

RESEARCH ARTICLE

# Crystal structure of the major quadruplex formed in the promoter region of the human c-MYC oncogene

Sascha Stump<sup>1</sup>, Tung-Chung Mou<sup>2</sup>, Stephen R. Sprang<sup>2</sup>, Nicholas R. Natale<sup>1</sup>, Howard D. Beall<sup>1</sup>\*

**1** Center for Environmental Health Sciences, Department of Biomedical and Pharmaceutical Sciences, University of Montana, Missoula, Montana, United States of America, **2** Center for Biomolecular Structure and Dynamics, Division of Biological Sciences, University of Montana, Missoula, Montana, United States of America

\* [howard.beall@umontana.edu](mailto:howard.beall@umontana.edu)



**OPEN ACCESS**

**Citation:** Stump S, Mou T-C, Sprang SR, Natale NR, Beall HD (2018) Crystal structure of the major quadruplex formed in the promoter region of the human c-MYC oncogene. PLoS ONE 13(10): e0205584. <https://doi.org/10.1371/journal.pone.0205584>

**Editor:** Maria Spies, University of Iowa, UNITED STATES

**Received:** August 23, 2018

**Accepted:** September 27, 2018

**Published:** October 12, 2018

**Copyright:** © 2018 Stump et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All atomic coordinates and structure factors are available from the RCSB Protein Data Bank (accession number: 6AU4).

**Funding:** The research described was supported by the National Institutes of Health grant #P20ES017670 (HDB, NRN), National Institute of General Medical Sciences of the National Institutes of Health CoBRE #P20GM103546 (SRS) and the ALSAM Foundation Skaggs Scholar Program grant: AIMS as Novel G-Quadruplex DNA Binding

## Abstract

The c-MYC oncogene mediates multiple tumor cell survival pathways and is dysregulated or overexpressed in the majority of human cancers. The NHE III<sub>1</sub> region of the c-MYC promoter forms a DNA quadruplex. Stabilization of this structure with small molecules has been shown to reduce expression of c-MYC, and targeting the c-MYC quadruplex has become an emerging strategy for development of antitumor compounds. Previous solution NMR studies of the c-MYC quadruplex have assigned the major conformer and topology of this important target, however, regions outside the G-quartet core were not as well-defined. Here, we report a high-resolution crystal structure (2.35 Å) of the major quadruplex formed in the NHE III<sub>1</sub> region of the c-MYC promoter. The crystal structure is in general agreement with the solution NMR structure, however, key differences are observed in the position of nucleotides outside the G-quartet core. The crystal structure provides an alternative model that, along with comparisons to other reported quadruplex crystal structures, will be important to the rational design of selective compounds. This work will aid in development of ligands to target the c-MYC promoter quadruplex with the goal of creating novel anticancer therapies.

## Introduction

Guanine-rich sequences of DNA and RNA can form a secondary nucleic acid structure known as a quadruplex. Quadruplex motifs have become the subject of significant interest due to their presence in human telomeres, 5'-untranslated regions of mRNA, and in gene promoter regions [1]. One such quadruplex-forming sequence is found in the promoter region of the human c-MYC oncogene. c-MYC is estimated to be dysregulated or overexpressed in approximately 70% of all human cancers and is responsible for mediating multiple pathways important in tumor cell survival [2]. Stabilization of the major quadruplex formed in the c-MYC promoter by various small molecules has been shown to inhibit transcription of c-MYC thereby reducing expression of the oncogene [3,4]. This reduction in c-MYC expression has

Anticancer Agents grant #1000233768 (HDB, NRN). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

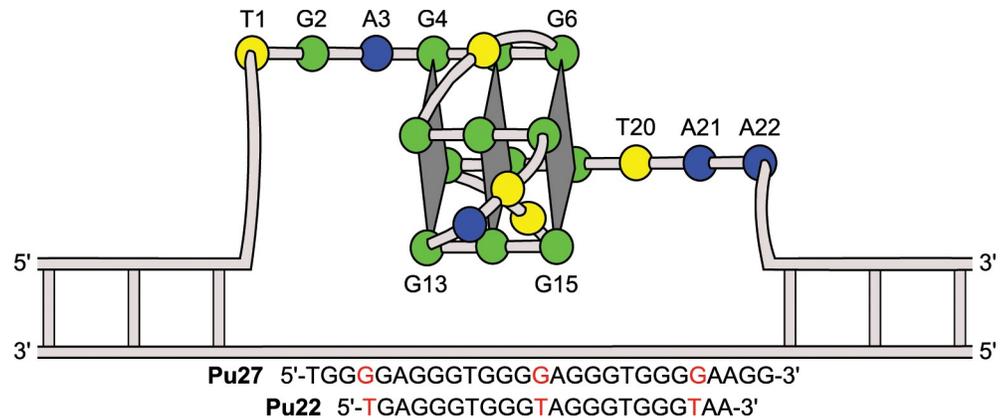
been demonstrated to induce apoptosis in multiple types of tumor cells [5,6]. Taken together, these findings suggest that the c-MYC promoter quadruplex is a promising antitumor target. Several research groups are designing small molecules to stabilize the c-MYC promoter quadruplex as a strategy to develop potential therapies for treatment of human cancers [7–9].

The general sequence motif that forms a quadruplex consists of several short guanine repeats (G), separated by short “loop” regions (L) comprised of other nucleotides with the overall general sequence of  $G_{3-5}L_{<7}G_{3-5}L_{<7}G_{3-5}L_{<7}G_{3-5}$  [1]. Quadruplexes can form as intramolecular or intermolecular arrangements, consisting of single or multiple nucleic acid strands, respectively. The basic unit of the quadruplex is the G-quartet, which is formed as a planar arrangement of four guanine residues held together through Hoogsteen bonding and stabilized by a central monovalent cation. Multiple G-quartets, usually three or more, stack upon each other to form the quadruplex secondary structure and are connected through external loop region nucleotides. The central channel cations are essential for quadruplex formation, with potassium generally preferred to sodium. Potassium cations are observed in a symmetric square antiprismatic coordination at the interface of two G-quartets, coordinated by the guanine O6 atoms, whereas the relatively smaller sodium atoms display square-planar coordination and are central to a single G-quartet. The species and abundance of these and other cations in solution can also serve to influence and stabilize the overall quadruplex topology [10]. Quadruplexes can be further categorized as parallel, anti-parallel, or hybrid by the types of loops formed and direction of the backbone in relation to the G-quartets (Fig 1) [1].

Parallel-type quadruplexes are found in promoter regions of c-MYC and several other oncogenes that are potential therapeutic targets for cancer including c-KIT, Bcl-2, VEGF, and HIF-1 $\alpha$  [3,11–16]. The human c-MYC promoter quadruplex is formed under negative supercoiling conditions in the nuclease hypersensitivity element III<sub>1</sub> region (NHE III<sub>1</sub>) [17]. NHE III<sub>1</sub> is 27 residues in length and responsible for the regulation of 80–90% of c-MYC oncogene transcription (Pu27) [18]. Duplex/quadruplex or single-stranded forms of the NHE III<sub>1</sub> region can be bound by transcription factors SP1 or CNBP/hnRNP, respectively, to increase c-MYC expression while formation of the quadruplex structure prevents transcription [19,20]. NHE III<sub>1</sub> contains 20 guanines grouped into five segments of 3 to 4 guanine nucleotides separated by one or two adenine or thymine nucleotides. This guanine-rich character allows formation of four possible quadruplex topologies, with the major form being a parallel intramolecular quadruplex comprised of the four guanine segments at the 3'-end of the sequence [18].

