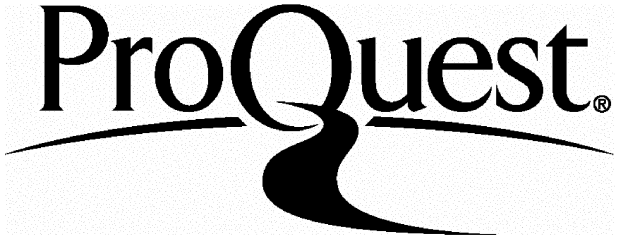
Crystallographic and Molecular Modelling Studies of G-Quadruplex-Ligand Complexes

A thesis submitted for the degree of
Doctor of Philosophy of the University of London

Nancy Husni Campbell

Cancer Research UK Biomolecular Structure Group
The School of Pharmacy
University of London

September 2009



This thesis describes research conducted at the School of Pharmacy, University of London between October 2005 and October 2008 under the supervision of Professor Stephen Neidle. I certify that the research described is original and that any parts of the work that have been conducted by collaboration are clearly indicated. I also certify that I have written all the text herein and have clearly indicated by suitable citation any part of the dissertation that has already appeared in publication.

Nancy Saddlad

Signature

14/2/2010

Date

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor Professor Stephen Neidle for providing me with the opportunity to discover the wonders of X-ray crystallography. Without Stephen's guidance and encouragement, I would not have had the chance to experience the thrill of being the first to set eyes onto a new structure. I am thankful to Stephen's patience and belief in me, for gently nudging me forwards, encouraging me to set challenging goals and striving to achieve them.

All members of the Cancer Research UK Biomolecular Structure Group, past and present, made the past few years at the School of Pharmacy a rewarding and enjoyable experience.

I am very grateful to my colleague Tony Reszka, who not only synthesised endless oligonucleotide preparations for my experiments but also synthesised the ligand BRACO19. Without Tony's contribution, this thesis would not have been complete.

I extend my thanks to Dr Sarah Burge for her help and support when I first started my PhD project. I also thank Dr Gary Parkinson for mounting my precious crystal when my hands were too shaky with excitement.

And to my long suffering friend, the most intelligent person I know, Dr Alan Todd, I can not express how happy I am to have you as my friend. In your analytical approach and crystal clear understanding of all matters, I found a role model and a tutor. Thank you for answering my endless questions.

My parents, Husni and Shahrazan, my sister Rana, my brothers Bashar and Hisham, thank you for loving me unconditionally.

To my husband Darius, thank you for making me happy. I could not have done this without your optimism, patience and support.

ABSTRACT

The majority of cancer cells are innately proficient at maintaining the length of telomeric DNA sequence present at the ends of their linear chromosomes. Telomerase, in its capping and illegitimate activation of its catalytic function, is central to this. Small molecules may interact with telomeric DNA sequences, inducing and/or stabilising the formation of four-stranded DNA quadruplex structures, which thus inhibit telomerase.

Detailed structural description of G-quadruplexes in their native and complex forms, enables the application of rational drug design approaches by developing small molecules that have high specificity for their unique structural features. This may lead to selective and differential targeting of the various forms of G-quadruplexes. The work in this thesis describes the detailed molecular structures of several quadruplex-small molecule complexes, of relevance to rational drug design.

The crystal structure is reported of a complex between the bimolecular human telomeric quadruplex formed by the sequence $d[TAGGGTTAGGGT]_2$ and the experimental anticancer drug BRACO19, to a resolution of 2.5 Å. This has revealed an intricate binding site, whereby BRACO19 is sandwiched between two consecutively-stacked parallel-stranded quadruplexes at the dimeric 5' - 3' interface; stacking onto a G-tetrad at the 3' face and a TATA-tetrad at the 5' face.

The co-crystal structures of the *Oxytricha nova* telomeric sequence $d[GGGGTTTTGGGG]_2$ in complex with a series of six 3,6-disubstituted aminoalkylamido acridine compounds are reported. All the structures are homologous with a published crystal structure in complex with a closely related ligand whereby the ligands are stacking between a G-tetrad and a diagonal T_4 loop albeit revealing a unique adaptation in the position of the 3- and 6- cyclic-amine end groups affected by their size and *ortho*/*para* substitutions.

Computational methods have been used, comprising molecular dynamics simulations and free energy calculations, enabling *in silico* comparison of ligand-quadruplex interactions where crystallisation experiments have not been attempted.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1 - INTRODUCTION

CHAPTER 2 - X-RAY STUDIES OF A HUMAN TELOMERIC QUADRUPLEX–BRACO19 COMPLEX

CHAPTER 3 - X-RAY STUDIES OF A SERIES OF SIX DISUBSTITUTED ACRIDINE LIGANDS IN COMPLEX WITH THE *OXYTRICHA NOVA* TELOMERIC QUADRUPLEX

LIST OF ABBREVIATIONS

$\Delta G_{binding}$	free energy of binding
A	adenine nucleotide
ALT	alternative lengthening of telomeres
AMBER	Assisted Model Building with Energy Refinement
AMOEBA	atomic multipole optimised energetics for biomolecular applications
bp	base pair
C	cytosine nucleotide
C-rich	cytosine-rich
CD	circular dichroism
d	deoxy
D loop	displacement loop
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
E_{MM}	molecular mechanical energy
esds	estimated standard deviations
ESP	electrostatic potential
E_{TOT}	total energy
FRET	fluorescence resonance energy transfer
G	guanine nucleotide
G-quadruplex	guanine-quadruplex
G-quartet	guanine-quartet
G-rich	guanine-rich
G-run	guanine-run
GAFF	general AMBER force field
GB	generalised Born
$G_{binding}$	binding free energy
G_{solv}	solvation free energy

HPLC	high performance liquid chromatography
hPOT1	human protection of telomeres 1
hTER	human telomerase RNA
hTERC	human telomerase RNA component
hTERT	human telomerase reverse transcriptase
hTR	human telomerase RNA
I	inosine
IC_{50}	concentration of compound needed to inhibit a given biological process by half e.g. enzyme, cell or microorganism
kbp	kilo base pairs
LLG	log likelihood gain
m-AMSA	4'-(9-acridinylamino)methanesulphon-*m*-anisidide
MD	molecular dynamics
MM-(PB/GB)SA	molecular mechanics - (Poisson-Boltzmann / Generalised Born) surface area
MMFF	Merck molecular force field
MPD	2-methyl-2,4-pentanediol
MR	molecular replacement
mRNA	messenger RNA
NHEJ	non-homologous end joining
NMR	nuclear magnetic resonance
nt	nucleotide
PB	Poisson-Boltzmann
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PFF	polarisable force field
PIP1	POT1-interacting protein 1
PME	particle mesh Ewald
PMPB	polarisable multipole Poisson-Boltzmann
POT1	protection of telomeres protein 1
QM	quantum mechanical
QSAR	quantitative structure activity relationship
Rap1	repressor/activator protein 1

RESP	restrained electrostatic potential
RFZ	rotation function z-score
RMSD	root-mean-square deviation
RNA	ribonucleic acid
RP	reverse phase
siRNA	small interfering RNA
SPR	surface plasmon resonance
T	thymine nucleotide
T loop	telomere loop
T_4	TTTT
TERT	telomerase reverse transcriptase
TFZ	translation function z-score
TIN2	TRF1-interacting nuclear factor 2
T_m (1 μM)	melting temperature at 1 μM ligand concentration
TMPyP4	5,10,15,20-tetra(*N*-methyl-4-pyridyl)porphine
TRAP	telomere repeat amplification protocol
TRF1	telomeric-repeat-binding factor 1
TRF2	telomeric-repeat-binding factor 2
TS	entropic component of the binding free energy
U	uracil nucleotide
UTR	untranslated region
UV	ultraviolet
VDW	van der Waals

CHAPTER 1

INTRODUCTION

1 INTRODUCTION

1.1 Chromosome maintenance and replication

As eukaryotic cells switched to linear chromosomes during evolution, they faced two main problems. The first was the "end-replication-problem" (Olovnikov, 1973), where the ability of DNA polymerase to build DNA (deoxyribonucleic acid) solely in the 5' to 3' direction, by extending existing polynucleotide chains, presented a difficulty in the replication of the 3' to 5' daughter strand called the "lagging strand". This was overcome by the use of small RNA primers which were employed to initiate the synthesis of small DNA pieces and then digested. Their places were filled by the DNA replication machinery, thus enabling the joining up of all the pieces. This mechanism nevertheless left a problem; the removal of the terminal primer at the 5' end which could not be replaced produced a slightly shorter daughter strand, resulting in progressive reduction of chromosomal DNA at the 3' ends during multiple cell cycles. The second problem was the need to distinguish healthy chromosome ends from DNA damage sites in need of intervention by the DNA repair machinery (McClintock, 1939; McClintock, 1941).

Both of these problems were solved in eukaryotes with the evolution of special structures at the extreme ends of chromosomes called "telomeres" which acted as a finite buffer against DNA loss due to replication. Telomeres enable cells to distinguish between natural chromosome ends and damaged DNA (de Lange, 2002). Also mechanisms evolved that compensated for the loss of DNA via the activation of a reverse-transcriptase enzyme called "telomerase" (Zahler & Prescott, 1988; Greider & Blackburn, 1985)

1.2 Telomeres

Telomeres are specialised DNA-protein structures (Blackburn, 2005; de Lange, 2005; De Boeck et al., 2009; de Lange, 2002) found at the end of linear chromosomes in eukaryotic cells and some prokaryotes, for example the bacteria species *Streptomyces* (Bendich & Drlica, 2000). Telomeres protect chromosome ends from recombination by serving to distinguish natural linear chromosome ends from damaged chromosome ends, and from degradation, by acting as a buffer to the shortening of chromosomes due to the end-replication-problem (Blackburn, 2005; Pisano et al., 2008; Verdun & Karlseder, 2007). The telomeres in eukaryotes and especially of mammalian origin have been extensively studied due to their role in cell senescence (Greider, 1998), genomic instability and cancer (Cosme-Blanco & Chang, 2008) .

1.3 Telomere structure, function and status

The structure of the telomere comprises a DNA component and a protein component (Blackburn, 2005; de Lange, 2002; De Boeck et al., 2009). The DNA component directs the formation of both a T loop (telomere loop) and a D loop (displacement loop). The protein component directs the formation of the shelterin complex (Grandin & Charbonneau, 2008; Greider, 1999; Griffith et al., 1999; de Lange, 2005; Raynaud et al., 2008).

1.3.1 The DNA component – structure, sequence and length

The DNA component comprises two parts:

- A double-stranded part
- A single-stranded part at the 3' end of chromosomes also known as the 3' overhang (Henderson & Blackburn, 1989; Klobutcher et al., 1981).

  The overhang is a conserved feature of telomeres (Henderson & Blackburn, 1989) and is formed of one G-rich strand that is longer

than its complementary cytosine-rich (C-rich) strand. Consequently, the 3' overhang extends beyond the double-stranded part and protrudes at each end of the chromosome (Klobutcher et al., 1981).

Structure: T loop and D loop

The telomeric DNA loop backs on itself forming a large terminal loop called a T loop. The 3' single-stranded overhang invades the double-stranded telomeric DNA forming a D loop. T loops were visualised using electron microscopy in telomeric DNA isolated from human and mouse cells (Figure 1.1) (Greider, 1999; Griffith et al., 1999).

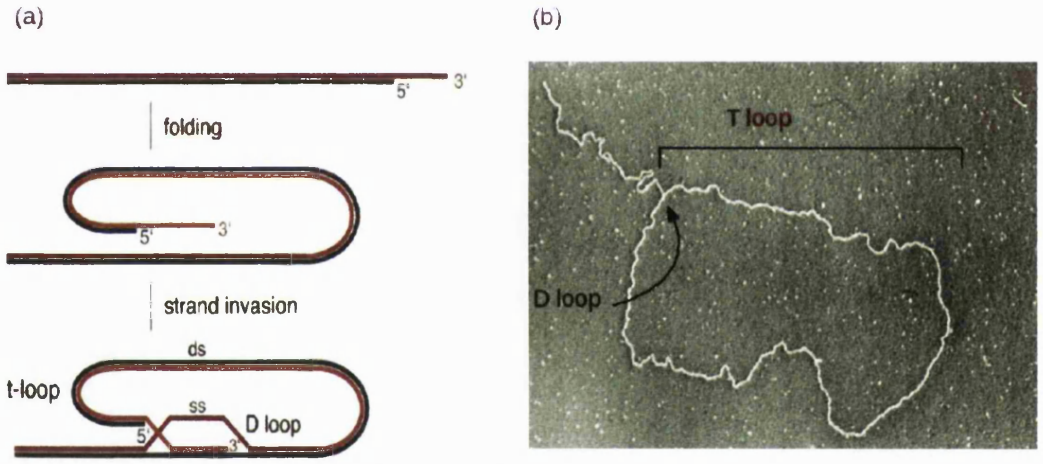


Figure 1.1: Structure of the telomere. (a) Schematic diagram of T loop and D loop (adapted from de Lange, 2005). (b) Visualisation of mammalian T loop and the junction where D loop is believed to form using electron microscopy (adapted from Griffith et al., 1999).

Sequence

The single-stranded DNA component is formed of a guanine-rich (G-rich) repetitive sequence. The sequence varies between species. For example, the sequence repeat TTAGGG forms telomeres in humans (Moyzis et al., 1988), TTGGGG in *Tetrahymena thermophila* (Blackburn & Gall, 1978) and TTTTGGGG in *Oxytricha nova* (Klobutcher et al., 1981).

Length

Length of telomeric DNA is affected by the following factors:

- *Species effects*

  The length of telomeric DNA is species specific (Klobutcher et al., 1981; Moyzis et al., 1988; Starling et al., 1990), varying for overall telomere length between 28 bp in the protozoan ciliate *Euplotes aediculatus* (Klobutcher et al., 1981), to 3 to 12 kbp in human chromosomes (Moyzis et al., 1988) and as high as 50 to 150 kbp in mice (Starling et al., 1990). Similarly, the length of the 3' overhang also depends on species; from 14 nt in the protozoan ciliate *Euplotes aediculatus* (Klobutcher et al., 1981), to around 200 nt in human telomeres (Wright et al., 1997). The size of the overhang is directly proportional to the rate of telomere shortening; from 150 nt in cells that lose 50 bp per cell division to 300 nt in cells that lose 100 bp per cell division (Huffman et al., 2000).

- *Cell type effects*

  Similarly, length of overall telomeric DNA varies with cell type, within a cell population and between chromosomes. For overall telomere length, sperm telomeres are 10 to 14 kbp long whilst telomeres in somatic cells are heterogeneous and several kbp shorter (Shiue et al., 1990).

- A broad range in the length of telomeric overhang exists in human cells as found in a study of 56 human cancerous cell types and five normal cell lines and a positive correlation exists between overhang length and telomere length (Lee et al., 2008), suggesting a dynamic regulation of overhang length. A study of 63 clinical endometrial cancer samples at different stages of cancer progression, demonstrated a significantly shortened overhang compared to normal endometrial tissue suggesting that erosion of 3' overhang length induces impaired

telomeric integrity and genomic instability. It also showed that the most aggressive subtypes of endometrial cancers have a significantly longer overhang than non-aggressive subtypes suggesting that cancer cells with long 3' overhang have a growth advantage due to stabilised telomere ends (Hashimoto et al., 2005).

1.3.2 The protein component

The protein component comprises a set of six telomere-specific proteins known as the shelterin complex. Not all proteins existing at chromosome ends form part of shelterin. The following criteria distinguish the shelterin complex proteins from non-shelterin proteins: shelterin proteins exist abundantly at telomeres only, they remain present at the telomere for the whole of the cell cycle and their known function is limited to telomeres (de Lange, 2005). Due to the complexity of proteins that generally associate with telomeres (Figure 1.2), the following discussion will concentrate on the shelterin complex components.

Shelterin complex

The term shelterin refers to the six core-proteins of the telomere protection complex (de Lange, 2005). The term telosome preferably refers to all protein components found at the telomere such as KU70, KU86, Tankyrase1 and Tankyrase2 (Grandin & Charbonneau, 2008). The shelterin complex is formed of six subunits (Figure 1.2): TRF1(**T**elomeric-**R**epeat-binding **F**actor **1**), TRF2, POT1(**P**rotection **O**f **T**elomeres **1**), Rap1 (**R**epressor/**a**ctivator **p**rotein **1**), TPP1 (formerly called TINT1/PTOP/PIP1) and TIN2 (**T**RF1-**I**nteracting **N**uclear Factor **2**) (de Lange, 2005; Verdun & Karlseder, 2007). Three shelterin subunits: TRF1, TRF2 and POT1, recognise the telomeric repeat TTAGGG. TRF1 and TRF2 bind to the double-stranded telomeric repeat TAGGGTT as shown by NMR and high-resolution crystal structures (Court et al., 2005; Hanaoka et al., 2005). POT1 binds the single-stranded telomeric repeat TTAGGG (Lei et al., 2004; Loayza et al., 2004). The remaining three subunits: Rap1, TPP1 and TIN2 interconnect the former three (de Lange, 2005).

The shelterin complex gives shape to the telomere enabling cells to distinguish healthy telomeres from DNA damage sites (de Lange, 2005). This protects the telomeres from inappropriate repair. Dysfunctional telomeres are subject to DNA repair by NHEJ (**N**on-**H**omologous **E**nd **J**oining) or homologous recombination (Verdun & Karlseder, 2007). POT1 and TRF2 protect telomeres from end-to-end chromosomal fusion and NHEJ (He et al., 2006). TRF1 and TRF2 are involved in the regulation of telomere length (Smogorzewska et al., 2000).

The apparent overlapping of functions of individual members of the shelterin complex indicates that the summation of the components forms its architectural functional unit (Verdun & Karlseder, 2007).

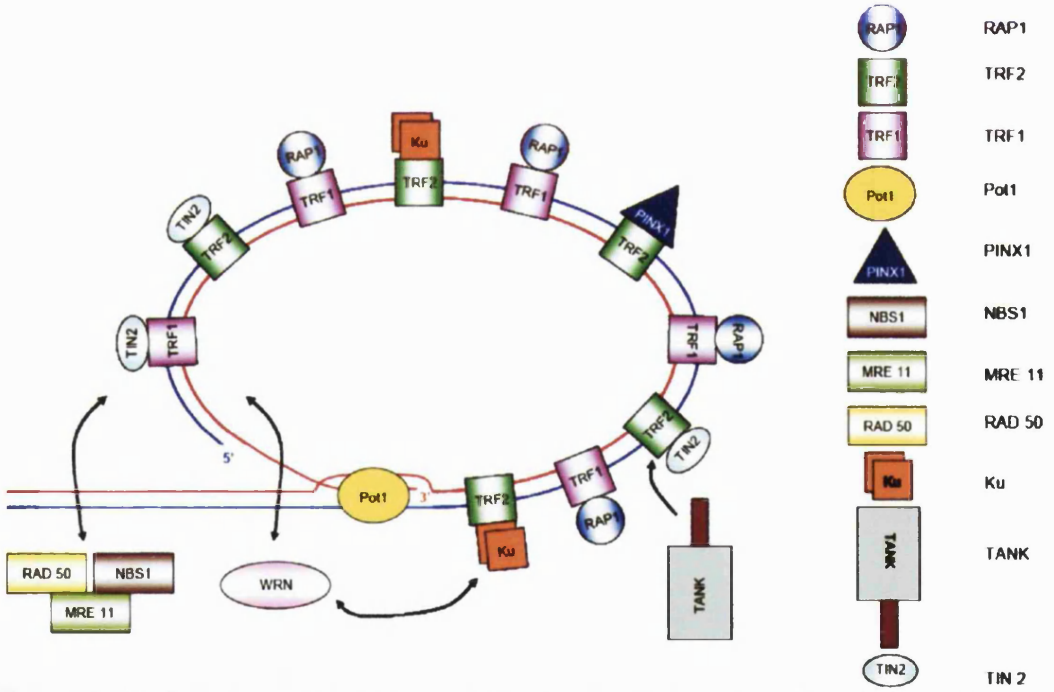


Figure 1.2: Protein components of the telomere. A schematic diagram of the T loop including the shelterin complex and a complex array of other associated proteins (taken from Raynaud et al. 2008).

1.3.3 Telomere status

The three states of the telomere are capped, uncapped and open and are discussed below (De Cian et al., 2008; de Lange, 2005; Verdun & Karlseder, 2007):

1. Capped: the shelterin complex forms a cap on the telomere protecting it from recognition and inappropriate repair by DNA repair machinery and by

telomerase for telomere elongation.

2. Uncapped: a dysfunctional telomere state where it is exposed to recognition as DNA damage site.
3. Open: the shelterin complex is intact but leaves the single-stranded 3' overhang accessible to telomerase for telomere elongation. DNA repair machinery does not recognise the 3' end as a DNA damage site.

1.4 Telomerase structure and function

Telomerase, an enzyme, is a ribonucleoprotein reverse transcriptase.

Normally, transcription involves the synthesis of DNA using RNA as template. Reverse transcription, as performed by telomerase, involves transcribing RNA into DNA. Telomerase, using its own RNA as a template, synthesises single-stranded DNA at the 3' overhang of telomeric ends (Figure 1.3) (Blackburn, 2005; Raynaud et al., 2008).

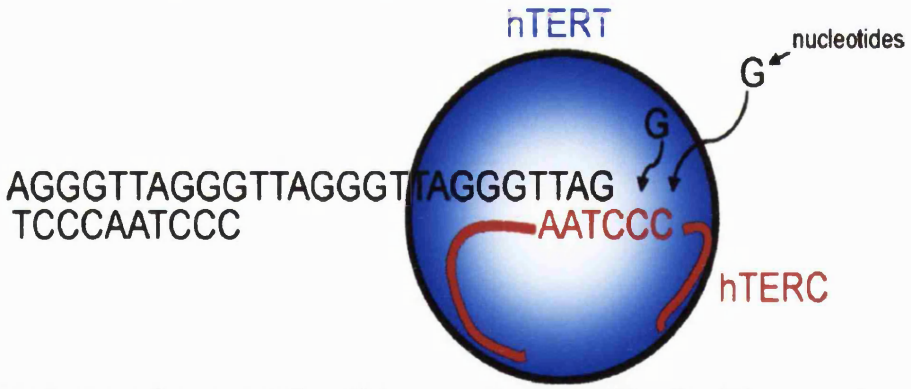


Figure 1.3: A schematic representation of telomerase structure and function. Telomerase, formed of hTERT and hTERC, uses its RNA component i.e. hTERC as a template to add telomeric repeats to the 3' overhang at the end of chromosomes (taken from Raynaud et al., 2008).

1.4.1 Telomerase components

Telomerase is formed of two parts: a protein part and an RNA part (Blackburn, 2005; Harrington et al., 1997).

1. *The protein component*

   The protein part is the catalytic subunit **h**uman **te**lomerase **r**everse **t**ranscriptase (**hTERT**). hTERT is evolutionary conserved (Lingner et al., 1997; Nakamura et al., 1997; Harrington et al., 1997).

2. *The RNA component*

   The RNA part is termed **h**uman **te**lomerase **RNA** **c**omponent (**hTERC**) (also known as hTR or TER). The sequence of the template part of the RNA component is 5'-CUAACCCUAAC and is complementary to the telomeric 3' overhang repeat TTAGGG. hTERC allows telomerase to add telomeric repeats directly to the 3' telomeric ends of chromosomes by serving as a template that simultaneously binds to the 3' single-stranded overhang, using it as a primer, and also acts as a reading template for single-stranded telomeric DNA synthesis (Feng et al., 1995; Blackburn, 2005).

The general view now is that both hTERT and hTERC are essential for the function of human telomerase and telomere maintenance (Cairney & Keith, 2008) which plays a critical role in the long-term proliferation of immortal cancer cells.

1.4.2 Telomerase and the end-replication-problem

Human cells lose 100 - 200 nucleotides of TTAGGG repeats per cell division (Counter et al., 1992; Harley et al., 1990). After a defined number of cell divisions, the end-replication-problem (Olovnikov, 1973), combined with the processing that occurs to create the 3' overhang (Dionne & Wellinger, 1996; Verdun & Karlseder, 2007; Wellinger et al., 1996), results in critical telomere shortening, which leads to cell senescence (Shay et al., 1991; Hara et al., 1991; Smogorzewska & de Lange, 2002; Deng & Chang, 2007) thus preventing the proliferation of ageing cells which are prone to replicative errors. The critical length of telomeres, termed the Hayflick limit (Hayflick, 1965) is dependent on the initial length of the shortest telomeres rather than mean telomere length (Hemann et al., 2001) and the rate of telomere shortening (Huffman et al., 2000).

However, where present (see below), telomerase synthesises DNA at the 3' telomeric ends compensating for the loss of DNA in replicating cells.

1.4.3 Telomerase expression

Telomerase is expressed in human germ and stem cells (Wright et al., 1996; Yui et al., 1998). Telomerase is not generally expressed in normal human non-replicating cells although small amounts have been detected for example in presenescent human fibroblasts (Masutomi et al., 2003). In normal somatic cells, the level of telomerase is either undetectable or barely detectable, and is insufficient to completely prevent telomere shortening. The function of telomerase in these cells is not completely understood but may be related to non-telomere-lengthening functions.

In contrast, 85% of cancer cells overexpress telomerase (Kim et al., 1994; Shay & Bacchetti, 1997; Hiyama & Hiyama, 2003). In normal cells, where telomerase is not expressed, the end-replication-problem causes the telomeric ends to shorten progressively and continuously with each replicative cycle. Once telomere length reaches a critical length, biochemical pathways in the cell, for example *p53*, are activated causing the cells to go into senescence. In contrast, at this critical stage in cancer cells, telomerase is activated, and as a result, telomeres are maintained at a suitable length and the cells become immortal. Immortality is one of the six hallmarks of cancer as defined by Hanahan and Weinberg (Hanahan & Weinberg, 2000).

1.4.4 Telomerase-independent telomere maintenance

Besides telomerase, other mechanisms exist to protect against chromosome degradation at telomeres such as chromosome circularisation, retrotransposition (where mobile genetic elements copy themselves in another position in the genome by first being transcribed to RNA, then transcribed back to DNA by a reverse transcriptase, and then being inserted at another position in the genome (Babushok et al., 2007)) and recombination (also known as **ALT** or **a**lternative **l**engthening of

telomeres) (Grandin & Charbonneau, 2008).

ALT is the subject of much attention at the moment because about 10% of cancer cells enlist this telomerase-independent method (Bryan et al., 1995), to maintain their telomeres (Cesare & Reddel, 2008). ALT involves homologous recombination-dependent replication mechanism of telomeric DNA (Bollmann, 2007; Cesare & Reddel, 2008).

It is not clear whether a correlation exists between ALT and tumour aggressiveness (Bollmann, 2007).

1.5 Dual role of telomere/telomerase in the initiation and suppression of tumorigenesis

1.5.1 Cancer cell transformation

Carcinogenesis is a multistep process, defined by stepwise accumulation of genetic abnormalities that drive the progressive transformation of normal cells into cancer cells (Hanahan & Weinberg, 2000).

Cell immortality is one of the hallmarks of cancer and involves telomere maintenance, either through the stabilisation of telomere length by telomerase (Bodnar et al., 1998), expressed in about 85% of cancer cells (Kim et al., 1994; Hiyama & Hiyama, 2003; Shay & Bacchetti, 1997), or through telomerase-independent pathway such as the ALT mechanism, involving homologous recombination, activated in about 10% of cancer cells (Bryan et al., 1995).

Immortality is achieved in human cells in which both telomerase is expressed and the pathways responsible for triggering non-replicative senescence as a response to short telomeres or loss of capping function are disabled (Counter et al., 1992; Shay & Wright, 2005; Deng & Chang, 2007; Bodnar et al., 1998). For example, virally transformed human cells that solely eliminate *p53* function escape crisis (also known as the M2 stage) at extremely low frequencies (Shay et al., 1993). Crisis is a stage in replicative senescence where loss of cell cycle checkpoints

leads to an extended life span accompanied by continued telomere shortening and characterized by genomic instabilities (Shay & Wright, 2005).

The process of cancer cell transformation involves the emergence of a sub-population that has overcome the senescence barrier normally induced by critically short telomeres.

When cells continue to proliferate beyond their normal replicative limit, they lose the telomeric protective effect and enter a stage termed crisis. Crisis is characterised by genomic instability and cell death. Eventually transformed cells emerge from the crisis stage. Most transformed cells have an upregulated telomerase activity, which maintains telomeres at a functional length, and inactivated *p53*- and RB1(Retinoblastoma 1)-dependent senescence pathways which act as checkpoints for monitoring telomere function in human cells. This allows cells with critically short telomeres to maintain their telomeres at a functional length allowing genomically unstable and immortal cancer cells to be established (Londoño-Vallejo, 2008; Verdun & Karlseder, 2007; Cosme-Blanco & Chang, 2008; Raynaud et al., 2008; Deng & Chang, 2007; Masutomi & Hahn, 2003; Dong et al., 2005; Shay & Wright, 2005).

1.5.2 Telomere shortening, genomic instability, senescence and tumorigenesis

Telomeres preserve the integrity of the genome. Telomeres of a suitable length are essential for a functional telomere structure. Cells with the shortest telomeres reach the point of critically short telomeres first, followed by cells of longer telomeres (Figure 1.4). Critically short telomeres cause a loss of telomere function and are associated with chromosome instability (Londoño-Vallejo, 2008; Raynaud et al., 2008; Cosme-Blanco & Chang, 2008).

The effects of critically short telomeres on the structural integrity of telomeres and how this induces replicative senescence pathways are not yet clear. An overhang of a suitable length may be necessary for T loop formation. It may be that critically short telomeres are no longer capable of participating in T loop structures leading to the loss of T loops and their telomeric protective function. However, due to the

difficulty of visualising T loops through electron microscopy, the answer is unclear (Cesare & Reddel, 2008).

Change in the protected status of shortened telomeres rather than the complete loss of telomeric DNA induces senescence (Karlseder et al., 2002). This study showed that overexpression of TRF2, a telomeric DNA binding protein, reduced the senescence setpoint - defined as telomere length at senescence - from 7 to 4 kilobases i.e. the exact length at which telomeres become critically short is not predefined.

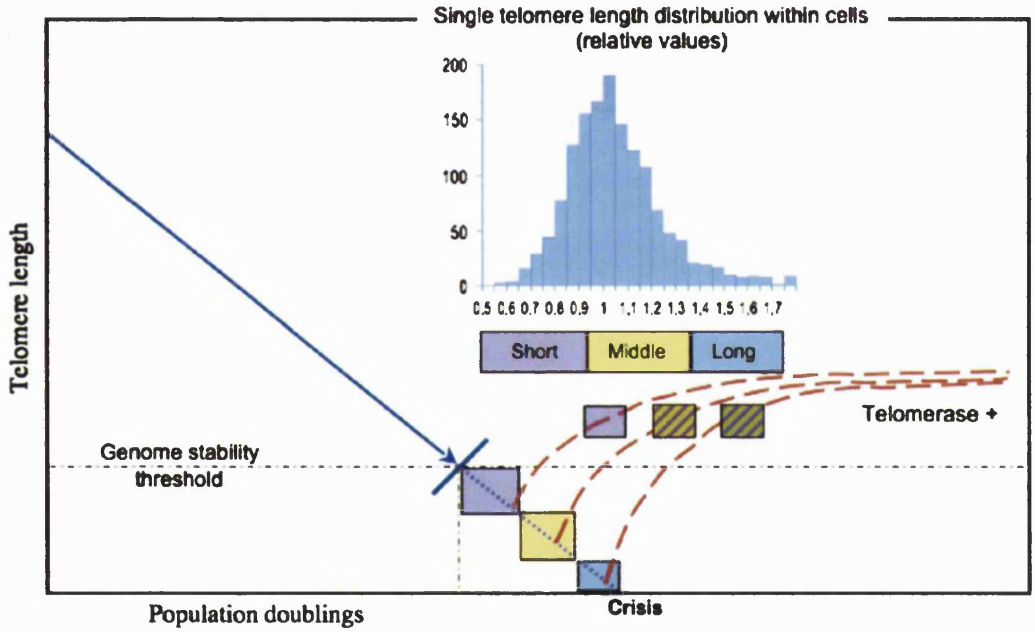


Figure 1.4: Telomere shortening and genomic instability. Critical telomere shortening occurs in cells proliferating beyond the senescence point i.e. telomere length is maintained below the threshold ensuring stability. Length of telomeres in a cell population follows a normal distribution resulting in cells with the shortest telomeres being first to become genomically unstable followed by the cells with the longer telomeres (taken from Londoño-Vallejo 2008).

Structural changes at the telomere caused by critically short telomeres lead to the induction of *p53*- and *p16*/retinoblastoma protein (*p16/pRB*)-dependent replicative senescence pathways (Shay et al., 1991; Hara et al., 1991; Smogorzewska & de Lange, 2002; Deng & Chang, 2007). This acts as a tumour suppressor mechanism (de Lange & Jacks, 1999) by preventing the proliferation of ageing cells which are prone to replicative errors.

Cells that stochastically lose *p53* function can very rarely escape the senescence checkpoints, thus enjoying an extended life span where they continue to proliferate and their telomeres continue to shorten, entering a stage of uncontrolled chromosomal instability called crisis, defined by chromosomal fusions (Counter et al., 1992). If accumulation of genomic instabilities also activates a telomere length maintenance mechanism, for example through telomerase expression, then the cells become immortal (Shay & Wright, 2005; Deng & Chang, 2007).

Telomere/telomerase hence have a dual role. If telomerase is activated before cells enter crisis, telomere length will be maintained enabling the cells to avoid critical telomere shortening and its consequent genomic instability, thus acting as a tumour suppressor. On the other hand, if telomerase is activated during crisis, then telomerase enables the proliferation of genomically unstable cells i.e. cancer cells (Figure 1.5) (Cosme-Blanco & Chang 2008; Masutomi & Hahn 2003).

Telomere shortening may be seen as a tumour suppressing mechanism whilst telomerase activation is a tumour promoting mechanism.

