding parameters are collected in Supplementary Table S2. The Soret peak of NMM displays hypochromicity of 2 ± 7% and a significant red shift of 18.3 ± 0.5 nm upon addition of SLC. Both values are consistent with those typical for NMM binding to parallel RHG4s (52), suggesting that the adjacent LH subunit does not affect the binding of NMM to the RH subunit.

The titration produced an isosbestic point at 392 ± 1 nm, suggesting that the binding between NMM and SLC is a two-state process. We confirmed this using an SVD analysis, which showed that only two spectral species, free and bound, were needed to characterize the family of spectra in Figure 5A. The amplitude vector from the SVD analysis was fit to a simple two-state binding model. The best-fitting binding model was determined to be 1:1 with a K_a value of 15 ± 2 μM^{-1} averaged over three trials. A 1:1 binding model is consistent with the crystal structure of the SLC–NMM complex (see below). In addition, other biophysical experiments also suggest a 1:1 binding. For example, we observed dramatic stabilization of SLC by NMM at one equivalent, with only a small increase at higher equivalents of NMM (Figure 4B). Similarly, one equivalent of NMM led to a new faster moving band on PAGE that did not change upon increase of NMM equivalents (Figure 4C). Titration of the 2R variant with NMM also yielded a 1:1 binding model with a similar K_a value of 15 ± 2 μM^{-1} and other binding parameters (Supplementary Figure S7 and Supplementary Table S2). These K_a values fall in the middle of the observed range of K_a values for NMM–RHG4 complexes (0.1–70 μM^{-1}; Figure 5C) (52).

Crystallographic characterization of SLC and SLC–NMM complexes

We have determined the crystal structures of native SLC and the SLC–NMM complex (Figure 6). The SLC crystal structure was solved in the *C*2 space group to a resolution of 2.39 Å, and the SLC–NMM crystal structure was solved in the $P4_12_12$ space group to a resolution of 1.45 Å. SLC adopts a hybrid LH/RHG4 topology in both structures. The two structures are very similar with a root mean square deviation (RMSD) of 1.43 Å for the G4s (Figure 6C, and Supplementary Figure S8). Interestingly, the LH subunits aligned to a greater extent as compared with the RH subunits, with an RMSD of 0.56 compared with 1.63 Å, respectively. The largest difference in the RH subunit was observed for the T_{20}-G_{21} stretch. The overall B-factor for the native SLC is significantly higher than that for the SLC–NMM complex (68.4 versus 42.3 $Å^2$; Supplementary Figure S9). Our analysis will focus on the major SLC–NMM structure due to its higher resolution and quality, unless we note otherwise.

Overall architecture of the native SLC and the SLC–NMM complex

All reported LHG4 and hybrid LH/RHG4 structures (listed in Supplementary Table S3) contain four tetrads divided into two subunits with parallel G4 topology (28–32). The native SLC and SLC–NMM structures determined here fit this pattern (Figure 6). To our knowledge, SLC–NMM is the first G4 DNA–ligand complex containing a LHG4 motif, albeit that NMM interacts with the RH subunit in this complex. Along with the native SLC, it is also the third reported DNA G4 structure with LH/RHG4 topology [two reported structures have PDB ids 6QJO and 6JCE (32)]. The observed topology agrees with the CD signature (Figures 2B and 4A). The LH subunit is formed exclusively by bases from the d[G(TGG)$_3$TG] LHG4 *Motif1* (31), whereas the RH subunit is formed by the d[(TGG)$_4$] sequence. The LH subunit has a pair of G-tetrad-contributing guanines which are not adjacent in sequence (G1/G12), a so-called split guanine pair required for LHG4 stabilization (28–32). The two subunits are linked by a single adenine (A13) oriented away from the structure (Figure 6A and B). This adenine is highly disordered in the