In a previous NMR study, it was found that the major biologically relevant conformation adopted by the c-MYC promoter quadruplex could be selected from other conformers by mutating residues G4, G14 and G23 to thymines and truncating the sequence to 22 nucleotides in length (Pu22) [18]. In a separate NMR study, the binding of a quindoline compound with Pu22 and the wild-type Pu27 sequence in solution was described [21]. The latter study revealed the small molecule bound in an “induced fit” manner at two sites on the quadruplex, with the 5' and 3' flanking sequences recruited to form a binding pocket for both the Pu27 and Pu22 sequences. To select the desired biologically relevant conformation, the same Pu22 sequence was further utilized in this work to obtain the reported crystal structure.

Our interest in obtaining the crystal structure of the c-MYC quadruplex formed by Pu22 stemmed from differences previously observed between solution and solid-state structures of other quadruplexes in the literature. Additionally, crystal structures often reveal other important structural information that may be absent in solution, such as the involvement of ions or water molecules important to the overall quadruplex topology, or potentially relevant quaternary interactions. On comparison of crystal and solution NMR structures of similar quadruplex sequences, differences are sometimes observed in the type of cation in the central channel and the overall topology of the quadruplexes. For example, multiple quadruplex topologies



**Fig 1. Topology of the c-MYC promoter quadruplex.** Diagram showing fully parallel topology of the c-MYC promoter quadruplex crystal structure with all reversal loops continuing in the right to left direction (5' to 3'). Thymines are represented in yellow, adenines in blue, and guanines in green. Sequence modifications in Pu22 are shown highlighted in red.

<https://doi.org/10.1371/journal.pone.0205584.g001>

have been demonstrated in studies using the sequence repeat found in human telomeres, d (GGGTTA). Solution NMR studies have revealed that an anti-parallel quadruplex is formed in the presence of sodium ions, and alternatively, a 3+1 hybrid-type quadruplex is observed in solution containing potassium ions, with three parallel strand-edges and one anti-parallel [22–24]. In contrast, the reported crystal structure of the same sequence adopts a fully parallel arrangement with potassium cations occupying the central channel [25]. This parallel topology was also shown to be favored in studies of solutions containing molecular crowding conditions [26,27]. However, this is not always the case: for example, the c-KIT oncogene promoter quadruplex in both solution NMR and crystal structures adopts a strictly parallel form containing potassium in the central channel, with the position of the nucleotides highly-conserved [28–30]. This led us to question whether the crystal structures of other oncogene promoter quadruplexes would be analogous to their NMR solution counterparts, or if they would differ significantly as is seen with the telomeric sequences. Knowledge of the distinct topology of these quadruplexes is crucial to the design of small molecules that stabilize their structures, and there is evidence that subtle differences in structural features could allow for selectivity between specific quadruplex targets [9,31].

Here, we report a high-resolution crystal structure of the major Pu22 quadruplex formed in the human c-MYC promoter and describe several features observed that are potentially important for small molecule binding and quadruplex stabilization. We compare the crystal structure of Pu22 with previously reported solution NMR structures in an effort to inform future design of quadruplex-targeted compounds [18,32,21,33]. In addition, we have examined the features of the Pu22 crystal structure in conjunction with other quadruplex crystal structures to probe for similar features including positions of ions, water molecules and quaternary interactions. This research aims to aid future development of novel quadruplex-targeted compounds and provide information helpful in co-crystallization studies with molecules designed to bind the c-MYC promoter and other quadruplex structures.

## Materials and methods

### Crystallization

The 22-residue DNA oligonucleotide (5' –TGAGGGTGGGTAGGGTGGGTAA–3') was synthesized and purified by Integrated DNA Technologies (standard desalting). The oligonucleotide

was diluted into a stock concentration of 10 mM in DNase/RNase free water and the concentration was determined using a Thermo Scientific NanoDrop spectrometer. The oligonucleotide was then diluted to 2.0 mM in 20 mM sodium cacodylate buffer at pH 6.5 containing 30 mM KCl and annealed by heating for 10 minutes at 95° C and cooled overnight at 4° C prior to crystallization experiments. Initial crystal screening was done in a 96-well plate format using an Art Robbins Instruments GRYPHON liquid-handling crystallization robot. Various precipitants (PEGs, 2-Methyl-2,4-pentanediol (MPD)) were screened along with varying concentrations of salts (NaCl, KCl, LiCl etc.) as previously outlined [34]. Crystals used for diffraction data collection were grown in a 24-well plate format using the hanging-drop vapor diffusion technique with 300 mM KCl, 50 mM LiCl, and 22.5% MPD in 50 mM sodium cacodylate at pH 6.5 with a 1:1 ratio of oligonucleotide to reservoir solution. Crystals were harvested using the reservoir solution or additional MPD (30%) as a cryo-protectant and flash cooled in liquid nitrogen for storage prior to data collection.

### Data collection and refinement

Initial diffraction screening was performed using a Rigaku MicroMax 007HF X-ray generator with VariMax HighFlux optic and R-AXIS IV image plate detector. Diffraction data for structure determination were taken at the Stanford Synchrotron Radiation Lightsource (SSRL) beam line 12–2. The native dataset was collected at a wavelength of 0.9793 Å over a 360° range with 0.2° rotation per image. Data were processed using XDS with autoxds script at SSRL [35]. These data were analyzed using Xtriage in Phenix prior to phasing [36]. Initial phases were obtained by molecular replacement using Phaser in Phenix using the truncated guanine decks of a CKIT-1 promoter quadruplex as a search model (PDBID: 4WO2) [29]. An improved, complete model was constructed through iterative cycles of refinement, phasing, and manual model building using Phenix and Coot, respectively [37]. Refinement was performed in Phenix refine, followed by submission to the PDB Redo webservice [38]. The structure obtained from the PDB Redo webservice was then re-refined in Phenix refine with simulated annealing and randomized atomic displacement parameters to reduce any potential bias in  $R_{\text{free}}$ . The final structure was refined in Phenix using data from 34.91–2.35 Å with a final  $R_{\text{work}}$ ,  $R_{\text{free}}$  of 0.220 and 0.245, respectively (Table 1). Atomic coordinates and structure factors are deposited in the RCSB Protein Databank with ID 6AU4 [39]. Visualization and RMSD calculations were performed using Pymol (<http://www.pymol.org>) [40]. Pearson correlation coefficient (CC) was calculated by randomly assigning the experimental reflection data to two half-datasets (x, y) as described by Karplus and Diederichs (Table 1) [41].

### Circular dichroism spectroscopy

Circular dichroism spectra were measured using a Jasco J-810 spectropolarimeter or on a Chirascan CD spectrophotometer at room temperature in a quartz cuvette with a 10 mm or 4 mm pathlength for sample A or B, respectively. The oligonucleotide sample was made at 2 μM concentration with 300 mM KCl, 50 mM LiCl, and 22.5% MPD in a 50 mM sodium cacodylate buffer at pH 6.5 to mimic the crystallization conditions (A) or at 5 μM in 10 mM potassium phosphate buffer pH 6 to replicate the conditions used in previous studies [21,42] (B). The solutions were annealed by heating to 95° C for 10 minutes and cooled overnight to 4° C prior to data collection.