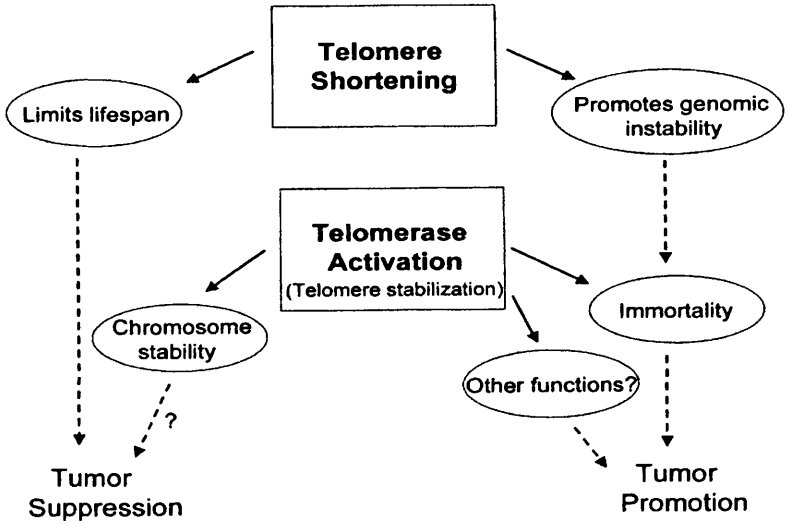


Figure 1.5: Dual role of telomerase and telomeres in tumorigenesis. Telomere and telomerase function as a mechanism for tumour suppression as well as playing an important part in tumorigenesis (taken from Masutomi & Hahn 2003).

1.5.3 The common biology of cancer and ageing

Telomere and telomerase dynamics play a critical role in both ageing and cancer.

During chronological ageing of stem and somatic cells, telomere shortening at chromosomal ends, oxidative stress and UV (ultraviolet) radiation induces accumulation of DNA damage leading to genomic instability. In the majority of ageing cells, this results in senescence or in apoptosis (programmed cell death), however, in a small minority, it produces pre-cancerous cells which proliferate rapidly, accumulating more genomic instabilities, hence more cells become senescent. Normally, only very few cells will eventually accumulate genomic changes that enable them to bypass senescence, favouring tumour progression to full malignancy (Figure 1.6) (Mimeault & Batra, 2008; Finkel et al., 2007).

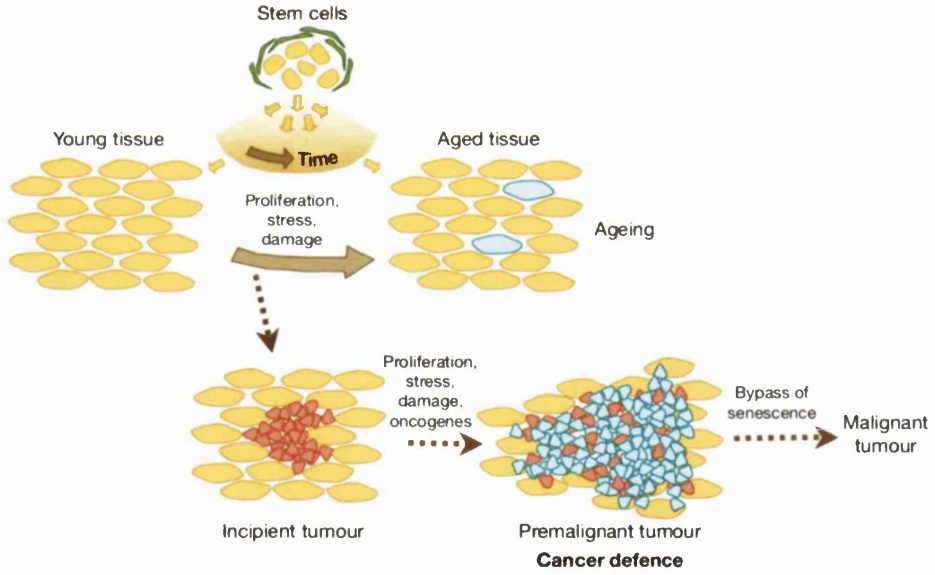


Figure 1.6: Common biology of cancer and ageing (taken from Finkel et al. 2007).

1.6 Telomere/telomerase as targets in cancer therapy

Targeting the telomere/telomerase system as anticancer therapeutics is based on two main observations:

1. Telomerase is expressed in about 85% of cancer cells (Hiyama & Hiyama 2003; Kim et al. 1994; Shay & Bacchetti 1997), with undetectable or very low

levels in normal cells (Kolquist et al., 1998) which are insufficient to maintain telomere length (Pan et al., 1997).

2. Telomeres are shorter in cancer cells than they are in normal somatic, germ and stem cells (Hashimoto et al., 2005; Lee et al., 2008; de Lange et al., 1990) making cancer cells significantly more sensitive to telomerase inhibition (Feldser & Greider, 2007).

These observations indicate that the telomere/telomerase system presents an ideal target for anti-cancer therapy where only cancer cells should be affected, sparing normal healthy cells.

The many approaches attempted in targeting the telomere/telomerase system stand as evidence to the attractiveness of this target (Table 1.1). Direct approaches involve targeting the components of telomerase. On the other hand, indirect approaches include telomerase inhibition, telomere disrupting agents, active immunotherapy and suicide gene therapy (Harley, 2008; Zimmermann & Martens, 2007; Phatak & Burger, 2007).

Table 1.1: Approaches to targeting the telomere/telomerase system (Harley, 2008; Zimmermann and Martens, 2007).

Approach	**Mechanism**	**Examples**
Telomerase inhibition (slow) ● direct	Inhibition of telomere elongation by inhibiting substrate-template recognition by interfering with the *substrate* by making 3' overhang inaccessible or the *template* by making hTERC unavailable	 G-quadruplex targeting ligands e.g. BRACO19, RHPS4 Oligonucleotide template antagonist e.g. GRN163L (a lipid -nucleotide conjugate with sequence TAGGGTTAGACAA)
● indirect	By blocking telomerase expression at the translational or transcriptional level	siRNA (small interfering RNA) targeting hTERT mRNA or hTERC
Telomere targeting (fast)	Disruption of telomere integrity by uncapping	G-quadruplex targeting ligands e.g. BRACO19
Active immunotherapy	By using hTERT as a tumour-associated antigen	hTERT peptide p540-548
Suicide gene therapy	By using hTERT promoter as an instrument to guide the expression of therapeutic genes in cancer cells	Herpes simplex thymidine kinase (HSTK) expression

G-quadruplex (Guanine-quadruplex) targeting agents, which are relevant to this study, act by causing the loss of telomere function resulting from either alteration in

telomere capping function, or from progressive loss of telomeric repeats necessary to maintain proper telomeric structure (Figure 1.7). Dysfunctional telomeres initiate *p53*-dependent cellular senescence or apoptosis to suppress tumorigenesis (Phatak & Burger, 2007; Deng & Chang, 2007).

The dual mechanism of G-quadruplex targeting agents overcomes the lag time encountered with pure telomerase inhibitors which require sufficient time to allow several cycles of cell proliferation to take place, leading to progressive telomere shortening, before reaching the critical length necessary for initiating senescence or apoptosis (Phatak & Burger, 2007). The two mechanisms are (Figure 1.7):

1. Direct inhibition of telomerase (also known as pure/slow telomerase inhibition) by causing the 3' single-stranded overhang fold into a G-quadruplex, making it inaccessible for recognition by the RNA template of telomerase (normally requiring an unfolded overhang) and inhibiting telomerase (Zahler et al., 1991; Sun et al., 1997). This results in shortened telomeres leading to senescence and cancer cell death (van Steensel et al., 1998).

2. Telomere targeting/disruption of telomere integrity by uncapping (fast).

   Normally, telomeres cap the ends of chromosomes and protect them from degradation and genomic instability. Telomeric proteins such as TRF2 and POT1 assemble less efficiently when the G-strand overhang is in the form of a G-quadruplex (which can be initiated/stabilised by G-quadruplex targeting agents) (Yanez et al., 2005) possibly interfering with T loop formation, which normally involves the insertion of 3' single-stranded telomeric overhang into double-stranded DNA (Pedroso et al., 2009). Losing the telomeric caps causes chromosomal end-to-end fusions making chromosomes unable to separate correctly during mitosis. This triggers DNA damage checkpoints and causes genomic instability (van Steensel et al., 1998; Yanez et al., 2005).

   It could be suggested that all cells, regardless of telomerase activation i.e. both healthy and cancer cells, will become uncapped upon exposure to

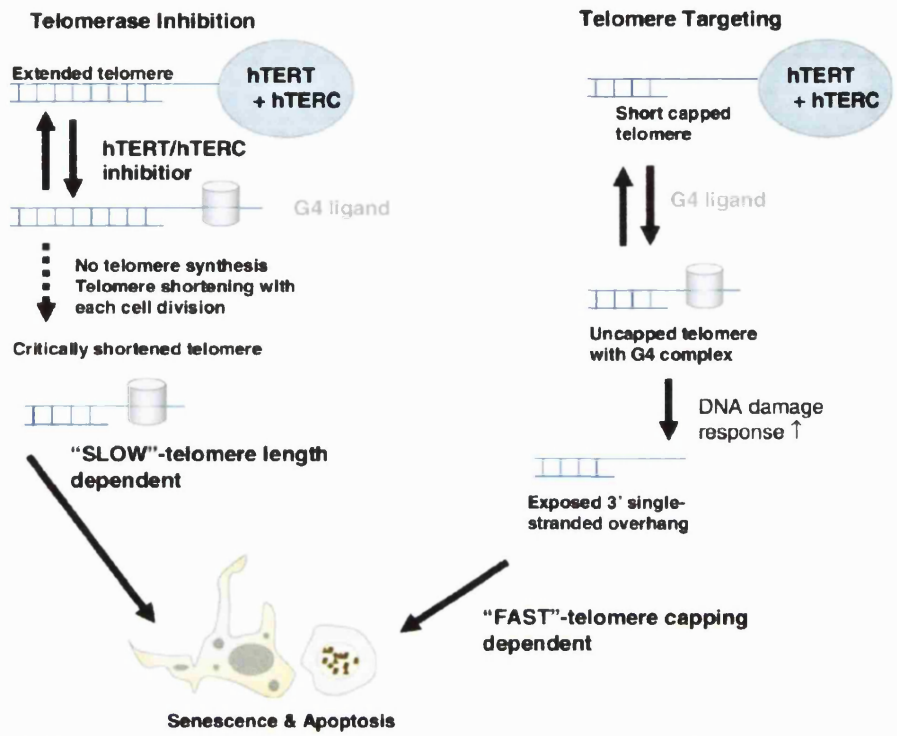


Figure 1.7: The two mechanisms of action for telomere targeting agents. Telomerase enzyme inhibition (slow-telomere length dependent) versus telomere targeting (fast-telomere capping dependent) (adapted from Phatak and Burger, 2007).

telomere targeting agents and consequently suffer the devastating effects of genomic instability brought on by chromosomal end-to-end fusion. However, recruitment of DNA repair proteins is not affected by G-quadruplexes, making it possible to overcome this obstacle in healthy cells (Yanez et al., 2005).

1.6.1 Advantages of targeting telomere/telomerase system

1. Increased sensitivity in cancer cells

   As normal cells have longer telomeres than cancer cells, more rounds of proliferation are required for their telomeres to shorten critically (Feldser & Greider, 2007), making cancer cells significantly sensitive to telomerase inhibition (Figure 1.8) and should provide tumour specificity resulting in telomerase-targeting drugs having a broad therapeutic window (Harley, 2008).

2. Universal targeting of cancer cells

   Telomerase is expressed in the majority of tumours from all cancer cell types (Hiyama & Hiyama, 2003; Shay & Bacchetti, 1997; Kim et al., 1994), and maybe considered as the first common tumour antigen (Cortez-Gonzalez & Zanetti, 2007), making it possible, at least in theory, to contemplate targeting almost all types of cancer with one drug.

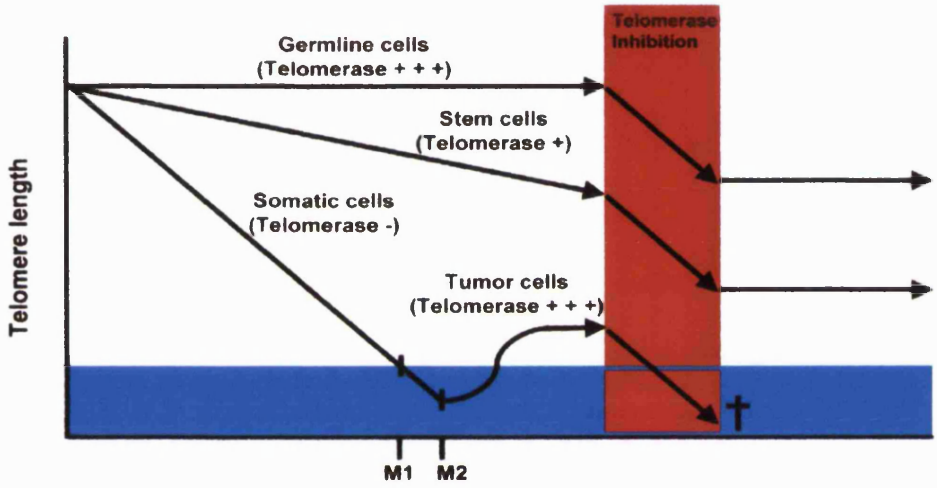


Figure 1.8: Telomerase targeting in telomerase-positive healthy normal cells versus cancer cells. Since cancer cells have shorter telomeres, telomerase inhibition for an adequate period of time (red bar) could selectively kill cancer cells (blue bar) whilst sparing normal telomerase-positive stem and germ cells (taken from Zimmermann and Martens 2007).

1.6.2 Resistance to telomere/telomerase-targeted anti-cancer therapy

ALT-dependent telomere length maintenance is unique in that it is independent of telomerase activation (Bryan et al., 1995) and consequently cancer types where ALT is activated, inherently or as a consequence of exposure to pure telomerase inhibitors, will be resistant this family of agents (Muntoni & Reddel, 2005). As noted earlier, only 10% of cancer cells are in this category.

1.7 G-quadruplex structures

G-quadruplex structures are higher order nucleic acid structures formed by four strands of G-rich (Guanine-rich) nucleic acid sequences present in the DNA of the

genome or its RNA-transcript (Patel et al., 2007; Neidle & Parkinson, 2008; Phan et al., 2006).

G-rich G-quadruplex-forming sequences are found in the DNA telomeric ends of chromosomes and its RNA-transcript, in the DNA promoter regions of proto-oncogenes and in the 5'-UTRs (5'-untranslated regions) of their mRNA (messenger RNA) product. Thus, G-quadruplex structure formation influences key biological processes involving telomere/telomerase integrity and function, expression and regulation of proto-oncogenes (normal gene that can become an oncogene due to mutations or increased expression) at the transcriptional level and at the translational level, respectively.

Targeting G-quadruplex structures with ligands that influence their formation is potentially key to influencing cancer prognosis, making G-quadruplex structures an attractive anti-cancer therapeutic target.

1.7.1 General background

The first description of the structural motif for G-quadruplex structures appeared in 1962 (Gellert et al., 1962). A G-quadruplex is a higher order four-stranded nucleic acid structure. Each of the four strands contributes one guanine to a central square planar arrangement called a Guanine-quartet (**G**-quartet and also known as a G-tetrad) (Figure 1.9). Each guanine is placed at a corner of the square. In this arrangement, each guanine forms hydrogen bonds with each of its two neighbouring guanines. The large planar surfaces of the G-quartets provide a large area for $\pi - \pi$ stacking interactions. The G-quartets stack on top of each other and a hole is formed in the centre. The walls of the hole are formed of the oxygen atoms (O6) belonging to the carbonyl in the guanines. The oxygen atoms give the central hole a negative charge. The shape and charge of the hole allows placing positively charged ions inside it. The positively charged ions counteract the negative charge of the oxygen atoms. The eight hydrogen bonds per G-quartet, the stacking interactions of the G-quartets and the positively charged ions contribute to the stability of G-quadruplex structures (Figure 1.9) (Burge et al., 2006; Patel et al.,

2007; Neidle & Parkinson, 2008; Phan et al., 2006).

Despite the advances in X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR), only a limited number of G-quadruplex structures are available (Burge et al., 2006; Neidle & Parkinson, 2008; Phan et al., 2006). These are discussed in detail elsewhere in this chapter.

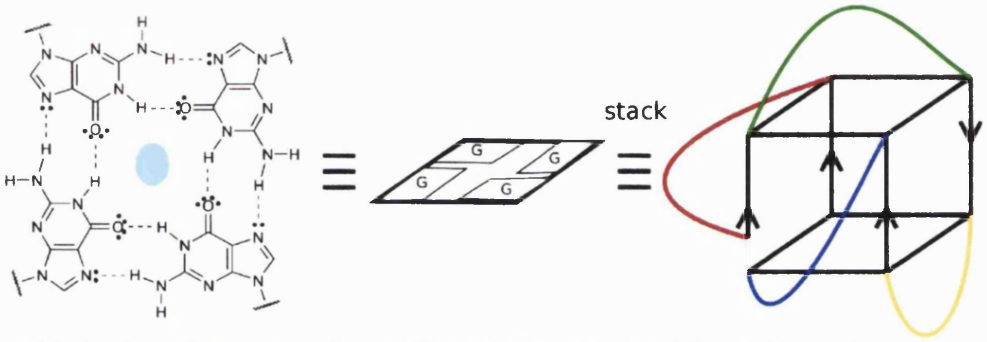


Figure 1.9: A schematic representation of a G-quadruplex structure. A G-quartet showing hydrogen bonds (ideal length equals 2.9 Å) (black dashed lines) and a central positively charged ion (blue circle) - (far left). Several G-quartets (middle) stack to form a G-quadruplex (far right). The G-quadruplex shown (far right) is an imaginary one containing all known loop types; red = V-shaped, green = diagonal, yellow=edgewise (or lateral) and dark blue = propeller (or double-chain reversal loop).

1.7.2 Prevalence in the genome

Experimental studies

G-quadruplex architecture is adopted in four key biological contexts and G-quadruplex forming sequences can be categorised as follows:

1. DNA telomeric ends of chromosomes (Chang et al., 2004; Schaffitzel et al., 2001). This interferes with telomerase function and telomere structure (Patel et al., 2007; De Cian et al., 2008; Kelland, 2005).
2. DNA purine-rich strands at the promoter elements of oncogenes (part of the double-stranded non-telomeric sequences). These G-rich strands can potentially form G-quadruplex structures when transient destabilisation of the duplex form takes place, a step associated with transcription and replication. Two recent reviews (Patel et al., 2007; Qin & Hurley, 2008) discuss some examples: *c-myc*, *c-kit* and *bcl-2* sequences. This

interferes with gene regulation (Verma et al., 2009; Rawal et al., 2006; Eddy & Maizels, 2006) at the transcriptional level (Du et al., 2008; Qin & Hurley, 2008).

3. Within RNA 5'-untranslated regions (UTRs) (Huppert et al., 2008; Patel et al., 2007; Kumari et al., 2007). A bioinformatic search (Huppert et al., 2008) of the human genome revealed an excess of G-quadruplex motifs in the 5'-ends of the 5'-UTRs. This puts G-quadruplex motifs in close proximity to translation start sites suggesting that 5'-UTRs are linked to translational control. Indeed, a study showed that an RNA G-quadruplex (as monitored by circular dichroism) in the 5'-UTR of the *NRAS* proto-oncogene modulates translation (Kumari et al., 2007). This provides the possibility of 5'-UTRs as therapeutic targets and the potential of designing small molecules that stabilise 5'-UTR G-quadruplex formation, thereby influencing the translation of oncogenes. This interferes with gene expression at the translational level (Huppert et al., 2008; Kumari et al., 2007).

4. RNA telomeric transcript: Recent studies showed that mammalian telomeres were transcribed into telomeric-repeat containing RNA (Schoeftner & Blasco, 2008; Azzalin & Lingner, 2008). Telomeric RNA molecules ranging between 100 to 9000 nucleotides in length and containing the telomeric repeat UUAGGG were found in nuclear fractions (Azzalin et al., 2007). A very recent study using NMR (Martadinata & Phan, 2009) revealed a parallel topology adopted by the RNA-containing telomeric sequence UAGGGUUAGGGU. This interferes with RNA-mediated mechanisms involved in telomere architecture and integrity (Azzalin & Lingner, 2008; Azzalin et al., 2007; Schoeftner & Blasco, 2008), although much of the biology of telomeric RNA remains to be discovered.

Bioinformatics and in silico studies

In silico studies and bioinformatic searches for G-rich tracts in the human genome showed that putative G-quadruplex-forming sequences are prevalent in the human genome. Two independent studies predicted that the genome contains approxim-

ately 370 000 G-quadruplex forming sequences (Huppert & Balasubramanian, 2005; Todd et al., 2005).

Putative G-quadruplex-forming sequences are predicted in promoter regions of genes in humans (Huppert & Balasubramanian, 2007) and prokaryotes (Rawal et al., 2006), implicating them in gene regulation and expression (Huppert et al., 2008; Todd & Neidle, 2008; Du et al., 2008; Rawal et al., 2006).

A positive correlation exists between putative G-quadruplex-forming sequences and proto-oncogenes (a normal gene that can become an oncogene due to mutations or increased expression) (Eddy & Maizels, 2006) implicating a link to cancer.

1.7.3 Structural features

G-quadruplex structures have unique structural features that may be targeted with ligands that exhibit complementary features (Phan et al., 2006; Ou et al., 2008; Neidle & Parkinson, 2008).

G-quadruplex structures have the following structural features:

1. A G-quartet which forms a large flat aromatic core capable of participating in π–π stacking interactions (with a ligand for example).
2. Loops which comprise nucleotides formed of nucleic acid bases and a negatively charged nucleic acid backbone. Thus loops are capable of participating in π–π stacking interactions and hydrogen bonding.
3. Grooves which have floors formed of the edges of guanine bases and walls formed of the negative nucleic acid backbone. The grooves have the following potential binding interactions:
    - N2 and N3 (N=nitrogen) from the edges of guanine bases can participate in hydrogen bonds with a suitable ligand.
    - OP1, OP2 and O4' (O=oxygen and P=phosphorus) from the negative nucleic acid backbone can participate in electrostatic interactions with positively charged ligands.

Types of loops: Propeller, Diagonal, Edge, V-shaped

Four types of loops are reported (Figure 1.9) (Patel et al., 2007; Phan et al., 2006; Burge et al., 2006) and are summarised below with examples listed in tables 1.2 and 1.3):

1. Propeller (also called a double-chain-reversal-loop): the term "propeller" loop was originally used to describe the visual effect produced by the particular arrangement of loops and G-quartets observed in the first reported crystal structure of the human telomere (Parkinson et al., 2002). This arrangement resembled a propeller when viewed facing the flat surface of the quartets.

   The term double-chain-reversal-loop describes the reversal of direction in the G-runs connected by the loop from anti-parallel to parallel i.e. both G-runs are oriented in the same 5' to 3' direction i.e. parallel.

   This loop straddles the quadruplex groove and connects a guanine nucleotide from a top G-quartet to a guanine in a bottom quartet so as the two connected G-runs are oriented in the same 5' to 3' directions.

2. Diagonal: this describes a loop that connects two guanines on opposite corners of the same G-quartet. This reverses the polarity of the connected G-runs so as one G-run is oriented 5' to 3' and the connected G-run is oriented 3' to 5' i.e. anti-parallel.

3. Edge (or lateral): this describes a loop that connects two adjacent G-runs in a quadruplex by linking guanine residues belonging to adjacent corners in the same G-quartet. This reverses the polarity of the connected strands as in the diagonal loop i.e. anti-parallel.

4. V-shaped: in a similar manner to the propeller loop, a V-shaped loop connects two adjacent runs of guanines, straddling the groove in the middle, and reversing the polarity of the linked strands. The only difference is that in a V-shaped loop a guanine from a top tetrad is connected to a guanine in a central G-quartet i.e. a quartet that is not external to the quadruplex.

Conformation about the glycosidic bond

The conformation about the glycosidic bond can be either *syn* or *anti*.

The conformation is described as *syn* when H8 and H1' are closer to each other (2.5 Å) compared to a longer distance (3.75 Å) in the *anti* conformation, a feature that can be monitored with NMR methods (Figure 1.10) (Patel et al., 1982).

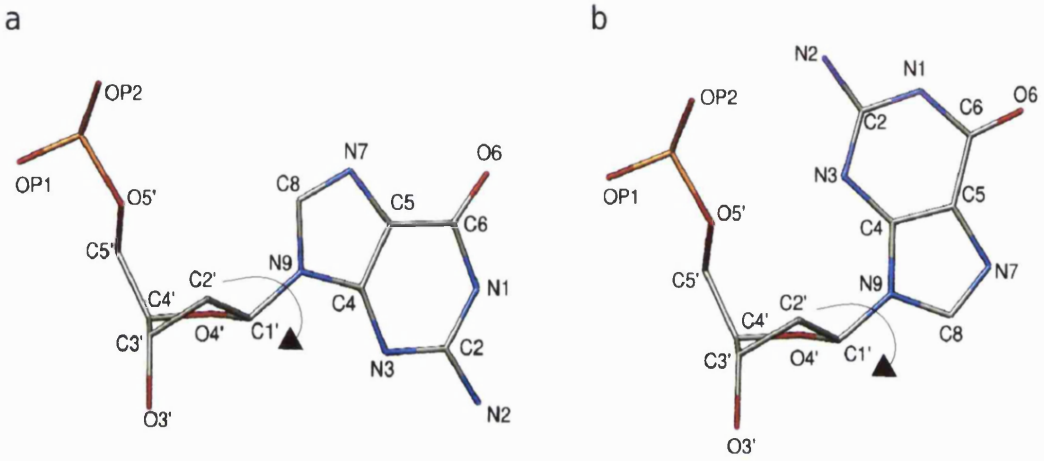


Figure 1.10: Conformation about the glycosidic bond. The glycosidic bond is indicated with an arrow. (a) anti and (b) syn conformations.

G-quadruplex topology

The overall topology of G-quadruplex structures is determined by the following three principal factors:

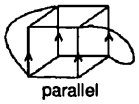
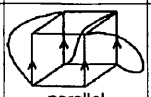
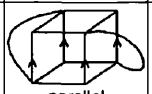
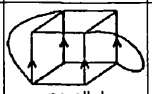
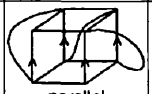
1. Polarity of the strands of the G-runs relative to each other.
2. Type of loops connecting the G-quartets.
3. Nature of the counter ion.

Various combination of these three factors result in a wide variety of quadruplex topologies. Topologies which have been encountered experimentally are:

1. Parallel (Table 1.2): for example the crystal structures of the native bimolecular (PDB ID 1K8P) and unimolecular (PDB ID 1KF1) human telomeric quadruplexes (Parkinson et al., 2002) and in complex with ligands;

with porphyrin (PDB ID 2HRI) (Parkinson et al., 2007) and with a naphthalene diimide (PDB IDs 3CCO and 3CDM) (Parkinson et al., 2008).

Table 1.2: X-ray crystal structures of the human telomeric sequence.

PDB ID	Topology	Salt content	Sequence	Length	Type	Native or ligand-complex	Reference
1K8P	 parallel	50mM NaCl 50mM KCl 50mM Li_2SO4 50mM potassium cacodylate *Annealing buffer=50mM potassium cacodylate with 30mM KCl*	$d[U^*AG_3U^*TAG3T]_2$ U*=5-bromouridine	12 mer	bimolecular	native	Parkinson et al. 2002
1KF1	 parallel	300mM KI 50mM potassium cacodylate *Annealing buffer=50mM potassium cacodylate with 30mM KCl*	$d[AG_3(T_2AG_3)_3]$	22 mer	unimolecular	native	Parkinson et al. 2002
2HRI	 parallel	80mM Li_2SO4 80mM NaCl 20mM potassium cacodylate *Annealing buffer=20mM potassium cacodylate with 50mM KCl*	$d[TAG_3T_2AG_3]_2$	11 mer	bimolecular	complex	Parkinson et al. 2007
3CCO	 parallel	150mM NaCl 50mM sodium cacodylate *Annealing buffer=20mM potassium cacodylate with 50mM KCl*	$d[TAG_3T_2AG_3T]_2$	12 mer	bimolecular	complex	Parkinson et al. 2008
3CDM	 parallel	5mM MgCl2 25mM NaCl 25mM KCl 50mM potassium cacodylate *Annealing buffer=20mM potassium cacodylate with 50mM KCl*	$d[TAG_3(T_2AG_3)_3]$	23 mer	unimolecular	complex	Parkinson et al. 2008

Also, more recently, the structure of the bimolecular form of the RNA transcript of the human telomeric sequence was studied using NMR methods. This also revealed a parallel structure (PDB ID 2KBP) (Martadinata & Phan, 2009).

2. Anti-parallel (Table 1.3): in contrast to the crystal structures of the human telomeric sequence, those studied using NMR methods exhibit different topologies. Four topologies were reported for a variety of modified human telomeric sequences suggesting a sensitivity to the nature of flanking nucleotides. These topologies are:

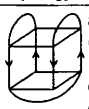
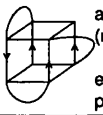
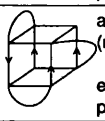
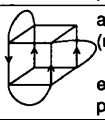
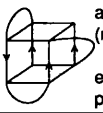
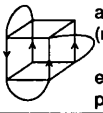
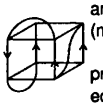
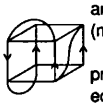
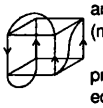
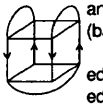
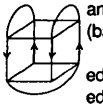
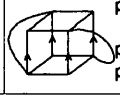
    a. all anti-parallel with edge-diagonal-edge loops; containing three

G-quartets (PDB ID 143D) (Wang & Patel, 1993) and containing two G-quartets (PDB IDs 2KF7 and 2KF8) (Lim et al., 2009).

b. mixed parallel/anti-parallel (also known as 3+1 which describes the polarity of the strands where three are oriented in the same 5' to 3' direction whilst the fourth is oriented in the opposite direction i.e. 3' to 5'). This category is further divided into two subtypes based on the order of loops when traced in the 5' to 3' direction.

    - edge-edge-propeller (PDB IDs 2GKU (Luu et al., 2006) 2HY9 (Dai et al., 2007), 2JSL and 2JSQ (Phan et al., 2007) and 2JPZ (Dai et al., 2007)).
    - propeller-edge-edge (PDB IDs 2JSM, 2JSK (Phan et al., 2007) and 2E4I (Matsugami et al., 2007)).

Table 1.3: NMR structures of DNA and RNA human telomeric sequences.

PDB ID	Topology	Sequence	Length	Type	Native or complex	References
143D Na^+ solution	 anti-parallel (basket) edge-diagonal-edge	d[$AG_3(T_2AG_3)_3$]	22 mer	unimolecular	native	Wang & Patel 1993
2AQY K^+ solution	 anti-parallel (mixed) ?-edge-edge	d[GIG(T_2AG_3)$_2$T] +d[TAG_3U] I=inosine, U=uridine	16 mer +6 mer	bimolecular	native	Zhang et al. 2005
2GKU K^+ solution	 anti-parallel (mixed) edge-edge-propeller	d[$T_2G_3(T_2AG_3)_3$A]	24 mer	unimolecular	native	Luu et al. 2006
2HY9 K^+ solution	 anti-parallel (mixed) edge-edge-propeller	d[$A_3G_3(T_2AG_3)_3A_2$]	26 mer	unimolecular	native	Dai et al. 2007 a
2JSL K^+ solution	 anti-parallel (mixed) edge-edge-propeller	d[$TAG_3(T_2AG_3)_3T_2$]	25 mer	unimolecular	native	Phan et al. 2007
2JSQ K^+ solution	 anti-parallel (mixed) edge-edge-propeller	d[$TA(G_3T_2A)_2G^*GGT_2AG_3T_2$] G*=8-bromoguanosine	25 mer	unimolecular	native	Phan et al. 2007
2JPZ K^+ solution	 anti-parallel (mixed) edge-edge-propeller	d[$T_2AG_3(T_2AG_3)_3T_2$]	26 mer	unimolecular	native	Dai et al. 2007 b
2JSM K^+ solution	 anti-parallel (mixed) propeller-edge-edge	d[$TAG_3(T_2AG_3)_3$]	23 mer	unimolecular	native	Phan et al. 2007
2JSK K^+ solution	 anti-parallel (mixed) propeller-edge-edge	d[$TAG_3T_2AG_3T_2AGG^*GT_2AG_3$] G*=8-bromoguanosine	23 mer	unimolecular	native	Phan et al. 2007
2E4I K^+ solution	 anti-parallel (mixed) propeller-edge-edge	d[$AG^*G_2T_2AG^*G_2T_2AG^*G^*GT_2AG^*G_2$] G*=8-bromoguanosine	22 mer	unimolecular	native	Matsugami et al 2007
2KF7 K^+ solution	 anti-parallel (basket) edge-diagonal-edge	d[$G_3T_2AG^*GG(T_2AG_3)_2$T] G*=8-bromoguanosine	22 mer	unimolecular	native	Lim et al 2009
2KF8 K^+ solution	 anti-parallel (basket) edge-diagonal-edge	d[$G_3(T_2AG_3)_3$T]	22 mer	unimolecular	native	Lim et al 2009
2KBP* *=RNA K^+ solution	 parallel propeller-?-propeller	[$UAG_3U_2AG_3U$]$_2$	12 mer	bimolecular	native	Martadinata & Phan 2009

Groove types

In a G-quadruplex structure, the floor of the grooves is formed of the edges of guanine bases and the walls are formed of the nucleic acid negatively charged backbone. Groove types are defined, based on the width of the grooves, as narrow, medium and wide as shown in figure 1.11.

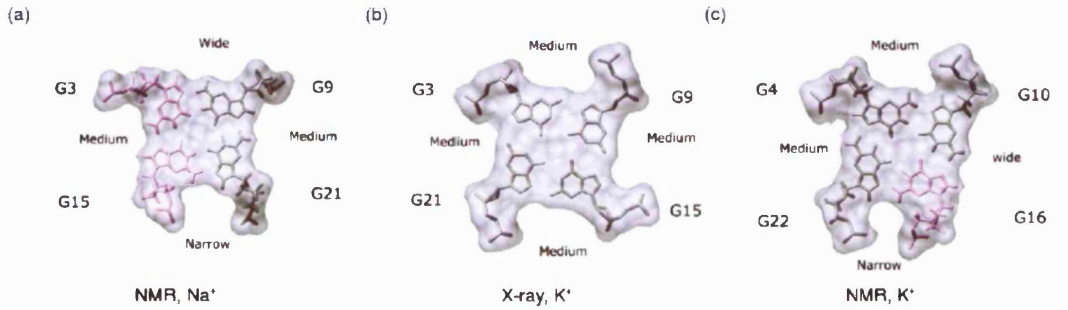


Figure 1.11: Groove types: wide, medium and narrow. (a+c) G-quadruplex structure of the human telomeric sequence as solved by NMR and (b) in the crystal (taken from Campbell & Parkinson 2007). For clarity, only the middle quartet is shown. Grey guanines are in anti conformation and pink guanines are in syn conformation. Surfaces are in purple (taken from Campbell and Parkinson, 2007).