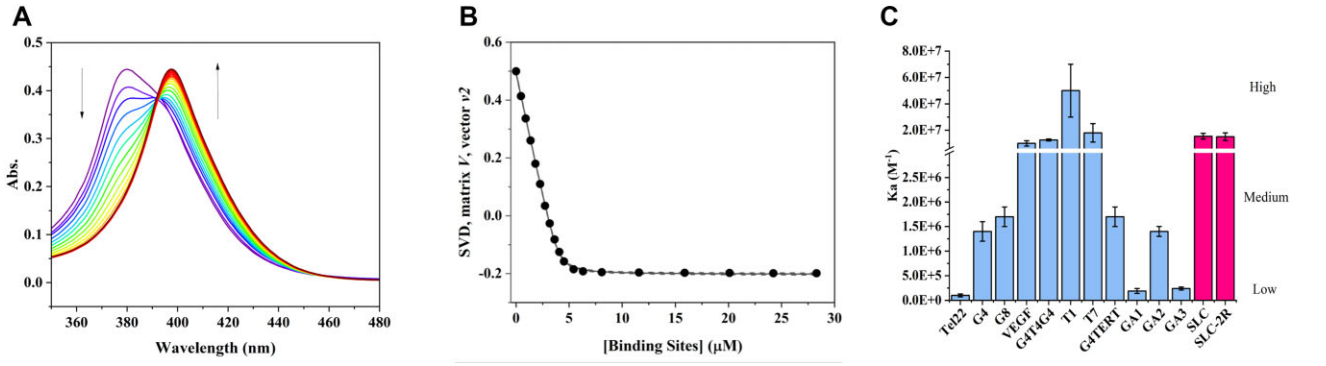


Figure 5. Determination of K_a for the SLC–NMM complex via UV-vis titration. (**A**) Representative UV-vis titration of 3.06 μM NMM with 457 μM SLC to a final [DNA]/[NMM] ratio of 9.25 at 20°C. (**B**) Fit of titration data to a 1:1 binding model with a floating [NMM] differing by 30% from that determined via UV-vis. [Binding sites] corresponds to [SLC]. The 95% confidence interval is shown as dashed lines. (**C**) Comparison of SLC–NMM and R2–NMM binding constants to those for other G4 DNA structures. The K_a values were determined via fluorescence (for G4, G8, T4G4T4 and T1) or UV-vis (all other sequences) in our lab. K_a was determined in the 100K buffer for SLC and in 10 mM Lithium cacodylate pH 7.2, 5 mM KCl, 95 mM LiCl buffer for all other sequences. Data for G4TERT (5′-AGGGGAGGGGCTGGGAGGGC-3′) and VEGF (5′-GGGAGGGTTGGGGTGGG-3′) are unpublished. Data for Tel22 are from (54), data for G4, G8 and T4G4T4 are from Ref (55) and data for T1 and T7 are from Ref (56). The binding strength is indicated on the right.

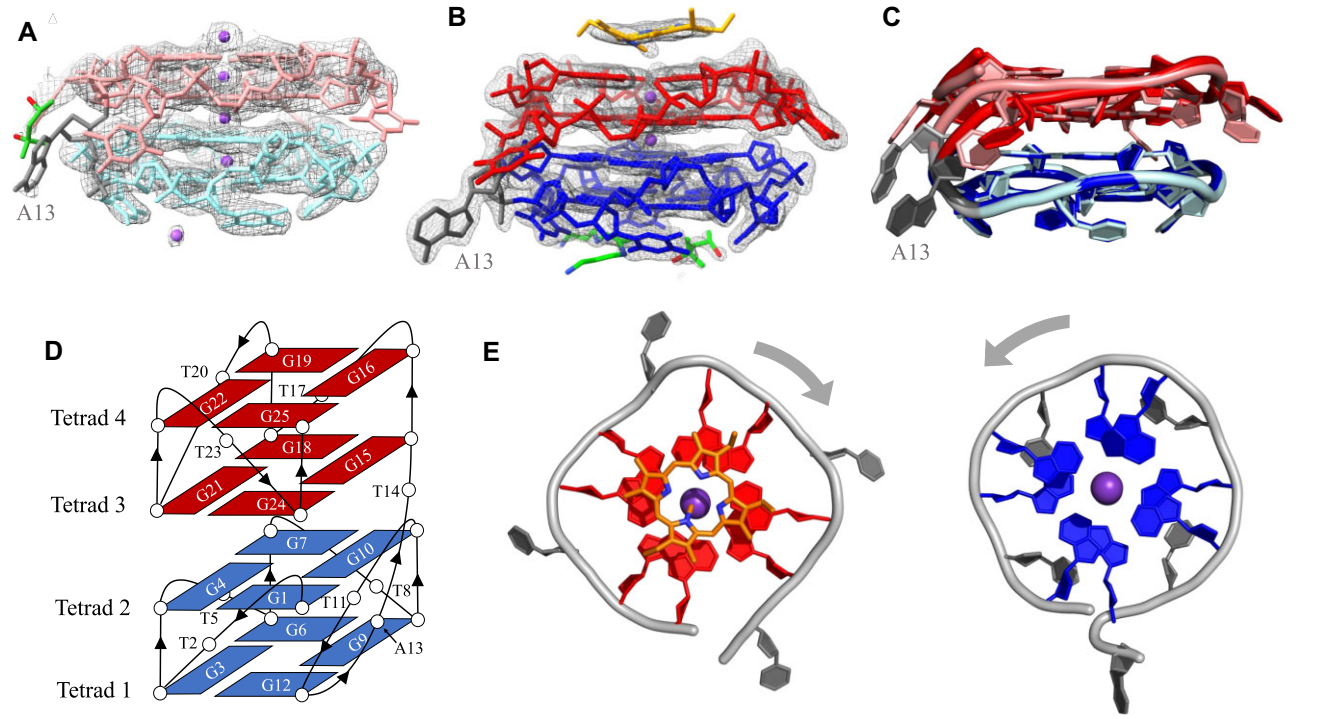


Figure 6. Architecture of the native SLC and the SLC–NMM complex structures. (**A**) The native SLC structure with electron density shown at $I/\sigma = 1$. The LH subunit is colored cyan and the RH subunit is in salmon. (**B**) The SLC–NMM structure with electron density shown at $I/\sigma = 1$. The LH subunit is colored blue and the RH subunit is in red. The adenine linker is colored gray. Potassium ions are represented by purple spheres, NMM carbon atoms are colored orange and MPD/spermine carbon atoms are colored forest green. (**C**) Overlay of native SLC and SLC–NMM (major conformation) structures. (**D**) Schematic representation of both structures. All bases are numbered, and the arrowed lines indicate backbone progression. (**E**) Individual RH (red) and LH (blue) subunits with the progression of the backbone from 5′ to 3′ indicated with gray arrows. Both subunits are viewed from the 3′ end, highlighting their RH and LH nature, as well as the position of thymine loops.

native SLC structure but is well defined in the SLC–NMM structure.