## Results

The 22 residue c-MYC promoter sequence Pu22 crystallized in the P2<sub>1</sub>2<sub>1</sub>2 space group in a stacked dimer formation with two parallel, single-stranded quadruplex structures per

**Table 1. Data collection and refinement statistics.**

Sequence	5' -TGAGGGTGGGTAGGGTGGGTAA-3'
<b>Data Collection</b>	
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2
Unit cell dimensions (a, b, c) (Å) α, β, γ (°)	65.7 69.8 33.0 90 90 90
Wavelength (Å)	0.9793
Resolution (Å)*	34.91–2.35 (2.43–2.35)
Total reflections*	85280 (8398)
Unique reflections*	6774 (652)
Multiplicity*	12.6 (12.9)
Completeness (%)*	99.4% (99.7%)
I/σ*	33.9 (4.4)
Wilson B-factor (Å <sup>2</sup> )	58.8
R <sub>meas</sub> <sup>†</sup> *	0.042 (0.605)
CC <sub>1/2</sub> <sup>‡</sup> *	1.000 (0.966)
<b>Refinement</b>	
Resolution (Å)*	34.91–2.35 (2.53–2.35)
R <sub>work</sub> <sup>§</sup> *	0.220 (0.319)
R <sub>free</sub> <sup>§</sup> *	0.245 (0.322)
Macromolecules	2
DNA Residues	44
Total Atoms	956
Potassium ions	6
Waters	16
Average Overall B-factor (Å <sup>2</sup> )	80.2
Average B-factor DNA Residues (Å <sup>2</sup> )	80.7
Average B-factor Potassium Ions (Å <sup>2</sup> )	53.1
Average B-factor Waters (Å <sup>2</sup> )	59.7
RMS (bonds) (Å)	0.009
RMS (angles) (°)	1.08
TLS groups	2
<b>PDB ID</b>	6AU4

\*Statistics in parentheses are for the high-resolution shell

<sup>†</sup>  $R_{meas} = \frac{\sum_{hkl} \sqrt{\sum_{i=1}^n |I_i(hkl) - \bar{I}(hkl)|}}{\sum_{hkl} \sum_{i=1}^n I_i(hkl)}$ , where  $I_i(hkl)$  is the  $i$ th observation of the intensity of the reflection  $hkl$  and  $n$  is the multiplicity.

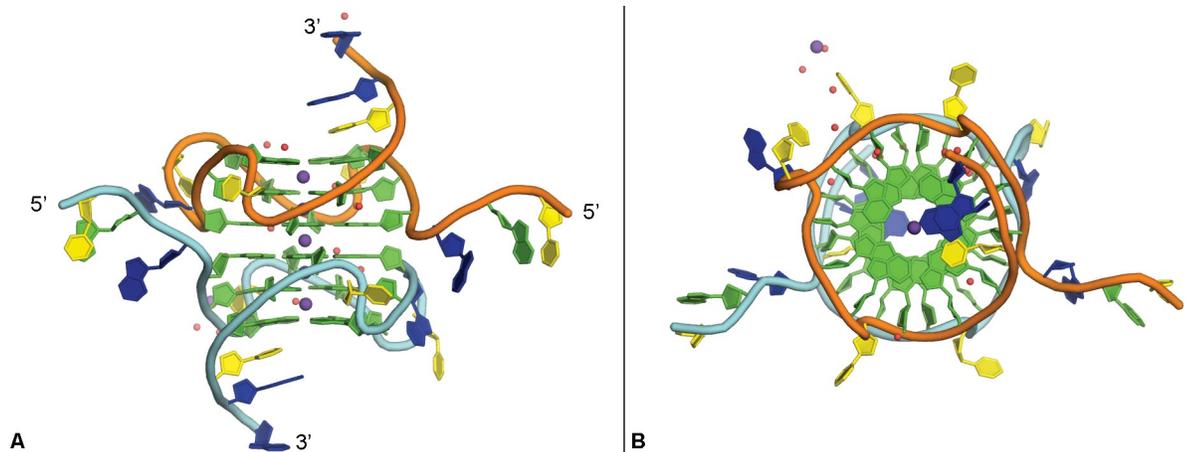
<sup>§</sup>  $R_{work} = \frac{\sum_{hkl} ||F_{obs}| - |F_{calc}||}{\sum_{hkl} |F_{obs}|}$ , where  $F_{obs}$  and  $F_{calc}$  are the observed and calculated structure-factor amplitudes for each reflection  $hkl$ .  $R_{free}$  was calculated with 10% of the diffraction data that were selected randomly and excluded from refinement.

<sup>‡</sup>CC<sub>1/2</sub> is the intra-dataset Pearson correlation coefficient (CC) calculated by randomly assigning the experimental reflection data to two half-datasets (x, y).  $CC =$

$$\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

<https://doi.org/10.1371/journal.pone.0205584.t001>

asymmetric unit (Table 1 and Fig 2). The two independent quadruplexes (A, B) are structurally similar with the majority of the core guanine-residue positions being maintained between the strands (G4-G6<sub>A/B</sub>, G8-G10<sub>A/B</sub>, G13-G15<sub>A/B</sub>, and G17-G19<sub>A/B</sub> RMSD = 0.190 Å). Each individual quadruplex contains three loops and a 5'-head and 3'-tail region. The 5'-head region consists of residues T1-G2-A3 and is observed stretched away from the central G-quartets with the residues roughly orthogonal relative to the core guanines. The two related 5'-head regions appear in a quasi-mirror-related arrangement with the related atomic positions



**Fig 2. Crystal structure of c-MYC quadruplex.** Side-view (A) and top-view (B) of the crystal structure of the c-MYC promoter quadruplex formed by Pu22, guanine bases in green, thymine in yellow, adenine in blue. Potassium ions in purple (two shown positioned behind strand backbones), waters in red. Orange/cyan backbone coloring represents strand A and B, respectively.