G-quadruplex topology: effects of ions, methodology and sequence

Ion type, type of method used whether NMR or X-ray crystallography and nature of human telomeric sequence influence G-quadruplex topology.

Crystal structures of the human telomeric sequence consistently revealed a parallel topology, irrespective of ion content and flanking nucleotides. Crystallisation conditions contained a variety of monocations such as sodium, lithium and potassium, as well as dications such as magnesium (Table 1.2).

When studied using NMR methods, structures of the human telomeric sequence revealed a more complex nature and appear to be influenced by both; flanking nucleotides and ion type (Table 1.3).

Flanking nucleotides are nucleotides that can be added/mutated/deleted at the 5' and 3' tails of the parent human telomeric sequence to produce a sufficient population of a stable species resulting in interpretable NMR spectra or in crystal formation. Indeed flanking nucleotides are seen in structures determined in solution

and in the crystal. Adenine and thymine nucleotides are commonly used with various examples shown in tables 1.2 and 1.3.

Examples of the effect of ion type, method used and flanking nucleotide on quadruplex topology are discussed below in more detail. This is done by grouping identical sequences then comparing and contrasting the various topologies reported within each group in the context of experimental salt conditions and method used in the study i.e. X-ray crystallographic or NMR methods.

The reported quadruplex structures formed by identical human telomeric sequences, including identical flanking nucleotides, varied according to ion content even when the same method was used. For example, the solution structure reported by NMR methods for the sequence d[$AG_3(T_2AG_3)_3$] was an anti-parallel quadruplex with a basket subtype in sodium-containing conditions (Table 1.4) (PDB ID 143D (Wang & Patel, 1993)) and a mixed parallel/anti-parallel quadruplex in conditions containing potassium (PDB ID 2E4I (Matsugami et al., 2007)). An identical sequence studied by X-ray crystallographic methods, and in conditions containing potassium, revealed an all parallel topology (PDB ID 1KF1 (Parkinson et al., 2002)).

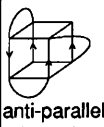
Another example is the sequence d[$TAG_3(T_2AG_3)_3$]. In this instance, conditions containing potassium were common to both topologies, however, each topology was obtained by a different method. The X-ray crystallographic method reported a parallel structure (PDB ID 3CCO (Parkinson et al., 2008)) and the NMR solution structure revealed an anti-parallel fold with a mixed parallel/anti-parallel (PDB IDs 2JSM and 2JSK (Phan et al., 2007)).

Current data shows that G-quadruplex structures in solution are sensitive to flanking nucleotides whilst those in the crystal structure are not. This seems to be related to the edge loop which is found in quadruplexes in solution but does not exist in the crystal structures. Flanking nucleotides can form base pairs and/or hydrogen bond interactions and/or π–π stacking interactions with edge-type loops, adding overall stability. In table 1.4, general observations of currently available solution structures, albeit still very limited in number, enable topologies to be

categorised in the context of the flanking nucleotides added to the parent sequence. However, until a sufficient number of structures becomes available and data are corroborated by other methods of study, trends can not be firmly established and must be viewed with caution. The three general categories are:

1. Sequences with flanking nucleotides at the 3' tail only, or the 5' tail only combined with sodium-containing conditions, exhibit an anti-parallel topology with basket subtype e.g. structures having PDB IDs 143D, 2KF7 and 2KF8.
2. Sequences with flanking nucleotides at both the 5' and 3' tails exhibit a mixed anti-parallel topology with loop order edge-edge-propeller e.g. structures with PDB IDs 2GKU, 2HY9, 2JSL, 2JSQ and 2JPZ.
3. Sequences with flanking nucleotides solely at the 5' tail combined with potassium-containing conditions exhibit a mixed anti-parallel topology with loop order propeller-edge-edge e.g. structures having PDB IDs 2JSM, 2JSK and 2E4I.

Table 1.4: Topologies of the human telomeric G-quadruplex structures as studied by NMR (in potassium-containing conditions unless indicated otherwise).

Topology	PDB IDs	Sequence	Length	References
 anti-parallel (basket) edge-diagonal-edge	143D* *=sodium solution	d[$AG_3(T_2AG_3)_3$]	22 mer	Wang & Patel 1993
	2KF7	d[$G_3T_2AG^*GG(T_2AG_3)_2T$] G*=8-bromoguanosine	22 mer	Lim et al 2009
	2KF8	d[$G_3(T_2AG_3)_3T$]	22 mer	Lim et al 2009
 anti-parallel (mixed) edge-edge-propeller	2GKU	d[$T_2G_3(T_2AG_3)_3A$]	24 mer	Luu et al. 2006
	2HY9	d[$A_3G_3(T_2AG_3)_3A_2$]	26 mer	Dai et al. 2007 a
	2JSL	d[$TAG_3(T_2AG_3)_3T_2$]	25 mer	Phan et al. 2007
	2JSQ	d[$TA(G_3T_2A)_2G^*GGT_2AG_3T_2$] G*=8-bromoguanosine	25 mer	Phan et al. 2007
	2JPZ	d[$T_2AG_3(T_2AG_3)_3T_2$]	26 mer	Dai et al. 2007 b
 anti-parallel (mixed) propeller-edge-edge	2JSM	d[$TAG_3(T_2AG_3)_3$]	23 mer	Phan et al. 2007
	2JSK	d[$TAG_3T_2AG_3T_2AGG^*G\ T_2AG_3$] G*=8-bromoguanosine	23 mer	Phan et al. 2007
	2E4I	d[$AG^*G_2T_2AG^*G_2T_2AG^*G^*GT_2AG^*G_2$] G*=8-bromoguanosine	22 mer	Matsugami et al 2007

The size of the cation determines its position in the central channel of the quadruplex. For example, sodium, due to its smaller size has more freedom of movement and can be found either in the plane of the G-quartet or in between two consecutive G-quartet planes (PDB ID 2O4F) (Creze et al., 2007). Potassium on the other hand, a larger cation, encounters a higher energy barrier to its movement in the central channel with steric hindrance contributing to the high energy barrier (van Mourik & Dingley, 2005).

1.7.4 Structures of G-quadruplex-ligand complexes

The number of available structures for quadruplex-ligand complexes is limited. All complexes are listed in table 1.5 and are discussed further below.

All of the reported structures for complexes containing the human telomeric quadruplex reveal a parallel topology. These comprise:

1. The X-ray crystal structure of the bimolecular human telomeric quadruplex in complex with porphyrin (TMPyP4) (PDB ID 2HRI (Parkinson et al., 2007)).
2. The X-ray crystal structures of the unimolecular and bimolecular human telomeric quadruplexes in complex with a naphthalene diimide ligand (PDB IDs 3CDM and 3CCO respectively (Parkinson et al., 2008)).
3. The solution structure solved by NMR of the tetramolecular quadruplex in complex with the ligand RHPS4 (PDB ID 1NZM ((Gavathiotis et al., 2003))).

An X-ray crystal structure was reported for a complex containing a tetramolecular quadruplex formed of the sequence d[TG_4T] which also revealed a parallel topology. Here, three molecules of the ligand daunomycin were observed stacking onto the 5' face of the quadruplex (PDB ID 1O0K (Clark et al., 2003).

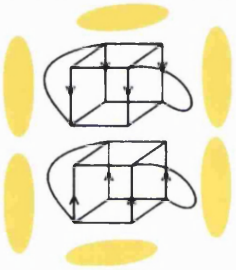
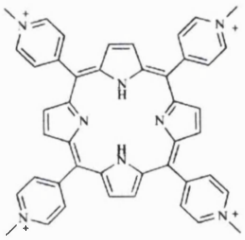
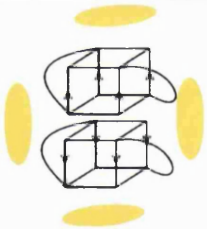
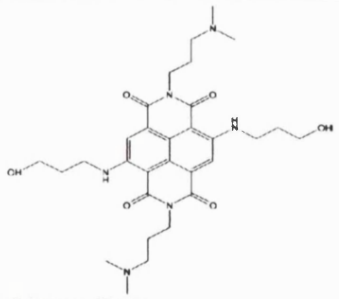
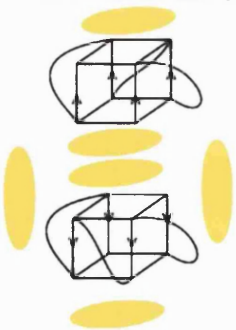
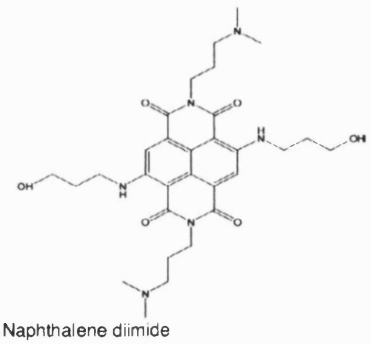
A quadruplex-ligand complex is also reported for the anti-parallel telomeric quadruplex of the ciliated protozoan *Oxytricha nova* complexated with a disubstituted acridine ligand (PDB ID 1L1H (Haider et al., 2003)).

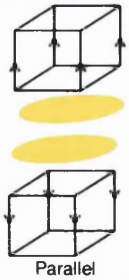
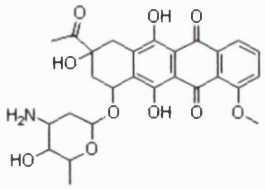
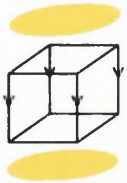
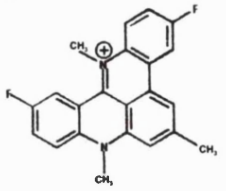
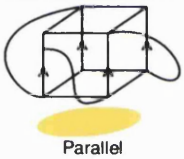
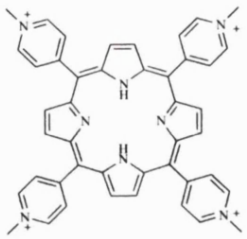
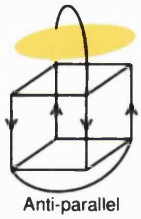
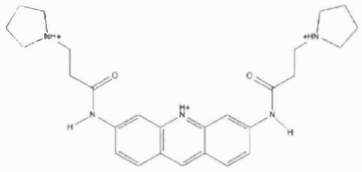
Moreover, the parallel quadruplex formed of the non-telomeric sequence of the human *c-myc* promoter gene was reported in complex with porphyrin (TMPyP4)

(PDB ID 2A5R (Phan et al. 2005)).

As shown in table 1.5 (below) the ligands in the complexes have a variety of chemical scaffolds. However, these scaffolds represent only a small number of current G-quadruplex-targeting-ligands (discussed elsewhere).

Table 1.5: Current structures of G-quadruplex-ligand complexes. Arrows indicate strand polarity. Yellow blob represents one ligand except in the complex with daunomycin (PDB ID 1O0K) where it represents three ligands.

PDB ID	Quadruplex stacking	Sequence	Ligand	References
2HRI K^+	 parallel 3' to 3' or end-to-end stacking	$d[TAG_3T_2AG_3]_2$ 11 mer bimolecular	 Porphyrin	Parkinson et al., 2007
3CCO K^+	 parallel 5' to 5' or face-to-face stacking	$d[TAG_3T_2AG_3T]_2$ 12 mer bimolecular	 Naphthalene diimide	Parkinson et al., 2008
3CDM K^+	 parallel 5' to 5' or face-to-face stacking	$d[TAG_3(T_2AG_3)_3]$ 23 mer unimolecular	 Naphthalene diimide	Parkinson et al. 2008 continued .....

PDB ID	Quadruplex stacking	Sequence	Ligand	References
1O0K	 Parallel 5' to 5' or face-to-face stacking	$d[TG_4T]_4$ 6 mer tetramolecular	 Daunomycin	Clark et al., 2003
1NZM* *=NMR	 Parallel	$d[TAG_3T]_4$ 6 mer tetramolecular	 RHPS4	Gavathiotis et al., 2003
2A5R* *=NMR	 Parallel	d[TGAGGGTGG IGAGGGTGGG GAAGG] (I=Inosine nucleotide)	 Porphyrin	Phan et al., 2005
1L1H	 Anti-parallel	$d[T_4G_4T_4]_2$	 BSU6039	Haider et al., 2003

1.8 G-quadruplexes are attractive targets for anti-cancer therapy

G-quadruplex structures emerged as attractive targets for anti-cancer therapy because of their connection to telomeres and telomerase. The role of telomere and telomerase in cancer was discussed earlier in this chapter in sections 1.4 and 1.5.

G-quadruplex structures are attractive targets for anti-cancer therapy for the following reasons:

1. The unfolded form of a telomeric G-quadruplex forming sequence is the substrate of telomerase. Folding of telomeric G-rich sequences into G-

quadruplexes inhibits telomere elongation by telomerase (Zahler et al., 1991).

Suitable small molecules, such as the first reported 2,6-diamidoanthraquinone compound (Figure 1.12), inhibit telomerase activity through interaction with G-quadruplex structures (Sun et al., 1997).

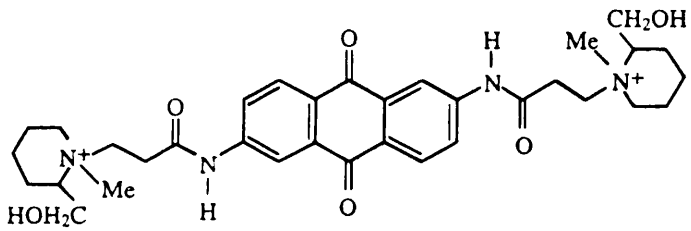


Figure 1.12: First compound shown to inhibit telomerase activity by binding to and stabilising G-quadruplex formation.

2. Telomerase is expressed in the majority of tumours from almost all cancer cell types (Shay & Bacchetti 1997; Hiyama & Hiyama 2003; Kim et al. 1994).
3. Activation of telomerase is required in essentially all tumours for the immortalisation of a subset of cells and transforming them into cancer cells (Masutomi & Hahn, 2003; Cosme-Blanco & Chang, 2008; Shay & Wright, 2005; Deng & Chang, 2007; Verdun & Karlseder, 2007; Raynaud et al., 2008).
4. The unique structural features of G-quadruplexes make it possible to target them with suitable ligands (Phan et al., 2006; Ou et al., 2008; Patel et al., 2007; Neidle & Parkinson, 2008).

1.8.1 Physiological role of G-quadruplex structures

Despite the readiness with which G-rich sequences form G-quadruplex structures *in vitro* (Gellert et al., 1962), evidence for G-quadruplex formation *in vivo* and their physiological relevance has only started to accumulate recently. As discussed earlier, G-rich sequences are abundant in a variety of functional parts of the genome which suggests that G-quadruplex formation could be a means for

telomere regulation, gene function, transcription, replication and translation (Paeschke et al., 2005; Paeschke et al., 2008; Duquette et al., 2004; Schaffitzel et al., 2001; Granotier et al., 2005; Lipps & Rhodes, 2009; Oganesian & Bryan, 2007; Maizels, 2006; Eddy & Maizels, 2006; Rawal et al., 2006; Verma et al., 2009).

Recently, the formation of a G-quadruplex structure in a 16-base pair G-rich sequence in the antigenically variable pilin locus of the human pathogen *Neisseria gonorrhoeae* was found to be necessary for recombination to take place. This site is involved in initiating recombination to direct gene conversion to a specific chromosomal locus. Introducing individual mutations that inhibit quadruplex formation prevented the formation of nicks on the G-rich strand resulting in loss of antigenic variability (Cahoon & Seifert, 2009).

Some of the prominent studies supporting G-quadruplex formation *in vivo* are discussed further below:

1. The discovery of telomere end-binding proteins that control the formation of G-quadruplex structures *in vivo* (Paeschke et al., 2005; Paeschke et al., 2008),as well as specific proteins that unwind G-quadruplex structures such as helicases from the RecQ family; for example BLM in humans (BLM is a gene that encodes a protein that belongs to the RecQ family of helicases and is defective in Bloom's Syndrome) (Sun et al., 1998).
2. Detection of G-quadruplex structures by direct visualisation of specific G-quadruplex-binding moieties bound to telomeric ends of chromosomes.
    - A study used a DNA-binding fluorescent carbazole derivative with preference to binding to G-quadruplex structures to detect the presence of G-quadruplex structures in chromosomal DNA extracted from human cells. The fluorescent emissions upon binding of the carbazole derivative to quadruplex DNA is characterized around 575 nm and at 545 upon binding to duplex DNA. The study showed binding at telomeric ends and also to other places in the chromosome. The latter suggested the formation of quadruplexes in promoter regions (Chang et al.,

2004).

- In another study, *in vitro* generated antibodies specific for G-quadruplex structures were shown to react specifically with G-quadruplex containing macronuclei of *Stylonychia lemnae* and sparing the G-quadruplex free micronuclei (Schaffitzel et al., 2001).
- A specific G-quadruplex ligand was shown to interact preferentially with the terminal ends of human chromosomes in cells cultured with the ligand (Granotier et al., 2005).

3. Visualisation of G-loops using electron microscopy. G-loops are novel DNA structures formed by G-rich sequences in double-stranded DNA during intracellular transcription. Duplex DNA is transiently denatured during transcription when the G-rich strand can form G-quadruplex structures and the C-rich (cytosine-rich) strand can form a DNA-RNA hybrid (Duquette et al., 2004). DNA supercoiling is also implicated in facilitating the formation of secondary DNA structures in both the G-rich and C-rich strands (Sun & Hurley, 2009).

1.8.2 Targeting G-quadruplex structures

Polymorphism

As discussed earlier, NMR and X-ray crystallographic studies revealed that the human telomeric sequence can adopt different topologies depending on the nature of flanking nucleotides, method used in the study and ion content. This makes the process of designing a suitable ligand to target the human telomeric quadruplex for therapeutic purposes a challenge.

Various suggestions have been made in the context of long quadruplex multimers whereby ligand-induced conformational changes in quadruplex folding result in changes in multimer compaction (De Cian et al., 2008). This is based on the observation that suitable ligands that bind to preformed quadruplexes, do in

addition act as molecular chaperones for G-quadruplex formation (De Cian & Mergny, 2007). This suggests that it maybe possible to use structural information of one conformation to design suitable ligands assuming that ligand binding will induce a conformational change towards that targeted G-quadruplex structure. However, it is not possible to verify this yet as there is no experimental structural data for G-quadruplex multimers (Neidle & Parkinson, 2008).

It is of note that some sequences on the other hand, for example the *c-kit* sequence, adopt only one topology in solution, producing interpretable spectra in NMR studies without the need for any mutations/deletions of the wild type sequence. Interestingly, the structure of the *c-kit* sequence exhibits an unprecedented G-quadruplex scaffold where an isolated guanine is involved in G-quartet formation, despite the presence of four three-guanine tracts (Phan et al., 2007). It is not clear yet or possible to anticipate the tendency to form polymorphic topologies for any given G-quadruplex-forming sequence.

G-quadruplex structures and molecular crowding conditions

The term molecular crowding in biological systems, also known as the excluded volume effect, describes the non-specific effect resulting from pure steric repulsion due to the very high concentration of macromolecules in the cell so that a substantial volume is physically occupied (Chebotareva et al., 2004; Ellis, 2001). Adding PEG (polyethylene glycol), ethanol or increasing the concentration of the solute in the experimental conditions are methods regularly used to mimic molecular crowding in the intracellular environment where approximately 40% of the total volume is occupied by intracellular macromolecules. Studying G-quadruplexes (and biomolecules in general) in conditions that mimic the *in vivo* environment can reveal information with regards to their behaviour and structure in the crowded environment in the cell (Miyoshi & Sugimoto, 2008).

The effect of molecular crowding is generally underappreciated and most often ignored in *in vitro* studies which are usually conducted in dilute conditions. However, studies of the structure and behaviour of G-quadruplex structures in general, and more specifically for the human telomeric sequence are starting to

accumulate using conditions that mimic intracellular conditions. Under molecular crowding conditions i.e. containing 40% PEG 200, the human telomeric G-quadruplex: adopted the parallel-stranded conformation (Xue et al., 2007), formed from duplex preferentially at 37° (Zhou et al., 2009), was induced to form under salt-deficient conditions (Kan et al., 2006) and with the complementary strand in close vicinity (Kan et al., 2007). In molecular crowding conditions simulated by adding ethanol (at 37° in 50% ethanol) or by increasing the concentration of the DNA, a transition in the human telomeric quadruplex to the parallel form has been demonstrated (Renciuk et al., 2009).

Crystals contain concentrated amounts of G-quadruplex molecules (closer to conditions of molecular crowding) whilst NMR solutions contain dilute concentrations of about 0.5 to 5mM (Zhang et al., 2005; Phan et al., 2005; Lim et al., 2009).

The physiologically relevant conformation of the human telomeric sequence remains the subject of ongoing debate (Patel et al., 2007; Neidle & Parkinson, 2008; Phan et al., 2006; Dai et al., 2008).

The parallel-stranded form of human telomeric G-quadruplex

The relevance of the parallel-stranded form of the human telomeric G-quadruplex to ligand targeting is supported by the following factors:

1. All reported structures of the human telomeric G-quadruplex in complex with various ligands exhibit a parallel fold in all the available crystal structures, as well as, in the one available solution structure (PDB ID 1NZM). All native forms of the crystal structures are also parallel.
2. The human telomeric sequence forms a parallel-stranded G-quadruplex structure under solution conditions containing potassium and 40% (w/v) PEG 200 (Kan et al., 2007). Adding PEG to the experimental conditions imitates the *in vivo* state in the cell by simulating intracellular molecular crowding conditions.
3. A recent study revealed the parallel-stranded structure of RNA human telomeric G-quadruplex in potassium solution in a study using NMR

methods. The sequence is identical to the one used in X-ray crystallographic studies i.e. UAGGGUUAGGGU (Martadinata & Phan, 2009). This is in agreement with a previous study in sodium solution, which used a combination of NMR, circular dichroism (CD), mass spectrometry and gel electrophoresis to demonstrate a parallel RNA telomeric G-quadruplex (Xu et al., 2008). This is also in accordance with the crystal structures for its DNA counterpart and with previous NMR studies of a fragment of the human telomeric sequence d[TTAGGGT] in potassium solution (Wang & Patel, 1992). The discovery that human telomeric DNA is not transcriptionally silent is a recent concept and the physiological relevance of RNA telomeric transcript remains to be determined. It may be involved in telomere architecture.

1.9 G-quadruplex targeting agents

Background

G-quadruplex targeting agents potentially initiate/stabilise the formation of G-quadruplex structures at any G-quadruplex-forming sequence in the genome.

At first, the idea of targeting G-quadruplex structures with suitable ligands was conceived - and remained restricted to - G-quadruplex formation at telomeric ends (Zahler et al., 1991; Sun et al., 1997).

The unanticipated discovery that the G-quadruplex scaffold is more widely biologically relevant i.e. promoters, 5'-UTRs, telomeric RNA transcript, resulted in rapid growth in the study of G-quadruplex structures, but it also presents new challenges. Previously, targeting G-quadruplex structures consisted of designing suitable ligands that bind to G-quadruplex structures over duplex DNA. Now, it involved targeting one unique G-quadruplex structure over other G-quadruplex structures (Neidle & Parkinson, 2008; Tan et al., 2008). This is important as the non-selective targeting of G-quadruplex structures is likely to result in uncontrolled or unpredictable overlapping of quadruplex targeting leading to complications in

application in the clinic.

General features

G-quadruplex targeting agents share general features which reflect their common target.

A G-quadruplex contains G-quartets, loops and grooves. The G-quartet is a large planar aromatic feature, providing an ideal platform for π–π stacking interactions with a complementary ligand. The negatively charged backbone forming the walls of the grooves and in the loops, where it connects the bases, provides opportunities for hydrogen bond interactions with positively charged moieties suitably positioned in a complementary ligand. The bases in the loops are also available for stacking interactions.

G-quadruplex targeting ligands are designed to reflect these features. They contain a large aromatic core and invariably positive charges which are sometimes part of flexible side chains. Indeed, structural data shows that ligands stack on the G-quartet face and some, namely porphyrin (PDB ID 2HRI) (Parkinson et al., 2007) and a naphthalene diimide (PDB IDs 3CCO and 3CDM) (Parkinson et al., 2008), interact with the loops by means of aromatic stacking (Table 1.5).

Compounds of interest in the literature

There is currently a wide variety of G-quadruplex-targeting-agents covering a wide range of scaffolds including diamidoanthraquinones (Figure 1.12) (Sun et al., 1997), cationic porphyrins such as TMPyP4 (Wheelhouse et al., 1998), trisubstituted acridines such as BRACO19 (Read et al., 2001), the natural compound telomestatin (Shin-ya et al., 2001), triazines such as 12459 (Riou et al., 2002) and pentacyclic acridines such as RHPS4 (Gowan et al., 2001) (Figure 1.13). First generation G-quadruplex targeting agents, such as disubstituted acridines, were generally lacking in selectivity for G-quadruplex structures versus duplex DNA and exhibited similar inhibitory and toxicity values (Harrison et al., 2003).

The ability of G-quadruplex ligands to inhibit telomerase *in vitro* is measured directly

using methods such as FRET (fluorescence resonance energy transfer), competition dialysis and (SPR) surface plasmon resonance (Tan et al., 2008). The TRAP (telomere repeat amplification protocol) method measures IC_{50} (concentration of ligand required to produce 50% inhibition in enzyme activity) in cell culture studies (De Cian et al., 2008). The more recent TRAP-LIG, is a modified form of the TRAP method where an extra step is added to remove the ligand which otherwise interferes with the PCR (polymerase chain reaction) step of the assay resulting in an overestimation of ligand activity (Reed et al., 2008).

BRACO19 and a series of disubstituted acridines are used in this work and will be discussed in more detail in the chapters 2 and 3.

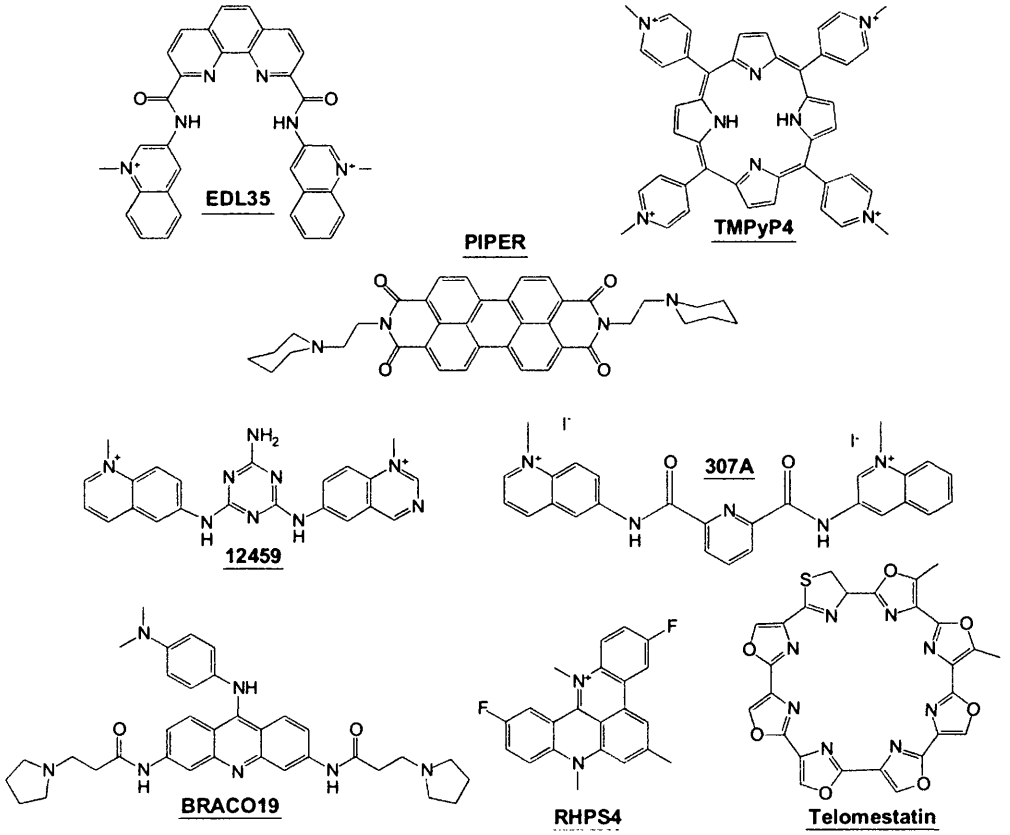


Figure 1.13: Chemical formulae of some G-quadruplex targeting agents. Variable types of chemical scaffolds are shown (adapted from De Cian et al. 2008).

CHAPTER 2

X-RAY STUDIES OF A HUMAN TELOMERIC QUADRUPLEX-BRACO19 COMPLEX

2 X-RAY STUDIES OF A HUMAN TELOMERIC QUADRUPLEX–BRACO19 COMPLEX

2.1 Background

BRACO19 (Figure 2.1), a 9-[4-(N,N-dimethylamino)phenylamino]-3,6-bis(3-pyrrolodinopropionamido) acridine (Read et al., 2001), is a G-quadruplex binding agent. It exhibits anticancer activity *in vitro* in cancer cell culture (Gunaratnam et al., 2007; Incles et al., 2004; Read et al., 2001; Harrison et al., 2003; Schultes et al., 2004; Moore et al., 2006) and *in vivo* in tumour xenografts (Gowan et al., 2002; Burger et al., 2005) through telomerase inhibition and telomere uncapping consistent with telomere shortening associated with G-quadruplex formation (Burger et al., 2005; Gowan et al., 2002; Gunaratnam et al., 2007; Incles et al., 2004).

BRACO19 is a lead compound for a "second-generation" family of quadruplex-binding ligands. Initially, molecular modelling studies were employed to develop disubstituted amidoanthracene-9,10-dione (also known as anthraquinone) and 3,6-disubstituted acridine chromophore families to investigate whether compounds possessing a planar aromatic chromophore may indeed inhibit telomerase activity via stabilisation and binding to a folded G-quadruplex structure. The study predicted that the acridine chromophore was comparable to the anthraquinone family in terms of G-quadruplex binding affinity (Read et al., 1999). Additionally, it contains a heterocyclic central nitrogen atom that has the ability to be protonated at physiological pH, enabling it to act as as pseudocation thus complementing the ion channel that runs in the centre of quadruplex structures. Protonation of this nitrogen would increase electron deficiency through the chromophore, enabling enhanced π–π stacking interactions with the G-quartet platform and improved quadruplex binding affinity as well as improving water solubility in comparison to

anthraquinones (Harrison et al., 1999).

The initial studies were pursued further and the 3,6,9-trisubstituted acridine family was designed (of which BRACO19 is a member) using the same approach, but with the added assumption that a ligand containing an extra substitution would possess improved quadruplex affinity over and above duplex DNA as each of the three substituents occupied a groove in a G-quadruplex (Read et al., 2001).

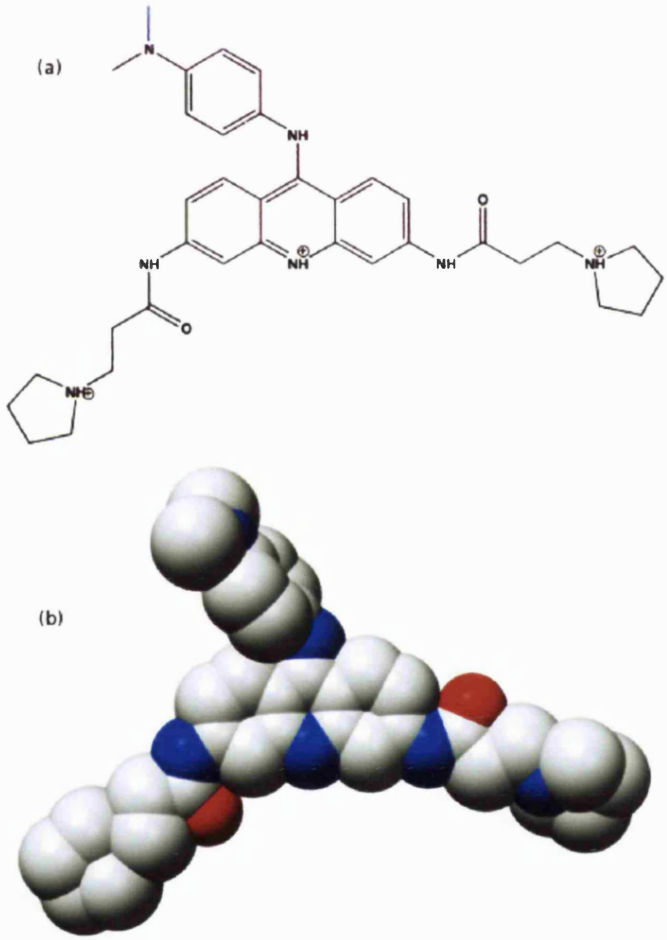
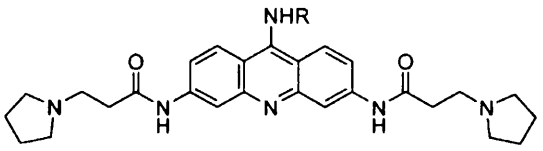


Figure 2.1: BRACO19 (a) chemical structure and (b) space-filling representation - extracted from the crystal structure PDB ID 3CE5 (oxygen in red, nitrogen in blue and carbon in grey). This colour scheme is used throughout this chapter unless stated otherwise.

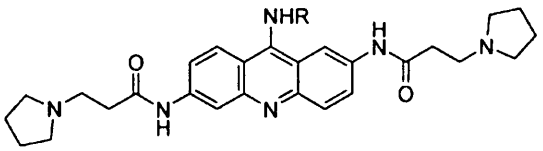
A variety of regioisomers of the trisubstituted acridine family (Figure 2.2) were investigated using molecular modelling, surface plasmon resonance, *in vitro* telomerase inhibition and cytotoxicity studies (Read et al., 2001; Harrison et al., 2003). The 3,6,9-trisubstituted compounds were the most potent telomerase

inhibitors (50% inhibitory concentration ranging from 20 to 115 nM) with modest acute cytotoxicity (50% inhibitory concentration ranging from 0.1 to 15.8 μM) (Harrison et al., 2003).

