

MORE THAN FORTY YEARS OF NUCLEIC ACID STRUCTURAL SCIENCE

Gary N. Parkinson ^a, Helen Berman ^b

Gary N. Parkinson

gary.parkinson@ucl.ac.uk

Helen M. Berman

berman@rcsb.rutgers.edu

^{a*} Department of Pharmaceutical and Biological Chemistry, University College London School of Pharmacy, London, WC1N 1AX, UK.

^b Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey, 08854 USA.

Abstract

Keywords

DNA drug, Quadruplex, Hydration, Minor groove binder, Intercalation

1. Introduction

The authors of this review are fortunate to have collaborated with Stephen Neidle at various stages of his career. Helen Berman met Stephen Neidle in the 1970s at a meeting when they were both young faculty members. She was at the Institute for Cancer Research in Philadelphia, and he was in the Biophysics Department at King's College London - the place where the data for the double helix was collected by Rosalind Franklin and Maurice Wilkins¹ and the place where Struther Arnott determined very high-quality structures of nucleic acid fibers.² At that time in her career, Berman had a strong interest in nucleic acids having worked on a small fragment of RNA with colleagues at MIT and Columbia.³ She was delighted when Stephen proposed to visit her laboratory to work on a nucleic acid structure. This visit marked the beginning of an exciting collaboration on nucleic acid-drug complexes that lasted for many years.

Gary Parkinson became aware of Stephen Neidle's work while he was training in Berman's laboratory at Rutgers University. When he interviewed for a post in Neidle's group, he was excited to meet a well-established research team, fully resourced with the latest x-ray crystallography facilities, revealing the importance that Neidle placed on the use of structural data to support the development of small molecule molecular entities within drug discovery. Thus, began a long and fruitful collaboration on quadruplex DNA/RNA structures.

In order to understand the fundamental processes in which nucleic acids were involved including replication, transcription, and translation, it is necessary to know the structures. In the 1970's it became possible to synthesize defined sequences of DNA and RNA and thus be able to determine their structures at near atomic resolution. When Neidle entered the field in the 1970's, he had a particularly strong interest in how drugs interact with nucleic acids and how they disrupt the processes in which they are involved. From the very beginning Stephen focused his research in this area. Over the forty years of his research his group determined 119 structures, most of which are duplex and quadruplexes DNA's, with more than 60% of these structures bound to drugs. By every standard Stephen Neidle is a leader in the field and is highly cited with two papers on quadruplexes having more than 2000 citations each.

In this review, we try to capture the essence of Stephen Neidle's many contributions to nucleic acid structural science, as summarized in Fig. 1. We focus on exemplars (Table S1) of his broad ranging approach to understanding nucleic acid drug interactions as a method to develop new therapeutic agents.

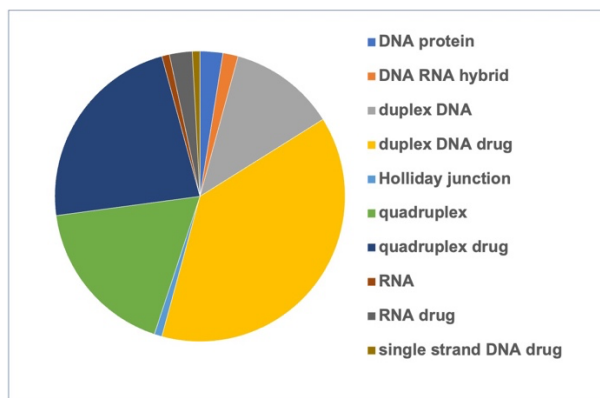


Fig. 1. Pie chart showing diversity of nucleic acid containing structures determined by the Neidle's group. The chart represents 118 PDB entries.

2. Double helical DNA-drug complexes

In the 1970s Berman worked with Neidle on small fragments of nucleic acids that were complexed with proflavines (Fig. 2). The structure of a complex between CpG and proflavine (CG_PF) presented large technical challenges.⁴ Because standard structure determination methods did not work, they used high-resolution data to determine the positions of the phosphorus atoms, as had been done in the determination of the structure of UpA.³ To their great surprise, the C3' endo – C2' endo patterns of sugar pucker that had been seen in other similar complexes were not observed. Instead, both sugars exhibited the C3' endo conformation. In the structure of the complex between dCpG and proflavine (dCG_PF) one chain had mixed sugar pucker and the other had all C3' endo pucker. Based on these structures, they demonstrated that upon intercalation, the conformation angles around the C5'-O5' bond and the glycosidic bond at the 3' end of DNA increase. These structures also demonstrated the versatility of proflavine interactions. In addition, the dCG_PF structure showed an intriguing pattern of water structures that had never been seen before^{5,6} (Fig. 3). It contained interlocking pentagons with a semi clathrate arrangement. This structure became the basis for Monte Carlo simulations to further understand what was at the time an unusual hydration pattern.⁷ The importance of hydration in drug nucleic acid complexes has subsequently been verified in other structures.

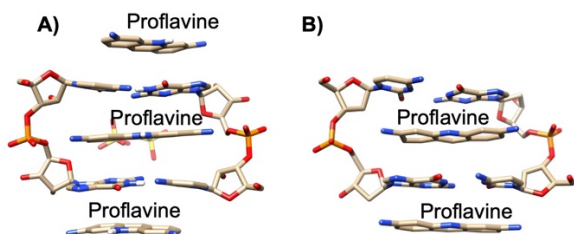


Fig. 2. A) The crystal structure CpG proflavine.⁴ One proflavine is intercalated and two are stacked above and below the base pairs. There are also two sulphate groups bound to the intercalated proflavine. **B)** The crystal structure of dCpG proflavine. One proflavine is intercalated and one is stacked.

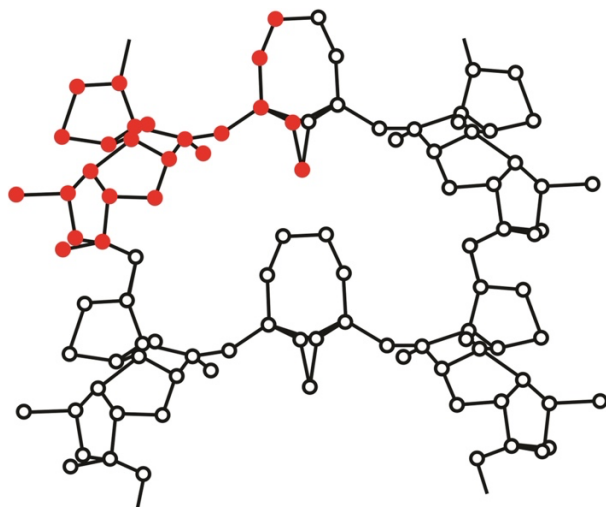


Fig. 3. The extended water network in the crystal structure of dCpG proflavine.⁵ The unique waters in the crystal are shown in red. This is the earliest example of pentagon water networks in a nucleic acid structure.