All guanines adopt an ‘anti’ conformation, which is typical for parallel RHG4s. Torsion angle analysis of the DNA backbone indicates a small number of outliers (Supplementary Figure S10) predominantly associated with the thymine loops and the adenine linker. The G-tetrads are overall relatively planar (out-of-plane deviation, D_{OOP} is 0.7–0.8 Å for both structures) with the external G-tetrads having somewhat greater nonplanarity. The G-tetrad nonplanarity of other four-tetrad two-subunit structures is similar but somewhat higher (Supplementary Table S4). The average distances between G-tetrads are 3.3 Å, similar to those in a variety of other LHG4 and LH/RHG4 structures, 3.2–3.4 Å

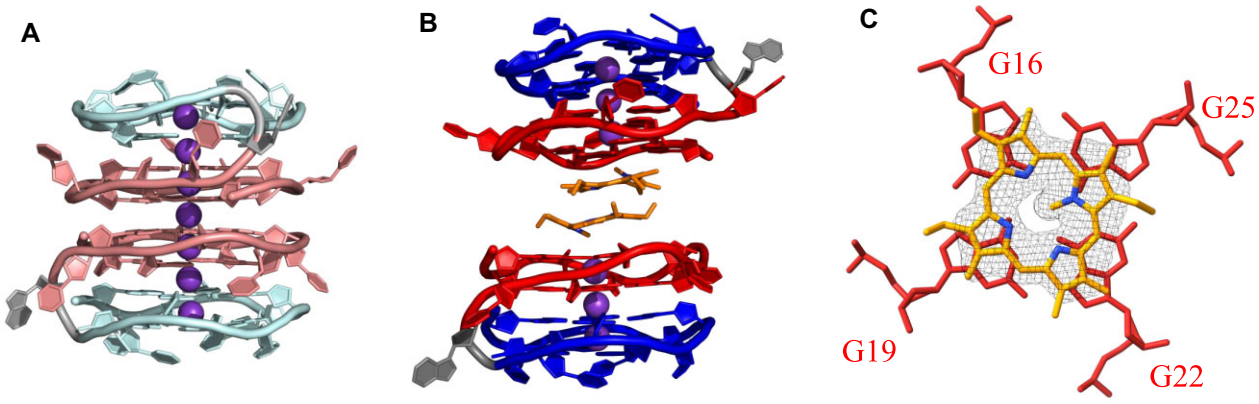


Figure 7. SLC dimer organization and NMM binding site. (**A**) Side view of the SLC dimer in the native SLC structure. (**B**) Side view of the SLC–NMM complex. (**C**) Top view of overlap between NMM and the terminal RH G-tetrad in the SLC–NMM complex. Electron density for NMM is displayed at $I/\sigma = 1$. Coloring scheme is the same as in Figure 6.

(Supplementary Table S5). Both subunits of the native SLC structure and SLC–NMM complex are parallel and display short medium-size grooves, as expected for the parallel topology (56). Groove widths are 14.7–15.1 Å measured from C5'-C5' (Supplementary Table S6). The helical twist between two G-tetrads in the LH subunit of the SLC–NMM complex is $-26.3 \pm 0.8°$ while that for the RH subunit is $32.5 \pm 0.4°$, underlying their different handedness. The helical twist between the two subunits (i.e. between G-tetrads 2 and 3) is significantly smaller, $-10 \pm 0.4°$ (Supplementary Table S7). The values for native SLC are similar.

Both subunits have single thymine loops—T2, T5, T8 and T11 in the LH subunit, and T14, T17, T20 and T23 in the RH subunit. All thymine bases in the RH subunit are oriented outward, leaving the 3′ terminal G-tetrad exposed and available for ligand binding in the SLC–NMM complex (Figure 6E) or stacking with a symmetry generated G4 in the native SLC structure (Figure 7A). In contrast, thymines in the LH subunit fold in and cap the 5′ G-tetrad (Figure 6E, and Supplementary Figure S11). This thymine arrangement is stabilized by the hydrogen bond between O4' of thymine's sugar and amine group (i.e. N2) of the nearby tetrad guanine, as was also reported earlier (28). As a result, the average B-factor for the thymine loops in the LH subunit is significantly lower than that for the RH subunit (Supplementary Figure S9). Thymine capping is present in all previously reported LHG4 structures and has been shown to be essential for LHG4 stabilization along with the presence of split guanines (28,29,31,32). The thymine capping likely prevents NMM binding to the LH subunit of SLC, while other smaller ligands (e.g. spermine and MPD in our case) can still fit in the remaining space (Supplementary Figure S11).

The two subunits share a central ion channel, a feature essential for G4 stabilization (Supplementary Figure S12). Native SLC forms a dimer with its symmetry mate, and its central ion channel spans both monomers and is populated by seven K^+ ions (Figure 7A). In the SLC–NMM structure, the ion channel is occupied by three K^+ ions. These are spaced by 3.3–3.6 Å and are found in the square antiprismatic arrangement of eight carbonyl oxygens. The average K–O distances are similar in both structures and in both subunits, 2.7–3.0 Å. The K–O and K–K distances are similar to those in canonical RHG4 structures (45,47).

The dimer interface in the native SLC structure

Two SLC molecules dimerize across their exposed outer RH G-tetrads (G-tetrad 4, Figure 7A). Such an interface is not present in the SLC–NMM complex because NMM binds at that location (Figure 7B). The G-tetrads at the dimer interface are spaced 3.42-Å apart, a somewhat longer distance than that determined for G-tetrads within each G4 (i.e. 3.30 Å; Supplementary Table S5). There is a K^+ ion present at the symmetry position, forming a continuous ion channel between the two molecules and stabilizing the dimer (Figure 7A). The G-tetrads at the dimer interface interact via π-stacking of the six and five membered rings of guanines with the six–six membered rings overlap being more extensive. This interface is similar to that between the two subunits in both SLC and SLC–NMM structures (Supplementary Figure S13).