BRACO19, a promising candidate, was put forward for *in vivo* antitumour activity studies. It combines potent cell-free inhibition of human telomerase (50% inhibitory concentration of 115 nM (Harrison et al., 2003)) with approximately a 90-fold differential to concentration causing non-specific acute cytotoxicity (mean 50% inhibitory concentration for acute cell kill across seven human tumour cell lines of 10.6 μM (Gowan et al., 2002)).



2,6,9-tri-substituted



2,7,9-tri-substituted

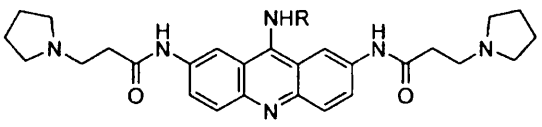


Figure 2.2: Trisubstituted acridine regioisomers.

BRACO19 showed significant antitumour activity *in vivo* both in combination with the antitumour agent paclitaxel (commercially known as Taxol®) in human tumour

xenografts in mice (Gowan et al., 2002) and as a single agent (Burger et al., 2005).

BRACO19 produced cancer cell senescence (Incles et al., 2004; Gunaratnam et al., 2007) and end-to-end chromosomal fusions (Incles et al., 2004) linked to telomere uncapping and telomere shortening (Incles et al., 2004; Gunaratnam et al., 2007). BRACO19 and a fluorinated-derivative were chosen as potential candidates for clinical studies in human cancer (Moore et al., 2006; Martins et al., 2007).

A mechanistic *in vitro* study by Gunaratnam et al. (Gunaratnam et al., 2007) demonstrated that BRACO19 produced cell growth arrest in a variety of cancer cell lines following long-term exposure at sub-cytotoxic concentrations. This was concurrent with time-dependent decrease in telomerase activity which agrees with previous single-time point observations (Read et al., 2001; Harrison et al., 2003; Moore et al., 2006). This supports the idea that BRACO19 binds to the telomeric single-stranded overhang, consistent with quadruplex formation. Furthermore, it showed that telomerase inhibition is only partial and coincides with the partial displacement of the protein hPOT1 from the overhang. This is in agreement with previous observations by Zaug et al. (Zaug et al., 2005) which showed that hPOT1 binding to the single-stranded DNA is detrimental to quadruplex formation and similarly associated with partial displacement of hPOT1.

Having a minimal length of telomeric DNA repeat tract appears necessary to serve as a molecular scaffold capable of nucleating a higher-order DNA-protein complex that can protect the telomere (discussed elsewhere). It was therefore expected that BRACO19 will result in the disruption of the functionality of the telomeric complex through telomere shortening as a result of direct telomerase inhibition associated with quadruplex formation. This was tested in cancer cell culture. However results did not meet the expectation that senescence would occur after several growth cycles comprising several failed telomere lengthening cycles sufficient to critically shorten the overhang which then initiates cell growth arrest. Instead, BRACO19 (Incles et al., 2004) and indeed many other quadruplex-targeting agents (Riou et al., 2002; Pennarun et al., 2005; Cookson et al., 2005; Rodriguez et al., 2008; Brassart et al., 2007; Phatak et al., 2007; Gomez et al., 2006) induce senescence after an apparently "insufficient" short-term exposure (one week for BRACO19).

This was discovered to be associated with telomere uncapping.

Biophysical and biochemical data collected from qualitative structure-activity correlation studies on the effect of 3,6- and 9-end substituent size and side chain length in trisubstituted acridine on telomerase inhibition activity and cancer cell kill can be summarised as follows (Moore et al., 2006):

1. Extending an uncharged substituent at the 9-position side chain to comprise more than two or three carbons has highly deleterious effects on quadruplex binding and telomerase potency respectively.
2. Extension of a positively charged 9-substituent does not have a major effect on telomerase inhibition but can counterbalance the adverse effects of extended 3- and 6-side chains.

A biopharmaceutical study of BRACO19 appeared to show that it has good solubility at 2mg/ml in water and physiological buffers with very poor permeability across biological barriers. This would require suitable formulation for efficient delivery in potential future clinical evaluation (Taetz et al., 2006). Similarly, BRACO19 has been reported to decompose at physiological pH and temperature, with products possessing decreased activity, suggesting that stability will have an important influence on potential future dosage form preparation (Taetz et al., 2008). However, these findings are at variance with many of the studies outlined above.

2.2 Aims

The aim of the structural work presented here was to attempt to gain insight at the atomic level to the intricate structural features of quadruplex-ligand binding. This was done through the study of a particularly promising ligand (BRACO19) which possesses significant specificity and binding affinity for quadruplex over duplex DNA coupled with *in vivo* activity. This is significant especially in view of the very small number of structures reported for quadruplex complexes.

2.3 Materials and methods

2.3.1 Crystallisation

Oligonucleotide synthesis, purification and annealing

The Eurogentec RP (reverse phase)-cartridge-purified human telomeric deoxyribo-oligonucleotide sequence d[TAGGGTTAGGGT] was purchased from Eurogentec Ltd and used without further purification. The oligonucleotide was annealed at 3 mM (quadruplex concentration where one quadruplex is formed of two strands of DNA) before use by incubation in a heating block at 80°C for 15 minutes in 20 mM potassium cacodylate buffer at pH 6.5 and left overnight to cool gradually to room temperature.

The ligand BRACO19 synthesis and purification

The ligand BRACO19 was synthesised in-house by Anthony Reszka in the laboratory of the Biomolecular Structure Group at the School of Pharmacy, University of London, using previously published procedures (Harrison et al., 2003) and purified by reverse phase HPLC. The final sample, as the free base, was analytically pure (> 95%). A 20 mM stock solution was prepared in 100% DMSO (dimethyl sulfoxide) and divided into 20 µl aliquots and kept at -20°C. Repetitive freezing/thawing of the sample was avoided by thawing the required amount at room temperature immediately prior to setting up crystallisation drops.

Complex preparation

Immediately prior to setting up the crystallisation experiment, the complex was prepared by the addition of ligand solution to the annealed quadruplex solution to make a solution of the complex at 2mM quadruplex and 2mM ligand. The reagent solution was then added to the complex to produce the initial crystallisation conditions mentioned in the following crystallisation step.

Preparation of library file for ligand

The ligand was constructed in the SYBYL 7.3 suite (Tripos, Inc., St. Louis, USA), with charges and atom types applied using the MMFF94 method (Halgren, 1996; Halgren, 1996). Stepwise minimisations using the MMFF94s force field (Halgren, 1996) were subsequently carried out to a convergence of 0.02 kcal.mol^{-1}/Å over 5000 steps using the Powell method (Powell, 1977) with a cutoff of 11 Å.

The final coordinates of the ligand were saved in PDB format and used as input for the Dundee PRODRG2 server (Schuettelkopf & van Aalten, 2004). The output file (in CIF format (Brown & McMahon, 2002)) was checked and modified based on the relevant crystal structure parameters for *m*-AMSA, also known as amsacrine, (4'-(9-acridinylamino)methanesulphon-*m*-anisidide) (Cambridge Structural Database, CSD, ref. Code ACRAMS) and *N*-(2-Carboxybenzoyl)-L-leucine methyl ester (Cambridge Structural Database, CSD, ref. Code CIBPEI).

Crystallisation conditions

The hanging-drop vapour-diffusion method was used. The 24-well VDX plates and 22 mm circular siliconised glass cover slips (Hampton Research Corporation) were used in the setup.

Initial crystallisation conditions in the 5 µl drop were: 1.6 mM quadruplex DNA, 1.6 mM BRACO19, 200 mM ammonium sulphate, 20 mM sodium chloride, 20 mM potassium chloride, 20 mM lithium sulphate and 36 mM potassium cacodylate buffer at pH 6.5. This was equilibrated against a reservoir well solution containing 600 µl of 1.6 M ammonium sulphate. Crystals grew as yellow square cuboids after 3 months at 12°C (285.15 K). The dimensions of the crystal used for data collection were approximately 0.1 x 0.1 x 0.2 mm (Figure 2.3).

Mounting the crystal

Prior to data collection, the crystal was briefly soaked in a cryoprotectant solution containing mother liquor and 25% glycerol then mounted on a fibre loop and flash-frozen in a dry nitrogen stream at 105 K. Glycerol-containing solution acts as a

cryoprotectant and is designed to vitrify preventing the formation of crystalline ice i.e. no ordered ice formation; instead a vitreous glass is formed which does not disrupt crystal order or interfere with the diffraction. Water contained within the crystal would otherwise expand during freezing, resulting in a range of adverse effects that can compromise data collection from cracked crystals to crystal dissolution (Garman, 1999; Garman, 2003; Garman & Owen, 2006; Garman & Schneider, 1997).

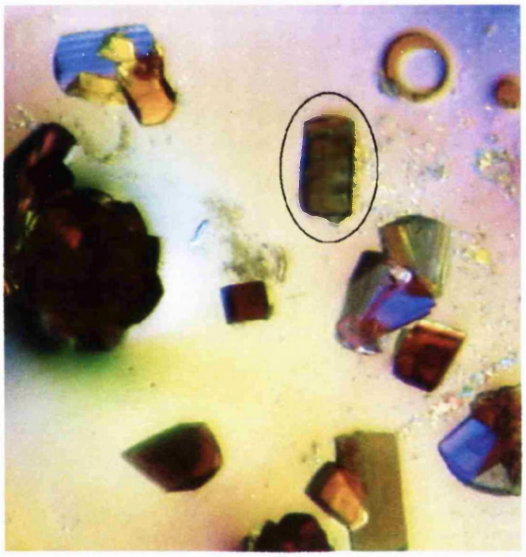


Figure 2.3: Crystals of the human telomeric bimolecular quadruplex in complex with BRACO19 (circled).

2.3.2 Data collection and processing

Data were collected in-house using a Rigaku R-AXIS IV image plate detector at wavelength 1.5418 Å (CuK_α radiation). A single crystal was used and diffraction extended to 2.5 Å (Figure 2.4). Indexing and data processing were carried out using the d*TREK (Pflugrath, 1999) part of the CrystalClear software package (Rigaku Corporation© 1997-2002) The space group was determined as I4 with cell dimensions of $a = 70.34 = b = 70.34$, $c = 34.29$ Å and $\alpha = \beta = \gamma = 90.00°$. Values for crystallographic data are shown (Table 2.2).

2.3.3 Structure solution and verification

The program Phaser (McCoy et al., 2007) from the CCP4 package (Collaborative Computational Project, 1994) was used to solve the structure by molecular replacement (MR). MR is defined by Philip Evans and Airlie McCoy as a term which describes "the use of a known model to solve the unknown crystal structure of a related molecule" where it "... enables the solution of the crystallographic phase problem by providing initial estimates of the phases of the new structure from a previously known structure" (Evans & McCoy, 2008).

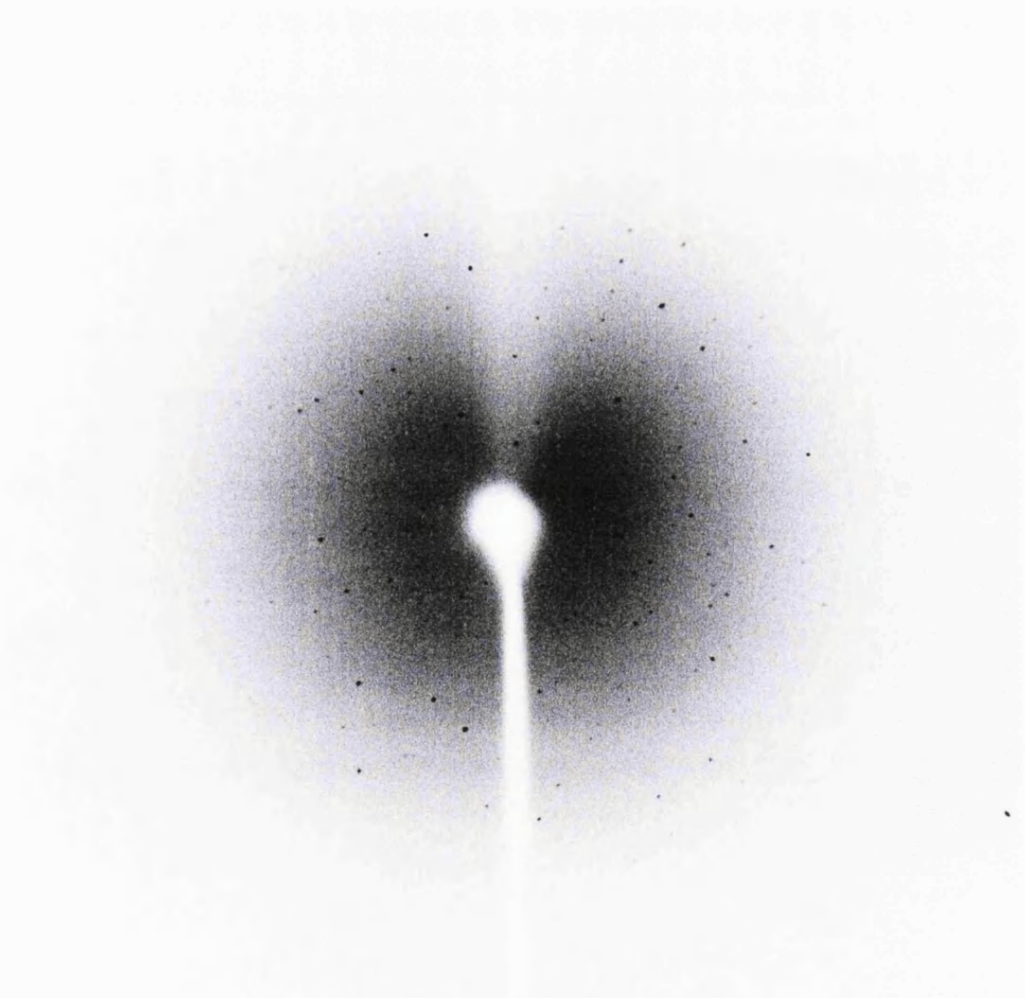


Figure 2.4: Diffraction image for the complex formed between the human telomeric quadruplex and BRACO19.

The crystal structure of the native bimolecular human telomere quadruplex formed of 12 nucleotides d[TAGGGTTAGGGT]$_2$ and determined to 2.4 Å (PDB ID 1K8P (Parkinson et al., 2002)) was used as a search model. Prior to use, the starting model was stripped of flanking thymine and adenine residues, ions and waters.

Phaser (McCoy et al., 2007) was run with the default settings. Two solutions were produced with the following similar scores (a score is the number of standard deviations above the LLG [Log Likelihood Gain] value) for RFZ (rotation function z-score) of 4.3, TFZ (translation function z-score) of 6.2 and LLG of 68 for the first solution and 4.7, 8.0 and 68 for the second solution, respectively. Visualisation of the two solutions showed that they displayed identical orientation and were translated by approximately 6.8 Å. In the I4 space group, the position of the model relative to the z axis is arbitrary. The first solution was selected for model building and refinement.

The unit cell contains eight asymmetric units. Each asymmetric unit contained one quadruplex bound to one BRACO19 molecule.

Solution verification

The solution was verified by checking the following in accordance with guidelines in Eleanor Dodson's paper (Dodson, 2008):

1. that the model makes chemical sense:

   The electron density maps were calculated using Refmac version 5.2.0019 (Murshudov et al., 1997) and visual inspection with Coot version 0.5 (Emsley & Cowtan, 2004) confirmed the position of the G-quartets, propeller TTA loops and two central ions in the quadruplex channel. It also showed that no clashes were present between symmetry copies.

2. that the model can be refined confirming that the solution generates amplitudes which agree with the observed ones:

   As model building progressed by adding the ligand, the remodelled loop, the flanking nucleotides and finally the water molecules, the R_{factor} and R_{free}

decreased stepwise with every new addition, starting from the initial R_{factor} and R_{free} values of 40.5% and 44.4% and decreased to 18.4% and 21.3% in the final model (Table 2.1).

3. if the electron density maps show features that are not part of the model:

   This was verified by the appearance of electron density volume comparable in shape and size to missing or omitted parts of the structure such as: the omitted flanking thymines and adenines, the ligand BRACO19 stacked onto the 3' G-quartet and the ions in the central quadruplex channel.

Table 2.1: R_{factor} and R_{free} decrease with model building.

Step	R_{factor} (%)	R_{free}(%)
Starting model	40.5	44.4
Fitting the potassium ions	35.8	39.9
Fitting the ligand	34.2	37.2
Fitting remodelled loop	26.0	30.4
Fitting flanking nucleotides	23.5	27.4
Fitting waters	18.4	21.3

2.3.4 Model building and structure refinement

Cycles of model building and structure refinement were carried out using Coot version 0.5 (Emsley & Cowtan, 2004) for visualisation and Refmac for refinement (Murshudov et al., 1997) programs. The default Refmac restraints dictionary was used for the standard residues (Taylor & Kennard, 1982; Saenger, 1983; Vagin et al., 2004). Two potassium ions were visible in the electron density map and were fitted into the central channel in the quadruplex.

In BRACO19, bond angle values (=128°) for the C(acridine core)-N-C(phenyl) junction at the 9th position and the C(acridine)-N(amide)-C(amide) at the 3- and 6-positions were derived from the crystal structure of *m*-AMSA (Cambridge Structural Database, CSD, ref. Code ACRAMS) and *N*-(2-Carboxybenzoyl)-L-leucine methyl ester (Cambridge Structural Database, CSD, ref. Code CIBPEI).

The maps revealed a significant electron density volume stacking on the 3' face of

the quadruplex, which corresponded in shape and size to a single BRACO19 molecule. The maps also showed the remodelled positions of the nucleotides in one of the TTA loops and were re-fitted. The flanking nucleotides forming the TATA quartet (T = thymine, A = adenine) were visible and fitted. Water molecules were initially placed automatically using the "Find Waters facility" in Coot version 0.5 (Emsley & Cowtan, 2004) using the default settings, while in later stages solvent molecules were added and removed manually. Criteria for keeping and/or adding water molecules were: the electron density volume was of spherical shape, water-sized, presented no clashes with neighbouring atoms and formed hydrogen-bonding interactions with nucleic acid atoms, ligand or water networks. The model was refined with all solvent molecules included, using all the data in the range 35.17 - 2.50 Å, to R_{factor} and R_{free} values of 18.3% and 21.3%.

Crystallographic data are shown in table 2.2.

2.3.5 Deposition in the Protein Data Bank Database

The structure solved in this work was deposited in the Protein Data Bank Database on 28th February 2008 and can be accessed using PDB ID 3CE5.

Visualisation and Image production

Coot version 0.5 (Emsley & Cowtan, 2004), Chimera version 1.3 (Pettersen et al., 2004) and PyMOL version 0.99rc6 (DeLano, 2008) were used in model visualisation and image production in this thesis.

Table 2.2: Crystallographic data.

Crystal	
Data collection	
Number of reflections	8677
Number of unique reflections	2994 (291)**
Space group	I4
Cell dimensions	
a, *b*, *c* (Å)	70.34, 70.34, 34.29
α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	35.17 - 2.50 (2.59 – 2.50)**
R_{merge}	0.065 (0.310)**
I / σ	9.8 (2.4)**
Completeness (%)	99.7 % (100.0%)**
Redundancy	2.9 (2.8)**
Refinement	
Resolution (Å)	35.18 - 2.50
Number of reflections (observed)	2992
Number of reflections (R_{free})	140
R_{factor} / R_{free}	18.3% / 21.3%
Number of quadruplexes per asymmetric unit	1
Number of ligands in asymmetric unit	1
Number of asymmetric units in unit-cell	8
Number of asymmetric units in biological unit	2
Number of atoms	
Total number of atoms	605
DNA	505
Ligand	44
Ions	2
Water	54
Average B-factor values (Å^2)	
DNA	22.85
G-quartets	19.98
Loops	27.55
Ligand	16.04
Ions	32.97
Water	32.2
RMSD	
Bond lengths (Å)	0.024
Bond angles (°)	2.813
PDB ID	3CE5

**One crystal was used. **Values in parentheses are for highest-resolution shell 2.59 - 2.50 Å.*

2.4 Experimental results

2.4.1 Quality of the model

The agreement of the model with the data is shown by the low crystallographic R values (R_{factor} = 18.3 % and R_{free} = 21.3 %).

The electron density map ($2F_o$-F_c) is continuous and contains the model in its entirety and no unexplained electron density is seen (Figure 2.5). A stereo view of the complex is shown (Figure 2.6).

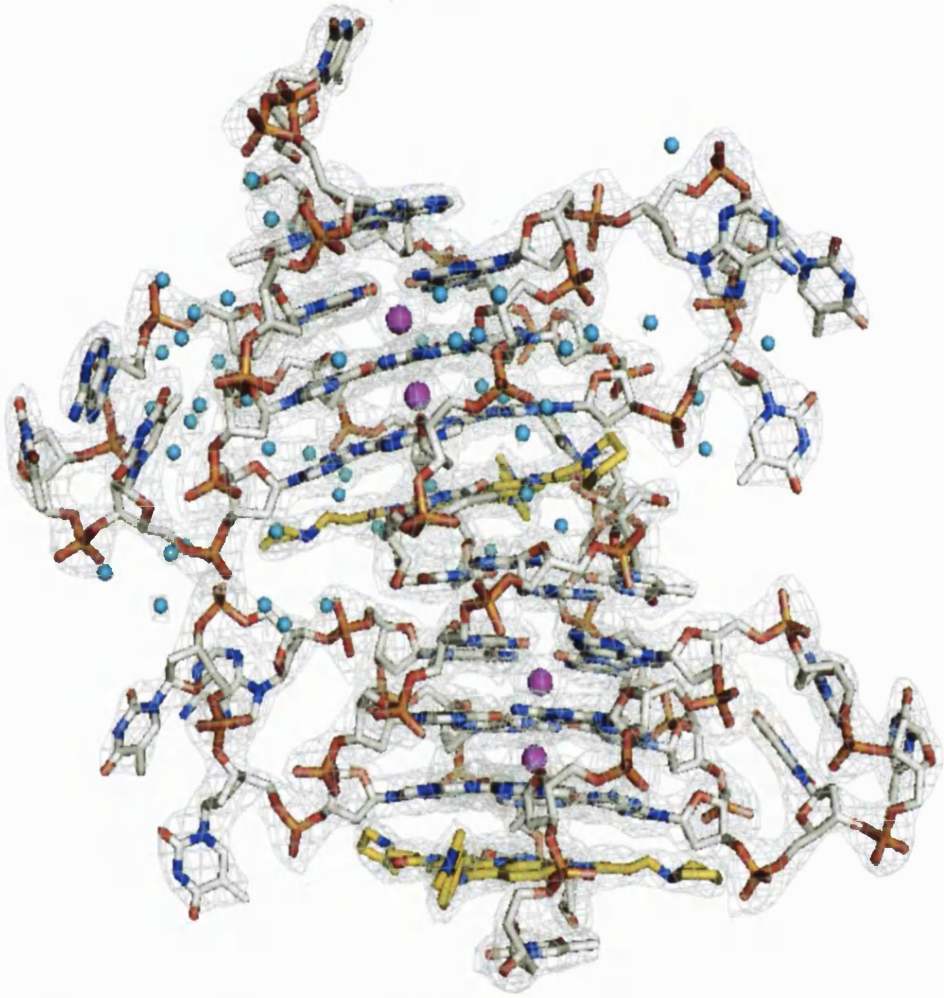


Figure 2.5: Two consecutive asymmetric units fitted into the $2F_o - F_c$ electron density map calculated at 1.5 sigma. Ligand is in yellow. Potassium ions are represented as pink spheres.

The DNA is described by a total of 505 atoms, of which 499 atoms are fully occupied, the remaining six atoms form three pairs (each atom in a pair has an occupancy of 0.5) of a phosphate group connecting the first nucleotide (thymine) of the first chain (chain A) to an adenine.

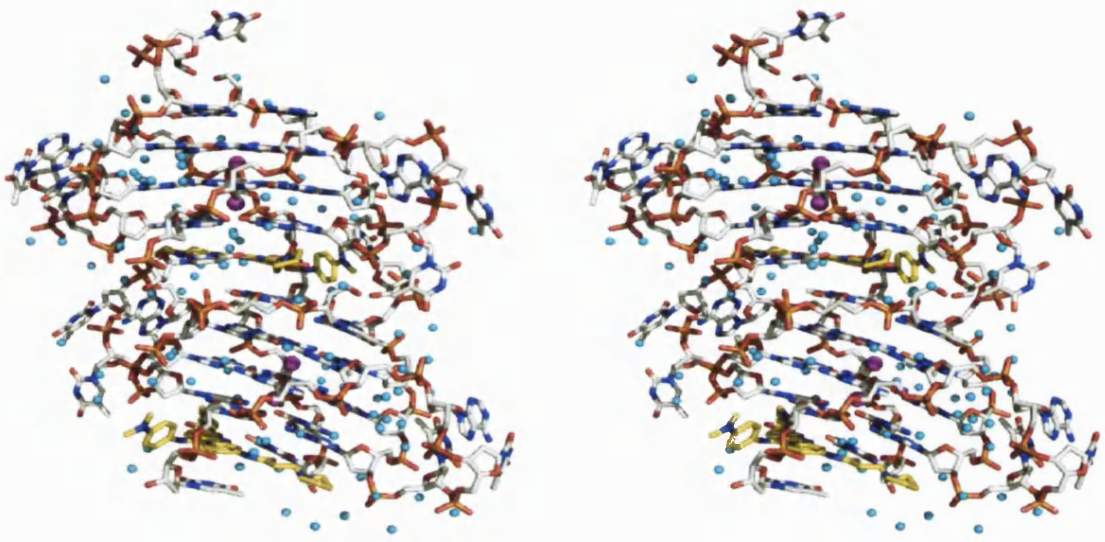


Figure 2.6: Stereo view of the bimolecular quadruplex-BRACO19 complex. The ligand is shown in yellow.

All of the ligand atoms (44 atoms), ions (2 atoms) and waters (54 atoms) are fully occupied. Average B-factor values for the DNA, G-quartets, loops, ligand, ions and waters are 22.85, 19.98, 27.55, 16.04, 32.97 and 32.2 Å^2 respectively (Table 2.2). As expected, B-factor values for the G-quartets and ligand are low indicating their static nature and average B-factor values for the loops are higher denoting their dynamic nature (Figure 2.7).

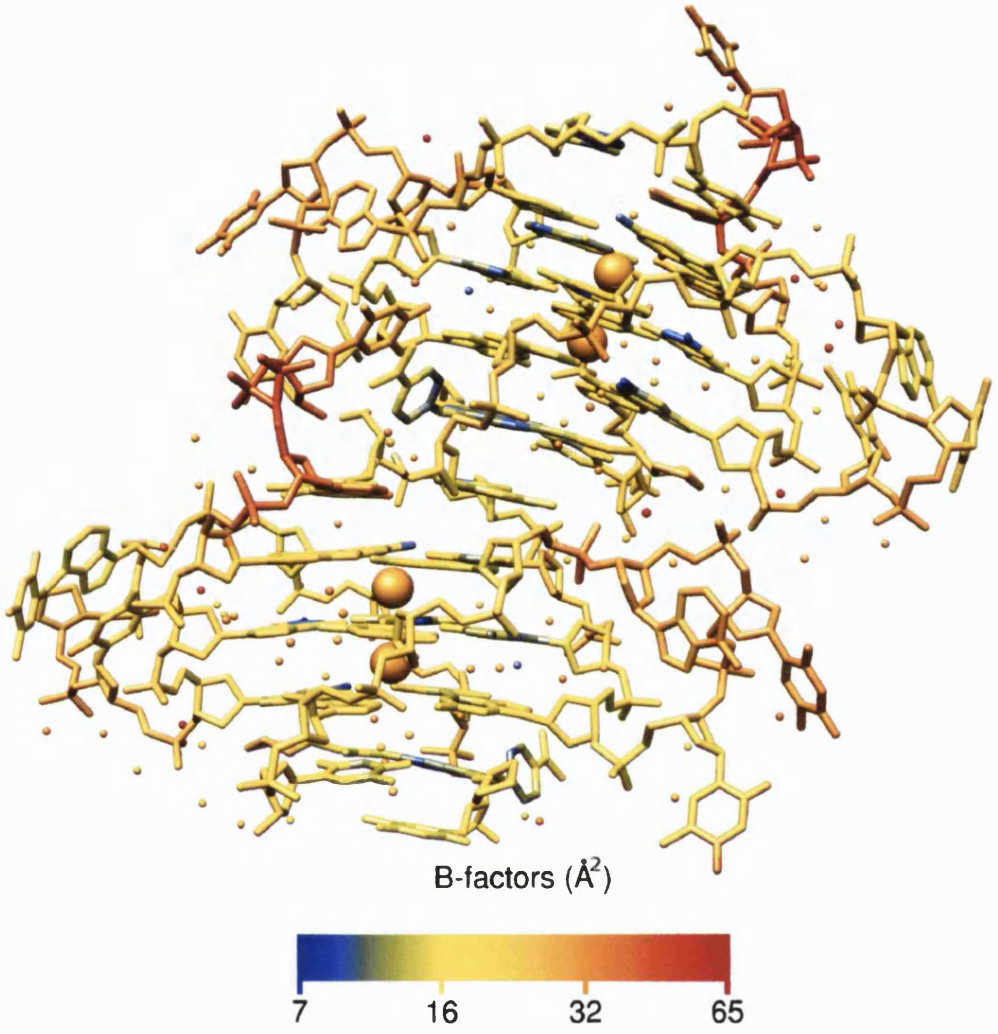


Figure 2.7: B-factor values for the complex formed between the human telomeric quadruplex and the ligand BRACO19. These are low for the G-quartets and ligand (shown in blue and yellow), increasing for the loop nucleotides (orange) and highest for some water molecules (red).

2.4.2 Overall structure description

The asymmetric unit is formed of one bimolecular quadruplex bound to one ligand. The quadruplex is formed of a dimer of the sequence d[TAGGGTTAGGGT] in a parallel-stranded conformation and propeller loops. Each DNA strand follows a right-handed helical trajectory.

Asymmetric units stack longitudinally forming quadruplex chains that extend throughout the crystal (Figure 2.8).

There are two loops, each formed of the sequence TTA, that project laterally in a propeller-type shape (also known as double-chain-reversal-loops) and interact with nearby loops throughout the crystal.

Longitudinal quadruplex stacking and lateral loop interactions comprise crystal packing interactions (discussed elsewhere in detail) (Figure 2.8).

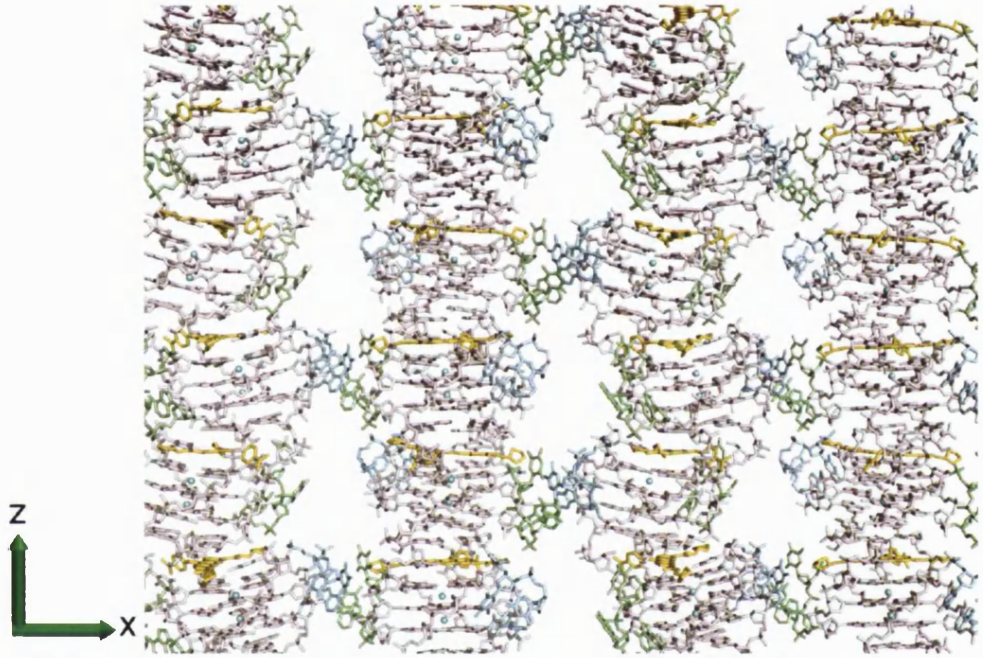


Figure 2.8: The continuous chains of quadruplexes in the crystal lattice. G-quartets are shown in pink. Ligand is shown in yellow. Lateral crystal packing interactions are formed between the loops (shown in green and blue).

The biological unit

The biological unit is formed of two parallel quadruplexes assembling in a dimer by stacking 5' to 3' i.e. the 5' face of one quadruplex stacks onto the 3' face of another, sandwiching one molecule of BRACO19 at the interface where it (as part of a coplane containing a thymine base) stacks directly onto a 3' G-quartet on one side and onto a 5' TATA-quartet on the other (Figure 2.9).

The quadruplexes in the biological unit are non-coaxially stacked resulting in the 3' G-quartet and 5' TATA-quartet at the interface being offset with regards to each other and are inclined by ca. 30° (Figure 2.10).