The Neidle laboratory has produced an extensive body of research on longer DNA sequences alone and complexed with drugs (Fig. 4). For example, in studies of how drugs bind to the minor groove, they examined how the anti-trypanosomal agent berenil, propamidine an analog of pentaamine which has anti HIV activity, and Hoechst 33258 whose analogs have been investigated as possible anti cancer agents, bind to two different sequences of DNA: dCGCGAATTCGCG (A2T2) and dCGCAAATTTTCGC (A3T3). The structures of these sequences bound to berenil,^{8, 9} propamidine¹⁰ and Hoechst 33258¹¹ were determined. The fine details of the DNA structures were examined as were the interactions of the drugs. In each case, the drugs bind differently to the different sequences. For example, berenil binds asymmetrically to A2T2 with water mediating interactions (Fig. 5a), whereas berenil binds symmetrically with A3T3 (Fig. 5b). In the case of the propamidine complexes the binding to A3T3 is asymmetric (Fig. 6a) and that to A2T2 is symmetric (Fig. 6b). Similarly, the binding of Hoechst 33258 to A2T2 is symmetric and to A3T3 asymmetric.¹² These studies support the concept that the DNA is not altered by the drug.

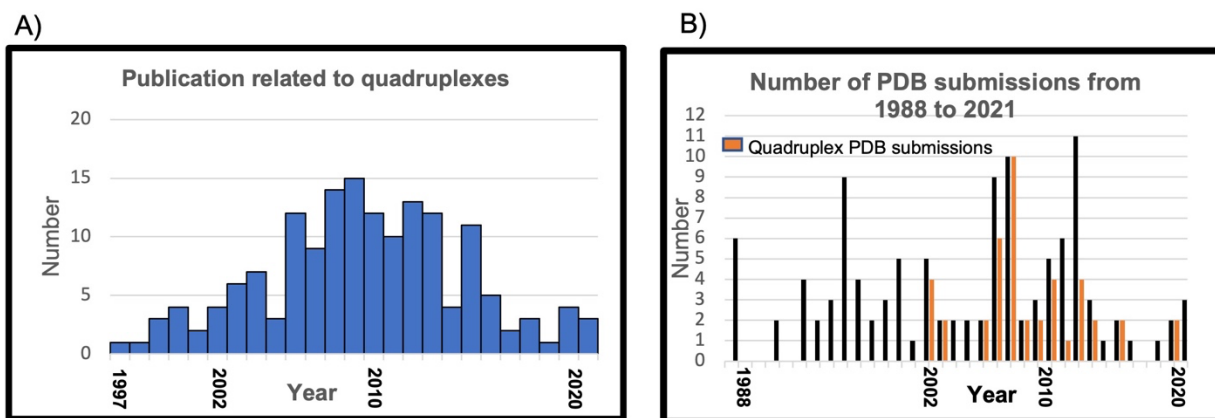


Fig. 4. A) Publications reporting research undertaken involving the Neidle group related to G4Qs since 1997. **B)** Individual structures deposited at the PDB involving the Neidle group since 1988. The black bars represent all structures submitted and the orange bars represent G4 quadruplex containing structures.

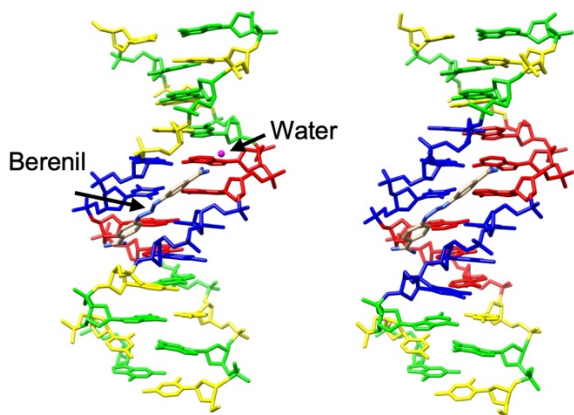


Fig. 5. A) Berenil bound to the CGCGAATTCGCG (A2T2) sequence (2DBE).⁸ A water molecule (shown in magenta) mediates the binding. **B)** Berenil bound to CGCAAATTCGCG (A3T3) (1D63).⁹ In this case the berenil is bound symmetrically in the minor groove and there is no water mediating the binding.