<https://doi.org/10.1371/journal.pone.0205584.g002>

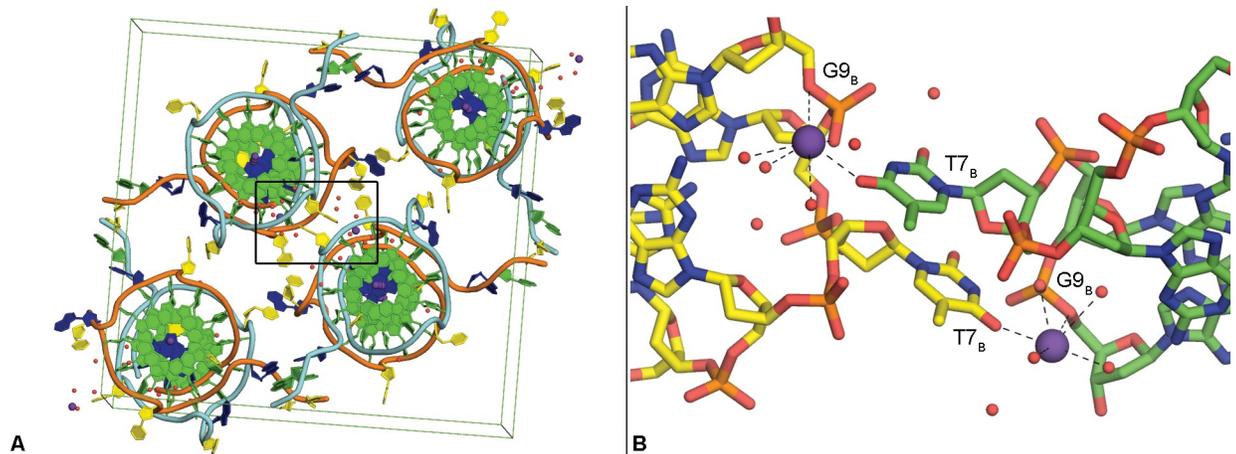
maintained (T1-G2-A3<sub>A/B</sub>, RMSD = 0.568 Å). The 3'-tail regions of the two quadruplexes in the asymmetric unit are comprised of residues T20-A21-A22 and are positioned below and approximately planar to the central G-quartets. The quadruplex structure also contains three double-chain reversal propeller-type loops, with two consisting of single thymine residues (T7, T16) that flank a third, two-residue loop (T11-A12). These regions display less similarity between quadruplex A and B, owing predominately to the loop formed by residues T11 and A12 (T7<sub>A/B</sub>, T16<sub>A/B</sub>, and T11-A12<sub>A/B</sub> RMSD = 2.182 Å). All of the loops, the head, and the tail regions are involved in interactions important for the crystal packing. Electron density is also observed for a non-channel potassium ion and several water molecules.

### Potassium ions

Each individual quadruplex contains two potassium ions at the centers of two stacked G-quartets and an additional potassium ion similarly positioned at the interface that forms the dimer in the asymmetric unit, for a total of five potassium ions in the central cavity (Fig 2). The positions of these potassium ions relative to the G-quartets is consistent with that reported for other quadruplex crystal and NMR structures [28,32]. The ions central to the G-quartets are observed in symmetric square antiprismatic coordination with the O6 atoms of the guanine residues, as expected, with an average bond distance of 2.65 Å. An additional potassium ion is positioned between two adjacent symmetry-related strand B quadruplexes, interacting with O4 oxygen of T7 of one and backbone phosphate of G9 of the other (Fig 3B). This ion is also observed in close proximity (< 3.7 Å) to three additional water molecules. This additional potassium ion was placed following careful consideration of each of the other components present in the crystallization conditions.

### Packing interactions

The Pu22 c-MYC promoter quadruplex structure displays several interesting packing interactions that induce a more elongated conformation than that observed for the same sequence by solution NMR. The two quadruplex strands, A and B,  $\pi$ - $\pi$  stack together at the G-quartets near their 5'-ends with the G-quartets containing residues G4, G8, G13, and G17 of each quadruplex interacting to create the extended dimer structure (Fig 2). Residues T1-G2-A3 from the



**Fig 3. Crystal packing and non-channel potassium ions.** Packing arrangement in the unit cell of c-MYC promoter quadruplex crystal structure formed by Pu22, coloring scheme same as Fig 2; black box indicates region shown in Fig 3B (A). Non-channel potassium ions interact with residues T7 and G9 (B).

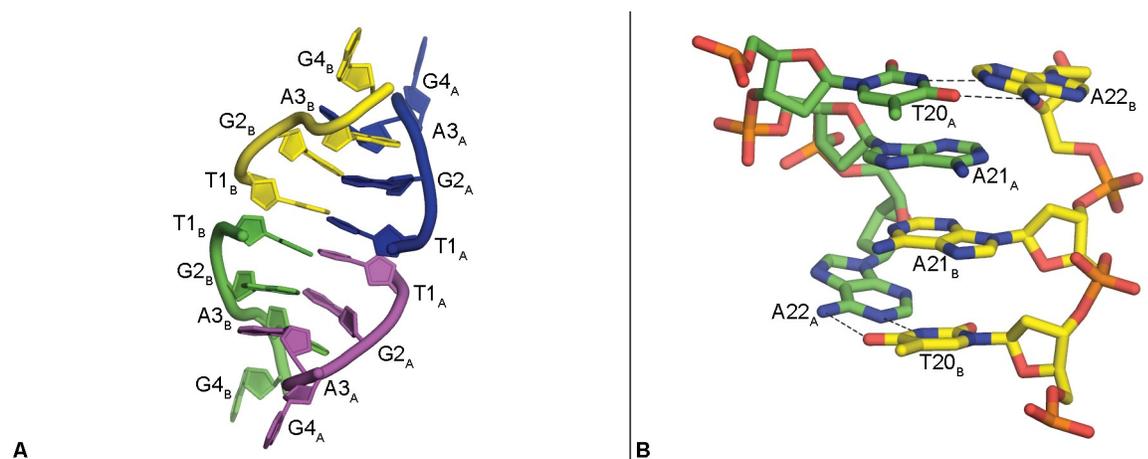
<https://doi.org/10.1371/journal.pone.0205584.g003>

5'-head of four quadruplexes in the lattice, two of A and two of B, form stacked structures stretched away from the central G-quartets, with base pairing between T1<sub>A</sub>/T1<sub>B</sub>, G2<sub>A</sub>/G2<sub>B</sub> and G3<sub>A</sub>/G3<sub>B</sub> (Fig 4A).

This exposes the G-quartet composed of residues G4, G7, G12, and G17, allowing the stacking of the two quadruplexes head-to-head in the asymmetric unit (Fig 2). Residues T20-A21-A22 from the 3'-tail of the two quadruplexes, A and B, interact forming a near-planar double helical structure (Fig 4B). Residue T20<sub>A/B</sub> and A22<sub>A/B</sub> form Watson-Crick base pairs with slightly elongated hydrogen bonds and A21<sub>A</sub> and A21<sub>B</sub> are involved in an apparent  $\pi$ -stacking interaction.

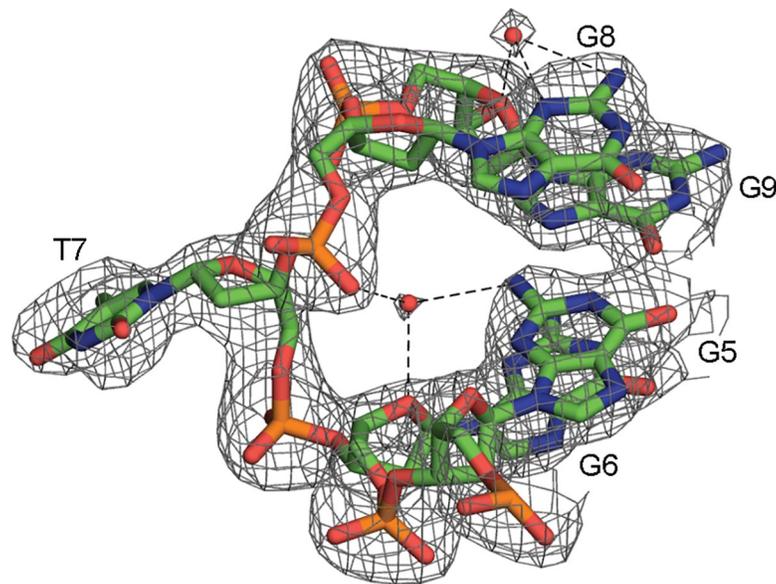
### Conserved water molecules

Two water molecules were observed in nearly identical positions in each of the independent quadruplexes formed by strand A and B (Fig 5). The conserved positions of these waters



**Fig 4. Packing interactions.** Quaternary interaction of four 5'-head regions (T1-G2-A3); each color represents a separate symmetry-related strand (A). Stacked helical structure formed between 3'-tail regions of (T20-A21-A22) in the crystal lattice (B).