Binding of NMM to the RH subunit of SLC

The SLC–NMM crystal structure displays unambiguous electron density for the NMM porphyrin ring and the N-methyl group. Peripheral substituents on NMM are disordered due to their high mobility and only one or two carbon atoms can be modeled. Our lab has solved several other G4-NMM structures (PDB ids 4FXM, 4G0F, 6PNK, 6P45, 8EBO and 8EDP). In all cases, peripheral substituents on NMM displayed no or weak density beyond the first carbon atom. The mobility of these groups stems from the fact that they are not engaged in any interactions with G4 DNA as we demonstrated for the case of human telomeric DNA (57). Interestingly the SLC–NMM structure is only the second case (after 4FXM) where clear density was observed for the N-methyl group, allowing us to unambiguously visualize its location. NMM is bound at the G-tetrad 4 in the RH subunit and is located at the expected π-stacking distance, even though its macrocycle does not efficiently overlap with the guanines of the terminal tetrad (Figure 7C). Rather, the out-of-plane N-methyl group of NMM protrudes into the ion channel, providing a 'lock-and-key' binding mode (Figure 7B and C). This mode has been observed in all published NMM–RHG4 structures, where G4s are parallel and contain 3–4 G-tetrads (40,50,51). The other face of NMM stacks with a symmetry-related copy of itself to form an NMM dimer (Figure 7B). Such a dimer has been previously observed twice, first in a

complex of NMM with the parallel RHG4 structure formed by a G-rich all-purine strand (45) and second, in a complex with a telomeric DNA RHG4 from *T. thermophila* (Yatsunyk *et. al.* unpublished observations). Interestingly, we believe that NMM dimer is also present in the recently published LH RNA G4 in complex with NMM (58). The authors did not model the second NMM in their structure, although its density is clear in the provided electron density maps. NMM is nonplanar due to the presence of the N-methyl group, but most of its nonplanarity is concentrated in the methylated pyrrole moiety. The two NMM molecules in the dimer are offset to allow for efficient overlap of their planar portions. SLC dimerization via NMM is likely caused by crystal packing forces as PAGE mobility of the SLC–NMM complex (Figure 4C) suggests that the DNA is present as a monomer in solution.

The SLC–NMM structure suggests a rationale for the selective binding of NMM to the RH subunit of SLC. Due to the outward direction of the thymine loops in the RH domain, its outer tetrad is available for NMM binding, while the thymine capping protects LH domain from binding to traditional G4 ligands. Another highly promising G4 ligand, PhenDC3, did not bind to the LHG4 due to the presence of thymine caps on both sides of the G4 (53). In contrast, NMM was shown to bind a LH RNA G4 with the sequence $r(GU)_{11}$ that possessed an exposed 5′-terminal G-tetrad (i.e. no thymine cap) (58). While this is just a single data point, it appears that the traditional G4 ligands can bind LHG4 that lack the thymine cap; perhaps it is also possible that some ligands may be able to remove or remodel the thymine cap to achieve binding. Finally, as we see in our structure, smaller ligands, such as MPD and spermine, can bind in the center of the thymine cap on the LH subunit (Supplementary Figure S11). Perhaps also, the presence of the thymine cap may enable differential targeting of LHG4 and RHG4 structures.

3D organization of DNA molecules in the two SLC structures

In the native SLC crystal structure, we observed only two sites of intermolecular contact, both involving loops: T5-T11' and T23-T23' (symmetry generated nucleotides are denoted by '). Both sets of thymines interact via π-stacking and display lower B-factors compared with other thymines, suggestive of lower mobility (Supplementary Figure S9).

The SLC–NMM crystal structure displays three sites of intermolecular contacts (Supplementary Figure S14). First, single-nucleotide thymine loops and the adenine linker participate in a perfect T8-A13'-T23" stack formed by nucleotides from three different molecules (one from the ASU and two generated by applying symmetry; Supplementary Figure S14A). Each molecule participates in three of those contacts bringing a total of seven molecules together in a well-organized 3D array. In addition, T11 hydrogen bonds with T14' via N3 and O2 atoms of T14' and O4 and N3 atoms of T11 (Supplementary Figure S14B). Finally, T17 stacks with T20' such that the sugar O4' of T20' forms hydrogen bonds with N1, N3 and O2 of T17 (Supplementary Figure S14C). The final important aspect of the 3D organization of the SLC–NMM complex is created by dimerization of NMM molecules bound to two adjacent SLC molecules as described in detail above.

Similarity between SLC and other LHG4 structures

All two-subunit G4 DNA structures with at least one LH domain [here and (28–32,49)] have a backbone that zig-zags around the circumference of the structure (Figure 8). In the all-LHG4 DNA structure (28), the native SLC structure, and in the SLC–NMM complex, the two backbones are symmetric, with their sinusoidal morphologies closely matching (Figure 8A–C). The previously reported native LH/RHG4 structure with PDB id 6QJO (32) has high backbone asymmetry (Figure 8D) likely a consequence of the 3′ end TGT base triplet, forcing the backbone of the RH domain into more pronounced loops.