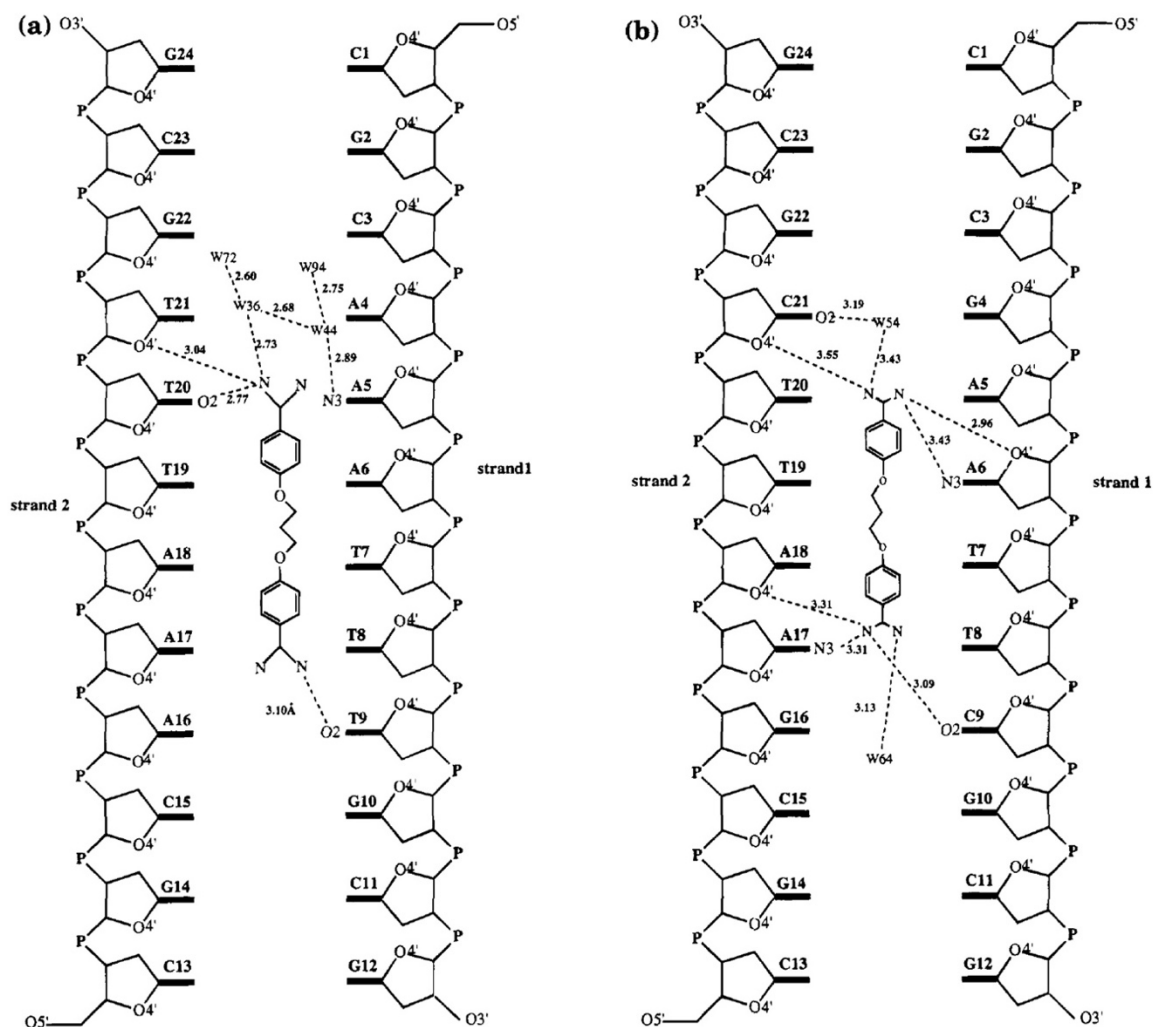


Fig. 6. Schematic diagrams of propamidine binding to two different dodecamer DNA sequences. Figure reprinted with permission from reference.¹⁰ Copyright 1995 American Chemical Society. **A)** Propamidine binding A3T3 DNA sequences (102D). **B)** Propamidine binding to A2T2 (1PRP).

To understand how heterocyclic diamines interact with DNA, a series of related drugs were investigated. One, DB 884, (2GYX)¹³ showed particularly strong binding to DNA. Extensive biophysical studies including calorimetry, SPR biosensor data, foot printing analysis, and CD - spectrometry were combined with x-ray crystallographic studies to explain the very strong binding. It was shown that the drug covers six base pairs with the central pyrrole of the drug forming an unprecedented central hydrogen bond that explains the very favorable binding enthalpy.

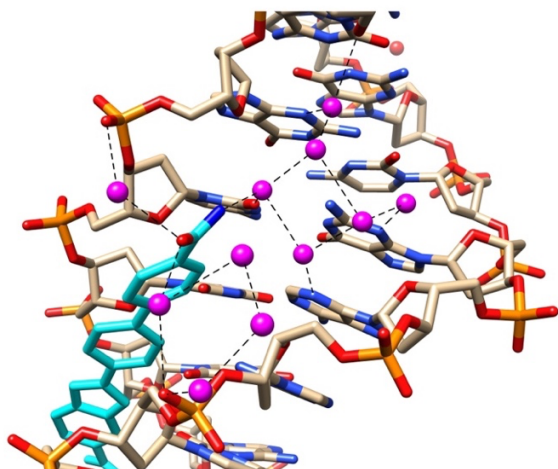


Fig. 7. Minor groove drug binding of DB 1963 (cyan) to DNA (3U08)¹⁴ showing an extensive water network that helps stabilize the complex. The conserved water molecules are shown in magenta. The dotted lines represent hydrogen bonds.

In another remarkable set of studies of three minor groove binders to the A2T2 sequence (3U08, 3U0U, 3U05),¹⁴ an extensive water network at the boundary of the AATT sequence and the G nucleotide (Fig. 7). It is posited that water adds the extra stabilization for the drug interaction. This structure comes full circle with the results first seen in the 80's in the dCG_PF complex which also demonstrated extensive networks of water.

In a more recent work, designed to understand GG mismatches, complexes were made of TTGGCGAA and two drugs, chromomycin and actinomycin.¹⁵ The chromomycin-containing complex shows the diversity of GG mismatches; in the asymmetric unit three helices form a pseudo continuous double helix with each helix incorporating a different GG mismatch (Fig. 8). The actinomycin-bound DNA complex forms a zig zag right-handed helix with G's flipping out. These studies highlight the versatility of guanine interactions as had been so well exemplified by the groundbreaking quadruplex research.

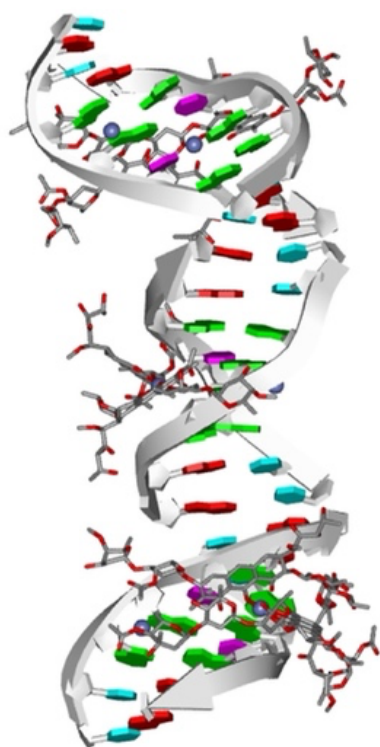


Fig. 8. The versatility of GG mispairs in a complex between chromomycin and DNA (6J0I).¹⁵ The asymmetric unit of the crystals contains 3 DNA duplexes forming a pseudo continuous double helix.