<https://doi.org/10.1371/journal.pone.0205584.g004>



**Fig 5. Conserved water molecules.** Water molecules in conserved positions between strand A and B of the c-MYC promoter crystal structure (only strand A shown as an example), 2Fo-Fc map shown contoured at 1.0  $\sigma$ .

<https://doi.org/10.1371/journal.pone.0205584.g005>

suggests their presence may be important in stabilizing the quadruplex secondary structure. The first water is seen interacting with O4' of residue G6, OP1 of G8 and N2 of G5. The second water is seen in interaction with N2/N3 of residue G8 and O4' of G9. These waters were placed based on the presence of positive density observed in the difference map. More bound waters of this type may be present, however, the high B-factor of the data did not allow for confident placement of additional atoms even at this near-atomic resolution.

### Comparison of the crystal and solution structures of Pu22

The overall topology of the crystal structure is consistent to that reported for the same sequence by solution NMR with several notable differences (Table 2, PDB:1XAV). The most striking difference between the solution and crystal structure is observed in the T1-G2-A3 region at the 5'-head of the quadruplexes (Fig 6). In the NMR structure, these residues lay stacked on top of the G-quartet composed of residues G4, G7, G12, and G17. ([18], PDB: 1XAV). In contrast, in the crystal structure this 5'-head region is observed extended and the top, nearest, G-quartet is the interface where the dimer is formed. The difference in this region in comparison to the NMR structure is also likely responsible for the high RMSD observed in the position of loop residue T7 and differences in the general shape of the phosphate backbone in this region. The 3'-tail region is also extended, albeit to a lesser degree, away from the G-quartet comprised of residues G6-G10-G15-G19 in the crystal structure in comparison to the NMR structure. The difference in the position of loop residue T16 can also likely be attributed to this extended conformation that is involved in crystal packing interactions.

To examine the effect of the crystallization conditions on the conformation of Pu22, we performed circular dichroism spectroscopy at conditions analogous to those used in solution NMR studies [21]. Parallel quadruplex structures display characteristic peaks at approximately positive 265 nm and negative 240 nm in their CD spectra [1]. The parallel topology was confirmed for the Pu22 oligonucleotide under both the conditions used in the solution NMR studies as well as at the crystallization conditions to provide additional evidence the conformation observed in the crystal structure is not an artifact of the crystallization environment (Fig 7).

**Table 2. RMSD of strand A of Pu22 (PDB: 6AU4) to published DNA quadruplex structures.**

RMSD (Å)	c-MYC (NMR, PDB: 1XAV)	c-KIT (X-ray, PDB: 4WO2)	HTelo (X-ray, PDB: 4FXM)
G-quartets*	1.36	6.65	1.03
T1-G2-A3 (5'-head)	4.16	-	-
T20-A21-A22 (3'-tail)	3.45	-	-
T7, T11-A12, T17 (loop)	4.34	-	-
Overall	7.31	10.38	2.78

\*G-quartets are comprised of G4-G5-G6, G8-G9-G10, G13-G14-G15, and G17-G18-G19.

RMSD comparisons were not made between distinct quadruplex structures (c-KIT, telomeric) for the loop, head and tail regions due to significant sequence differences.

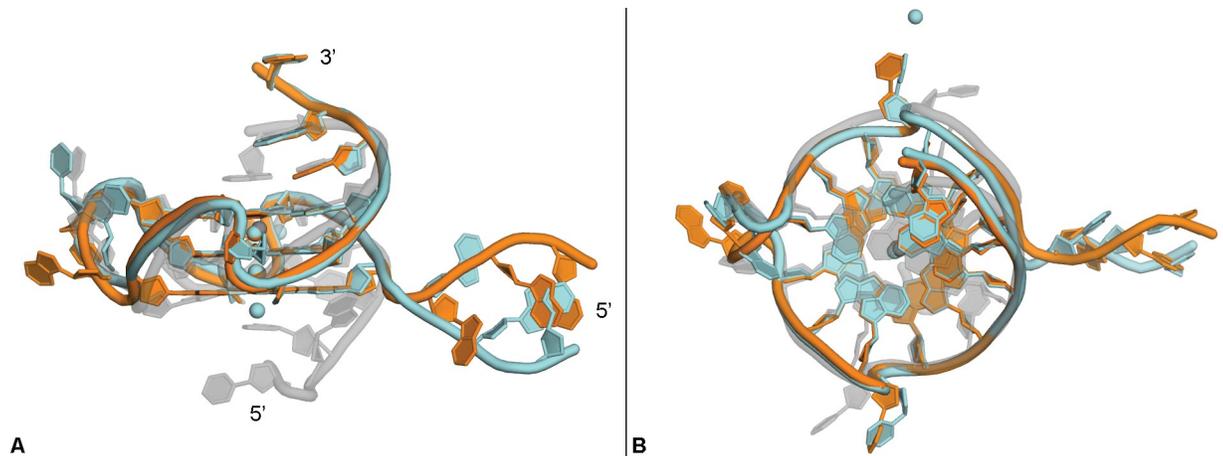
<https://doi.org/10.1371/journal.pone.0205584.t002>

### Comparison to other crystal structures

The human c-KIT promoter region also contains a quadruplex sequence motif, and its crystal structure has some distinct similarities and differences to the c-MYC quadruplex formed by Pu22 (Fig 8A and Table 2, PDB: 4WO2). In both crystal structures, a quadruplex dimer forms the asymmetric unit, with the G-quartets closest to the 5'-region observed in a stacked formation. The c-KIT strands (A & B) forming the dimer are rotated one guanine relative to one another (approximately 90°), whereas the c-MYC strands (A & B) are rotated two guanines relative to one other (approximately 180°). This causes the c-KIT quadruplex loop region nucleotide C9<sub>A</sub> to occupy the same space as loop nucleotide A12<sub>A</sub> of the c-MYC quadruplex, and this is regardless of the additional T11<sub>A</sub> present in this loop region of the c-MYC structure. The loop nucleotide C9<sub>B</sub> of c-KIT superimposes with loop nucleotide T7<sub>B</sub> of the c-MYC quadruplex. The c-MYC quadruplex structure contains a non-channel potassium ion interacting with residues T7 and G9. A non-channel potassium ion is also observed at an unrelated site in the c-KIT crystal structure interacting with residues A16 and G17.

Several crystal structures of the fully parallel quadruplex formed by the human telomeric repeat sequence in complex with a small molecule ligand have been reported, including an example of the intramolecular arrangement at 1.65 Å (Fig 8B, PDB: 4FXM) [43]. The telomeric quadruplex forms a dimer with both strands reported as identical in the asymmetric unit. The telomeric structure is remarkably similar to the Pu22 c-MYC quadruplex with the exception of the loop regions. In the c-MYC structure, these loops are formed by one or two nucleotides, whereas the telomeric structure contains loops made of three nucleotides each. However, even with this difference in loop region length, the c-MYC quadruplex can be superimposed on the telomeric quadruplex with an RMSD of only 2.78 Å (Table 2, PDB: 4FXM). Interestingly, the Watson-Crick packing interaction seen with the T20-A21-A22 tail of two c-MYC quadruplexes is also observed in the crystal structure of the telomeric quadruplex, but between two loop regions containing residues T11-A13 and A1, suggesting that a 3'-terminal TAA sequence may be useful for crystallization of other quadruplexes [43].