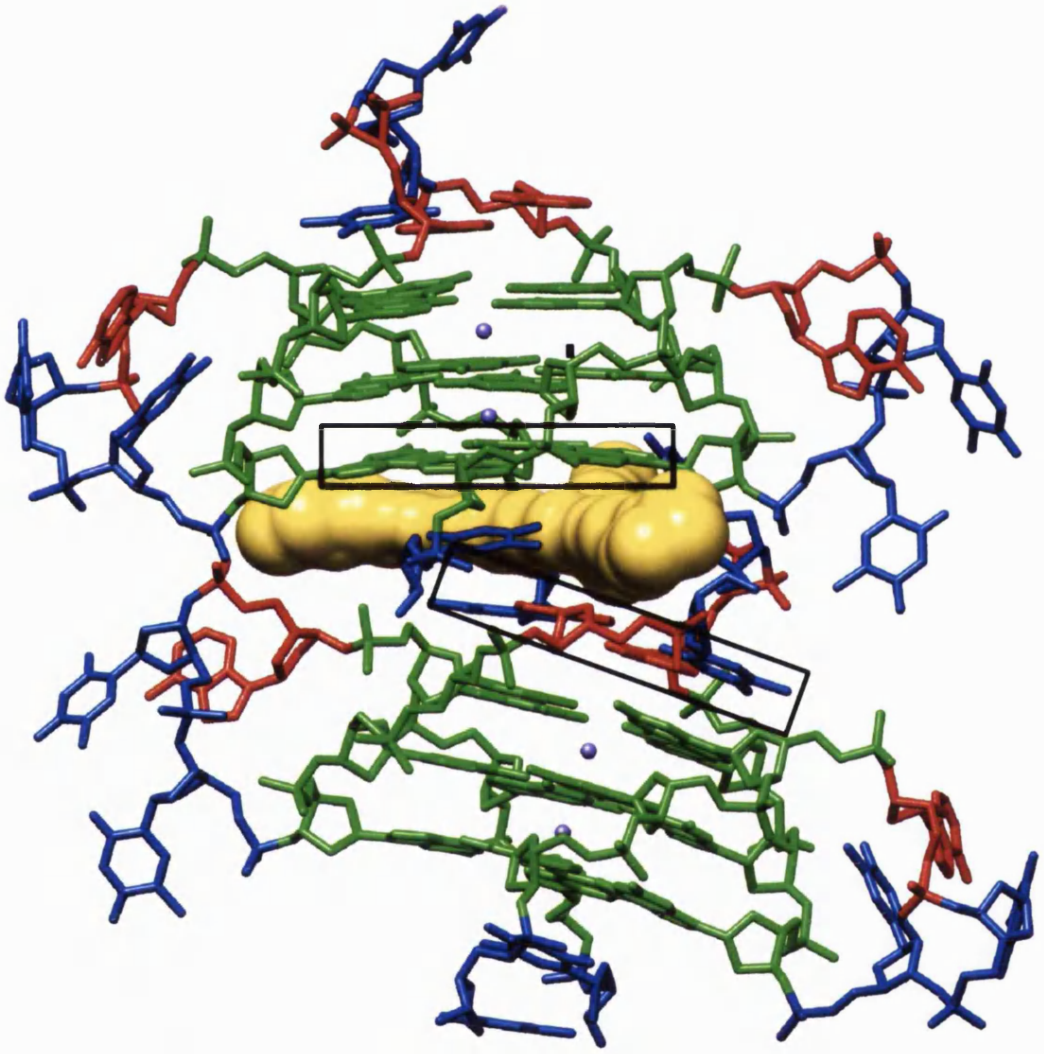


Figure 2.9: The biological unit of the complex formed between the human telomeric quadruplex and the ligand BRACO19. This is formed of two stacking 5' to 3' quadruplex structures sandwiching one BRACO19 molecule at the interface. BRACO19 (shown in surface representation in yellow) stacks directly onto a 3' G-quartet (shown in green in the black box above the ligand) and onto a 5' TATA-quartet on the other side (shown in blue and red in the black box below the ligand). Water molecules are omitted for clarity. Guanine = green, adenine = red and thymine = blue. Potassium ions are shown as pink spheres.

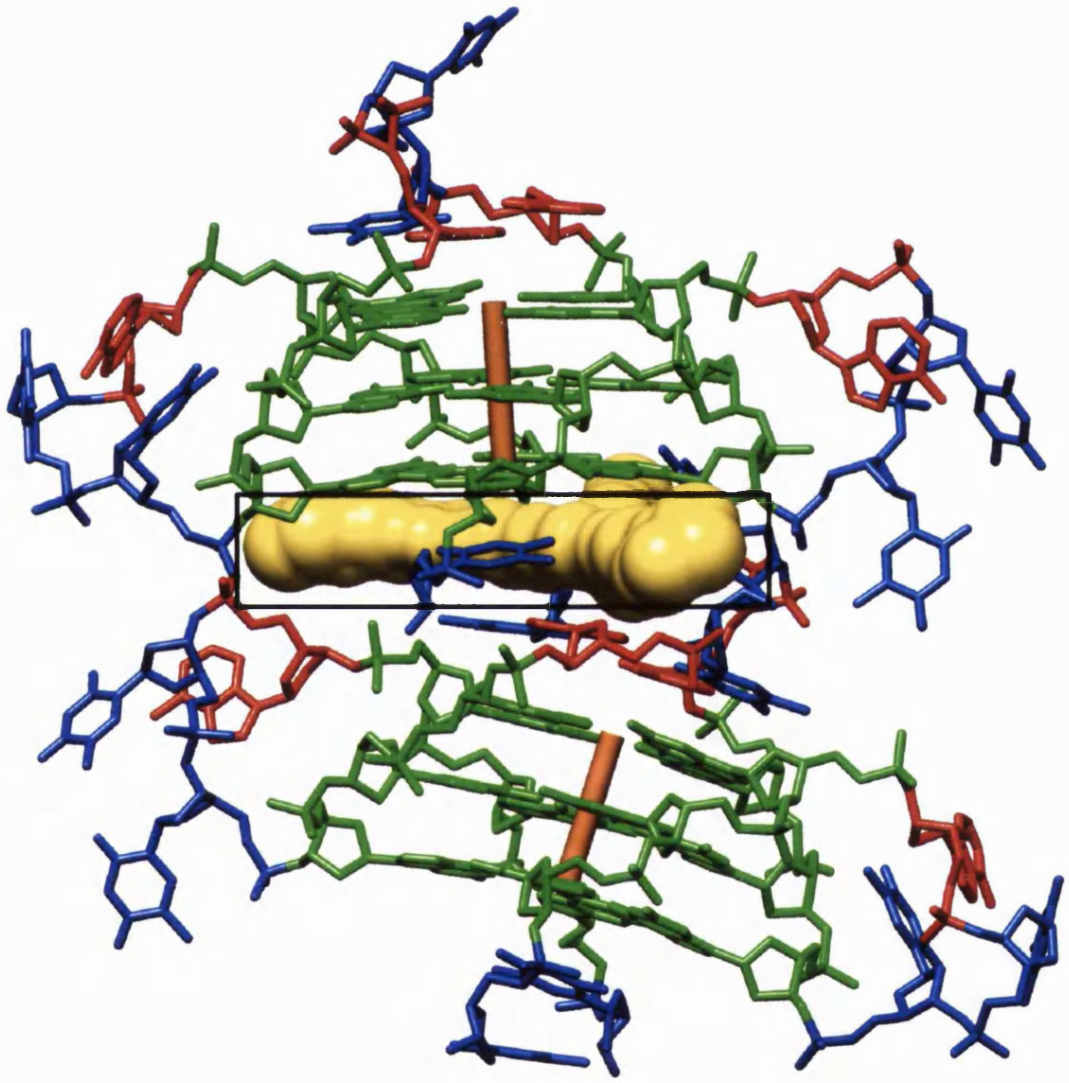


Figure 2.10: The two quadruplexes in the asymmetric unit are non-coaxially stacked. The 3' G-quartet and the 5' TATA-quartet are offset with respect to each other and inclined by ca. 30 ° (axes represented in orange cylinders). The ligand (in surface representation in yellow) and a flanking thymine (shown in blue sticks) from the 3' quadruplex face (black box) form a coplane stacked at the interface between the two quadruplexes.

Loops

One of the loops is identical in conformation to the loop structure observed in the native human telomere G-quadruplex crystal structure (PDB IDs 1K8P and 1KF1 (Parkinson et al., 2002)) and will be referred to henceforth as "traditional". The other loop assumes a "remodelled" conformation (Figure 2.11).

Each remodelled loop interacts with a traditional loop forming lateral crystal packing interactions perpendicular to the axis of the central channel of quadruplex chains throughout the crystal lattice (Figure 2.8 and Figure 2.13).

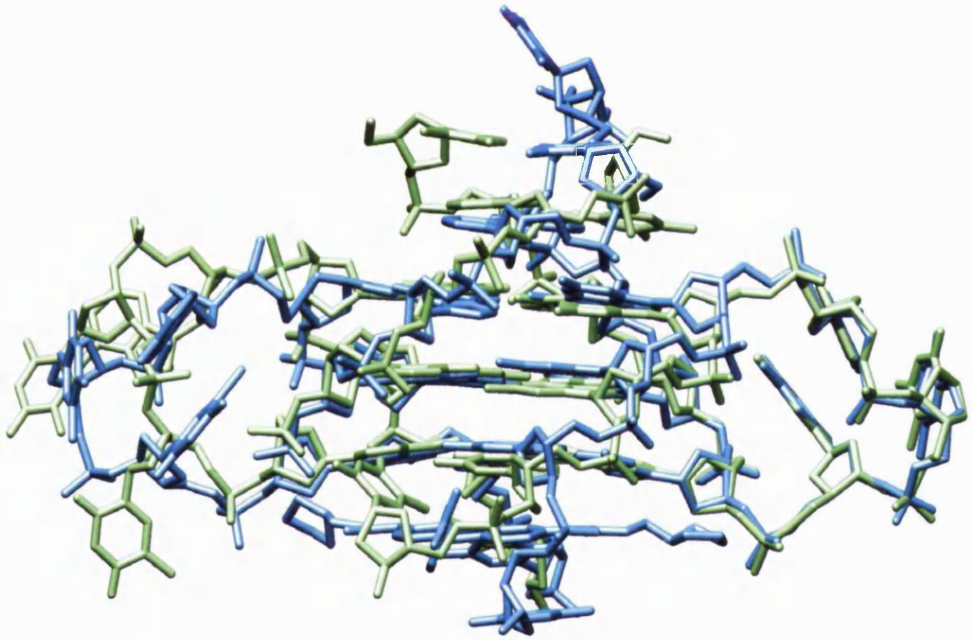


Figure 2.11: The remodelled loop. The aligned structures of 3CE5 (blue) and 1K8P (green) showing the perfectly superposed traditional loop structure (right) and the new conformation of the remodelled loop shown in green (left).

Loop 1 or the traditional loop:

The adenine (A8) swings back into position between the two thymines (T6 and T7) (Figure 2.12). The first thymine (T6) and adenine (A8) adopt the C2'-*endo* sugar pucker, whereas the second thymine (T7) adopts a C3'-*endo* sugar pucker. This loop interacts laterally with the remodelled loop of a symmetry-related quadruplex contributing to crystal packing interactions.

Loop 2 or the remodelled loop:

The first thymine (T18) flips out to stack onto the second thymine (T7) from the traditional loop (Figure 2.12). The adenine (A20) stacks onto the adenine (A8) which in turn stacks onto the thymine (T6) from the traditional loop.

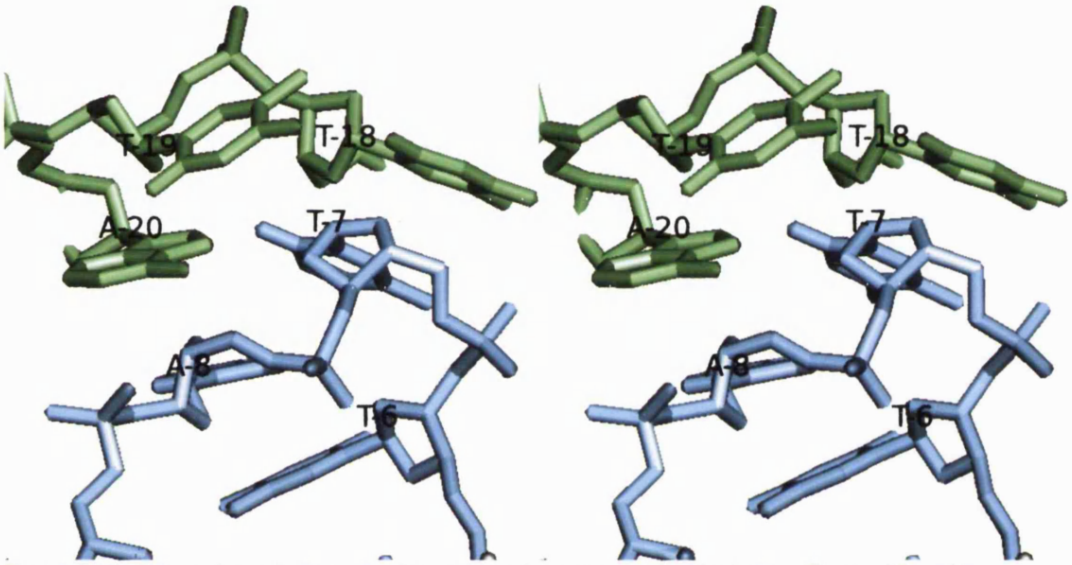


Figure 2.12: Stereo view of crystal packing interactions via quadruplex loops. The traditional loop formed of nucleotides T6, T7 and A8 (blue) and the remodelled loop formed of T18, T19 and A20 (green).

Crystal packing

Stacking asymmetric units pile up forming a continuous column of 5' to 3' stacked quadruplex chains repeating along the crystal lattice (Figure 2.8).

The flanking 5' and 3' thymine and adenine residues along with the ligand mediate longitudinal crystallographic contacts whilst the TTA loop nucleotides contribute to the lateral ones (Figure 2.8).

Throughout the crystal lattice each quadruplex interacts laterally with a symmetry-related quadruplex via each of its two loops, in an arrangement where each traditional loop interacts with a remodelled loop. The remodelled loop accommodates the new crystal packing arrangement described here (Figure 2.12 and Figure 2.13), whereas in the native human telomere quadruplex (PDB ID 1KF1(Parkinson et al., 2002)), the traditional loops interact with the G-quartet face

of a symmetry-related quadruplex.

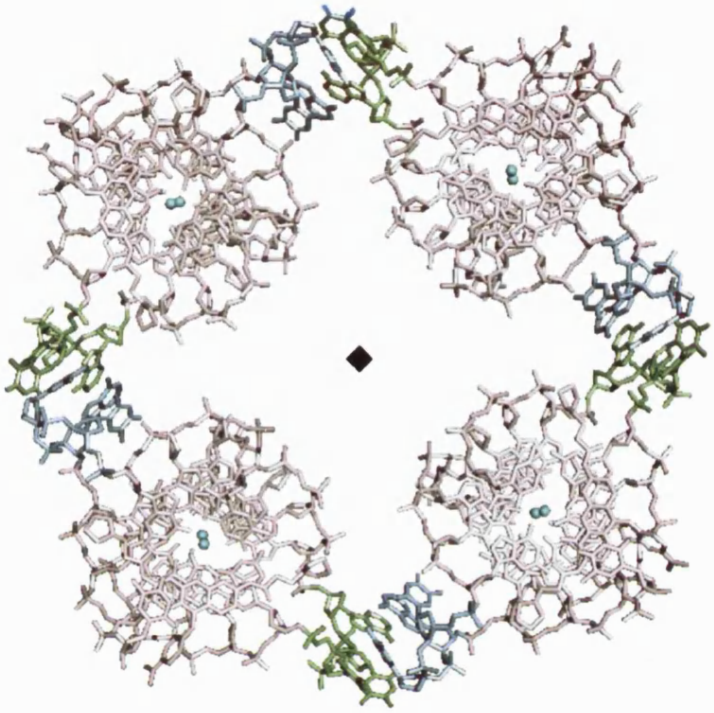


Figure 2.13: The 4-fold symmetry is shown in a projection down the c axis. Four quadruplexes interact via their lateral loops forming essential crystal packing interactions. Key: Guanine = pink, traditional loop = blue, remodelled loop = green and potassium ions = cyan spheres.

The flanking thymine and adenine residues along with the ligand contribute to longitudinal packing interactions. The ligand and one thymine (from the 3' quadruplex face) form a coplane (Figure 2.10) which stacks directly onto the 3' G-quartet on one side and onto a 5' TATA-quartet on the other.

Backbone nucleic acid torsion angles

The values for the backbone torsion angles alpha (α), beta (β), gamma (γ), delta (δ), epsilon (ε), zeta (ζ) and the glycosidic angle chi (χ) (Figure 2.14) (Schneider et al., 1997) were calculated using the programme 3DNA (Lu & Olson, 2003) (Table 2.3).

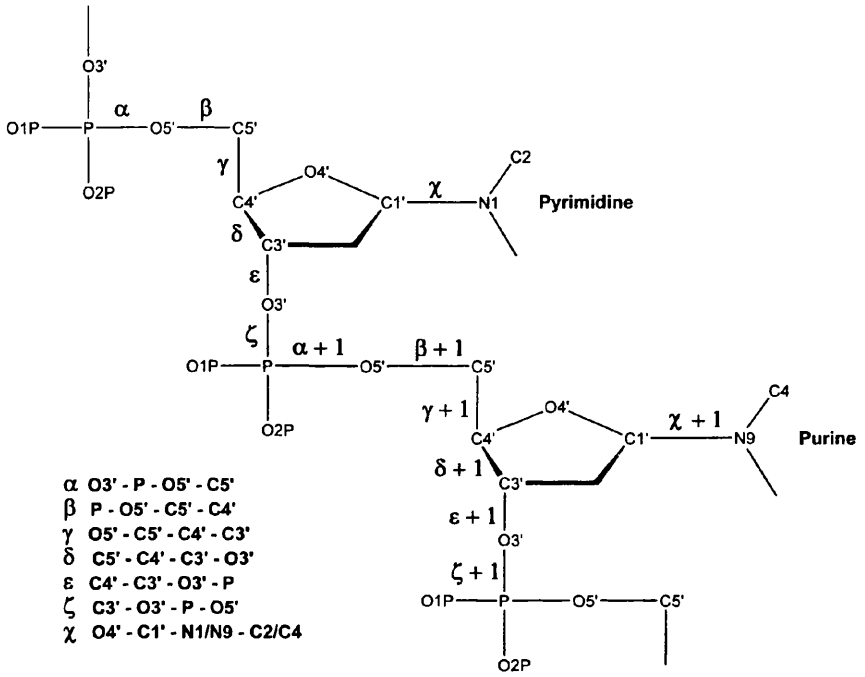


Figure 2.14: Backbone torsion angles (adapted from Schneider et al., 1997).

Throughout the discussion in this section, comparisons are made between the bimolecular quadruplex-BRACO19 complex (PDB ID 3CE5) reported in this work and both of; the native bimolecular (PDB ID 1K8P) and unimolecular (PDB ID 1KF1) quadruplexes (Parkinson et al., 2002). Also comparisons are made with the bimolecular quadruplex in complex with porphyrin (PDB ID 2HRI (Parkinson et al., 2007)) and in complex with a naphthalene diimide (PDB ID 2CCO (Parkinson et al., 2008)). Also, the unimolecular quadruplex – naphthalene diimide complex is included in the comparisons (PDB ID 3CDM (Parkinson et al., 2008)). To allow for easy and unambiguous comparisons, the non-common flanking nucleotides and the second loop in the unimolecular structure were not considered in the comparisons of backbone torsion angles. The three sequences were aligned and renumbered (Table 2.4).

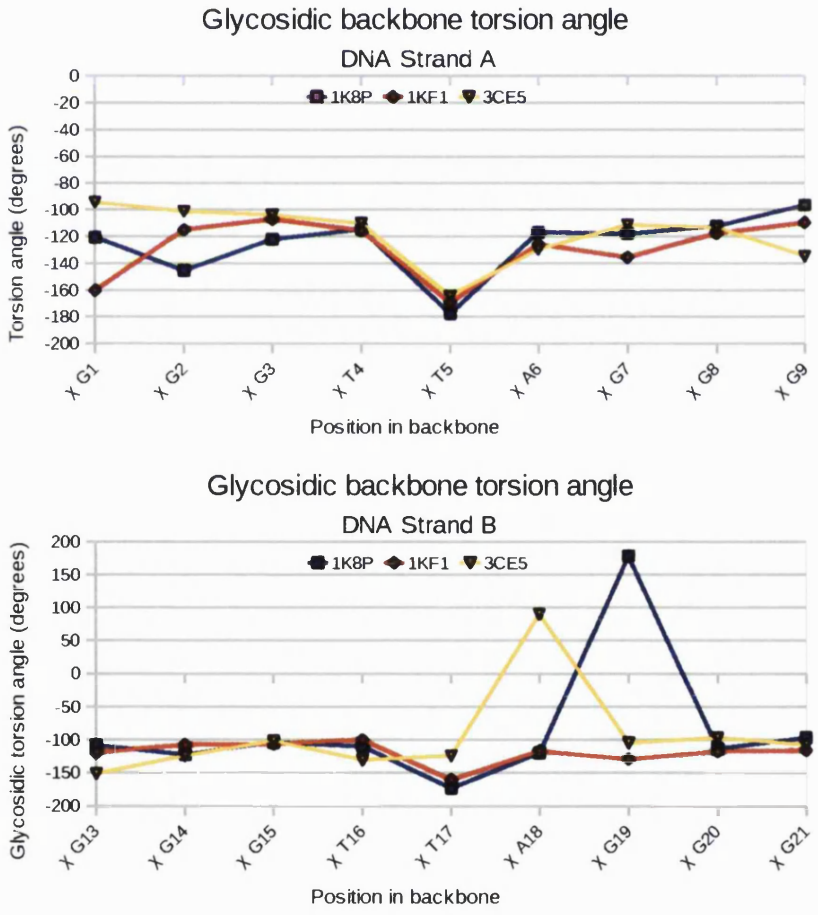
Table 2.3: Backbone torsion angles for the bimolecular quadruplex-BRACO19 complex calculated using the 3DNA software (Lu & Olson, 2003), esds estimated at 2-4°.

DB ID 3CE5		Backbone torsion angles in degrees (°)						
Sequence	Aligned	Alpha (α)	Beta (β)	Gamma (γ)	Delta (δ)	Epsilon (ε)	Zeta (ζ)	Chi (χ)
1 T	T-2	n/a	n/a	151	140	-46	85	-143
2 A	A-1	150	113	155	126	-57	125	-69
3 G	G1	-112	170	39	152	-180	-119	-94
4 G	G2	-37	-170	21	140	176	-114	-101
5 G	G3	-23	-177	14	140	-140	67	-104
6 T	T4	70	-173	-169	123	-115	-69	-110
7 T	T5	-130	153	6	82	-156	-56	-164
8 A	A6	-76	-161	63	146	-93	-81	-130
9 G	G7	-115	-154	-21	127	-177	-102	-111
10 G	G8	-61	-179	37	132	-168	-107	-113
11 G	G9	-68	-172	19	97	-154	-73	-135
12 T	T10	-57	-145	48	126	n/a	n/a	-95
13 T	T11	n/a	n/a	11	145	-133	148	-85
14 A	T12	103	-155	158	144	-161	-87	-161
15 G	G13	119	-176	-173	105	-177	-76	-151
16 G	G14	-75	-161	55	136	-173	-102	-124
17 G	G15	-53	-169	29	149	-122	72	-102
18 T	T16	84	154	34	148	-123	138	-131
19 T	T17	-75	169	50	144	-95	58	-124
20 A	T18	151	-176	140	144	-51	-133	90
21 G	G19	42	-131	58	142	166	-97	-105
22 G	G20	-36	-170	20	142	180	-108	-97
23 G	G21	-70	-174	46	139	-176	-96	-107
24 T	T+1	-24	-179	-16	89	n/a	n/a	-128

Table 2.4: Alignment and renumbering of residues in quadruplex structures.

PDB ID	Residue name and number as per PDB files downloaded from the PDB website: http://www.rcsb.org																						
1K8P	1BRU	2A	3G	4G	5G	6BRU	7T	8A	9G	10G	11G	12T	13BRU*	14A*	15G*	16G*	17G*	18BRU*	19T*	20A*	21G*	22G*	23G*	24T*
1KF1		1A	2G	3G	4G	5T	6T	7A	8G	9G	10G	11T	12T	13A	14G	15G	16G	17T	18T	19A	20G	21G	22G	
3CE5	1T	2A	3G	4G	5G	6T	7T	8A	9G	10G	11G	12T	13T*	14A*	15G*	16G*	17G*	18T*	19T*	20A*	21G*	22G*	23G*	24T*
2HRI	1T	2A	3G	4G	5G	6T	7T	8A	9G	10G	11G			12T*	13A*	14G*	15G*	16T*	17T*	18A*	19G*	20G*	21G*	
3CCO	1T	2A	3G	4G	5G	6T	7T	8A	9G	10G	11G	12T	1T*	2A*	3G*	4G*	5G*	6T*	7T*	8A*	9G*	10G*	11G*	12T*
3CDM	1T	2A	3G	4G	5G	6T	7T	8A	9G	10G	11G	12T	13T	14A	15G	16G	17G	18T	19T	20A	21G	22G	23G	24T
Aligned																								
	T-2	A-1	G1	G2	G3	T4	T5	A6	G7	G8	G9	T10	T11	A12	G13	G14	G15	T16	T17	T18	G19	G20	G21	T+1

The guanines participating in the G-quartets adopt rigid geometries due to the eight Watson-Crick and Hoogsteen hydrogen bonds. This is evident in the low variation observed in the measured values for the glycosidic backbone torsional angle chi (χ) (Figure 2.15). The crystal structure of the native unimolecular human telomeric quadruplex have a mean glycosidic angle for the G-quartets of -120±15° (Neidle & Parkinson, 2008). The mean for the complex is closely related at -112±17°. This is also close to the mean for a native tetramolecular quadruplex (PDB ID 1S47 (Cáceres et al., 2004)) which is not subject to the restraining effects of loops connecting the G-quartets at -117±20° (Neidle & Parkinson, 2008) .



The value for the glycosidic angle for A18 in the complex (yellow line - Figure 2.15) displays a large peak which reflects the adopted *syn* conformation for this

nucleotide in comparison to the *anti* conformation present in the unimolecular quadruplex and also coincides with the remodelled loop (formed of T16, T17 and A18) in the complex. The same figure also shows a peak for G19 in the native bimolecular quadruplex (blue line). However, as the occupancies for the atoms forming the loop connected to G19 in the native bimolecular quadruplex (PDB ID 1K8P) equal zero (Parkinson et al., 2002) and the observation that the geometry of the guanines directly connected to the loop nucleotides is affected by the loop conformation (Neidle & Parkinson, 2008), it is not possible to comment on this.

The backbone dihedral angles trends in the complex generally follow those observed in the native crystal structures for the strand containing the traditional loop (Figure 2.16). However, the backbone torsion angle values for the remodelled loop deviate from those of the native crystal structures for the loop forming nucleotides i.e. T16, T17 and A18 (Figure 2.17) and also for G19 which is directly connected to this loop. This effect on G19 is due to the local strain imposed by the changes in backbone conformation in the remodelled loop.

Comparisons were made between the values for the glycosidic angle chi (χ) for the loop structure of the BRACO19-quadruplex complex and those for the loops in the native (PDB ID 1KF1) and in other ligand-quadruplex listed below (Table 2.5).

Table 2.5: PDB IDs of structures used in comparisons.

PDB ID	Quadruplex	Ligand	Reference
1KF1	Unimolecular	n/a	Parkinson et al., 2003
2HRI	Bimolecular	Porphyrin	Parkinson et al., 2007
3CCO	Bimolecular	Naphthalene dlimide	Parkinson et al., 2008
3CDM	Unimolecular	Naphthalene diimide	

Comparison of chi angle values revealed that the glycosidic angle for the first and second thymine in the loops (traditional and remodelled) in the BRACO19-quadruplex complex at -131±1° and -124±1°, respectively, are closely related to those of the compared structures (range of -174 to -84° for T1 and -169 to -99° for T2). In contrast to the chi value for the third nucleotide in the remodelled loop (an acenine: A3). At 90°, it lies outside the range for the equivalent nucleotide in

compared crystal structures (-147 to -108°). This shows that the conformation of the first, second and third nucleotides in the loops of the compared structures have closely related chi angle values reflecting similar conformations. Only one nucleotide, A3 in the remodelled loop shows significant changes. This nucleotide adopts the *syn* conformation in contrast to the *anti* conformation adopted in all other loops (Figure 2.18).

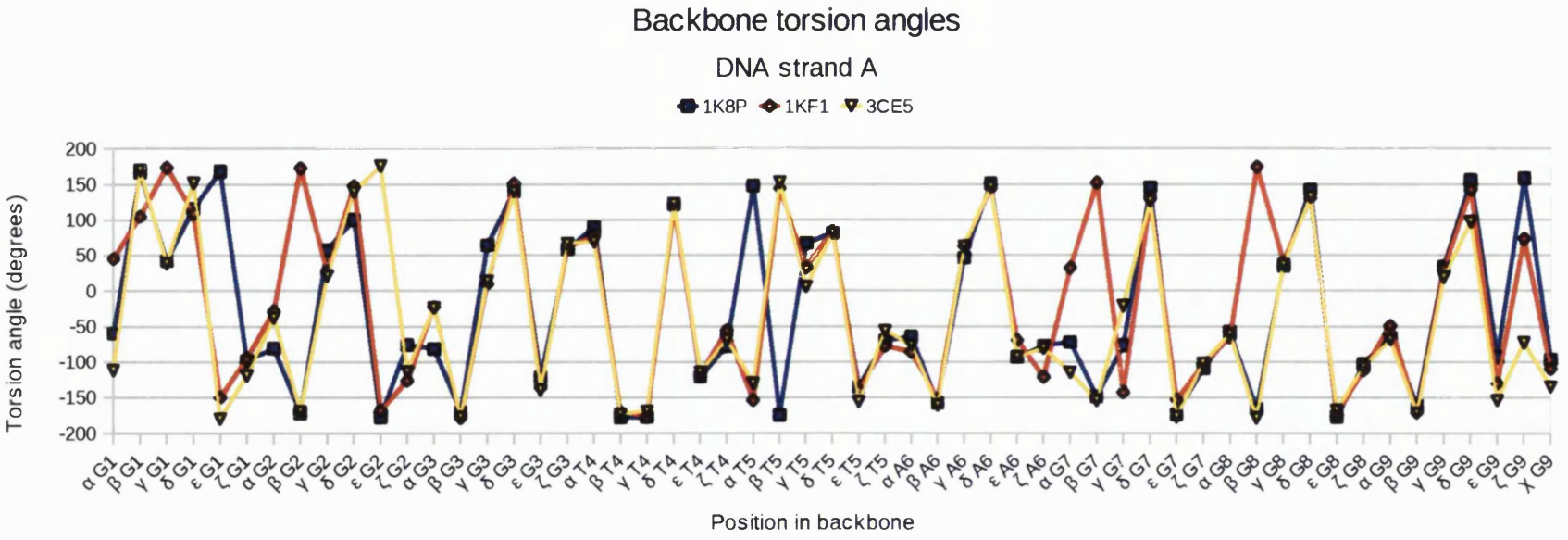


Figure 2.16: Comparison of backbone torsion angles for strand A. This strand contains the traditional loop and this is reflected in the closely related values for the torsion angles measure for the bimolecular quadruplex-BRACO19 complex (3CE5 shown in yellow) and for both of the native bimolecular and unimolecular quadruplex sequences (1K8P shown in blue and 1KF1 shown in red, respectively) for strand A.

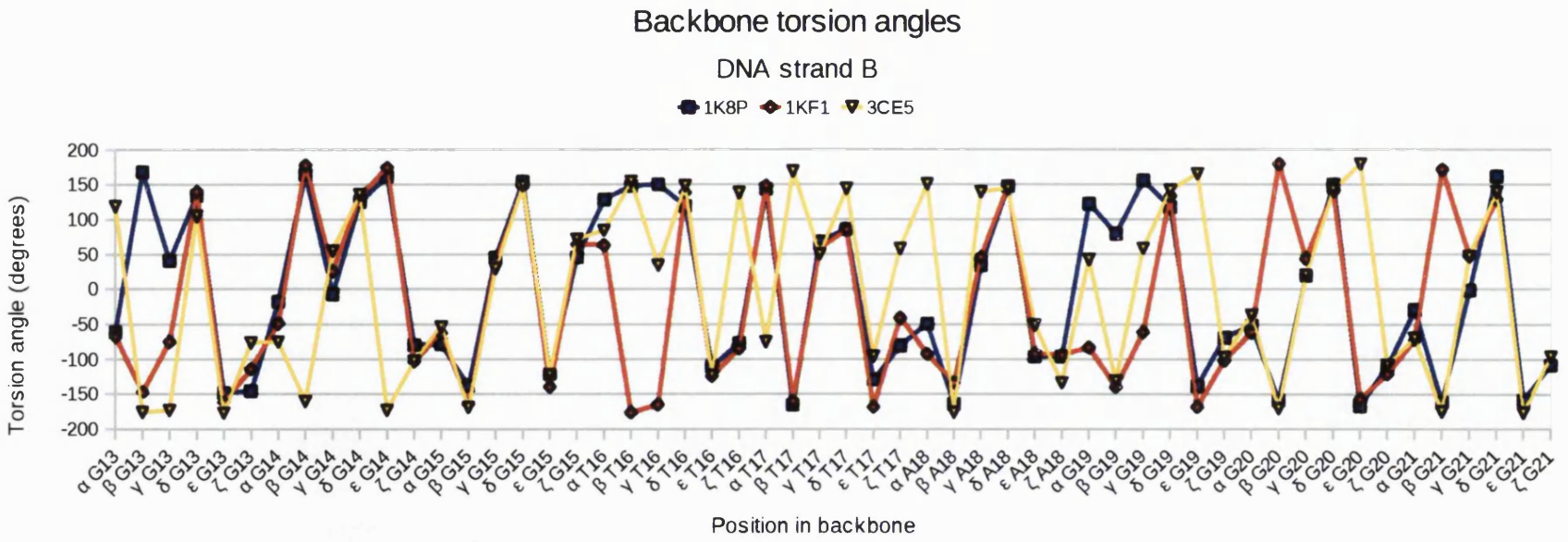


Figure 2.17: Comparison of torsion angle values for strand B. This strand contains the remodelled loop and this is reflected in the deviation of torsion angle values (yellow) from those for the same loop in the native structures (blue and red). This is especially evident for the three nucleotides forming the loop (T16, T17 and A18) and for the immediate nucleotide G19. G19 displays the expected deviation from the native structures due to the strains imposed by the change in the geometry of the remodelled loop.