3. Quadruplexes (G4Q) and their drug complexes

By the mid 1990s the Neidle group had synthesized an extensive library of ligands that targeted a variety of DNA motifs. This was to be expanded when a family of amidoanthraquinone derivatives was shown as the first non-nucleoside telomerase inhibitors of G4 quadruplexes (G4Q), where the anthraquinone stack and stabilize the G4Q.¹⁶ The integrated approach taken by an international multidisciplinary team that Stephen Neidle had help to assemble, included molecular modelling, x-ray-derived data analysis on relevant telomeric sequences d[TG_4T], 1-D NMR data on human telomeric sequences d[$\text{T}_2\text{AG}_3\text{T}$], biophysical thermal denaturation experiments and in-vitro assays. This approach put the Neidle group at the forefront of G4Q targeted structure-based drug discovery (SBDD) efforts for over 20 years (Fig. 4).

The late 90s was an opportune time to advance G4Q research. G4Q was validated as a new molecular target within telomeres and provided a novel therapeutic pathway to not only inhibit telomerase function but also interfere with chromosomal maintenance. While NMR methods had provided important structural data on telomeric G4Qs for structure-based drug design including the core human telomere sequence, it did not deliver the full picture of the diversity of secondary structures available to telomeric DNA and RNA. The structural determination of a parallel-stranded G4Q (139D)¹⁷ folded as an intermolecular quadruplex confirmed the importance of the

G-tetrad core in providing the necessary conformational stability while retaining the characteristic grooves and possible hydration structures seen in duplex DNA. Additionally, it highlighted the importance role of the K^+ stabilizing the central-tetrad core. The human telomeric $d(AG_3(T_2AG_3)_3)$ sequence folded in the presence of Na^+ revealed the TTA connecting loops in an intramolecular G4Q (143D)¹⁸ context, and revealed additional hints towards the key molecular pharmacophores needed for the rational design of small molecular entities targeting these sequences. The antiparallel arrangement of the DNA backbone results in a mix of wide and narrow groove widths that allow discrete hydration structures and opportunities for design of selective binders. However, these structures did not provide any clue towards possible alternative folded topologies that would ultimately be needed to fully explain the quantitative structure-activity relationships (QSAR) ligand data. There was a gap between the molecular modeling data based on the available NMR structures, and the data derived from the ligand libraries that specifically targeted human telomeric sequences. It was not until new crystallographic models became available in 2002 that this would be understood. Thus, Neidle and his collaborators relied on modeling using NMR-derived models and data, along with hit identification, assay development, high-throughput screening, and QSAR. In the meantime, the goal of obtaining structural data of the molecular target with ligands bound continued in the Neidle's group with the characterization of the structure of the ligand 1,4-bis-piperidino amidoanthraquinone, a molecule designed to target non-duplex nucleic acids here in complex with telomeric G4Q putative forming sequence. The ligand was observed bound to a G4Q in which four independent strands folded as an intermolecular parallel stranded quadruplex $d[(TG_4T)]_4$.¹⁹ However, these crystals did not provide diffraction data quality required for a full understanding of the molecular interactions at atomic resolution. In addition, this sequence lacked the all-important connecting loops, crucial within the context of the human telomeric sequence $d(TTAGGG)$.

When the Neidle group completed its move to the Chester Beatty Laboratories at the new ICR site, Fulham Road in 2000, they expanded their crystallographic investigations in support of drug discovery efforts. Newly developed ligand scaffolds were explored, and large libraries were generated that provided a wealth of new molecular entities to target telomeres; these included 2,6-disubstituted amidoanthracene-9,10-dione²⁰ and 3,6-disubstituted acridine chromophores. Perhaps the most significant advance was the generation of a family of 3,6,9-trisubstituted acridine ligands where the substitution at the 9 position provided the necessary selectivity over duplex DNA.²¹ The structure-activity relationships of these and other molecules are summarized in several important papers²²⁻²⁴.