During preparation of this manuscript, a 3.8 Å structure was reported of a sequence modified c-MYC promoter quadruplex in complex with the DEAH/RNA helicase DHX36 [44]. In the reported complex structure, it is suggested quadruplex destabilization by the helicase results in a one-residue shift of the nucleotides involved in formation of the three G-quartets, causing the quartet nearest to the 5'-head region to be reformed by G4, G8, A12, and T16. This finding is noteworthy, however, due to the low-resolution of the structure and rearrangement



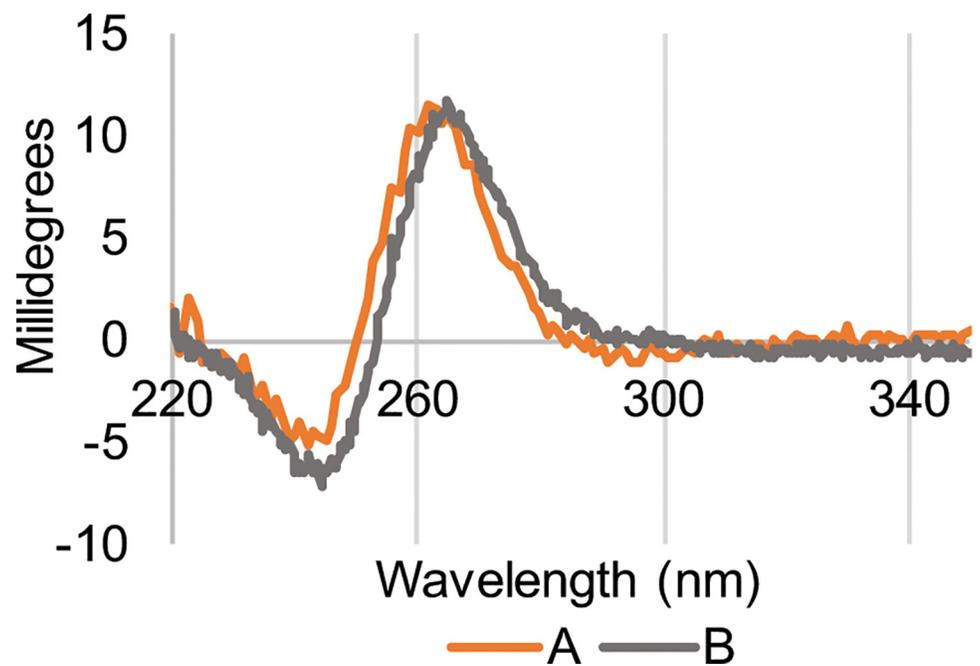
**Fig 6. Comparison to NMR solution structure.** Overlay of NMR solution structure (PDB:1XAV, grey) with strand A (orange) and strand B (cyan) of the c-MYC promoter crystal structure, side view (A), top view (B).

<https://doi.org/10.1371/journal.pone.0205584.g006>

observed in the presence of the helicase, it did not appear relevant for comparison with our crystal structure.

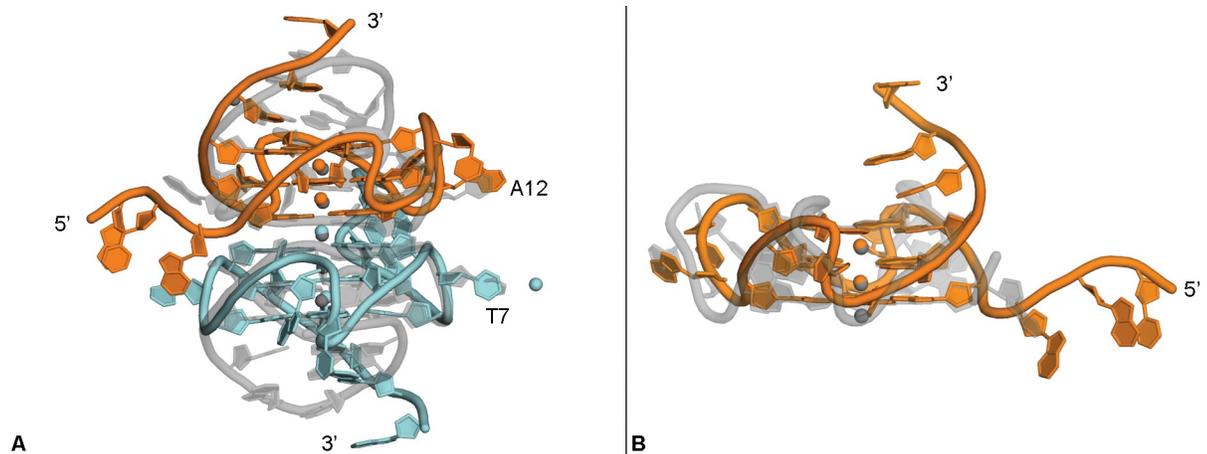
### Discussion

In this manuscript we report a high-resolution crystal structure of the major quadruplex formed in the promoter region of the human c-MYC oncogene. The overall topology is in general agreement with the previously published solution NMR structures, with the major differences occurring predominately in the head and tail regions of the sequence and not in the



**Fig 7. Circular dichroism spectroscopy.** CD spectra of the Pu22 oligonucleotide under crystallization conditions (A) or conditions previously reported to form the parallel c-MYC promoter quadruplex (B).

<https://doi.org/10.1371/journal.pone.0205584.g007>



**Fig 8. Comparison to other quadruplex crystal structures.** Overlay of c-KIT crystal structure (PDB:4WO2, grey) with strand A (orange) and strand B (cyan) of the c-MYC promoter crystal structure; loop residue labels correspond to c-MYC crystal structure (A). Overlay of human telomeric crystal structure (PDB: 4FXM, grey, ligand not shown) with strand A (orange) of the c-MYC promoter crystal structure (B).

<https://doi.org/10.1371/journal.pone.0205584.g008>

central G-quartet structure [18,21]. This is important as it strengthens the validity of previous and ongoing computational and synthetic studies that have used the NMR solution structures as a guide.

The more exposed G-quartets observed in the crystal structure, in comparison to those reported using solution NMR, confirm both the flexibility of the head and tail regions of the promoter sequence and the rigid nature of the G-quartets in this quadruplex. This could be biologically relevant as the flanking regions would not be terminal residues in genomic DNA and are likely able to project away from the G-quartets as is observed in the crystal structure (Fig 2). This will be important in future studies with the goal of developing small molecule ligands to bind the c-MYC quadruplex. For example, in the solution NMR study of a small molecule interaction with the c-MYC promoter quadruplex, it was observed that part of the ligand binding pocket involved nucleotides in the 5'-head (A3) and 3'-tail regions (T20, A21) (PDB: 2L7V, [21]). It is possible, as suggested by the c-MYC quadruplex crystal structure, that these residues are not in close proximity to provide such interactions in the biological setting as illustrated schematically in Fig 1. In this scenario, the reported "induced fit" mode of ligand binding would require major structural rearrangement and be less energetically favorable. Therefore, an alternative strategy would be to design ligands with a binding mode that takes advantage of the more linearized conformation of the head and tail regions, potentially affording tighter binding and selectivity for the c-MYC quadruplex. This alternative binding hypothesis will be a crucial consideration in the future design of c-MYC quadruplex-targeted small molecule ligands.