Hydration in the SLC–NMM complex

The resolution of the native SLC structure, 2.4 Å, precludes the observation of an extended water network. On the other hand, the SLC–NMM structure contains 62 water molecules. Most waters are found between LH and RH subunits (Figure 9A) and in the grooves. These waters bridge the phosphate backbones, tetrad guanines and sugar O4' atoms. Coordination of waters to phosphates is the most common. They form clusters rather than a continuous network, a consequence of the resolution limit of this analysis. Multiple waters stabilize the G-tetrads in the RH subunit by connecting N2 and N3 of guanines to phosphates but also to sugar O4' atoms directly or via another water molecule (Figure 9B). In contrast, the LH subunit is more tightly packed and only its G-tetrad 2 at the LH/RH interface interacts with four waters, one per each guanine. These waters connect N2 and N3 of guanines to sugar O4' and phosphates of the tetrad below (Figure 9C). A number of waters are also found outside of G-tetrad 1 on the LH subunit. These connect MPD and spermine molecules found there to the phosphate backbone of the symmetry generated molecule.

Conclusions

We have characterized a sequence d[$G(TGG)_3TGA(TGG)_4$] (named SLC; the linker between two domains is underlined) found within the promoter of the human SLC2A1 gene. This sequence is not conserved in lower mammals such as the mouse, although it is largely intact in one primate sequence examined here (Figure 1A). The SLC sequence folds *in vitro* into a hybrid LH/RHG4 structure whose geometry is only minimally affected by binding to the NMM ligand. Extending the SLC sequence by 5′-G or 5′-GG from its genomic context results in a fully RHG4 structure regardless of whether additional nucleotides were present or not at the 3′ end. These RHG4s likely have four G-tetrads arranged into two subunits in a similar fashion to the structure of an antiproliferative DNA sequence (49). Extending the SLC sequence by 3′-C or 3′-CA preserved the original LH/RHG4 fold so long as we did not add any nucleotides in the 5′ direction. Because in this report we adhered to the genomic context, at this point it is unclear whether it is the nature of the base or the direction of extension that dictates refolding of the LH subunit into a RH fold. We are currently investigating this question.

The hybrid LH/RHG4 formed by SLC interacts with NMM and binds to it with a relatively strong binding constant of 15 μM^{-1} and a stoichiometry of 1:1. Interestingly, NMM does not convert the LH subunit into a RH subunit as was observed previously for PhenDC3 (53). Rather it merely

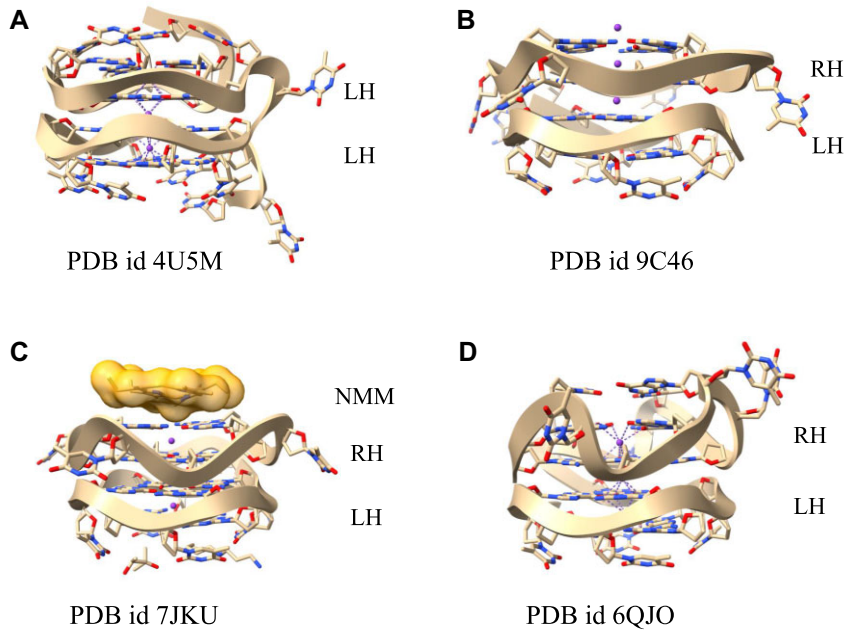


Figure 8. Cartoon views of the published (**A**) LHG4 (28), (**B**) the LH/RHG4 hybrid structure (32), (**C**) the native SLC structure and (**D**) the SLC–NMM structure. These have been drawn with the ChimeraX program (59).

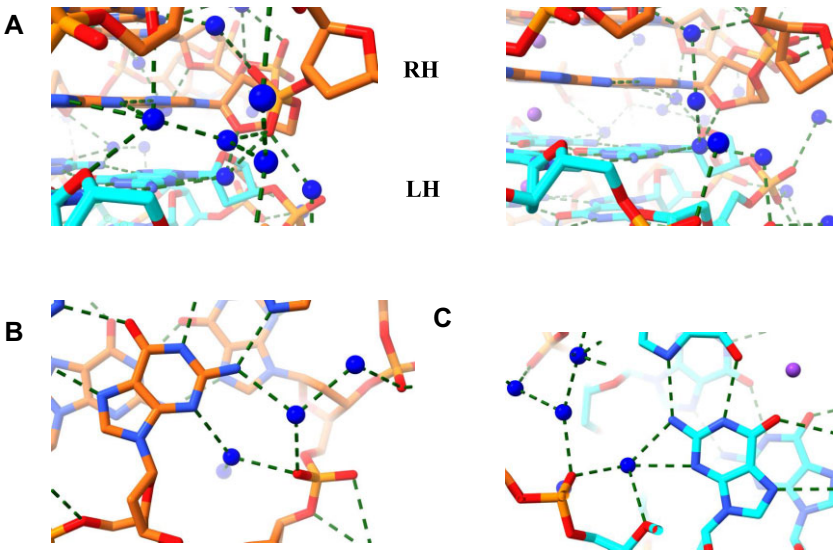


Figure 9. Water clusters in the SLC–NMM structure. Water molecules are depicted as blue spheres. LH subunit carbons are colored cyan and RH subunit carbons are mauve. (**A**) Water clusters between two subunits. Example of guanine–water contacts in (**B**) RH and (**C**) LH subunits.