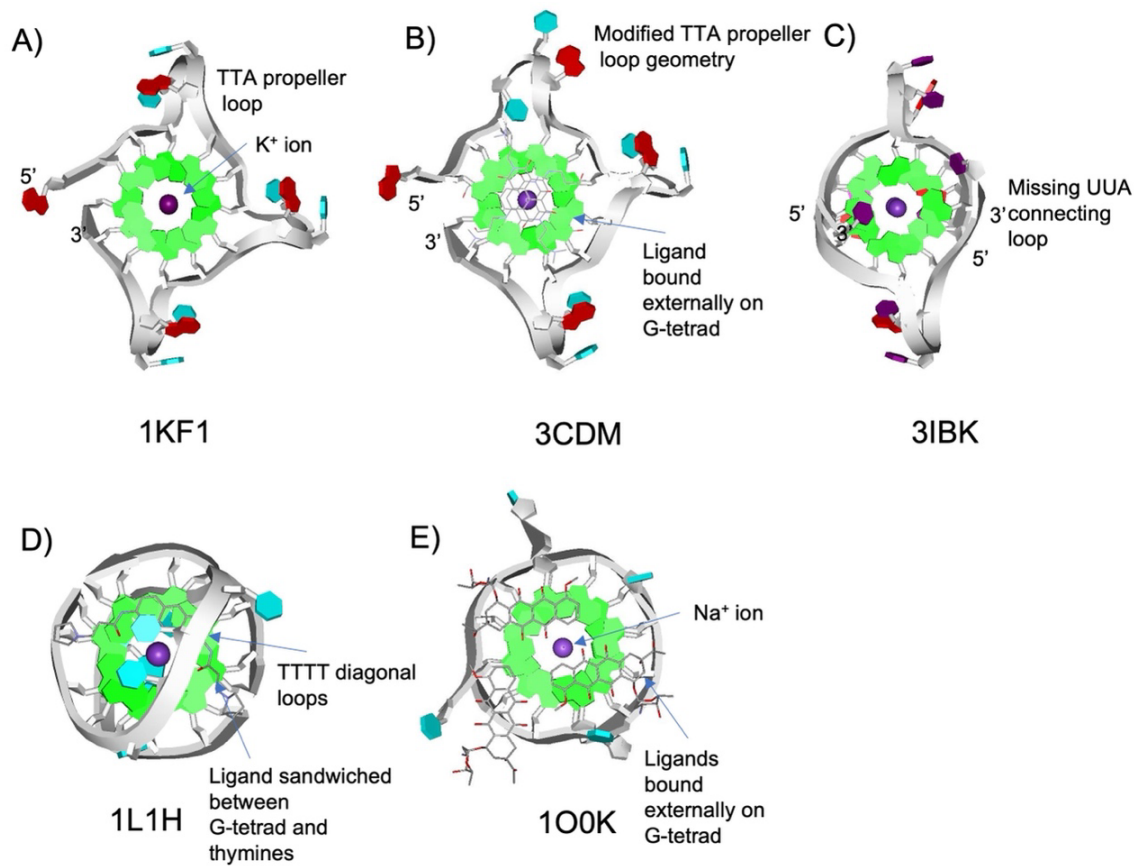


Fig. 9. Crystal structures of G4Qs determined in the Neidle Lab. Schematic DNA and RNA with ligands drawn as stick representations (ViewerLite).²⁵ **A)** An intramolecular G4Q formed from one strand containing four human telomeric repeats d(TTAGGG) and folded into three stacked tetrads with connecting chain-reversal loops. **B)** An intermolecular G4Q bound to a tetra-substituted naphthalene diamide quadruplex binding ligand. Ligands are able to stack on top of the external 5' or 3' ends tetrads without modifying the overall topology as the connecting chain-reversal loops sit external to the G4Q. **C)** A bimolecular G4Q formed from two RNA strands containing a human sequence containing two r(UUAGGG) (TERRA) repeats resulting in only two UUA loops. **D)** An intermolecular G4Q formed from *Oxytricha* telomeric DNA containing two d(TTTTGGGG) repeats folded into four stacked-tetrads with diagonal loops bound to a di-substituted aminoalkylamido acridine compound. The diagonal loops are opened-up in the presence of the di-substituted ligands, allowing the insertion of the ligands and the stacking of the acridine core over the G-tetrad similar to intercalation observed with dsDNA. **E)** The *Oxytricha* telomeric DNA bound to three daunomycins. The absence of connecting backbone loops and their steric constraints provides additional opportunities for ligand binding through π stacking over the G-tetrad core.

By 2002, the ground-breaking publication of two quadruplex crystal structures were reported revealing an unexpected and completely novel G4Q topology (1K8P, 1KF1).²⁶ The structures annealed in the presence of K⁺ were completely different from the NMR derived models annealed in presence of Na⁺ that had been used as the templates for modeling ligand interactions and provided the first glimpse of a G4Q of human telomeric sequence in a crystalline lattice. The

topology of these structures was unexpected in that they folded in an all-parallel arrangement that necessitates connecting chain-reversal loops and with three equal groove widths. The three parallel stacked G-tetrads and the novel chain-reversal loop arrangements (first described as propeller loops) answered many of the questions raised by the previously determined structural activity relationships (Fig. 9). Although the structure determinations provided the basis for SBDD, a model that included ligands was the true goal. A renewed effort was undertaken to generate crystallography-derived structural models of the interactions of ligands to the DNA G4Qs. The Neidle group reported on two *Oxytricha* telomeric DNA quadruplex structures (1JRN, 1JPQ)²⁷ that differed from the structures derived by x-ray methods using similar sequences. The crystallographic structural arrangements and topologies of the *Oxytricha* telomeric DNA were further validated by the first crystal structure of the quadruplex DNA-drug complex (1L1H)²⁸. In this structure, the diagonal loops are opened-up and the ligands are accommodated in an end-stacked conformation not easily predicted by molecular modeling. It became apparent that the plasticity of the connecting loops is an important element in ligand interactions and in controlling groove dimensions. The next ligand G4Q complex confirmed the terminal G-tetrads as key ligand binding sites for the polyaromatic core of the ligands but the sequence lacked the important connecting loops. The ligand, daunomycin is observed stacked above the unconstrained terminal G-tetrads formed from four parallel strands (1O0K)²⁹. Within the Neidle group the focus remained on understanding the connecting loops and a structural analysis was undertaken on the consequence of changing loop lengths between stacked G-tetrads. It was observed that the sequences (G₄T₃G₄) folded as bimolecular G4Qs with antiparallel strands in this case linked with lateral loops rather than diagonal loop (2AVJ, 2AVH).³⁰

The next breakthrough was the determination of the structure of a ligand, TMPyP4 bound to G4Q containing the human telomeric sequence (2HRI).³¹ The observation of the sequence folded in an all-parallel arrangement confirmed this crystallographically-determined topology as an important molecular target. This intermolecular all-parallel arrangement of the telomeric DNA sequence is stabilized in the presence of K⁺ in contrast to the NMR determined basket-type intramolecular G-quadruplex topology, first shown to form in the presence of Na⁺ ions of the same sequence¹⁸. At this time the topological diversity of NMR derived G4Q structures containing human telomeric sequences was becoming apparent with the publications of the hybrid-1 topology folded here in the presence of K⁺³² and the hybrid-2 topology.³³ In these NMR experiments the topological arrangements of the G4Qs could be selected for based on their sequence length and on the terminal sequences and on the selection of metal cations used to stabilize the G4Qs. This is different from single all-parallel G4Q topology as determined by crystallographic methods, observed alone or in the presence of ligands. The wealth of crystallographic data being generated by the Neidle group at this time affirmed the group's position in the field and established this alternative folded topology for the human telomeric sequence as a viable model for the design of highly selective ligands. The importance of this all-parallel topology was further supported by the work on the two G4Q structures containing human telomeric DNA bound to a ligand (3CE5³⁴, 3CDM³⁵). In the crystal lattice both these sequences fold as bimolecular parallel-stranded human telomeric quadruplex with 3CE5 shown in complex with an important 3,6,9-trisubstituted acridine molecule, BRACO19. In 2004, BRACO19 and AS1410 were important lead molecules that targeted telomeres. They were taken forward for development by the company Antisoma Plc as potential first in class agents. The crystal structures confirmed that the selectivity and specificity of these G4Q-targeting agents are consistent with a model in which DNA hybridization to the RNA templating region of the holoenzyme of telomeres is prevented. New molecular entities were also

being developed in the Neidle laboratory, such as the naphthalene diamides. The successful co-crystallization of these molecules with human telomeric sequences (3CCO, 3CDM)³⁵ expanded the understanding of G4Qs as flexible molecular targets. The adaptability of the chain-reversal loops of the G4Qs provides the necessary conformational space required for selective binding, while still retaining the core stacked tetrads and the associated rigidity, thus redefining the quadruplex-drug recognition interfaces. In all cases this molecular target retains the context of an all parallel G4Q folded motif either as a bimolecular or intramolecular folded arrangement, each containing four repeats of human telomeric hexamer sequences. Importantly, the chain-reversal loops topology keeps the external tetrad open for other G4-tetrad aromatic-aromatic interaction, as seen in the packed crystalline lattice or for ligands to bind and interact through π stacking. In solution the formation of alternative topologies such as hybrid-1 and -2 result in the nucleotides in the linking loops stacking and interacting together over these external G4-tetrads, sterically restricting ligand binding to the external G4-tetrad and reducing the opportunities for π stacking.