The packing interactions observed in the crystal highlight the importance of the loop, head and tail regions of the sequence to allow for crystal formation and will be generally helpful in optimizing quadruplex sequences for crystallization. The Watson-Crick bonding observed at the 3'-tail of the Pu22 sequence also suggests a preference for nucleotides not involved in the G-quartet core region to follow the normal B-DNA base pairing paradigm even in very close proximity to the G-quartets. However, there is previous evidence that these flanking regions may remain single-stranded under conditions of negative supercoiling [17].

Comparison of the c-MYC quadruplex crystal structure to the c-KIT quadruplex and the human telomeric quadruplex show the majority of differences occur in the head, tail and loop regions. Differences in the length of loop regions and position of the loop nucleotides will

provide useful tools for designing ligands that would be selective for specific quadruplex structures as has been suggested previously [9,45]. Overall, this work has provided the basis for ongoing co-crystallization studies with the Pu22 c-MYC promoter quadruplex and our set of novel quadruplex ligands, the anthracenyl isoxazole amides (AIMs) [33,46]. These findings will also be of benefit to other researchers in their efforts to target the c-MYC promoter and other therapeutically relevant quadruplex targets.

## Supporting information

**S1 Fig. G-quartet potassium ions.** Example of square antiprismatic coordinated potassium ions central to two stacked G-quartets in the crystal structure.

(TIF)

**S1 File. 6AU4 PDB validation report.**

(PDF)

**S2 File. 6AU4 molecular coordinates.**

(PDB)

**S3 File. 6AU4 x-ray data.**

(MTZ)

## Acknowledgments

We would like to acknowledge Dr. Bruce Bowler and Dr. Levi McClelland for their assistance with CD spectroscopy. We would also like to acknowledge Dr. Clyde Smith and Dr. Enrique Rudiño Piñera for their assistance with the data collection at SSRL.

Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

## Author Contributions

**Conceptualization:** Sascha Stump, Howard D. Beall.

**Data curation:** Sascha Stump, Tung-Chung Mou.

**Formal analysis:** Sascha Stump, Tung-Chung Mou.

**Funding acquisition:** Howard D. Beall.

**Investigation:** Sascha Stump.

**Methodology:** Tung-Chung Mou.

**Project administration:** Sascha Stump, Tung-Chung Mou, Stephen R. Sprang, Nicholas R. Natale, Howard D. Beall.

**Resources:** Stephen R. Sprang, Nicholas R. Natale, Howard D. Beall.

**Supervision:** Tung-Chung Mou, Stephen R. Sprang, Nicholas R. Natale, Howard D. Beall.

**Validation:** Sascha Stump, Tung-Chung Mou, Stephen R. Sprang, Howard D. Beall.

**Visualization:** Sascha Stump, Tung-Chung Mou.

**Writing – original draft:** Sascha Stump.

**Writing – review & editing:** Sascha Stump, Tung-Chung Mou, Stephen R. Sprang, Nicholas R. Natale, Howard D. Beall.

## References

1. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* 2006; 34: 5402–5415. <https://doi.org/10.1093/nar/gkl655> PMID: 17012276
2. Dang CV. MYC on the Path to Cancer. *Cell.* 2012; 149: 22–35. <https://doi.org/10.1016/j.cell.2012.03.003> PMID: 22464321
3. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. 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.* 2002; 99: 11593–11598. <https://doi.org/10.1073/pnas.182256799> PMID: 12195017
4. Ou T-M, Lu Y-J, Zhang C, Huang Z-S, Wang X-D, Tan J-H, et al. Stabilization of G-Quadruplex DNA and Down-Regulation of Oncogene c-myc by Quindoline Derivatives. *J Med Chem.* 2007; 50: 1465–1474. <https://doi.org/10.1021/jm0610088> PMID: 17346034
5. Su L, Zheng H, Li Z, Qiu J, Chen S, Liu J, et al. Mechanistic studies on the anticancer activity of 2,4-disubstituted quinazoline derivative. *Biochim Biophys Acta BBA—Gen Subj.* <https://doi.org/10.1016/j.bbagen.2014.07.004> PMID: 25018006
6. Brown RV, Danford FL, Gokhale V, Hurley LH, Brooks TA. Demonstration that Drug-targeted Down-regulation of MYC in Non-Hodgkins Lymphoma Is Directly Mediated through the Promoter G-quadruplex. *J Biol Chem.* 2011; 286: 41018–41027. <https://doi.org/10.1074/jbc.M111.274720> PMID: 21956115
7. Phan AT, Modi YS, Patel DJ. Propeller-Type Parallel-Stranded G-Quadruplexes in the Human c-myc Promoter. *J Am Chem Soc.* 2004; 126: 8710–8716. <https://doi.org/10.1021/ja048805k> PMID: 15250723
8. Dai J, Carver M, Hurley LH, Yang D. Solution Structure of a 2:1 Quindoline–c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *J Am Chem Soc.* 2011; 133: 17673–17680. <https://doi.org/10.1021/ja205646q> PMID: 21967482
9. Hu M-H, Wang Y-Q, Yu Z-Y, Hu L-N, Ou T-M, Chen S-B, et al. Discovery of a New Four-Leaf Clover-Like Ligand as a Potent c-MYC Transcription Inhibitor Specifically Targeting the Promoter G-Quadruplex. *J Med Chem.* 2018; <https://doi.org/10.1021/acs.jmedchem.7b01697> PMID: 29474069
10. Bhattacharyya D, Mirihana Arachchilage G, Basu S. Metal Cations in G-Quadruplex Folding and Stability. *Front Chem.* 2016; 4. <https://doi.org/10.3389/fchem.2016.00038> PMID: 27668212
11. Fernando H, Reszka AP, Huppert J, Ladame S, Rankin S, Venkitaraman AR, et al. A Conserved Quadruplex Motif Located in a Transcription Activation Site of the Human c-kit Oncogene. *Biochemistry (Mosc).* 2006; 45: 7854–7860. <https://doi.org/10.1021/bi0601510> PMID: 16784237
12. Rankin S, Reszka AP, Huppert J, Zloh M, Parkinson GN, Todd AK, et al. Putative DNA Quadruplex Formation within the Human c-kit Oncogene. *J Am Chem Soc.* 2005; 127: 10584–10589. <https://doi.org/10.1021/ja050823u> PMID: 16045346
13. Dexheimer TS, Sun D, Hurley LH. Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the bcl-2 P1 Promoter. *J Am Chem Soc.* 2006; 128: 5404–5415. <https://doi.org/10.1021/ja0563861> PMID: 16620112
14. Agrawal P, Lin C, Mathad RI, Carver M, Yang D. The Major G-Quadruplex Formed in the Human BCL-2 Proximal Promoter Adopts a Parallel Structure with a 13-nt Loop in K<sup>+</sup> Solution. *J Am Chem Soc.* 2014; 136: 1750–1753. <https://doi.org/10.1021/ja4118945> PMID: 24450880
15. Sun D, Guo K, Rusche JJ, Hurley LH. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* 2005; 33: 6070–6080. <https://doi.org/10.1093/nar/gki917> PMID: 16239639
16. De Armond R, Wood S, Sun D, Hurley LH, Ebbinghaus SW. Evidence for the Presence of a Guanine Quadruplex Forming Region within a Polypurine Tract of the Hypoxia Inducible Factor 1 $\alpha$  Promoter. *Biochemistry (Mosc).* 2005; 44: 16341–16350. <https://doi.org/10.1021/bi051618u> PMID: 16331995
17. Sun D, Hurley LH. The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J Med Chem.* 2009; 52: 2863–2874. <https://doi.org/10.1021/jm900055s> PMID: 19385599