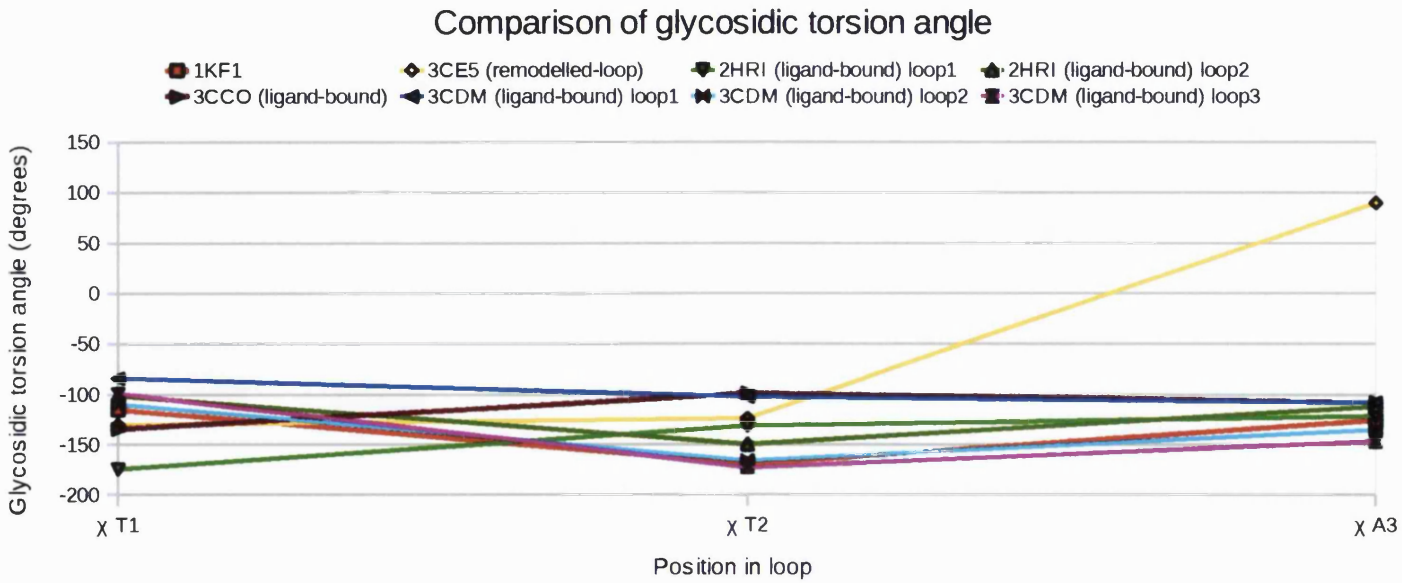


Figure 2.18: Comparison of the values for the glycosidic angle chi (χ) for the loop structure of the BRACO19-quadruplex complex and those for the loops in the native (Parkinson et al, 2003)(1KF1 shown in red) and in other ligand-quadruplex complexes; a bimolecular quadruplex in complex with porphyrin (Parkinson et al., 2007) (light green for loop1 and dark green for loop2), with one type of a naphthalene diimide ligand (dark red) and for another (Parkinson et al., 2008) in complex with a unimolecular quadruplex (magenta, light and dark blue). Chi for the third nucleotide A3 is 90° compared to a mean of -122°±15 for other complexes). A3 adopts the syn conformation in contrast to the anti conformation adopted in the native quadruplex and its ligand-complexes.

2.4.3 The G-quartets

A stack of eight G-quartets interspersed by one TATA-quartet and one ligand-thymine coplane constitute the framework of the biological unit. In each G-quartet, guanine bases assemble via Hoogsteen and Watson-Crick hydrogen bonds, are in the *anti* conformation and roughly follow a four-fold rotation symmetry along the helix axis (Figure 2.19).

Standard base-pair geometry descriptors can be determined for quadruplex structures such as quartet-quartet helical twist and helical rise using the programme 3DNA (Lu & Olson, 2003).

The guanine quartets have quartet-quartet helical twist angles of 26° and 28° for the first and second G-quartet steps in the 5' to 3' direction. This is in accordance with the reported values of approximately 28-30° (Neidle & Parkinson, 2008). The helical rise values are 3.1 and 3.2 Å, respectively. This is closely related to the reported values of 3.2 to 3.3 Å (Neidle & Parkinson, 2008).

This structure provides the first instance of a quadruplex crystal structure where 5' to 3' quadruplex stacking is observed, providing a working framework for long quadruplex chains i.e. more than one quadruplex in tandem, by carefully connecting the adjacent DNA strands.

Also, it is possible to examine other quartet step parameters. For example, the top quadruplex shifts by approximately 7 Å relative to the bottom quadruplex as the two quadruplexes incline relative to each other by about 30° effectively creating a one sided opening where the ligand fits.

The range for hydrogen bond distances present in the G-quartets is 2.4 Å to 3.2 Å (Figure 2.19).

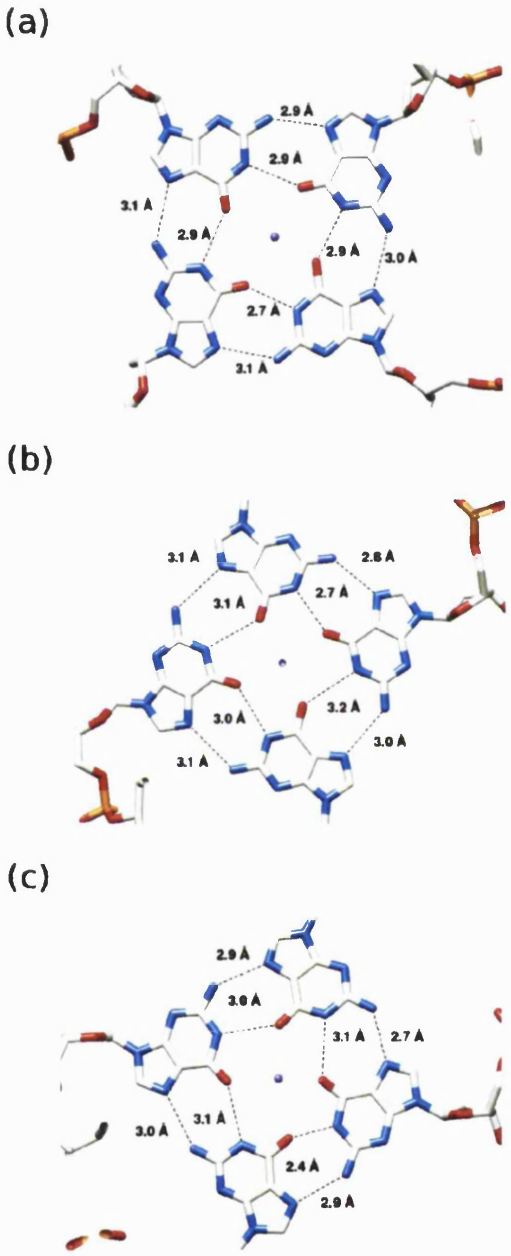


Figure 2.19: Top view showing the distances for hydrogen bonds (represented with black dashed lines) for the 5' top quartet (a), middle quartet (b) and 3' bottom quartet (c) . The twist of the quartets are shown relative to each other.

2.4.4 The TATA-quartet

The ligand-thymine coplane and the TATA-quartet plane form a two step transition at the interface connecting the 3' face of one quadruplex in the biological unit to the

5' face of the other quadruplex, resulting in uninterrupted planar stacking between the two quadruplexes. The TATA-quartet is formed of nucleotides belonging to both quadruplexes as follows:

1. a thymine and an adenine base belonging to the same strand from the 5' face of one quadruplex.
2. an adenine base belonging to the other DNA strand of the same 5' quadruplex face.
3. a thymine from the 3' quadruplex face.

Each thymine and adenine nucleotide forms one reverse Watson-Crick base pair (Figure 2.20).

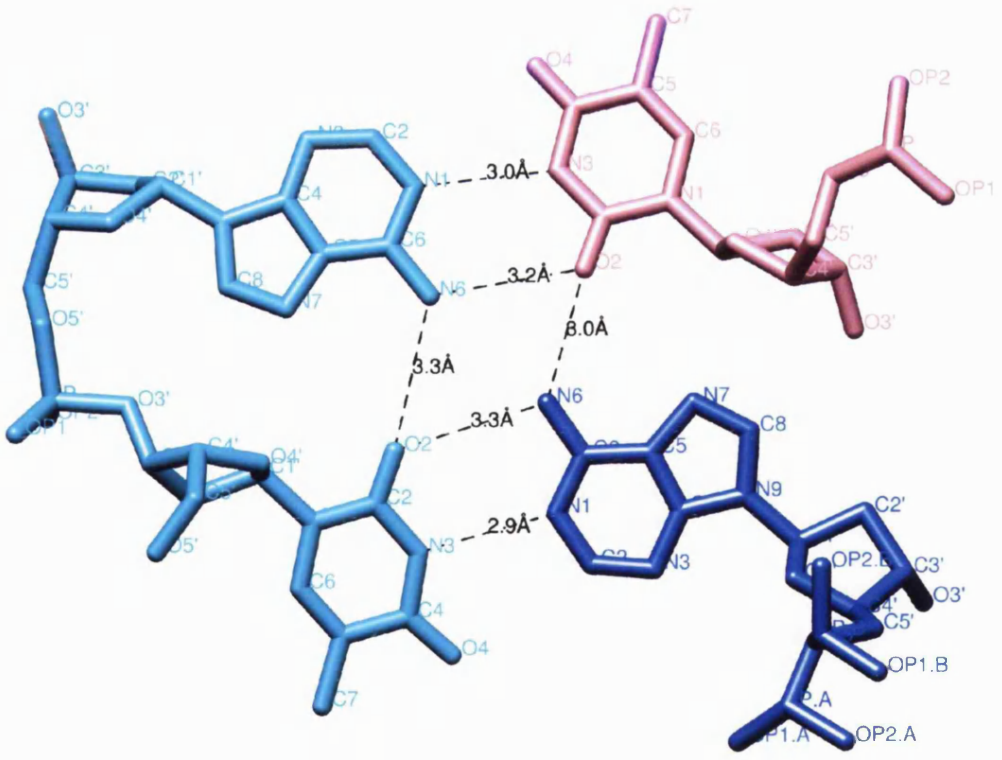


Figure 2.20: The TATA-quartet showing two reverse Watson-Crick T•A base pairs. Hydrogen bonds shown in black dashes and bond distances in Angstroms. Atom names are shown. Nucleotides from the same 5' DNA strand are shown in light blue. The adenine from the other DNA strand belonging to the same 5' face quadruplex is shown in navy. The thymine belonging to the 3' face quadruplex in shown in pink.

2.4.5 Groove types

Grooves can be divided into: wide, medium and narrow. Grooves can be also defined according to whether a particular groove is straddled by a loop or not. This gives two types of grooves: loop-bound and loop-free. The structure presented here contains two loop-bound grooves and two loop-free grooves.

A loop-bound groove describes a groove where a TTA loop connecting a guanine in the bottom G-quartet to a guanine in the top G-quartet. A loop-free groove describes a groove that does not contain such a loop and thus is open to the solvent.

Grooves are medium in size compared to grooves in the anti-parallel structure (PDB ID 143D) (Figure 2.21). The groove widths show little variation ranging between 9.7 to 10.9 Å (Figure 2.22). Groove width is measured as the shortest

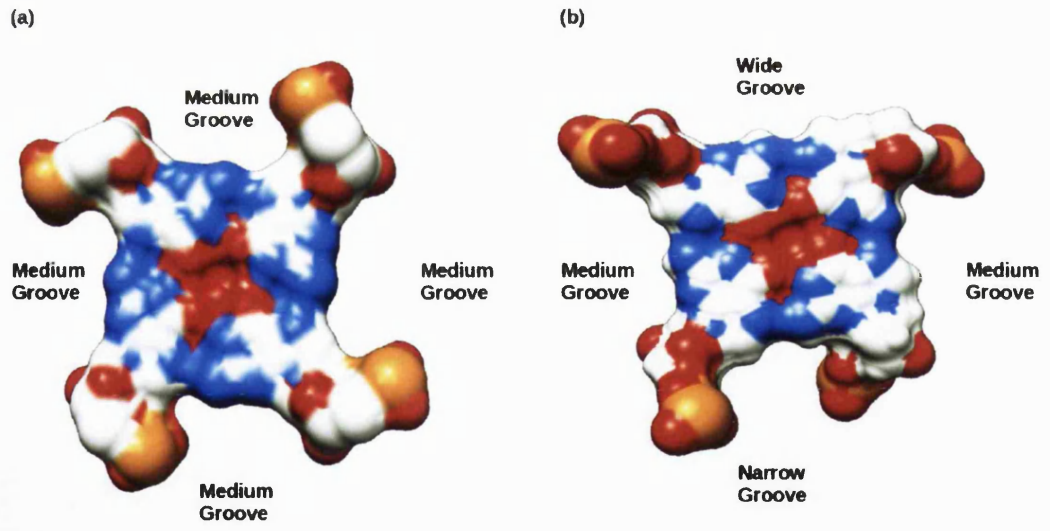


Figure 2.21: Comparison of groove widths. The equivalent four grooves in the BRACO19-quadruplex complex (a) compared to the narrow and wide grooves (b) seen in the anti-parallel quadruplex (PDB ID 143D (Wang & Patel, 1993)).

distance between backbone phosphate atoms that form the walls of grooves (Figure 2.22 and Figure 2.23).

The loop-free grooves are defined by three boundaries: two walls and one floor. Its

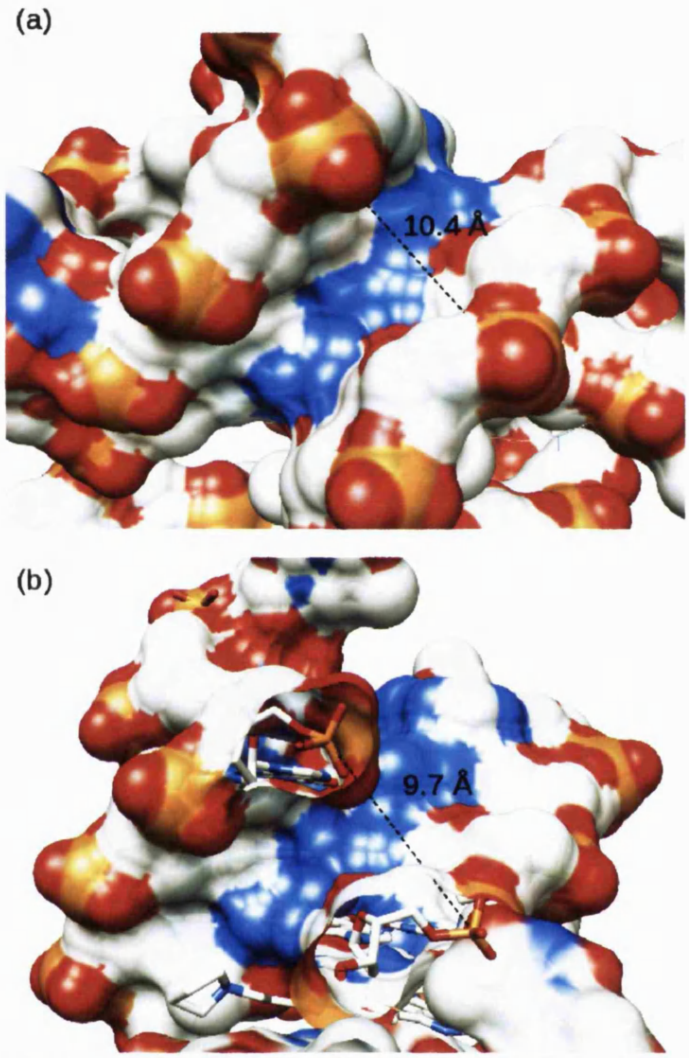


Figure 2.22: Groove types. Loop-free (a) and loop-bound grooves (b). The surface representation shows nitrogen atoms forming the floor of the grooves (blue) and backbone phosphate and oxygen atoms forming the walls (orange and red respectively). The loop in (b) is clipped for clarity. The groove width measurements are shown in dashed lines.

walls are formed of negatively charged DNA backbone (phosphate OP1, OP2 and sugar O4') and its floor of positively charged edges of the G-quartet-forming guanine bases (N2, N3 and N7) (Figure 2.22). The loop-bound grooves are defined by four boundaries: two walls, a floor and a boundary composed of the straddling loop. The straddling loop forms a relatively flexible boundary that caps the groove.

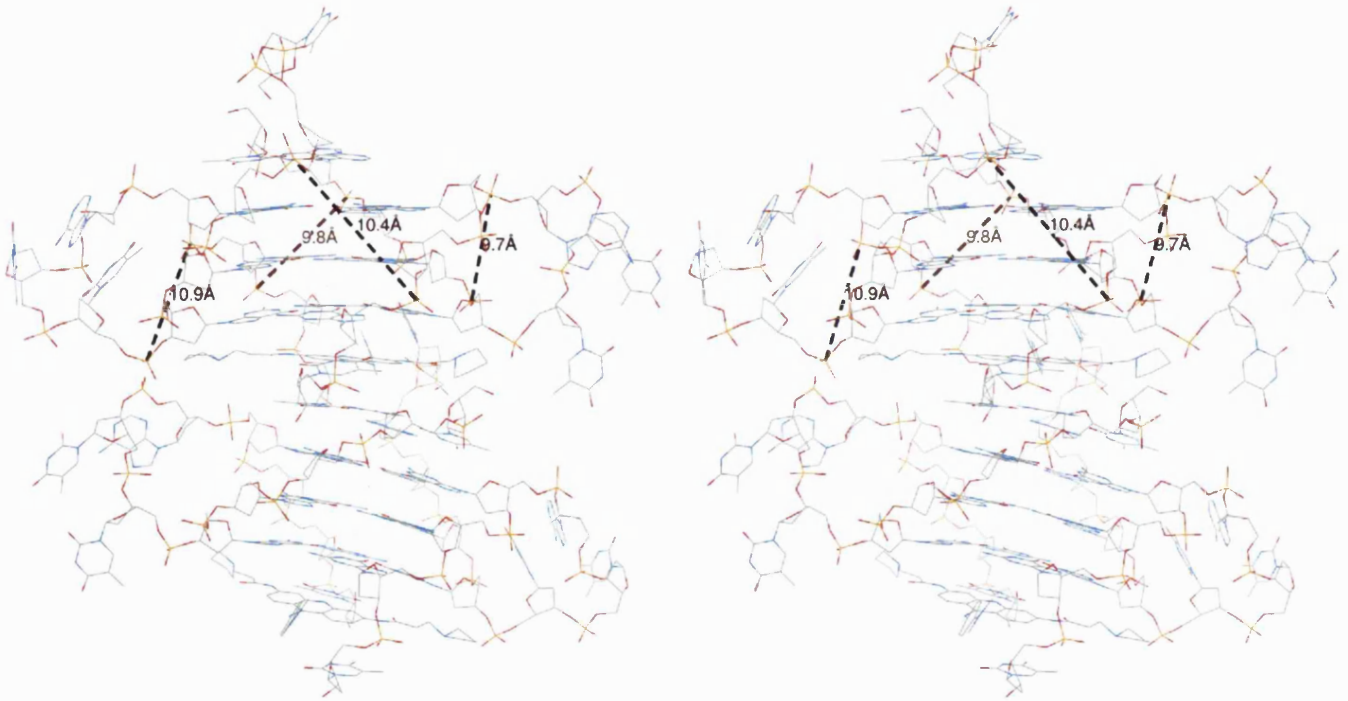


Figure 2.23: Stereo view of the complex showing the groove widths for all four grooves.

2.4.6 The asymmetry of the binding site

The binding site is formed at the 5'-3' dimeric interface of the stacked 5' TATA quartet, one BRACO19 molecule plus a thymine base, and the 3' G-quartet consecutively (Figure 2.24).

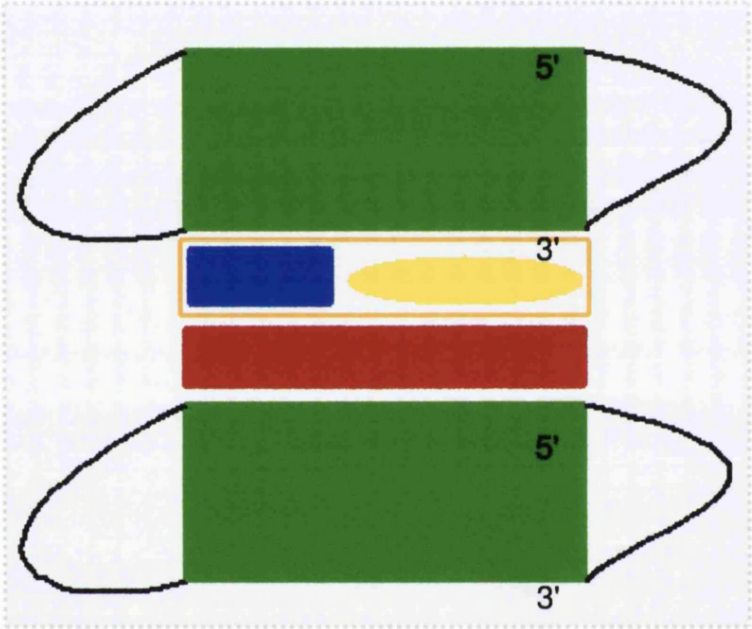


Figure 2.24: A schematic representation of the components of the binding site (grey box). Two quadruplexes (green rectangles) stack 5' to 3', sandwiching one ligand molecule at the interface (yellow oblong). The ligand forms a coplane (orange rectangle) with a thymine base (dark blue rectangle), which in turn stacks directly onto a 3' G-quartet on one side and onto a 5' TATA-quartet on the other.

The ligand is sandwiched asymmetrically between two quadruplexes that are non-coaxially stacked and inclined by ca. 30° resulting in the apparent discontinuity of the central potassium channel between consecutively stacked quadruplexes and a shift in the position of the G-quartet relative to the TATA-quartet at the binding site interface.

The effect of this shift is apparent in the differential positioning of the ligand relative to the flanking quartets, so as the 3- and 6- substitutions are placed at opposite edges of the 3' G-quartet onto which it stacks directly. At the same time, the same substitutions are placed towards neighbouring edges of the 5' TATA quartet. As a result, overall π–π stacking of the acridine core of BRACO-19 is onto two guanine

bases on one side and a reverse Watson-Crick A·T base pair on the other.

This also affects the position of the 3-, 6- and 9-substituents in BRACO19 relative to the grooves of the two quadruplexes. The terminal 3- and 6-pyrrolidino end hetero rings stacking onto the 3' face are placed towards the propeller loops (and their loop-bound grooves). At the same time, the 9-substituent is positioned near a loop-free groove (Figure 2.25). In contrast, at the 5' face G-quartet, one of the pyrrolidino rings is placed near a loop-free groove and the 9-substituent is placed above the TATA-quartet. The 9-substituent is at an angle (approximately 48˚) to the plane of the triacridine ring pointing away from the plane of the TATA-quartet (Figure 2.25).

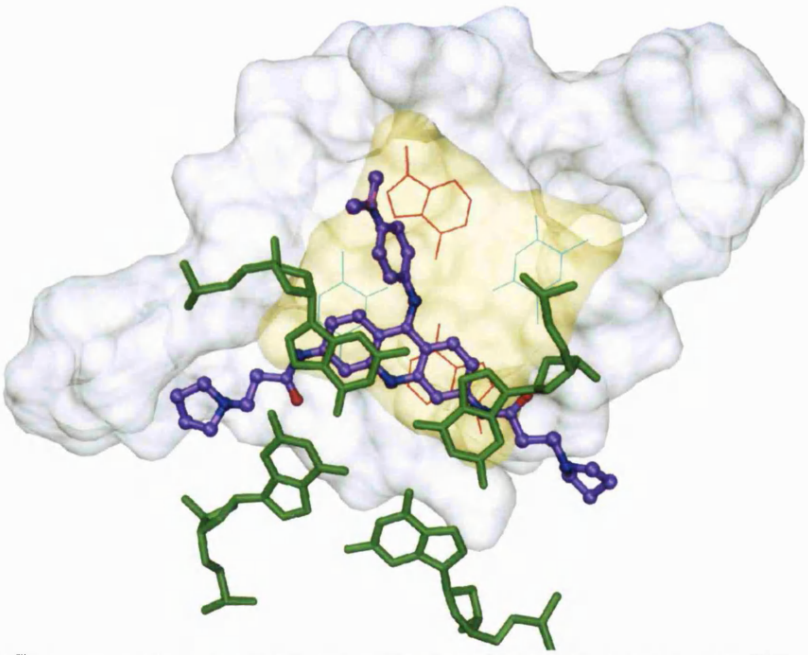


Figure 2.25: The asymmetry of the binding site. The ligand (shown in purple) is stacking onto two guanines (green) belonging to the 3' quadruplex face, and onto a T•A (red and blue, respectively) base pair on the 5' quadruplex face (bases of the TATA-quartet are shown in transparent yellow surface representation). The 3- and 6-substituents point towards a loop-bound groove and a loop-free groove straddling two adjacent sides of the TATA-quartet. At the same time, the same substituents point towards two loop-bound grooves relative to the 5' G-quartet (green).

The full extent of the available binding site "potential binding site"

In order to fully assess the extent of the volume and shape of the binding site, the programme DMS, part of the DOCK suite (Moustakas et al., 2006) was used to create a negative image of the available binding site and was visualised using

Chimera version 1.3 (Pettersen et al., 2004).

This showed that the ligand is indeed occupying a volume identified by DOCK as a binding site (Figure 2.26). However, it also showed that the ligand can extend further into the grooves where binding opportunities are still unexploited. For example, the 9-position substituent can extend further into a narrow and deep groove. The walls of this groove are formed of the negatively charged DNA backbone (in parallel orientation) and the floors of which are made of the edges of guanine bases. The presence of a water network molecules in this narrow groove emphasises the possibility of exploiting these interactions with the DNA backbone using a suitable longer substituent.

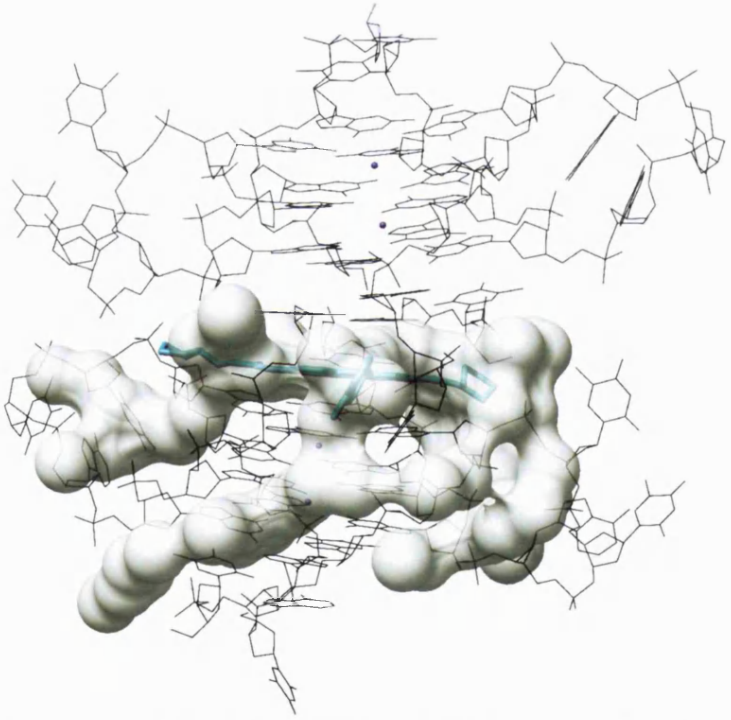


Figure 2.26: A volume representation (in grey) showing all three grooves available for additional potential binding.

As for the 3- and 6- position substituent, they also reside within the binding site as defined by DOCK, however, it also shows that the pyrrolidino groups have reached the edge of the potential binding volume and if any more interactions are to be probed, it would be in a downward direction into the shallow and wide grooves of the loops. This would increase the size of the ligand possibly making it less soluble.

2.4.7 Overall ligand description

The ligand has a unique orientation for the amido moieties at its 3- and 6-substituent positions so as the oxygen atoms belonging to each of the carbonyl groups face opposite directions as shown in figures 2.27 and 2.28. This flexibility in the side chains capitalises on the hydrogen bonding pattern of donors and acceptors in the binding site by accommodating as many of the complementary pairs as possible.

Of the eight donor-acceptor substituents in BRACO19, seven participate in hydrogen bonding in the binding site and are subdivided as follows (Figure 2.27 and Figure 2.28):

1. One direct hydrogen bond connecting the nitrogen in the amide group belonging to the 6-position side chain to O4 of the coplanar thymine.

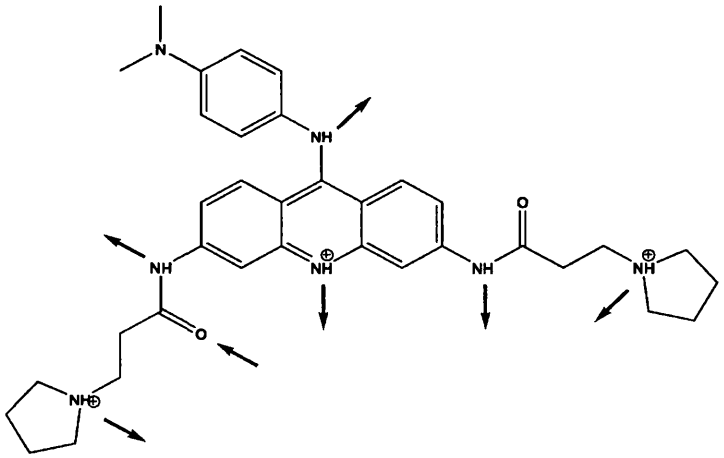


Figure 2.27: Structure of BRACO19. Arrows indicate directionality of hydrogen bonds found in the crystal structure reported here.

2. Two indirect (bridged through water) hydrogen bonds connecting BRACO19 to the coplanar thymine: one connecting the central nitrogen in the acridine core to N3 of the thymine base. The other connecting the oxygen of the carbonyl moiety in the amide group in the 3-position side chain to O2 of the thymine base.
3. One hydrogen bond connecting the nitrogen (N19) in the phenylamino

moiety at the 9-position substitution to a water molecule. However, the water molecules in the immediate neighbouring area are not well resolved.

4. Two hydrogen bonds with each one linking one of the positively charged nitrogens in the pyrrolidino rings at the ends of side chains to a network of water molecules that extend into the shallow wide grooves of the loops. These waters in turn make contact with the negatively charged backbone of

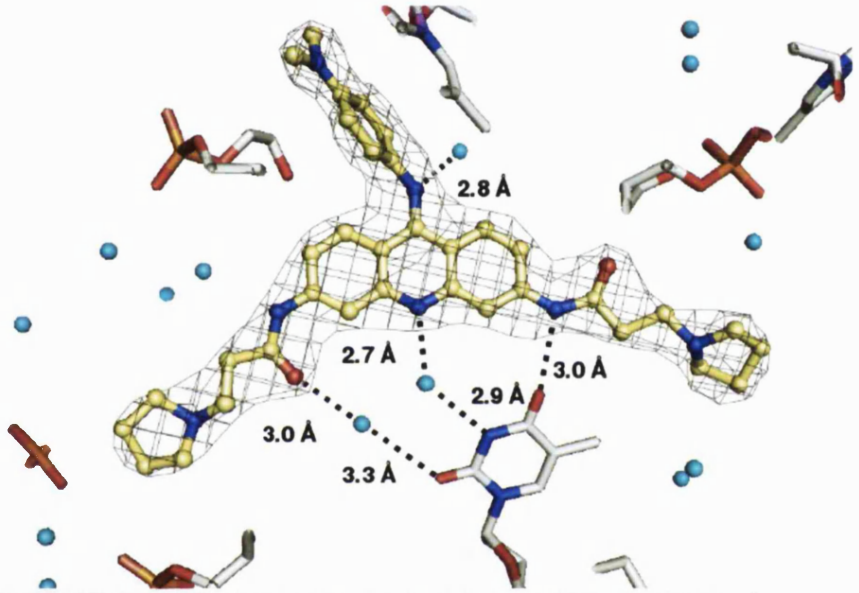


Figure 2.28: BRACO19 (in yellow) and a thymine base from the 3' quadruplex face form a coplane interacting through direct (3.0 Å) and water-bridged hydrogen bonds (2.7, 2.9, 3.0 and 3.3 Å) (hydrogen bond are represented with dashed black lines). Water molecules are shown as cyan spheres. BRACO19 is shown laid onto the 5' TATA quartet.

the DNA.

5. One hydrogen bond between the nitrogen of the amide group at the 3-position side chain and a network of water molecules that extend into the wide shallow groove of the loop ending up contacting the DNA backbone.