In the early 2000s, concurrent with the research on targeting telomeres, both the Neidle group and the Balasubramanian group independently undertook comprehensive bioinformatics studies of the prevalence of quadruplexes in the human genome.^{36, 37} They showed that putative G4Qs forming sequences are not randomly distributed; they are focused within promoters of genes involved in replication and cancer related genes. This, combined with the understanding that many helicases involved in unwinding G4Q topologies are compromised, led to structural investigations into promoter regions and the targeting of DNA sequences as therapeutic targets to modulate cell cycle progression and induce programmed cell death. The first structural determination in collaboration with Anh Tuân Phan and Dinshaw Patel was carried out with a monomeric putative G4Q folding sequence identified within the c-kit promoter region and published in 2007 (2O3M)³⁸ providing the first solution-based representation of this promoter target. These structures showed similar structural features seen in the human telomeric sequences with chain-reversal loops. Indeed, this relationship had already been observed in the NMR determined c-myc promoter structure in 2005 (1XAV).³⁹ That structure showed an uncanny resemblance to the human telomeric crystallography-determined structure that folded as an all parallel G4Q with three chain-reversals.²⁶ Subsequent collaborations with Shanker Balasubramanian brought together a G-rich sequence within the c-kit oncogene promoter that forms a parallel G-quadruplex having asymmetric G-tetrad dynamics determined by NMR methods (2KQG and 2KQH)⁴⁰. Subsequently, the equivalent crystal structure of a brominated c-kit-1 proto-oncogene promoter quadruplex DNA (3QXR)⁴¹ was determined by the Neidle group. The topologies of the promoter region determined in solution by NMR agree with the crystallography-derived structures; they both display double-chain-reversal loops, lateral loops, and a stem-loop. By 2014 the determination of a B-raf dimer DNA quadruplex was reported (4H29)⁴² using a sequence taken from the promoter region of the BRAF gene.

Work continued targeting telomeres and in 2009, a structural understanding of ligand interaction with G4Qs was advanced using the telomeric Oxytricha containing sequences and resulted in the publication of a suite of ligand/DNA structures (3EM2, 3EQW, 3ERU, 3ES0, 3ET8, 3EUM, 3EUI).⁴³ The G4Q Oxytricha nova telomeric sequences-based structures fold as bimolecular anti-parallel-stranded quadruplexes and are in complex with a family of 3,6-disubstituted acridines: BSU-6038; BSU-6042; BSU-6045; BSU-6048; BSU-6054; BSU-6066. This work progressed with the publication of a bimolecular anti-parallel-stranded Oxytricha Nova telomeric quadruplex in complex with a 3,6-disubstituted acridine ligand containing bis-3-fluoropyrrolidine end side

chains (3NYP, 3NZ7).⁴⁴ In 2012, the Neidle group reported the first crystal structure of a human telomeric G-quadruplex DNA bound to a metal-containing ligand (a copper complex) (3QSC).⁴⁵ By 2013 several crystal structures of an intramolecular human telomeric DNA G-quadruplex bound by the naphthalene diimide were determined. Sequences included a bimolecular forming G4Q and intra molecular G4Qs, bound to MM41, BMSG-SH-3, (3UYH, 4DAQ, 4DA3).⁴⁶

A new thread of investigation was developing in the Neidle group in the late 2000s, when it became known that in mammalian cells, non-coding DNA telomeric repeats are also transcribed into guanine-rich RNA sequences r(GGGUUA), telomeric repeat-containing RNA (TERRA). This prompted structural investigations into transcribed RNA based on their potential to fold as G4Qs. In 2010, the Neidle group published work on the structure determination of a parallel-stranded G4Q formed from the human RNA telomeric sequence, thus revealing the importance of the propeller-like folds, here with connecting UUA loops (3IBK).⁴⁷ This ground-breaking structure provided the first insight into a G4Q RNA containing a human sequence (TERRA). The RNA structure displays the same overall topology and structural arrangement as its DNA analogue, with the expected C3'-endo sugar puckers along with a modified hydration structure. Shortly thereafter, the first crystal structure of a telomeric RNA G-quadruplex complexed with an di-substituted acridine-based ligand was determined (3MIJ),⁴⁸ along with the same disubstituted acridine ligand complex bound to human telomeric DNA (3QCR). These structures revealed the adaptability of the UUA loops within the chain-reversal loops, where the O2' hydroxyl groups of the ribonucleotide sugars play a central role in defining an RNA specific binding environment for the acridine molecule complexed to the RNA quadruplex along with the additional stability associated with all-anti glycosidic within the G-tetrads.

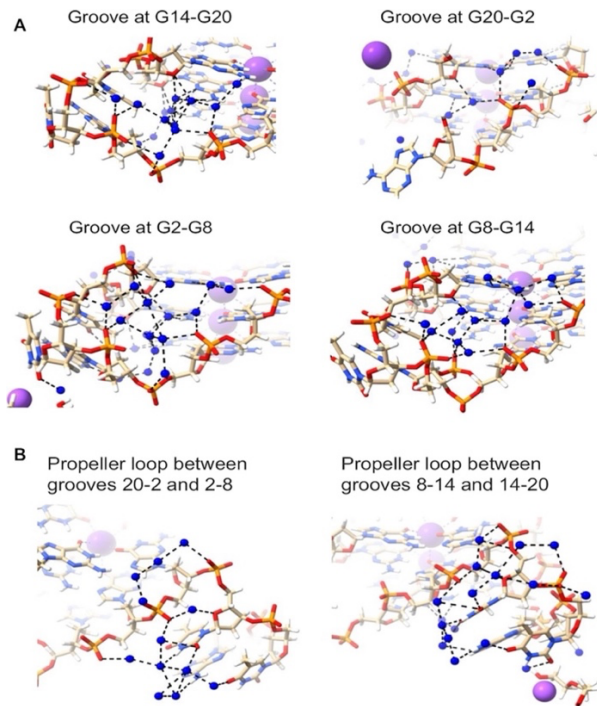


Fig. 10. Water cluster networks in the 7KLP crystal structure of an intramolecular G4Q formed from one strand containing four human telomeric repeats and folded into three stacked-tetrads with connecting chain-reversal T-T-A loops. Waters are shown as dark blue spheres, (A) in the grooves

and (B) in the chain-reversal loops. Potassium ions in the central core are shown as mauve spheres. Figure reproduced with permission from reference ⁴⁹ (CC BY).