stacks with the external G-tetrad of the existing RH subunit leading also to an increase in the homogeneity of the structure. The LH subunit is partially blocked from binding by a thymine cap, although binding of ligands smaller than NMM may be possible in the space that remains. Here, we observe MPD and spermine binding. Our results are consistent with recent findings that a fully LHG4 RNA binds to NMM via exposed terminal G-tetrad (58). NMM does not have extended cationic side chains, which would allow its interactions with phosphate groups forming the surfaces of the grooves, thus enhancing its affinity for G4 (15,18). Molecules with such features could bind more strongly to the RH domain and perhaps even to the LH domain if the side chains are able to extend into the grooves and loops, especially at the wider regions of the grooves. But such features also lead to decreased selectivity for G4 versus other DNA structures.

This study suggests that the propensity of the noncanonical LH motif identified to fold into LHG4s, d[G(TGG)$_3$TG] (*Motif1*) and d[(GGT)$_3$GTG] (*Motif2*), although rarely found in the human genome, is highly dependent on the nature of the flanking sequences. Biophysical experiments on the motifs in the absence of flanking sequences may thus give a misleading picture of LH propensity in the human genome. It also appears that the canonical G4 ligand NMM can stabilize structures that contain LH subunits by interacting with the RH subunit or even bind LH subunits in cases where the thymine cap is not present (58), or disrupted. To what extent LHG4 structures and their complexes with ligands are impediments

to helicase unwinding, transcription factor binding or polymerase precession (60–62), remains to be determined. Their rare prevalence in only a few genes, may make them attractive and highly selective therapeutic targets, such as that in the SLC2A1 promoter G4 examined here. This is especially true if LH-selective ligands can be found, for example, by screening of large chemical libraries. The rarity and the lack of evolutionary conservation found here for the SLC2A1 LHG4, suggests that LHG4s do not have innate biological functions, by contrast with conventional RHG4s (7,60–63).

Data availability

The SLC and SLC–NMM structures have been deposited with the PDB and assigned PDB ids 9C46 and 8TAA, respectively.

Supplementary data

Supplementary Data are available at NAR Online.

Funding

National Institutes of Health [1R15CA253134 to L.Y.]; Swarthmore [to L.Y.]; National Institute of General Medical Sciences [P30GM124165, P30GM133893]; NIH-ORIP [S10OD021527]; U.S. Department of Energy's Argonne National Laboratory [DE-AC02-06CH11357]; DOE Office of Science's Biological and Environmental Research [FWP # BO070]; US Department of Energy's Basic Energy Sciences [DE-SC0012704]. Funding for open access charge: Swarthmore College.

Conflict of interest statement

None declared.

References

1. Henderson,E., Hardin,C.C., Walk,S.K., Tinoco,I. and Blackburn,E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell*, **51**, 899–908.
2. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
3. Sundquist,W.I. and Klug,A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.
4. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
5. Bochman,M.L., Paeschke,K. and Zakian,V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.*, **13**, 770–780.
6. Varshney,D., Spiegel,J., Zyner,K., Tannahill,D. and Balasubramanian,S. (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.*, **21**, 459–474.
7. Spiegel,J., Adhikari,S. and Balasubramanian,S. (2020) The structure and function of DNA G-quadruplexes. *Trends Chem.*, **2**, 123–136.
8. Zyner,K.G., Mulhearn,D.S., Adhikari,S., Martínez Cuesta,S., Di Antonio,M., Erard,N., Hannon,G.J., Tannahill,D. and Balasubramanian,S. (2019) Genetic interactions of G-quadruplexes in humans. *eLife*, **8**, e46793.
9. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
10. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
11. Bedrat,A., Lacroix,L. and Mergny,J.-L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.*, **44**, 1746–1759.
12. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
13. Hänsel-Hertsch,R., Beraldi,D., Lensing,S.V., Marsico,G., Zyner,K., Parry,A., Di Antonio,M., Pike,J., Kimura,H., Narita,M., *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, **48**, 1267–1272.
14. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl Acad. Sci. U.S.A.*, **99**, 11593–11598.
15. Balasubramanian,S., Hurley,L.H. and Neidle,S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.*, **10**, 261–275.
16. Rigo,R., Palumbo,M. and Sissi,C. (2017) G-quadruplexes in human promoters: a challenge for therapeutic applications. *Biochim. Biophys. Acta Gen. Subj.*, **1861**, 1399–1413.
17. Lago,S., Nadai,M., Cernilogar,F.M., Kazerani,M., Domíniguez Moreno,H., Schotta,G. and Richter,S.N. (2021) Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat. Commun.*, **12**, 3885.
18. Criscuolo,A., Napolitano,E., Riccardi,C., Musumeci,D., Platella,C. and Montesarchio,D. (2022) Insights into the small molecule targeting of biologically relevant G-quadruplexes: an overview of NMR and crystal structures. *Pharmaceutics*, **14**, 2361.
19. Chen,E.V., Nicoludis,J.M., Powell,B.M., Li,K.S. and Yatsunyk,L.A. (2023) Crystal structure of a three-tetrad, parallel, K\sp +-stabilized human telomeric G-quadruplex at 1.35 Å resolution. *Acta Crystallogr. Sect. F*, **79**, 144–150.
20. Ou,A., Schmidberger,J.W., Wilson,K.A., Evans,C.W., Hargreaves,J.A., Grigg,M., O'Mara,M.L., Iyer,K.S., Bond,C.S. and Smith,N.M. (2020) High resolution crystal structure of a KRAS promoter G-quadruplex reveals a dimer with extensive poly-A π-stacking interactions for small-molecule recognition. *Nucleic Acids Res.*, **48**, 5766–5776.
21. Stump,S., Mou,T.-C., Sprang,S.R., Natale,N.R. and Beall,H.D. (2018) Crystal structure of the major quadruplex formed in the promoter region of the human c-MYC oncogene. *PLoS One*, **13**, e0205584.
22. Wei,D., Parkinson,G.N., Reszka,A.P. and Neidle,S. (2012) Crystal structure of a c-kit promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop conformation. *Nucleic Acids Res.*, **40**, 4691–4700.
23. Wei,D., Todd,A.K., Zloh,M., Gunaratnam,M., Parkinson,G.N. and Neidle,S. (2013) Crystal structure of a promoter sequence in the B-raf gene reveals an intertwined dimer quadruplex. *J. Am. Chem. Soc.*, **135**, 19319–19329.
24. Guédin,A., Lin,L.Y., Armane,S., Lacroix,L., Mergny,J.-L., Thore,S. and Yatsunyk,L.A. (2018) Quadruplexes in 'dicty': crystal structure of a four-quartet G-quadruplex formed by G-rich motif found in the dictyostelium discoideum genome. *Nucleic Acids Res.*, **46**, 5297–5307.
25. Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
26. Winnerdy,F.R. and Phan,A.T. (2020) Chapter Two - Quadruplex structure and diversity. In: Neidle,S. (ed.) *Annual Reports in Medicinal Chemistry*. Vol. **54**. Academic Press, Cambridge, Massachusetts. pp. 45–73.