The ligand: 3,6,9- substitutions

The positively charged pyrrolidino rings at the ends of the side chains at the 3,6-positions and the phenylamino substitution at the 9- position form hydrogen bonds with suitable donors/acceptors in the binding site and are as follows:

1. The 3-position: the pyrrolidino ring is placed so that its positively charged nitrogen atom participates through hydrogen bonding in a water network that extends above and below into the loop-bound grooves of both

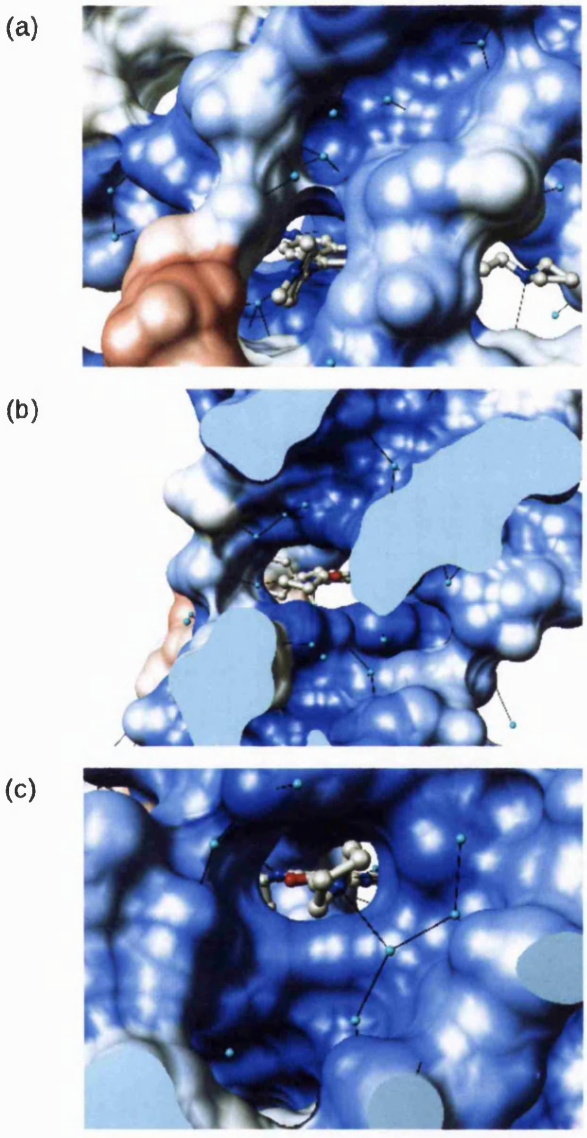


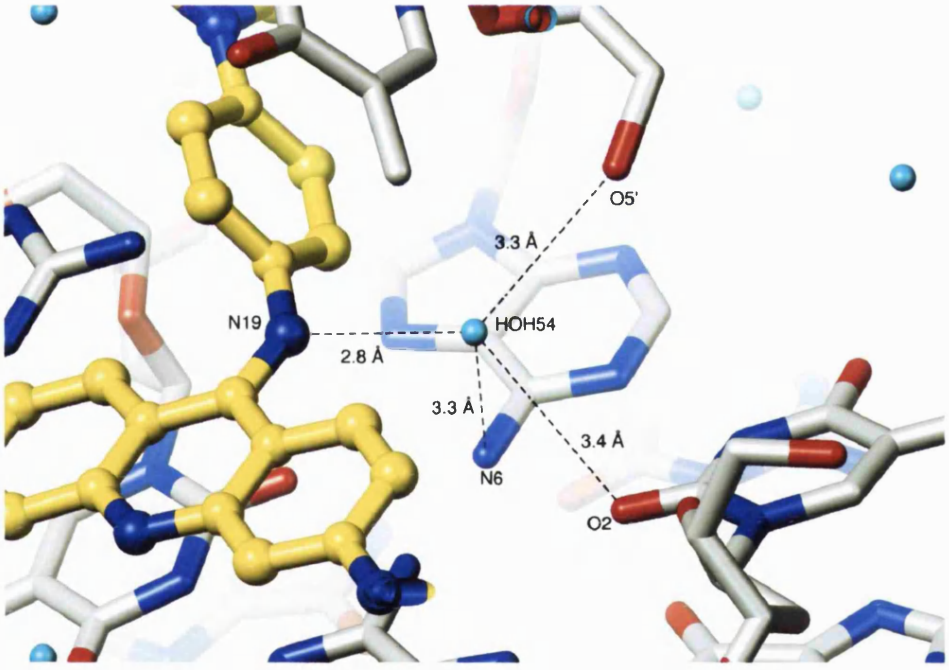
Figure 2.29: Hydrogen-bonding pattern for the 3,6,9-substituents. (a) The anilino moiety at the 9th position is positioned near a groove filled with a network of well-resolved water molecules. (b+c) The nitrogen atoms part of the terminal pyrrolidino rings at the 3rd and 6th positions participate directly in hydrogen bonding with the water network in their respective grooves. Water molecules are shown in cyan spheres, carbon in grey, oxygen in red and nitrogen in blue. Surface representation is shown for the DNA and colouration reflects nature of atoms.

quadruplexes in the biological unit. The water network bridges the ligand through hydrogen bonding to contact the negatively charged backbone of

the DNA. No direct contact between this substituent and the DNA is observed (Figure 2.29 (b)).

2. The 6-position: similar to the 3-position, the pyrrolidino ring at the 9-position also is placed so that it participates in a network of waters. However, here the water network extends only upwards towards the loop-bound groove of one of the quadruplexes i.e. 3' face finally contacting the DNA backbone (Figure 2.29 (a)).

3. The 9-position: The anilino moiety is almost perpendicular to the acridine plane and inclined so as to fit tightly into a binding pocket where it stacks



onto the 5' thymine of the neighbouring quadruplex (Figure 2.29 (a)). The O5' atom of the mentioned thymine hydrogen bonds to a water molecule (3.3 Å) that in turn hydrogen bonds (2.8 Å) to the linker nitrogen atom in BRACO19 (linking the anilino moiety to the acridine core) (Figure 2.30). This water is in-line with the potassium ion channel of the quadruplex and separated from it by the TATA-quartet (discussed elsewhere). The nearby binding pocket extends further into the quadruplex into a narrow and deep

loop-free groove. Several well-resolved water molecules are present in the hydrophilic part of this groove.

3' quadruplex face

BRACO19 stacks directly onto the G-quartet at the 3' quadruplex face. Both of the two 3' terminal thymines contribute directly to the binding site, each belonging to one of the DNA strands. One thymine extends down and away from the last G-quartet plane, effectively stepping down one plane, becoming coplanar with the ligand. This is reminiscent of the *Oxytricha nova*-ligand complex (PDB ID 1L1H). The thymine forms one direct hydrogen bond with the nitrogen atom, part of the amide bond in the side chain of the ligand, whereas the remaining two bonds are bridged through water to the central nitrogen in the acridine ring and to the oxygen atom in the carbonyl group belonging to the amide bond in the other long side chain.

The second thymine extends further down, effectively taking another step down the planes thus taking part in forming the TATA-quartet plane on the 5' face of the adjacent quadruplex in the asymmetric unit.

5' quadruplex face

This face is formed of the TATA-quartet formed of both quadruplexes in the asymmetric unit. As mentioned earlier, one thymine, extending from the former 3' face, forms part of the quartet. The remaining ATA comprise the 5' flanking nucleotides, one A belonging to one strand and TA to another.

2.4.8 Water structure

An extensive network of water molecules is observed and mostly belong to the first shell of hydration. This restriction to the first shell can be attributed to the relatively low resolution of the structure i.e. 2.5 Å. The water molecules satisfy different functions such as: forming bridging hydrogen bonding interactions between the ligand and the DNA, between the long side chains of the ligand (3- and 6-substituents) and the DNA backbone and filling the narrow grooves thus shielding

the negative charge of the atoms of the DNA backbone that forms the walls of the grooves.

The structure of the asymmetric unit i.e. one quadruplex bound to one BRACO19 molecule contains fifty-four fully occupied molecules of water occupying the four grooves of the quadruplex; two loop-free grooves and two loop-bound grooves.

The grooves contain a varied number of waters. The two loop-free grooves and the groove bounded by the remodelled loop (Figure 2.31) contain a similar number of waters (6 to 8 molecules of water each) whereas the groove bound by the traditional loop (loop 2 shown on the right in Figure 2.31) contains 21 molecules of water.

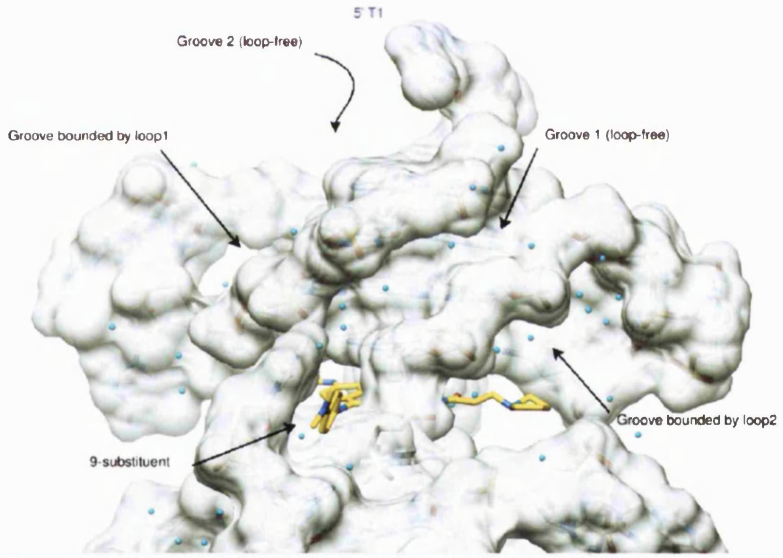


Figure 2.31: Water content of the grooves of the quadruplex. Groove 2 (loop-free) is on the far side of the quadruplex and is not visible to the viewer in this view. The remodelled loop (loop1) straddles the groove on the left while the traditional loop (loop2) straddles the groove on the right.

In the former group, the waters in groove 1 (loop-free) form mostly direct hydrogen bonding contacts with the DNA backbone (O3', OP1, OP2 and O4') and the edges of the guanine bases (N2 and N3) (Figure 2.32). The water molecules in groove 2 (loop-free) also contact the DNA backbone and guanine bases but also contact nearby water molecules (Figure 2.33). The groove bounded by loop 1 (remodelled loop) (Figure 2.34) contacts the DNA backbone and the nearby water molecules equally and also reveals one water molecule forming a hydrogen bond with the

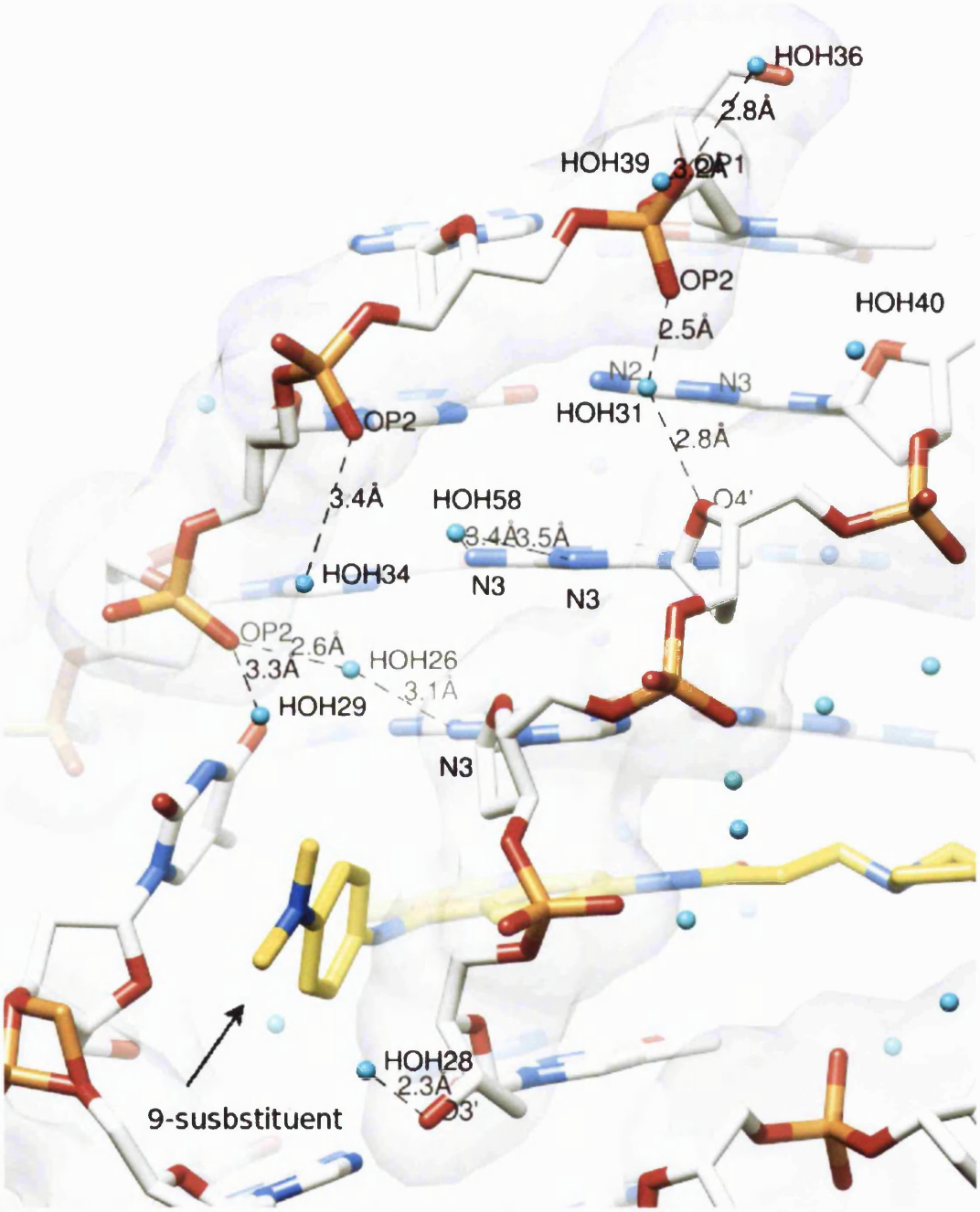


Figure 2.32: Water structure in groove 1 (loop-free). The water molecules contact the DNA backbone (OP1, OP2, O3' and O4') and the edges of the guanine bases (N3 and N2). Distances for some of the hydrogen bonds are shown in Å (dashed black lines). Ligand shown in yellow.

nitrogen atom in the end pyrrolidino ring in the side chain (3.0 Å).

On the other hand, the water network in the groove bound by the traditional loop is well-defined and reasonably structured. The groove is comparatively abundant with

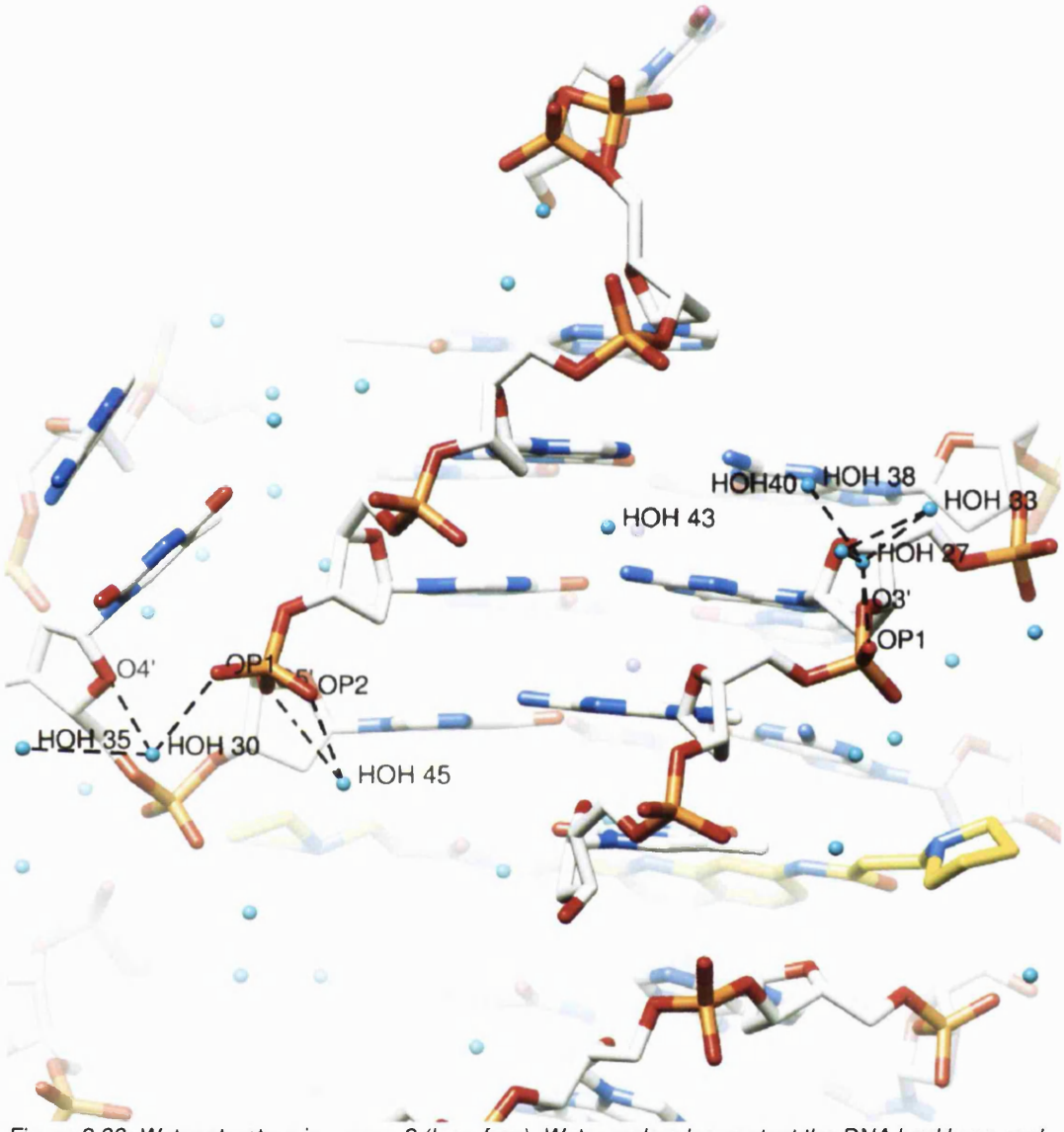


Figure 2.33: Water structure in groove 2 (loop-free). Water molecules contact the DNA backbone and also form part of localised water-water networks near the backbone atoms. The water network is limited and lacks definition. Ligand shown in yellow.

water (21 molecules). The waters are comparable to the ones seen in the structure of the native bimolecular quadruplex (PDB ID 1K8P) and contact the DNA backbone, the edges of the guanine bases and also form part of an extensive hydrogen-bonded network of waters (Figure 2.35).

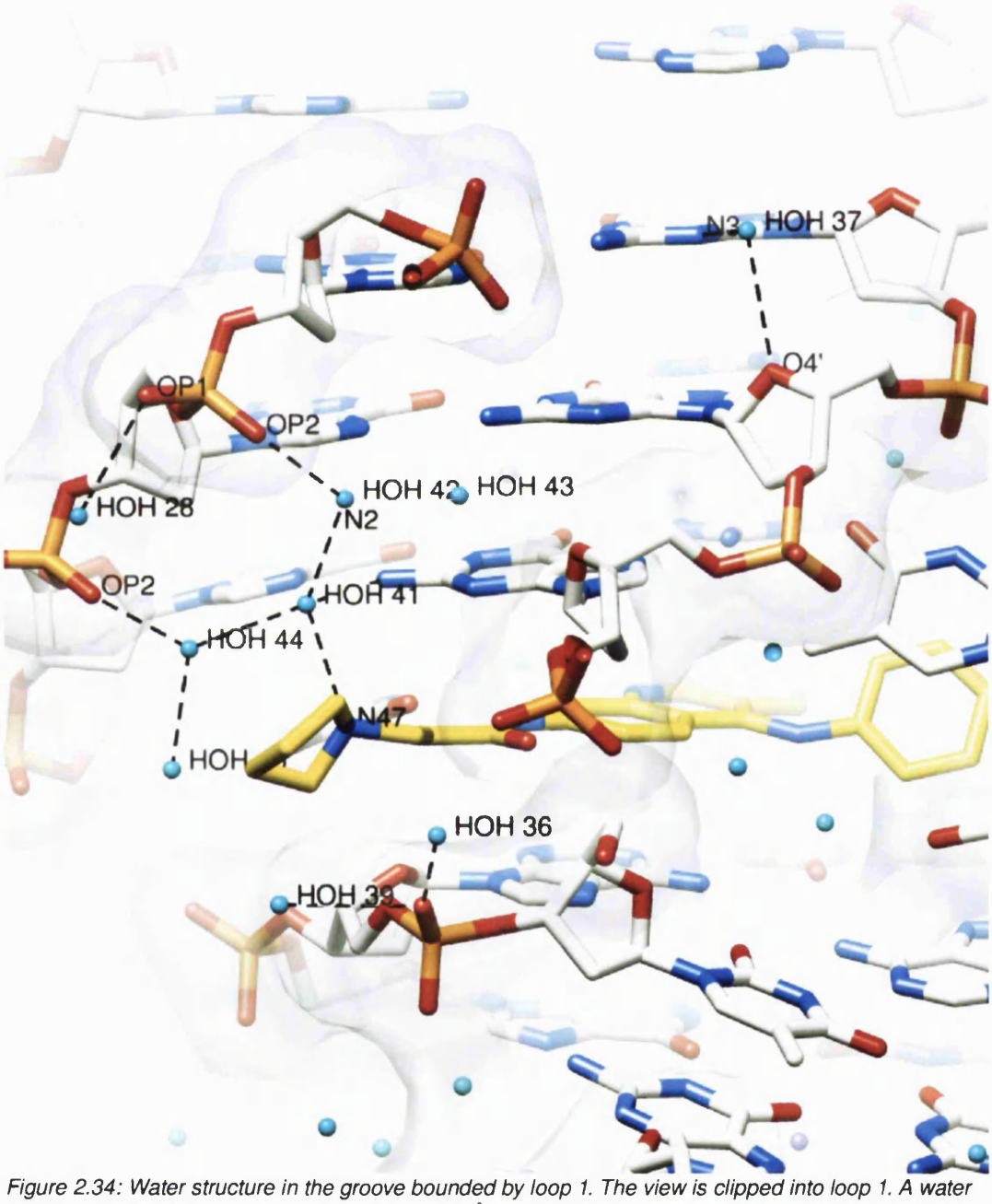


Figure 2.34: Water structure in the groove bounded by loop 1. The view is clipped into loop 1. A water molecule (HOH 41) forms a hydrogen bond (3.0 Å) with the nitrogen atom (N47) in the end pyrrolidino ring in BRACO19. Ligand shown in yellow.

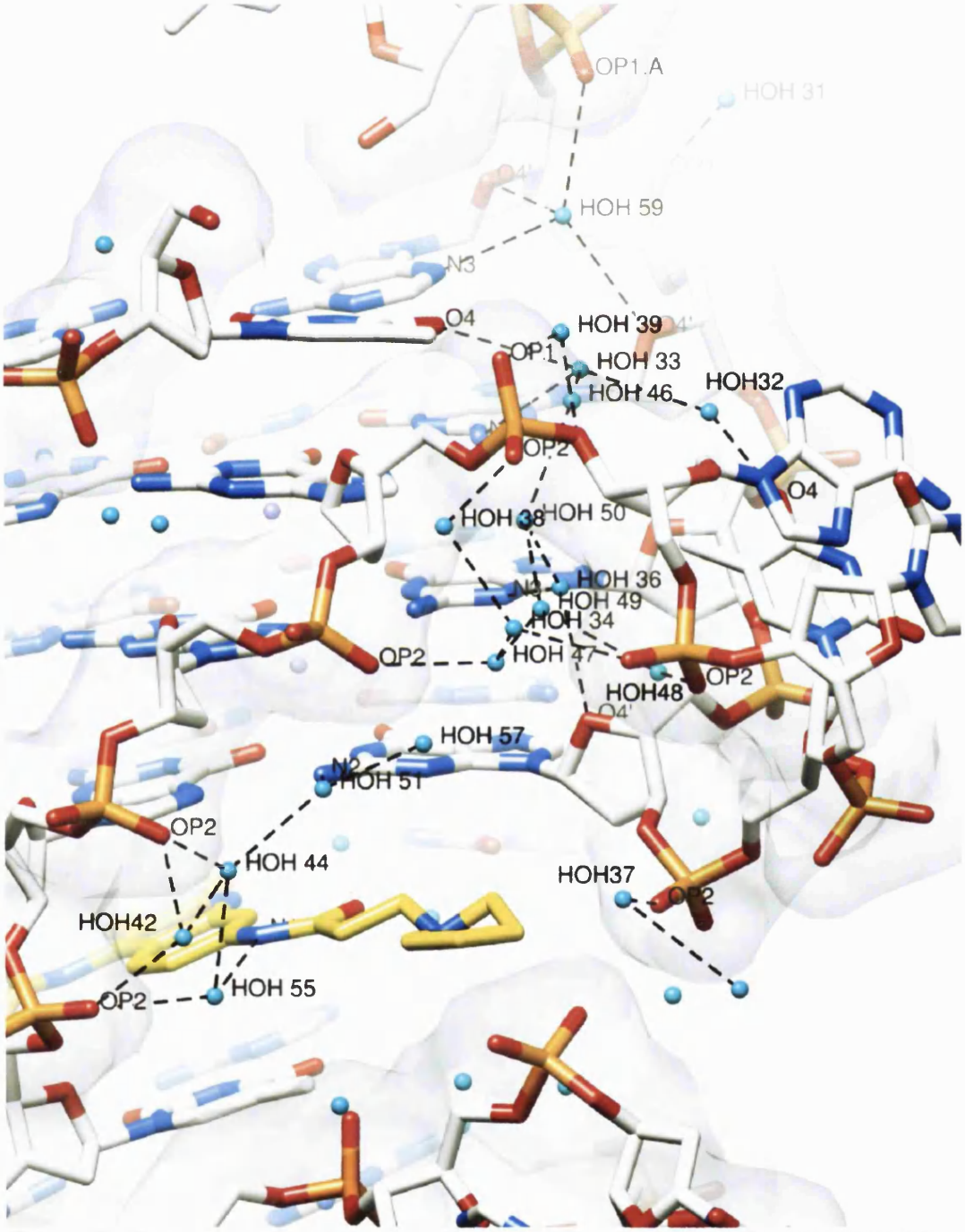


Figure 2.35: Water structure in the groove bounded by loop 2 (traditional loop). The water network is well-defined. Similar to the water structure seen in the native bimolecular quadruplex (PDB ID 1K8P) (Parkinson et al., 2003), a relatively extensive hydrogen-bonded network of water molecules is observed. Ligand shown in yellow.

2.5 Discussion

We have shown here a G-quadruplex binding trisubstituted acridine ligand BRACO19 fitting into a binding pocket at the interface between two 5' to 3' stacking quadruplexes. The pocket is formed of a 3' G-quartet, a 5' TATA-quartet, a coplanar thymine and a stacking thymine. The ligand is held in place by a combination of steric effects, hydrogen bonds and π-π stacking interactions.

The structure revealed that the hydrophobic 9-substituent anilino group plays a major role in ligand binding by fitting tightly into a narrow hydrophobic pocket whereas the 3- and 6- position flexible side chains extend towards the loops participating in a water network contacting the DNA backbone. This explained the key role of the 9-position anilino substituent in increasing quadruplex affinity by 10 folds. It also provided a rationale for quadruplex binding behaviour of BRACO19 analogues where increasing the length of the 3,6- side chains by -(CH_2)- groups results in progressive decrease of quadruplex binding for example: the binding constant is 3.1 x 10^7 M^{-1} for BRACO19 and it is 6 x 10^5 M^{-1} for BRACO19 + 3 [-(CH_2)-] (Moore et al., 2006). In contrast, extending the 9-anilino group into a -(C=O)-(CH_2)$_3$-pyrrolidino moiety increases the binding constant to 7.7 x 10^7 M^{-1}. The 3- and 6- side chains in BRACO19 extend just enough to place the positively charged nitrogen in the pyrrolidino group in a position to participate in local water networks contacting the DNA backbone. Increasing the length of the side chains further would put this nitrogen out of reach of the water network abolishing these contacts. Additionally, it would place hydrophobic alkyl moieties in the middle of a water network which can result in disturbing the established structural stability.

The tight fit of the ligand into the binding site is a result of a combination of factors. The quadruplexes' shift, tilt and roll relative to each other create an opening where the ligand fits, bounded by 5' and 3' flanking thymine and adenine nucleotides that flip into the binding site, thus holding the ligand tightly in place. Due to the the 5' to 3' stacking of the G-quadruplexes in this structure, the flanking nucleotides may belong to a loop connecting the two quadruplexes. Indeed, an *in silico* molecular

modelling study showed that such a loop, formed of the sequence TTA, can equally contribute to the binding site, thus stabilising the ligand, without affecting the integrity of the tandem repeats of quadruplexes (Haider & Neidle, 2009).

The complementarity between BRACO19 and its binding site is such that the pattern of hydrogen bond donors and acceptors in the ligand complements that of the binding site, thus enabling seven out of eight potential hydrogen bonds to be formed by the ligand with the quadruplexes or structural waters.

Furthermore, one of the side chains of the ligand BRACO19 is rotated by 180° so as the carbonyl oxygen (part of the amide bond) points in the opposite direction to its equivalent in the other side chain. This allows the nitrogen in the amide bond to form a direct hydrogen bond with the coplanar thymine whilst the carbonyl oxygen in the other amide bond forms a water-bridged hydrogen contact with the same thymine. Structures of complexes formed between the anti-parallel bimolecular quadruplex of the *Oxytricha nova* telomeric sequence and a series of closely related disubstituted acridines showed both of the the carbonyl oxygen atoms pointing in the same direction.

The demonstrated flexibility of the "receptor" (G-quadruplex) and "ligand" (BRACO19) to fit in a tight manner is reminiscent of the induced-fit concept encountered in protein targeting (Valente et al., 2006; Johnson, 2008). DNA-induced fit also occurs in the case of DNA-ligand binding, albeit it has been shown for duplex DNA. For example, intercalation of ligands in between base pair steps in duplex DNA is attributed to induced-fit between the ligand and the DNA (Todd et al., 1999).

The structure also shows that the features of human telomeric crystal structures remain conserved i.e. parallel-stranded quadruplex with propeller loops even though the drop conditions contained both sodium and potassium ions, whereas conformation seems to be ion sensitive in NMR studies (discussed elsewhere). It also demonstrates the flexibility of the loops as is evident by their remodelling to accommodate novel favourable crystal packing interactions or ligand interactions which is in accord with previous observations (Parkinson et al., 2008; Parkinson et

al., 2002; Parkinson et al., 2007).

These results suggest that 9-substituent may be a good choice for optimisation to capitalise on contacts in the loop-free narrow deep groove. The water network in this groove is finite, the groove is tube-like with walls and floors that are fixed in their shape and charge pattern (compared to the loop-bound grooves) – reminiscent of the minor groove of duplex DNA. The latter is successfully targetted with ligands (Cai et al., 2009). In contrast, the loop-bound grooves are shallow and wide containing an extended and seemingly non-directional water network. Traditionally, a network of structural waters provides information regarding the pattern of hydrogen bond donors and acceptors within a binding site, which can be successfully mimicked by a complementary ligand. In the case of quadruplex structures with regards to targeting the loop-bound grooves, practical accessibility to interactions with the loop bases may be elusive due to the shallow and wide character of the grooves, the flexibility of the loops and the large extended water network present. All of this creates a binding site with constantly varying dimensions. Furthermore, the water network is widely extended and designing a ligand that capitalises on the correct hydrogen bonding pattern may be challenging. On the other hand it may be possible to design ligands that interact with the loops in an induced-fit manner where the flexibility of the loop-bound groove may result in favourable outcome.

This structure adds a new level of complexity in targeting G-quadruplex structures with small molecules. Previously, the binding site was simply represented by a G-quartet platform available for stacking interactions with a complementary chromophore in the ligand. The side chains of the ligand were envisaged to extend and/or interact with the backbone of the DNA of the loops. Now, it seems that the binding of ligands to quadruplex structures encompasses a fully defined tightly bounded binding site, the formation of which is dependent on the conformational flexibility on part of both the ligand and the G-quadruplex i.e. loops, to produce the required complementarity for tight ligand fit. However, these results may be ligand specific and thus it may not be possible to draw general conclusions or a set of rules for quadruplex-ligand binding, especially, in the absence of structural data

regarding tandem repeats of quadruplexes.

2.6 Conclusion

The complementarity in shape and pattern of hydrogen-bond donors and acceptors between BRACO19 and the binding site causing the ligand to fit snuggly explains the experimental results and QSAR (Quantitative Structure Activity Relationship) data which thus far had been lacking. However, some of the binding features may be acridine specific e.g. the coplanar thymine and exact hydrogen bonding pattern, or sequence specific e.g. the formation of the TATA-quartet is flanking sequence dependent. At the same time, it is showing a continuous stack of quadruplexes with a ligand molecule bound at the interface. This indicates the possibility of a similar arrangement for longer sequences of DNA which are capable of forming tandem repeats of telomeric quadruplexes. The latter reflects the physiological conditions as the length (i.e. number of repeats of guanine runs) of the single-stranded G-rich telomeric sequence can easily accommodate several quadruplexes in tandem. This is especially informative since these structures are currently elusive to both X-ray and NMR techniques. It also displays the flexibility of the target to accommodate suitable ligands e.g. the coplanar thymine is a recurrent feature in the di- and tri-substituted acridines in complex with anti-parallel and parallel-stranded quadruplexes (Haider et al., 2003; Campbell et al., 2008) whereas it is absent in porphyrin and diimide containing complexes (Parkinson et al., 2007; Parkinson et al., 2008). In the former, the coplanar thymine extends the ligand plane to match the dimensions of the G-quartet whilst in the latter the ligand is large enough to do this sufficiently.

Overall, this structure demonstrates the 3-dimensional flexibility of the quadruplex and the potential of an induced fit scenario in quadruplex-ligand binding. However, access to structures of tandem repeats of quadruplexes is crucial to realistically evaluate this possibility. Structural data for multi tandem repeats of quadruplexes is yet to become available. The crystal structure also shows that none of the grooves are fully probed, this is true for the acridine family and most likely for other

quadruplex binding agents. This means that, at least in principle, it is possible to selectively target a specific quadruplex over other quadruplexes through the design of ligands with up to four substituents, each designed to fit into a groove. However, the flexible nature of the loop-bound grooves combined with the chemical complexity involved in the design and synthesis of large aromatic multi substituted ligands, whilst maintaining drug-like characteristics e.g. solubility, is a challenging prospect.

CHAPTER 3

X-RAY STUDIES OF A SERIES OF SIX DISUBSTITUTED ACRIDINE LIGANDS COMPLEXED WITH THE *OXYTRICHA NOVA* TELOMERIC QUADRUPLEX

3 X-RAY STUDIES OF A SERIES OF SIX DISUBSTITUTED ACRIDINE LIGANDS COMPLEXED WITH THE *OXYTRICHA NOVA* TELOMERIC QUADRUPLEX

3.1 Background

The family of 3,6-disubstituted acridine derivatives are G-quadruplex targeting agents, that were rationally designed as telomerase inhibitors, on the basis that inhibition of telomerase occurs by stabilising G-quadruplex structure formation in the single-stranded telomeric ends of chromosomes. The most potent inhibitors have an IC_{50} against telomerase activity of between 1.3 and 8 μM with comparable binding for duplex and quadruplex DNA (Read et al., 1999; Read et al., 2001; Harrison et al., 1999).

History

An amidoalkyl-functionalised anthracene-9,10-dione (also known as 9,10-anthraquinone) was shown, using NMR studies, to stabilise quadruplex structures against thermal denaturation (Sun et al., 1997). This compound belonged to a family of ligands which were originally studied as duplex DNA-interactive agents (Agbandje et al., 1992; Collier & Neidle, 1988), with preferential binding to triplex DNA (Keppler et al., 1999). A series of disubstituted regioisomeric derivatives showed activity against telomerase between 1.3 to 17.3 μM (Perry et al., 1998; Perry et al., 1998). Fluorenone derivatives showed reduced activity of 8 to 12 μM with decreased cellular activity, which was attributed through computational molecular modelling studies to the effect produced by the five-membered central ring which rendered the core crescent-shaped (Perry et al., 1999). Further molecular modelling investigations predicted that a tricyclic acridine chromophore

would be comparable to the anthraquinone core (Read et al., 1999). Additionally, the nitrogen atom in the central ring, which when protonated, would increase electron deficiency through the core, and thus enhance π–π stacking interactions with the G-quartets in quadruplex structures (Harrison et al., 1999). Indeed, a series of 3,6-disubstituted acridine derivatives were synthesized and showed activity against telomerase between 1.3 and 8 μM, equivalent to their cytotoxicity in ovarian cancer cell lines (Harrison et al., 1999). These studies suggested that the acridine derivatives were suitable leads as G-quadruplex binding ligands based on a di-substituted (and chemically accessible) acridine core.