Coming full circle from his original studies of dCG₂PF which showed extensive water networks in a drug nucleic acid complex,⁵ Neidle has recently analyzed water networks in G-quadruplex crystal structures and structured waters that mediate small molecule binding to G-quadruplexes.^{49, 50} These analyses of high-resolution crystal structures with varying topologies that have different groove widths and connecting loop arrangements, show extended spines of hydration that are distinct from A/T-rich regions in duplex DNAs. It is anticipated this research will be important in identifying specific networks of hydration that small molecule groove binding molecules can exploit to mediate key interactions between ligand and DNA (Fig. 10).

The many years of research on quadruplexes and their interactions with drugs has recently led to a licensing agreement for a new and first in class treatment for pancreatic cancer - QN-302 one of the family of naphthalene diamides initially developed at UCL SoP targeting G4Qs. Neidle's goal of confirming G4Qs as viable therapeutic targets may soon be reached.

4. References

- [1] R.E. Franklin, R.G. Gosling, Molecular configuration in sodium thymonucleate, *Nature*, 171 (1953) 740-741.
- [2] S. Arnott, D.W. Hukins, Refinement of the structure of B-DNA and implications for the analysis of x-ray diffraction data from fibers of biopolymers, *J Mol Biol*, 81 (1973) 93-105.
- [3] J.L. Sussman, N.C. Seeman, S.H. Kim, H.M. Berman, Crystal structure of a naturally occurring dinucleoside phosphate: uridylyl 3',5'-adenosine phosphate model for RNA chain folding, *J Mol Biol*, 66 (1972) 403-421.
- [4] S. Berman, W. Carrell, H. L. Glusker, J. P., Neidle, S., Taylor, G., & Achari, A., Molecular and crystal structure of an intercalation complex: Proflavine-cytidylyl-(3',5')-guanosine., *Biopolymers*, 18 (1979) 2405-2429.
- [5] S. Neidle, H.M. Berman, H.S. Shieh, Highly structured water network in crystals of a deoxydinucleoside---drug complex, *Nature*, 288 (1980) 129-133.
- [6] H.S. Shieh, H.M. Berman, M. Dabrow, S. Neidle, The structure of drug-deoxydinucleoside phosphate complex; generalized conformational behavior of intercalation complexes with RNA and DNA fragments, *Nucleic Acids Res*, 8 (1980) 85-97.
- [7] M. Mezei, D.L. Beveridge, H.M. Berman, J.M. Goodfellow, J.L. Finney, S. Neidle, Monte Carlo studies on water in the dCpG/proflavin crystal hydrate, *J Biomol Struct Dyn*, 1 (1983) 287-297.

- [8] D.G. Brown, M.R. Sanderson, J.V. Skelly, T.C. Jenkins, T. Brown, E. Garman, D.I. Stuart, S. Neidle, Crystal structure of a berenil-dodecanucleotide complex: the role of water in sequence-specific ligand binding, *EMBO J*, 9 (1990) 1329-1334.
- [9] D.G. Brown, M.R. Sanderson, E. Garman, S. Neidle, Crystal structure of a berenil-d(CGCAAATTTGCG) complex. An example of drug-DNA recognition based on sequence-dependent structural features, *J Mol Biol*, 226 (1992) 481-490.
- [10] C.M. Nunn, S. Neidle, Sequence-dependent drug binding to the minor groove of DNA: crystal structure of the DNA dodecamer d(CGCAAATTTGCG)₂ complexed with propamidine, *J Med Chem*, 38 (1995) 2317-2325.
- [11] A.A. Wood, C.M. Nunn, A. Czarny, D.W. Boykin, S. Neidle, Variability in DNA minor groove width recognised by ligand binding: the crystal structure of a bis-benzimidazole compound bound to the DNA duplex d(CGCGAATTCGCG)₂, *Nucleic Acids Res*, 23 (1995) 3678-3684.
- [12] N. Spink, D.G. Brown, J.V. Skelly, S. Neidle, Sequence-dependent effects in drug-DNA interaction: the crystal structure of Hoechst 33258 bound to the d(CGCAAATTTGCG)₂ duplex, *Nucleic Acids Res*, 22 (1994) 1607-1612.
- [13] M. Munde, M. Lee, S. Neidle, R. Arafat, D.W. Boykin, Y. Liu, C. Bailly, W.D. Wilson, Induced fit conformational changes of a "reversed amidine" heterocycle: optimized interactions in a DNA minor groove complex, *J Am Chem Soc*, 129 (2007) 5688-5698.
- [14] D. Wei, W.D. Wilson, S. Neidle, Small-molecule binding to the DNA minor groove is mediated by a conserved water cluster, *J Am Chem Soc*, 135 (2013) 1369-1377.
- [15] R. Satange, C.Y. Chuang, S. Neidle, M.H. Hou, Polymorphic G:G mismatches act as hotspots for inducing right-handed Z DNA by DNA intercalation, *Nucleic Acids Res*, 47 (2019) 8899-8912.
- [16] D. Sun, B. Thompson, B.E. Cathers, M. Salazar, S.M. Kerwin, J.O. Trent, T.C. Jenkins, S. Neidle, L.H. Hurley, Inhibition of human telomerase by a G-quadruplex-interactive compound, *J Med Chem*, 40 (1997) 2113-2116.
- [17] Y. Wang, D.J. Patel, Solution structure of a parallel-stranded G-quadruplex DNA, *J Mol Biol*, 234 (1993) 1171-1183.
- [18] Y. Wang, D.J. Patel, Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex, *Structure*, 1 (1993) 263-282.
- [19] M.A. Read, S. Neidle, Structural characterization of a guanine-quadruplex ligand complex, *Biochemistry*, 39 (2000) 13422-13432.
- [20] P.J. Perry, A.P. Reszka, A.A. Wood, M.A. Read, S.M. Gowan, H.S. Dosanjh, J.O. Trent, T.C. Jenkins, L.R. Kelland, S. Neidle, Human telomerase inhibition by regioisomeric disubstituted amidoanthracene-9,10-diones, *J Med Chem*, 41 (1998) 4873-4884.
- [21] S.M. Gowan, J.R. Harrison, L. Patterson, M. Valenti, M.A. Read, S. Neidle, L.R. Kelland, A G-quadruplex-interactive potent small-molecule inhibitor of telomerase exhibiting in vitro and in vivo antitumor activity, *Mol Pharmacol*, 61 (2002) 1154-1162.
- [22] S. Neidle, M.A. Read, G-quadruplexes as therapeutic targets, *Biopolymers*, 56 (2000) 195-208.
- [23] J.L. Li, R.J. Harrison, A.P. Reszka, R.M. Brosh, Jr., V.A. Bohr, S. Neidle, I.D. Hickson, Inhibition of the Bloom's and Werner's syndrome helicases by G-quadruplex interacting ligands, *Biochemistry*, 40 (2001) 15194-15202.