27. Parkinson,G.N. and Collie,G.W. (2019) X-ray crystallographic studies of G-quadruplex structures. *Methods Mol. Biol.*, **2035**, 131–155.
28. Chung,W.J., Heddi,B., Schmitt,E., Lim,K.W., Mechulam,Y. and Phan,A.T. (2015) Structure of a left-handed DNA G-quadruplex. *Proc. Natl Acad. Sci. U.S.A.*, **112**, 2729–2733.
29. Das,P., Winnerdy,F.R., Maity,A., Mechulam,Y. and Phan,A.T. (2021) A novel minimal motif for left-handed G-quadruplex formation. *Chem. Commun.*, **57**, 2527–2530.
30. Das,P., Ngo,K.H., Winnerdy,F.R., Maity,A., Bakalar,B., Mechulam,Y., Schmitt,E. and Phan,A.T. (2021) Bulges in left-handed G-quadruplexes. *Nucleic Acids Res.*, **49**, 1724–1736.
31. Bakalar,B., Heddi,B., Schmitt,E., Mechulam,Y. and Phan,A.T. (2019) A minimal sequence for left-handed G-quadruplex formation. *Angew. Chem. Int. Ed Engl.*, **58**, 2331–2335.
32. Winnerdy,F.R., Bakalar,B., Maity,A., Vandana,J.J., Mechulam,Y., Schmitt,E. and Phan,A.T. (2019) NMR solution and X-ray crystal structures of a DNA molecule containing both right- and left-handed parallel-stranded G-quadruplexes. *Nucleic Acids Res.*, **47**, 8272–8281.
33. Cao,S., Chen,Y., Ren,Y., Feng,Y. and Long,S. (2021) GLUT1 biological function and inhibition: research advances. *Future Med. Chem.*, **13**, 1227–1243.
34. Ancey,P., Contat,C. and Meylan,E. (2018) Glucose transporters in cancer – from tumor cells to the tumor microenvironment. *FEBS J.*, **285**, 2926–2943.
35. Ren,J. and Chaires,J.B. (1999) Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry*, **38**, 16067–16075.
36. Gray,R.D. and Chaires,J.B. (2011) Analysis of multidimensional G-quadruplex melting curves. *Curr. Protoc. Nucleic Acid Chem.*, **45**, 17.4.1–17.4.16.
37. Qu,X. and Chaires,J.B. (2000) Analysis of drug–DNA binding data. *Methods Enzymol.*, **321**, 353–369.
38. Mergny,J.-L. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.
39. Viladoms,J. and Parkinson,G.N. (2014) HELIX: a new modular nucleic acid crystallization screen. *J. Appl. Crystallogr.*, **47**, 948–955.
40. Vonrhein,C., Flensburg,C., Keller,P., Sharff,A., Smart,O., Paciorek,W., Womack,T. and Bricogne,G. (2011) Data processing and analysis with the *autoPROC* toolbox. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 293–302.
41. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.-W., Kapral,G.J., Grosse-Kunstleve,R.W., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D*, **66**, 213–221.
42. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of ıt Coot. *Acta Crystallogr. Sect. D*, **66**, 486–501.
43. Kabsch,W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
44. Farag,M., Messaoudi,C. and Mouawad,L. (2023) ASC-G4, an algorithm to calculate advanced structural characteristics of G-quadruplexes. *Nucleic Acids Res.*, **51**, 2087–2107.
45. Ye,M., Chen,E.V., Pfeil,S.H., Martin,K.N., Atrafi,T., Yun,S., Martinez,Z. and Yatsunyk,L.A. (2023) Homopurine guanine-rich sequences in complex with N-methyl mesoporphyrin IX form parallel G-quadruplex dimers and display a unique symmetry tetrad. *Bioorg. Med. Chem.*, **77**, 117112.
46. Del Villar-Guerra,R., Gray,R.D. and Chaires,J.B. (2017) Characterization of quadruplex DNA structure by circular dichroism. *Curr. Protoc. Nucleic Acid Chem.*, **68**, 17.8.1–17.8.16.
47. Beseiso,D., Chen,E.V., McCarthy,S.E., Martin,K.N., Gallagher,E.P., Miao,J. and Yatsunyk,L.A. (2022) The first crystal structures of hybrid and parallel four-tetrad intramolecular G-quadruplexes. *Nucleic Acids Res.*, **50**, 2959–2972.