18. Ambrus A, Chen D, Dai J, Jones RA, Yang D. Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human c-MYC Promoter. Implications for G-Quadruplex Stabilization. *Biochemistry (Mosc)*. 2005; 44: 2048–2058. <https://doi.org/10.1021/bi048242p> PMID: 15697230
19. Brooks TA, Hurley LH. The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat Rev Cancer*. 2009; 9: 849–861. <https://doi.org/10.1038/nrc2733> PMID: 19907434
20. Raiber E-A, Kranaster R, Lam E, Nikan M, Balasubramanian S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res*. 2012; 40: 1499–1508. <https://doi.org/10.1093/nar/gkr882> PMID: 22021377
21. Dai J, Carver M, Hurley LH, Yang D. Solution structure of a 2:1 quindoline-c-MYC G-quadruplex: insights into G-quadruplex-interactive small molecule drug design. *J Am Chem Soc*. 2011; 133: 17673–17680. <https://doi.org/10.1021/ja205646q> PMID: 21967482
22. Wang Y, Patel DJ. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*. 1993; 1: 263–282. [https://doi.org/10.1016/0969-2126\(93\)90015-9](https://doi.org/10.1016/0969-2126(93)90015-9) PMID: 8081740
23. Luu KN, Phan AT, Kuryavyi V, Lacroix L, Patel DJ. Structure of the Human Telomere in K+ Solution: An Intramolecular (3 + 1) G-Quadruplex Scaffold. *J Am Chem Soc*. 2006; 128: 9963–9970. <https://doi.org/10.1021/ja062791w> PMID: 16866556
24. Phan AT, Kuryavyi V, Luu KN, Patel DJ. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K+ solution. *Nucleic Acids Res*. 2007; 35: 6517–6525. <https://doi.org/10.1093/nar/gkm706> PMID: 17895279
25. Parkinson GN, Lee MPH, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*. 2002; 417: 876–880. <https://doi.org/10.1038/nature755> PMID: 12050675
26. Xue Y, Kan Z, Wang Q, Yao Y, Liu J, Hao Y, et al. Human Telomeric DNA Forms Parallel-Stranded Intramolecular G-Quadruplex in K+ Solution under Molecular Crowding Condition. *J Am Chem Soc*. 2007; 129: 11185–11191. <https://doi.org/10.1021/ja0730462> PMID: 17705383
27. Heddi B, Phan AT. Structure of Human Telomeric DNA in Crowded Solution. *J Am Chem Soc*. 2011; 133: 9824–9833. <https://doi.org/10.1021/ja200786q> PMID: 21548653
28. Phan AT, Kuryavyi V, Burge S, Neidle S, Patel DJ. Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-kit Promoter. *J Am Chem Soc*. 2007; 129: 4386–4392. <https://doi.org/10.1021/ja068739h> PMID: 17362008
29. Wei D, Parkinson GN, Reszka AP, Neidle S. 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*. 2012; 40: 4691–4700. <https://doi.org/10.1093/nar/gks023> PMID: 22287624
30. Wei D, Husby J, Neidle S. Flexibility and structural conservation in a c-KIT G-quadruplex. *Nucleic Acids Res*. 2015; 43: 629–644. <https://doi.org/10.1093/nar/gku1282> PMID: 25452341
31. Balasubramanian S, Hurley LH, Neidle S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat Rev Drug Discov*. 2011; 10: 261–275. <https://doi.org/10.1038/nrd3428> PMID: 21455236
32. Yang D, Hurley LH. Structure of the Biologically Relevant G-Quadruplex in The c-MYC Promoter. *Nucleosides Nucleotides Nucleic Acids*. 2006; 25: 951–968. <https://doi.org/10.1080/15257770600809913> PMID: 16901825
33. Weaver MJ, Kearns AK, Stump S, Li C, Gajewski MP, Rider KC, et al. AIMing towards improved antitumor efficacy. *Bioorg Med Chem Lett*. 2015; 25: 1765–1770. <https://doi.org/10.1016/j.bmcl.2015.02.063> PMID: 25782743
34. Campbell NH, Parkinson GN. Crystallographic studies of quadruplex nucleic acids. *Methods*. 2007; 43: 252–263. <https://doi.org/10.1016/j.ymeth.2007.08.005> PMID: 17967696
35. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr*. 2010; 66: 125–132. <https://doi.org/10.1107/S0907444909047337> PMID: 20124692
36. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010; 66: 213–221. <https://doi.org/10.1107/S0907444909052925> PMID: 20124702
37. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of *Coot*. *Acta Crystallogr D Biol Crystallogr*. 2010; 66: 486–501. <https://doi.org/10.1107/S0907444910007493> PMID: 20383002
38. Joosten RP, Long F, Murshudov GN, Perrakis A. The *PDB\_REDO* server for macromolecular structure model optimization. *IUCrJ*. 2014; 1: 213–220. <https://doi.org/10.1107/S2052252514009324> PMID: 25075342
39. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000; 28: 235–242. <https://doi.org/10.1093/nar/28.1.235> PMID: 10592235

40. Schrodinger. The PyMOL Molecular Graphics System, Version 1.8. 2015.
41. Karplus PA, Diederichs K. Linking crystallographic model and data quality. *Science*. 2012; 336: 1030–1033. <https://doi.org/10.1126/science.1218231> PMID: 22628654
42. Paramasivan S, Rujan I, Bolton PH. Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods*. 2007; 43: 324–331. <https://doi.org/10.1016/j.ymeth.2007.02.009> PMID: 17967702
43. Nicoludis JM, Miller ST, Jeffrey PD, Barrett SP, Rablen PR, Lawton TJ, et al. Optimized End-Stacking Provides Specificity of N-Methyl Mesoporphyrin IX for Human Telomeric G-Quadruplex DNA. *J Am Chem Soc*. 2012; 134: 20446–20456. <https://doi.org/10.1021/ja3088746> PMID: 23181361
44. Chen MC, Tippana R, Demeshkina NA, Murat P, Balasubramanian S, Myong S, et al. Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*. 2018; 558: 465–469. <https://doi.org/10.1038/s41586-018-0209-9> PMID: 29899445
45. Neidle S. Quadruplex nucleic acids as targets for anticancer therapeutics. *Nat Rev Chem*. 2017; 1: 0041.
46. Han X, Li C, Mosher MD, Rider KC, Zhou P, Crawford RL, et al. Design, synthesis and biological evaluation of a novel class of anticancer agents: Anthracenylisoxazole lexitropsin conjugates. *Bioorg Med Chem*. 2009; 17: 1671–1680. <https://doi.org/10.1016/j.bmc.2008.12.056> PMID: 19167892