- [24] M. Read, R.J. Harrison, B. Romagnoli, F.A. Tanious, S.H. Gowan, A.P. Reszka, W.D. Wilson, L.R. Kelland, S. Neidle, Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors, *Proc Natl Acad Sci U S A*, 98 (2001) 4844-4849.
- [25] ViewerLite, Accelrys ViewerLite 4.2, Molecular Simulations Inc. , 2008, pp. Graphics Programme.
- [26] G.N. Parkinson, M.P. Lee, S. Neidle, Crystal structure of parallel quadruplexes from human telomeric DNA, *Nature*, 417 (2002) 876-880.
- [27] S. Haider, G.N. Parkinson, S. Neidle, Crystal structure of the potassium form of an Oxytricha nova G-quadruplex, *J Mol Biol*, 320 (2002) 189-200.
- [28] S.M. Haider, G.N. Parkinson, S. Neidle, Structure of a G-quadruplex-ligand complex, *J Mol Biol*, 326 (2003) 117-125.
- [29] G.R. Clark, P.D. Pytel, C.J. Squire, S. Neidle, Structure of the first parallel DNA quadruplex-drug complex, *J Am Chem Soc*, 125 (2003) 4066-4067.
- [30] P. Hazel, G.N. Parkinson, S. Neidle, Topology variation and loop structural homology in crystal and simulated structures of a bimolecular DNA quadruplex, *J Am Chem Soc*, 128 (2006) 5480-5487.
- [31] G.N. Parkinson, R. Ghosh, S. Neidle, Structural basis for binding of porphyrin to human telomeres, *Biochemistry*, 46 (2007) 2390-2397.
- [32] K.N. Luu, A.T. Phan, V. Kuryavyi, L. Lacroix, D.J. Patel, Structure of the human telomere in K⁺ solution: an intramolecular (3 + 1) G-quadruplex scaffold, *J Am Chem Soc*, 128 (2006) 9963-9970.
- [33] J. Dai, M. Carver, C. Punchihewa, R.A. Jones, D. Yang, Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K⁺ solution: insights into structure polymorphism of the human telomeric sequence, *Nucleic Acids Res*, 35 (2007) 4927-4940.
- [34] N.H. Campbell, G.N. Parkinson, A.P. Reszka, S. Neidle, Structural basis of DNA quadruplex recognition by an acridine drug, *J Am Chem Soc*, 130 (2008) 6722-6724.
- [35] G.N. Parkinson, F. Cuenca, S. Neidle, Topology conservation and loop flexibility in quadruplex-drug recognition: crystal structures of inter- and intramolecular telomeric DNA quadruplex-drug complexes, *J Mol Biol*, 381 (2008) 1145-1156.
- [36] A.K. Todd, M. Johnston, S. Neidle, Highly prevalent putative quadruplex sequence motifs in human DNA, *Nucleic Acids Res*, 33 (2005) 2901-2907.
- [37] J.L. Huppert, S. Balasubramanian, Prevalence of quadruplexes in the human genome, *Nucleic Acids Res*, 33 (2005) 2908-2916.
- [38] A.T. Phan, V. Kuryavyi, S. Burge, S. Neidle, D.J. Patel, Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter, *J Am Chem Soc*, 129 (2007) 4386-4392.
- [39] A. Ambrus, D. Chen, J. Dai, R.A. Jones, D. Yang, Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization, *Biochemistry*, 44 (2005) 2048-2058.
- [40] S.T. Hsu, P. Varnai, A. Bugaut, A.P. Reszka, S. Neidle, S. Balasubramanian, A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics, *J Am Chem Soc*, 131 (2009) 13399-13409.
- [41] D. Wei, G.N. Parkinson, A.P. Reszka, S. Neidle, 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 (2012) 4691-4700.

- [42] D. Wei, A.K. Todd, M. Zloh, M. Gunaratnam, G.N. Parkinson, S. Neidle, Crystal structure of a promoter sequence in the B-raf gene reveals an intertwined dimer quadruplex, *J Am Chem Soc*, 135 (2013) 19319-19329.
- [43] N.H. Campbell, M. Patel, A.B. Tofa, R. Ghosh, G.N. Parkinson, S. Neidle, Selectivity in ligand recognition of G-quadruplex loops, *Biochemistry*, 48 (2009) 1675-1680.
- [44] N.H. Campbell, D.L. Smith, A.P. Reszka, S. Neidle, D. O'Hagan, Fluorine in medicinal chemistry: beta-fluorination of peripheral pyrrolidines attached to acridine ligands affects their interactions with G-quadruplex DNA, *Org Biomol Chem*, 9 (2011) 1328-1331.
- [45] N.H. Campbell, N.H. Karim, G.N. Parkinson, M. Gunaratnam, V. Petrucci, A.K. Todd, R. Vilar, S. Neidle, Molecular basis of structure-activity relationships between salphen metal complexes and human telomeric DNA quadruplexes, *J Med Chem*, 55 (2012) 209-222.
- [46] M. Micco, G.W. Collie, A.G. Dale, S.A. Ohnmacht, I. Pazitna, M. Gunaratnam, A.P. Reszka, S. Neidle, Structure-based design and evaluation of naphthalene diimide G-quadruplex ligands as telomere targeting agents in pancreatic cancer cells, *J Med Chem*, 56 (2013) 2959-2974.
- [47] G.W. Collie, S.M. Haider, S. Neidle, G.N. Parkinson, A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex, *Nucleic Acids Res*, 38 (2010) 5569-5580.
- [48] G.W. Collie, S. Sparapani, G.N. Parkinson, S. Neidle, Structural basis of telomeric RNA quadruplex--acridine ligand recognition, *J Am Chem Soc*, 133 (2011) 2721-2728.
- [49] K. Li, L. Yatsunyk, S. Neidle, Water spines and networks in G-quadruplex structures, *Nucleic Acids Res*, 49 (2021) 519-528.
- [50] S. Neidle, Structured Waters Mediate Small Molecule Binding to G-Quadruplex Nucleic Acids, *Pharmaceuticals (Basel)*, 15 (2021).