48. Largy,E., Mergny,J.-L. and Gabelica,V. (2016) Role of alkali metal ions in G-quadruplex nucleic acid structure and stability. In: Sigel,A., Sigel,H. and Sigel,R.K.O. (eds.) *The Alkali Metal Ions: Their Role for Life. Metal Ions in Life Sciences*. Vol. **16**. Springer, Cham, pp. 203–258.
49. Do,N.Q., Chung,W.J., Truong,T.H.A., Heddi,B. and Phan,A.T. (2017) G-quadruplex structure of an anti-proliferative DNA sequence. *Nucleic Acids Res.*, **45**, 7487–7493.
50. Gray,R.D., Buscaglia,R. and Chaires,J.B. (2012) Populated intermediates in the thermal unfolding of the Human telomeric quadruplex. *J. Am. Chem. Soc.*, **134**, 16834–16844.
51. Hendler,R.W. and Shrager,R.I. (1994) Deconvolutions based on singular value decomposition and the pseudoinverse: a guide for beginners. *J. Biochem. Biophys. Methods*, **28**, 1–33.
52. Yett,A., Lin,L.Y., Beseiso,D., Miao,J. and Yatsunyk,L.A. (2019) N-methyl mesoporphyrin IX as a highly selective light-up probe for G-quadruplex DNA. *J. Porphyr. Phthalocyanines*, **23**, 1195–1215.
53. Das,P. and Phan,A.T. (2022) Tetrad-binding ligands do not bind specifically to left-handed G-quadruplexes. *Chem. Commun.*, **58**, 11264–11267.
54. Nicoludis,J.M., Miller,S.T., Jeffrey,P.D., Barrett,S.P., Rablen,P.R., Lawton,T.J. and Yatsunyk,L.A. (2012) Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. *J. Am. Chem. Soc.*, **134**, 20446–20456.
55. Sabharwal,N.C., Savikhin,V., Turek-Herman,J.R., Nicoludis,J.M., Szalai,V.A. and Yatsunyk,L.A. (2014) N-methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes. *FEBS J.*, **281**, 1726–1737.
56. Lin,L.Y., McCarthy,S., Powell,B.M., Manurung,Y., Xiang,I.M., Dean,W.L., Chaires,B. and Yatsunyk,L.A. (2020) Biophysical and X-ray structural studies of the (GGGTT)3GGG G-quadruplex in complex with N-methyl mesoporphyrin IX. *PLoS One*, **15**, e0241513.
57. Nicoludis,J.M., Barrett,S.P., Mergny,J.-L. and Yatsunyk,L.A. (2012) Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.*, **40**, 5432–5447.
58. Roschdi,S., Yan,J., Nomura,Y., Escobar,C.A., Petersen,R.J., Bingman,C.A., Tonelli,M., Vivek,R., Montemayor,E.J., Wickens,M., *et al.* (2022) An atypical RNA quadruplex marks RNAs as vectors for gene silencing. *Nat. Struct. Mol. Biol.*, **29**, 1113–1121.
59. Goddard,T.D., Huang,C.C., Meng,E.C., Pettersen,E.F., Couch,G.S., Morris,J.H. and Ferrin,T.E. (2018) UCSF ChimeraX: meeting modern challenges in visualization and analysis: UCSF ChimeraX Visualization System. *Protein Sci.*, **27**, 14–25.
60. Esain-Garcia,I., Kirchner,A., Melidis,L., Tavares,R.D.C.A., Dhir,S., Simeone,A., Yu,Z., Madden,S.K., Hermann,R., Tannahill,D., *et al.* (2024) G-quadruplex DNA structure is a positive regulator of *MYC* transcription. *Proc. Natl Acad. Sci. U.S.A.*, **121**, e2320240121.
61. Chen,Y., Simeone,A., Melidis,L., Cuesta,S.M., Tannahill,D. and Balasubramanian,S. (2024) An upstream G-quadruplex DNA structure can stimulate gene transcription. *ACS Chem. Biol.*, **19**, 736–742.
62. Wulfridge,P. and Sarma,K. (2024) Intertwining roles of R-loops and G-quadruplexes in DNA repair, transcription and genome organization. *Nat. Cell Biol.*, **26**, 1025–1036.
63. Wang,G. and Vasquez,K.M. (2023) Dynamic alternative DNA structures in biology and disease. *Nat. Rev. Genet.*, **24**, 211–234.

Received: July 5, 2024. **Revised:** November 18, 2024. **Editorial Decision:** November 19, 2024. **Accepted:** November 21, 2024