The crystal structure of the native form of the bimolecular quadruplex formed by the *Oxytricha nova* telomeric sequence d[GGGGTTTTGGGG] was reported in 2002 (PDB IDs 1JPQ and 1JRN (Haider et al., 2002)) and in complex with a 3,6-disubstituted acridine BSU6039 in 2003 (PDB ID 1L1H (Haider et al., 2003)). The

(a) (b)

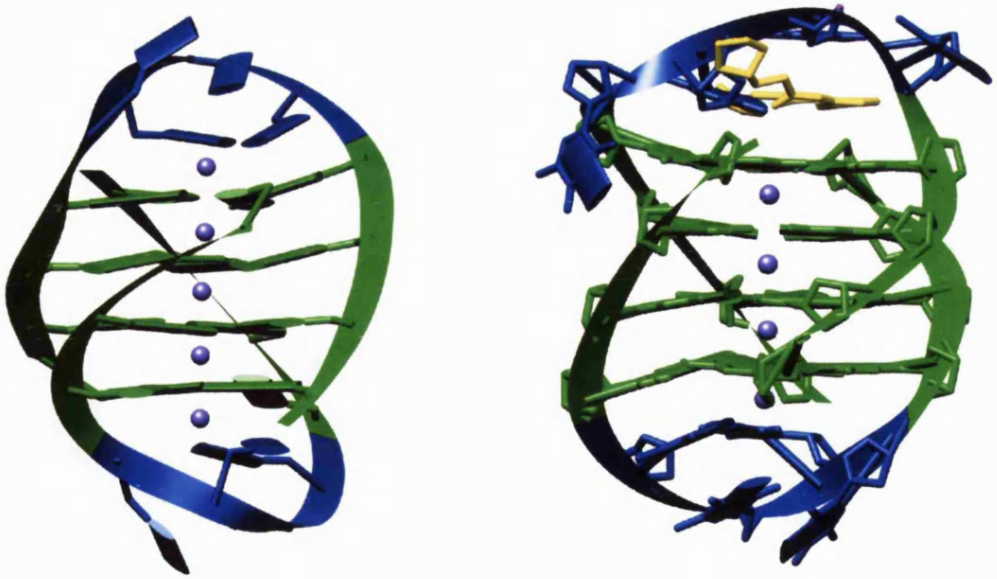


Figure 3.1: Structures of the native and ligand-bound bimolecular Oxytricha nova quadruplexes. In the native structure (PDB ID 1JPQ (Haider et al., 2002)), the diagonal loop adopts a closed conformation (a) in contrast to the open conformation (b) in the ligand-bound form (PDB ID 1L1H (Haider et al., 2003)). Thymine and guanine nucleotides are shown schematically with the backbone shown in ribbon representation, the nucleic acid bases in slab form and the central potassium ions in purple spheres. (Thymine = blue, guanine = green and ligand=yellow).

anti-parallel quadruplex contained two diagonal TTTT (T_4) loops, each loop straddling the end G-quartet externally at opposite ends of the G-quartet stack. In the native form, the two loops were identical in conformation, with the thymine bases held in place via hydrogen bond interactions, stacking interactions and coordination to a potassium ion that formed part of the continuous central potassium ion channel. The structure of the complex revealed an identical quadruplex with one ligand molecule bound to one of the diagonal loops. The ligand disrupted the closed conformation of the diagonal loop by displacing two thymine bases and ejecting the coordinated potassium ion within. BSU6039 formed a coplane with one thymine base belonging to the loop, and the coplane stacked in turn onto the end G-quartet (Figure 3.1). This was the first crystal structure of a quadruplex-ligand complex and it revealed the stacking binding mode of ligands to quadruplex structures. Furthermore, it showed the flexibility of the loops and also indicated their active role in enabling ligand binding, seen in the contribution of loops in forming the binding site itself and direct stabilising interactions with the ligand i.e. coplane formation via two hydrogen bonds (shown in (a) in Figure 3.2).

Loop flexibility was also observed in the propeller loops in human telomeric bimolecular and unimolecular ligand-quadruplex complexes; with porphyrin (PDB ID 2HRI (Parkinson et al., 2007)) and two naphthalene diimide complexes (PDB IDs 3CCO and 3CDM (Parkinson et al., 2008)). These structures revealed flexible loops amenable to remodelling to enable favourable novel crystal packing interactions and ligand binding. Additionally, the complex formed between the bimolecular human telomeric quadruplex and the trisubstituted acridine BRACO19 (PDB ID 3CE5 (Campbell et al., 2008)) revealed a coplane comprising a thymine nucleotide and the ligand, reminiscent of the structure of the complex formed between the telomeric quadruplex of *Oxytricha nova* and the disubstituted acridine BSU6039, albeit via three hydrogen bonds, two of which were bridged through water (Figure 3.2). In this instance, the thymine was a flanking nucleotide which could, due to the 5' to 3' stacking of the quadruplexes in the structure, belong to loop nucleotides in a multi-quadruplex (in tandem) arrangement. Indeed, a molecular modelling study showed that the nucleotides belonging to the loop used to connect the two stacked

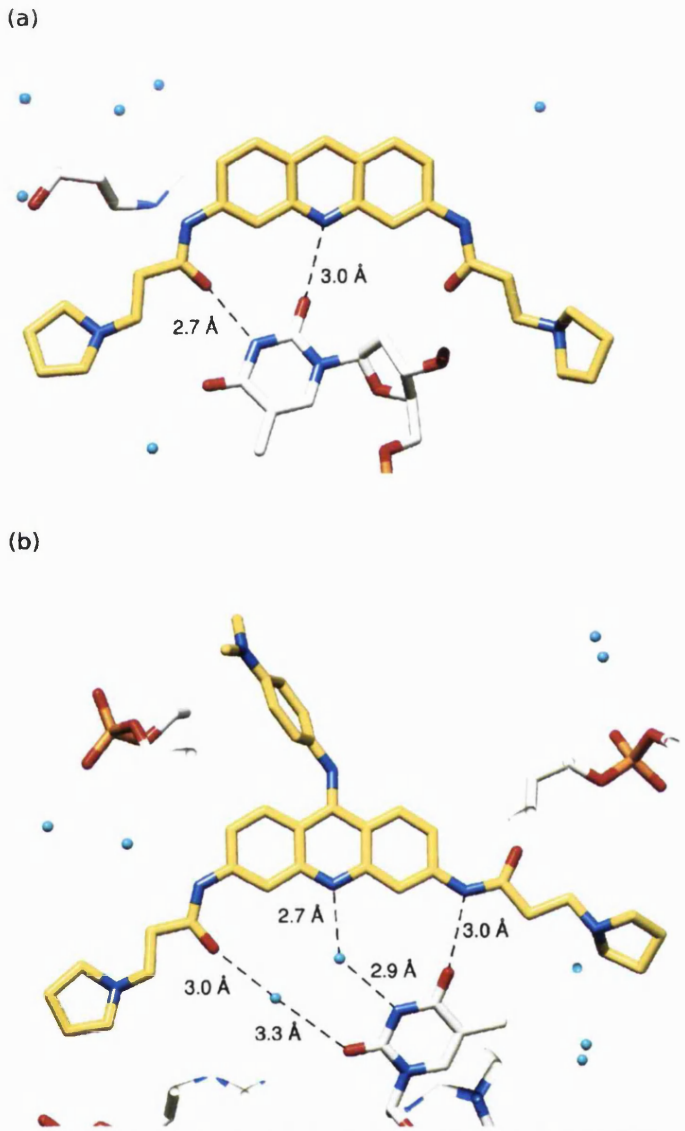


Figure 3.2: A coplane is formed between a thymine base and the ligand. (a) BSU6039 in complex with the Oxytricha nova telomeric quadruplex. (b) BRACO19 in complex with the human telomeric quadruplex. Hydrogen bond distances are represented in dashed black lines. Carbons in the ligands are in yellow and in the DNA in white. Nitrogen = blue, oxygen = red and phosphorus = orange. Waters are shown as cyan spheres.

quadruplexes was flexible to flip into the binding site and interact with the ligand and within a stable quadruplex multimer (Haider & Neidle, 2009). Additionally, experimental data showed loop nucleotides flipping into a stacking position onto the end G-quartets in the porphyrin-quadruplex complex (PDB ID 2HRI (Parkinson et al., 2007)).

In summary, all quadruplex-ligand complexes reported to date showed evidence of

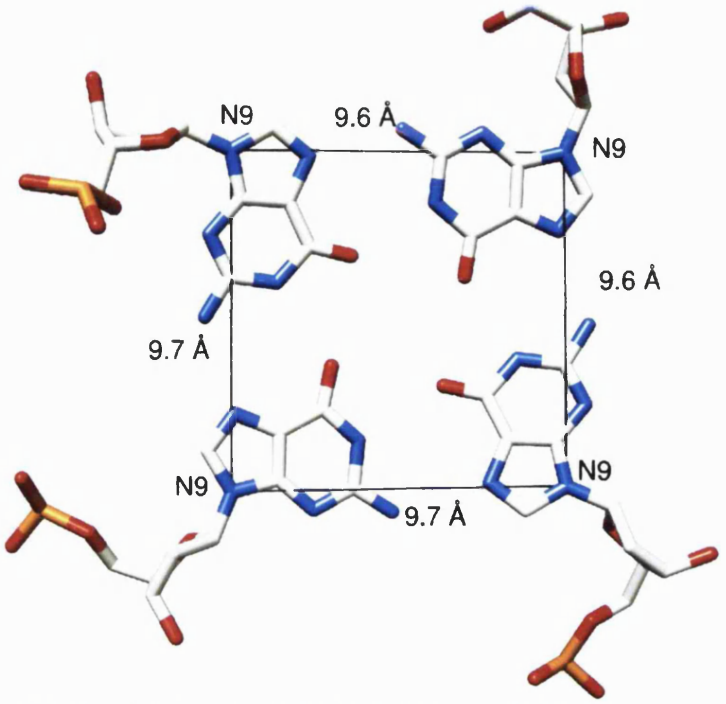


Figure 3.3: Dimensions of a G-quartet. A square represents the distances between the N9 atoms of adjacent guanine bases in a G-quartet plane.

loop remodelling, whether to accommodate a ligand (either by direct binding or by flipping onto end G-quartets) or simply to enable novel crystal packing interactions.

Finally, the formation of a coplane is seen in instances where the ligand has a surface area that is too small to fully take advantage of the available stacking interactions with the relatively large G-quartet surface (surface area of the guanine bases in a G-quartet in the parallel human telomeric quadruplex is equivalent to a square with a side length of approximately 9.7 Å (Figure 3.3)). Such is the case in the human telomeric BRACO19-quadruplex complex (PDB ID 3CE5 (Campbell et al., 2008) and in the *Oxytricha nova* telomeric BSU6039-quadruplex complex (PDB ID 1L1H (Haider et al., 2003)). With larger ligands such as porphyrin and the naphthalene diimides (Parkinson et al., 2007; Parkinson et al., 2008), a similar arrangement is simply not necessary as these ligand are sufficiently large to interact fully with the G-quartet without contribution from the loop bases.

Aim

The aim of this work is to use X-ray crystallographic structural studies to gain further insights into the detailed structural aspects of interaction between a series of

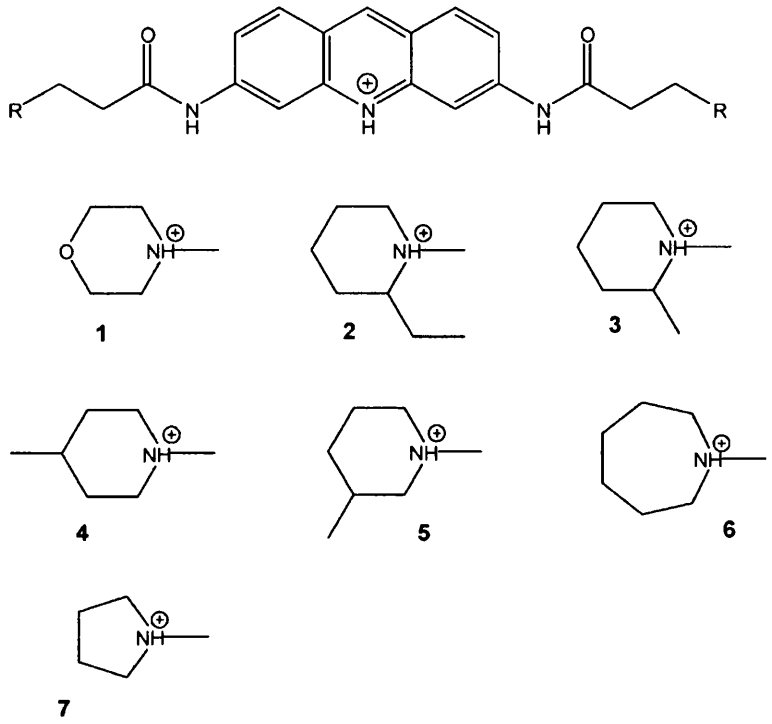


Figure 3.4: Disubstituted acridine series. 1=BSU6038, 2=BSU6042, 3=BSU6045, 4=BSU6048, 5=BSU6054, 6=BSU6066 and 7=BSU6039.

modified disubstituted acridines (Figure 3.4) - to which the ligand BSU6039 belongs (compound 7) - and the anti-parallel bimolecular quadruplex formed by the telomeric sequence of *Oxytricha nova*, with the view of developing a qualitative structure activity relationship between the various complexes and the activity of the bound ligands against telomerase.

3.2 Materials and methods

3.2.1 Crystallisation

Oligonucleotide synthesis, purification and annealing

The Eurogentec RP (reverse phase)-cartridge-purified *Oxytricha nova* telomeric deoxyribo-oligonucleotide sequence d[GGGGTTTTGGGG] was purchased from Eurogentec Ltd and used without further purification. The oligonucleotide was annealed at 3 mM (quadruplex concentration where one quadruplex is formed of two strands of DNA) before use by incubation in a heating block at 90°C for 15 minutes in 20 mM potassium cacodylate buffer at pH 6.5 and 50 mM potassium chloride and left to cool gradually to room temperature overnight.

Syntheses and purification of ligands in the study

The ligands were synthesized in-house by Anthony Reszka, in the laboratory of the Biomolecular Structure Group at the School of Pharmacy, University of London, using previously published procedures (Harrison et al., 1999).

A total of six ligands in the series were used. The hydrochloride salt of five of the ligands was used. The free base of compound 1 (Figure 3.4) was used.

A 20 mM stock solution was prepared in 100% DMSO (dimethyl sulfoxide) and kept at -20°. Immediately prior to setting up the experiment, the ligand was thawed and the required amount added to the DNA.

Complex preparation

Immediately prior to setting up the crystallisation experiment, the complex was prepared by adding ligand solution to the annealed quadruplex solution to make a solution of the complex at 1mM quadruplex and 2mM ligand. The Reagent solution was then added to the complex to produce the initial crystallisation conditions.

3.2.2 Crystallisation conditions

The hanging-drop vapour-diffusion method was used to crystallise all complexes. The 24-well VDX plates (Hampton Research Corporation) and 22 mm circular siliconised glass cover slips (Hampton Research Corporation) were used in the setup.

Initial crystallisation conditions in the 2 µl drop were: 0.25 mM DNA, 0.5 mM ligand, 40 mM KCl, 5 mM $MgCl_2$, 4.1 mM spermine and 5% v/v 2-methyl-2,4-pentanediol (MPD). This was equilibrated against 600 µl of 35% (v/v) MPD. The DNA:ligand ratio was 1:2.

Crystals grew as yellow cubic blocks after 1 week at 12°C. The dimensions of a typical crystal were approximately 0.3 x 0.3 x 0.2 mm.

Mounting the crystal

The crystals of the complexes were flash-frozen using glycerol cryoprotectant and were maintained at 105 K in a dry nitrogen stream using an Oxford Cryosystem (Oxford Cryosystems Ltd.).

3.2.3 Data collection and processing

X-ray diffraction data were collected on an in-house Rigaku R-AXIS IV image plate detector at wavelength 1.54 Å (CuKα radiation) emitted from a Rigaku RU200 rotating anode generator (100 mA, 50 kV) focussed by an Osmic focussing mirror system (Rigaku/MSC).

Indexing and data processing were carried out using the d*TREK part of the CrystalClear 1.3.6 software package (Rigaku Corporation© 1997-2002) by Gary Parkinson in this group.

X-ray diffraction data for the crystal of the complex formed with the ligand BSU6066 i.e. compound 6 (Figure 3.4) were collected at the laboratories of Oxford Diffraction Ltd. in Oxford, England, using the Xcalibur Nova system. An Oxford Diffraction Onyx CCD detector was used and the wavelength 1.54056 Å (CuKα radiation)

emitted from an Oxford Diffraction Nova Enhanced Ultra X-ray source (100 mA, 50 kV). This X-ray source incorporates a micro-focus sealed tube integrated with high performance multi-layer optic. A diffraction image for this complex is shown (Figure 3.5).

3.2.4 Structure Solution and verification

The program Phaser (McCoy et al., 2007) part of CCP4 (Collaborative Computational Project, 1994) package programme suite was used to solve the structures by molecular replacement (MR). The structure of the isomorphous *Oxytricha nova* telomeric quadruplex in complex with a closely related disubstituted acridine BSU6039 (PDB ID 1L1H (Haider et al., 2003)) was used as a search model after it was first stripped of the ligand, ions and waters. Phaser (McCoy et al., 2007) was run with the default settings.

One solution was produced for all complexes formed in the smaller unit cell (space group $P2_12_12$) and for example the scores for the complex formed with the ligand BSU6042 were: RFZ=7.3, TFZ=20 and LLG=680. The solution produced a model that agrees with the previously reported structures for the native quadruplex (PDB IDs 1JPQ and 1JRN (Haider et al., 2002)) and in the complex with BSU6039 (PDB ID 1L1H (Haider et al., 2003)). The unit cell and space group ($P2_12_12$) were identical in the complexes to the previously reported complex (PDB ID 1L1H). The asymmetric unit contained one ligand bound in one of the diagonal loops in the quadruplex. The unit cell dimensions and angles for each complex are tabulated (Table 3.1).

Similarly, PHASER produced one solution for the the complex formed with the ligand BSU6042 in the large unit cell (space group $P2_12_12_1$) and the scores were: RFZ=6.0, TFZ=34.7 and LLG=1010. The asymmetric unit contained two quadruplexes each bound to one ligand.

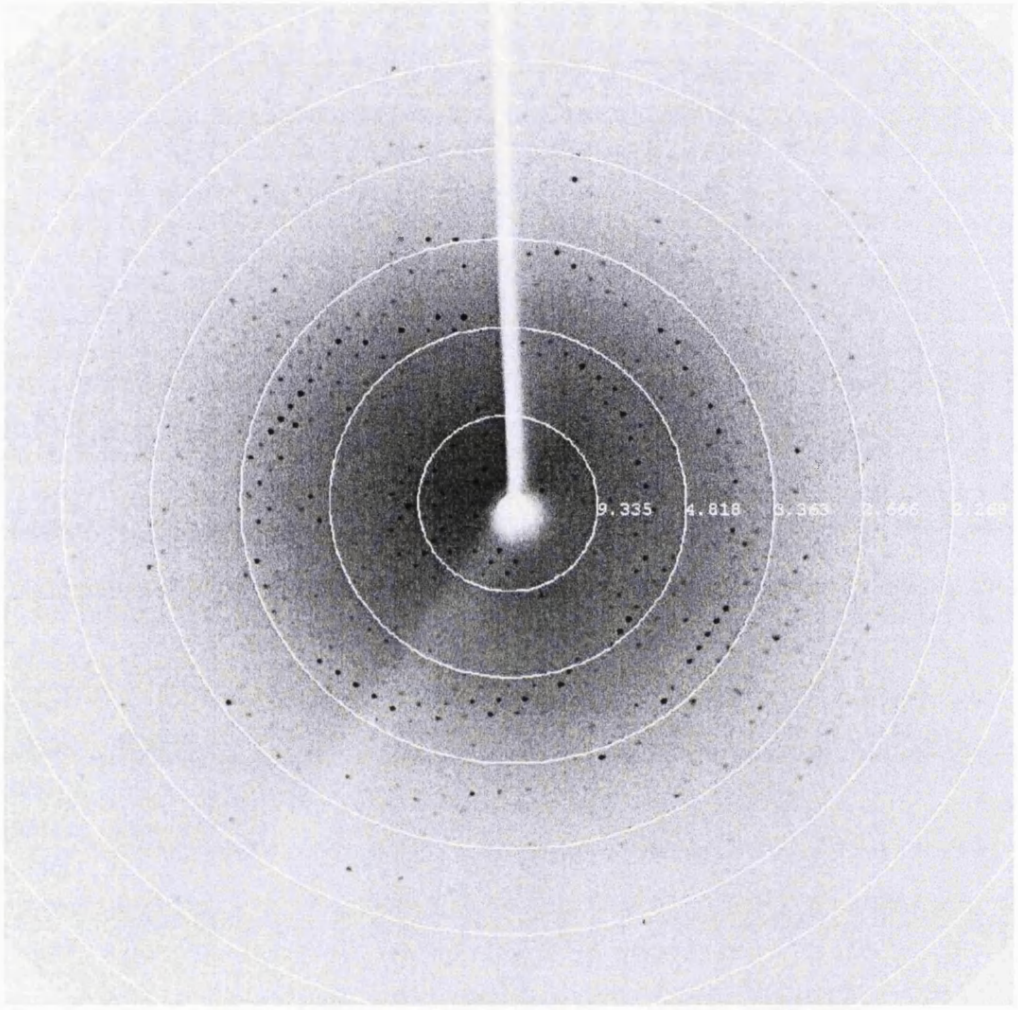


Figure 3.5: Diffraction image for the crystal of the complex formed with ligand BSU6066.

Solution verification

The solution was verified by checking the following in accordance with guidelines in Eleanor Dodson's paper (Dodson, 2008):

1. that the model makes chemical sense:

   The electron density maps were calculated using Refmac version 5.2.0019 (Murshudov et al., 1997) and visual inspection with Coot version 0.5 (Emsley & Cowtan, 2004) confirmed a. the correct position of the quartet-forming guanines b. position and direction of the diagonal T_4 loops c. that no clashes

were present between symmetry copies and d. the presence of potassium ions in the central quadruplex channel.

2. that the model can be refined confirming that the solution generates amplitudes which agree with the observed ones:

   This is shown by by the refinement of the initial R_{factor} and R_{free} values from the initial range of values of [33.9% - 26.3%] (R_{factor}) and [39.6% - 29.2%] (R_{free}), to the final range of values of [23.7% - 19.7%] (R_{factor}) and [31.4% - 24.5%] (R_{free}) (Table 3.1).

Table 3.1: R_{factor} and R_{free} decrease as model fitting progresses from the initial stage of structure solution (only DNA) to the final fitted model (DNA, ligand and waters).

		Initial		Final	
PDB ID	Ligand	R_{factor} (%)	R_{free}(%)	R_{factor} (%)	R_{free}(%)
3EM2	BSU6038	28.1	30.6	20.7	29.3
3EQW	BSU6042-small	33.9	39.6	22.6	31.4
3EUI	BSU6042-large	31.6	35.2	23.7	28.5
3ERU	BSU6045	25.1	27.9	20.1	24.9
3ES0	BSU6048	26.3	29.2	19.7	24.5
3ET8	BSU6054	26.6	32.0	21.1	27.7
3EUM	BSU6066	26.4	32.9	22.4	27.0

3. if the electron density maps show features that are not part of the model:

   This is verified by the appearance of a volume of electron density identical to the corresponding disubstituted acridine ligand in shape and size, and similarly for the ions in the central quadruplex channel. Additionally, with regards to the BSU6042-quadruplex complex in space group $P2_12_12_1$ i.e. in the larger unit cell, a volume of electron density revealed a new position for the last thymine in the ligand-bound T_4 loop. In its new environment, this thymine is contacting the edges of a guanine base in another quadruplex via two hydrogen bonds (shown in (b) in Figure 3.12). This is in contrast to the equivalent thymine in the $P2_12_12$ space group which does not contact DNA atoms. In the new space group i.e. $P2_12_12_1$, this thymine is flipped into a new

position enabling the formation of novel direct crystal packing interactions.

3.2.5 Model building and structure refinement

Overall

Cycles of model building using Coot version 0.5 (Emsley & Cowtan, 2004) and structure refinement using Refmac version 5.2.0019 (Murshudov et al., 1997) were carried out. As in chapter 2, the standard nucleic acid dictionary in refmac was used (Vagin et al., 2004; Saenger, 1983; Taylor & Kennard, 1982).

Four potassium ions were visible in the electron density map and were fitted into the central channel in the quadruplex. The structure of the complex formed between the quadruplex and BSU6042 in space group $P2_12_12_1$ showed the last thymine in the ligand-containing loop to flip back into a new position and was refitted (discussed elsewhere). Also, the first thymine in the complex formed with the ligand BSU6066 flipped into a new position and was refitted.

For each ligand, the maps revealed a significant electron density volume present in only one of the T_4 loops and stacking onto the top G-quartet. The volume matched the shape and volume of its corresponding ligand and was subsequently fitted. One ligand was bound to one quadruplex in all instances.

Water molecules were initially placed automatically with the *Find Waters* facility in Coot version 0.5 (Emsley & Cowtan, 2004) using the default sigma cut-off of 2.5. The $2F_o$-F_c and F_o-F_c maps were generated and inspected visually to accept, delete or add water molecules. The added water molecules were examined and accepted according to the following criteria: density of spherical shape, water-sized, present no clashes with neighbouring atoms, form hydrogen-bonding interactions with suitable donor/acceptor neighbouring atoms and standard distance and geometry. The rejected water molecules were deleted. This cycle of building new water molecules and refinement followed by rejecting failed waters was repeated until no water molecules could be added.

$P2_12_12_1$

Flipped thymine In this space group, the position of the last thymine in the ligand containing loop was ambiguous i.e. it did not fit into the $2F_o$-F_c density compared to the rest of the structure at the same sigma level. The F_o-F_c map confirmed the presence of a volume similar in shape and size to a thymine flipped into a new conformation. This thymine nucleotide was omitted and an omit map was calculated. The thymine was fitted into the omit map.

Fitting the ligand When examining any remaining unmodelled density in the maps, significant density for a third ligand molecule was observed. This volume of electron density was similar in shape and volume to a spermine molecule and was subsequently fitted.

3.2.6 Deposition in the Protein Data Bank Database

The structures solved in this work were deposited in the Protein Data Bank Database on with PDB IDs 3EM2, 3EQW, 3EUI, 3ERU, 3ES0, 3ET8 and 3EUM. (corresponding ligands are listed in table 3.2).

3.3 Experimental results

A total of six ligands were co-crystallised with the *Oxytricha nova* telomeric DNA sequence d[GGGGTTTTGGGG] forming a bimolecular quadruplex structure that is isomorphous with the native structure (PDB IDs 1JPQ and 1JRN (Haider et al., 2002)) and the previously reported complex with compound 7 (Figure 3.4). All the complexes crystallised in space group $P2_12_12$, (PDB ID 1L1H (Haider et al., 2003)). A second form of crystals were obtained with compound 2 (Figure 3.4). These contained two independent quadruplex complexes in the asymmetric unit. These two complexes are closely similar to the other complexes.

3.3.1 Quality of the models

The agreement of the models with the experimental crystallographic data is shown

by the low crystallographic R_{factor} and R_{free} values (Table 3.2).

Table 3.2: Crystallographic data

	1	2a	2b	3	4	5	6
PDB ID	**3EM2**	**3EQW**	**3EUI**	**3ERU**	**3ES0**	**3ET8**	**3EUM**
Ligand	**BSU6038**	**BSU6042 - small**	**BSU6042 - large**	**BSU6045**	**BSU6048**	**BSU6054**	**BSU6066**
Data collection							
Total number of reflections collected	13302	16162	24705	17542	13641	9322	12155
Number of unique reflections	3093	3454	5854	4422	3549	2602	5372
Space Group	$P2_12_12$	$P2_12_12$	$P2_12_12_1$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Cell dimensions: a, b and c (Å)	55.23 42.71 26.84	55.28 42.73 26.62	55.356 42.542 48.636	55.22 42.63 26.85	55.53 42.46 27.30	55.40 42.59 27.14	55.62 42.31 27.08
Angle (degrees)							
α	90	90	90	90	90	90	90
β	90	90	90	90	90	90	90
γ	90	90	90	90	90	90	90
Maximum resolution (Å)	2.09	1.73	2.09	1.92	2.00	2.45	1.78
R_{merge}	0.057	0.092	0.087	0.059	0.041	0.084	0.079
I/σI	6.5	9.5	7.1	12.1	15.1	8.3	15.6
Completeness (%)	96.3	79.7	92.8	90.0	99.8	99.7	81.7
Redundancy	3.71	3.31	3.61	3.77	2.91	3.58	2.3
Refinement							
Resolution range used in refinement (Å)	27.62 - 2.30	22.59 - 2.20	27.68 - 2.20	22.72 - 2.00	24.50 - 2.20	27.15 - 2.45	24.35 - 1.78
Number of unique reflections used in refinement	3000	3454	5575	4215	3549	2471	5106
Completeness (%)	97.0	99.6	94.5	95.7	99.6	99.4	81.9
R_{factor}	20.3	22.2	23.4	19.9	19.5	20.7	22.2
R_{free}	29.3	31.4	28.5	24.9	24.5	27.7	27.0
Number of G-quadruplexes per asymmetric unit	1	1	2	1	1	1	1
Number of ligands per asymmetric unit	1	1	2	1	1	1	1
Number of asymmetric units per unit cell	4	4	4	4	4	4	4
Number of atoms							
DNA	506	506	1012	506	506	506	506
Ligand	36 n/a	40 n/a	80 14	42 n/a	38 n/a	38 n/a	38 n/a
Potassium ions	4	4	8	4	4	4	4
Water	64	66	162	71	56	51	52
Average B-factor values (Å²)							
DNA	16.11	11.72	18.73	16.06	9.86	15.58	15.04
G-quartets	15.43	11.20	17.66	15.56	9.26	14.27	13.73
Loops	17.59	12.85	21,04	17.60	11.16	18.43	17.86
Ligand **Spermine**	18.02 n/a	17.19 n/a	24.08 29.70	18.73 n/a	10.52 n/a	17.27 n/a	16.68 n/a
Ions - Potassium	18.07	14.11	18.84	15.11	10.94	22.58	13.02
water	23.13	19.16	31.23	27.02	21.11	25.71	31.40

Average B-factor values for the DNA, G-quartet, loops, ligand, ions and water are listed (Table 3.2). As expected, B-factor values for the G-quartets are lower than those for the loops indicating their static nature and the loops' dynamic nature.

For all the models, the electron density map ($2F_o$-F_c) is continuous and contains the model in its entirety and no unexplained electron density is observed. For example, the electron density maps are shown for the complex formed with the ligand BSU6048 and BSU6042 in space groups $P2_12_12$ and $P2_12_12_1$, respectively (Figure 3.6).

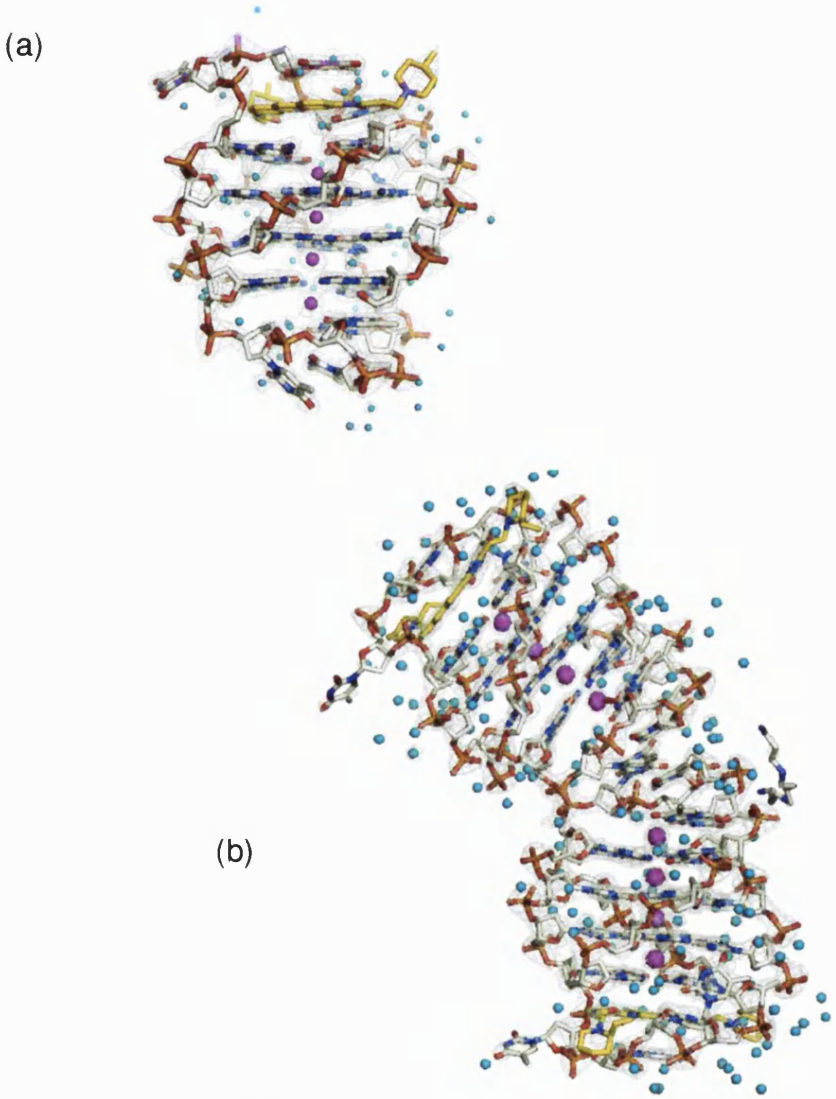


Figure 3.6: Electron density maps shown at 1.5 sigma. (a) The complex formed with ligand BSU6048. (b) The complex formed with ligand BSU6042 in the $P2_12_12_1$ space group.

Also, individual ligands are shown fitted into the corresponding omit maps at 1.5 sigma (Figure 3.7).

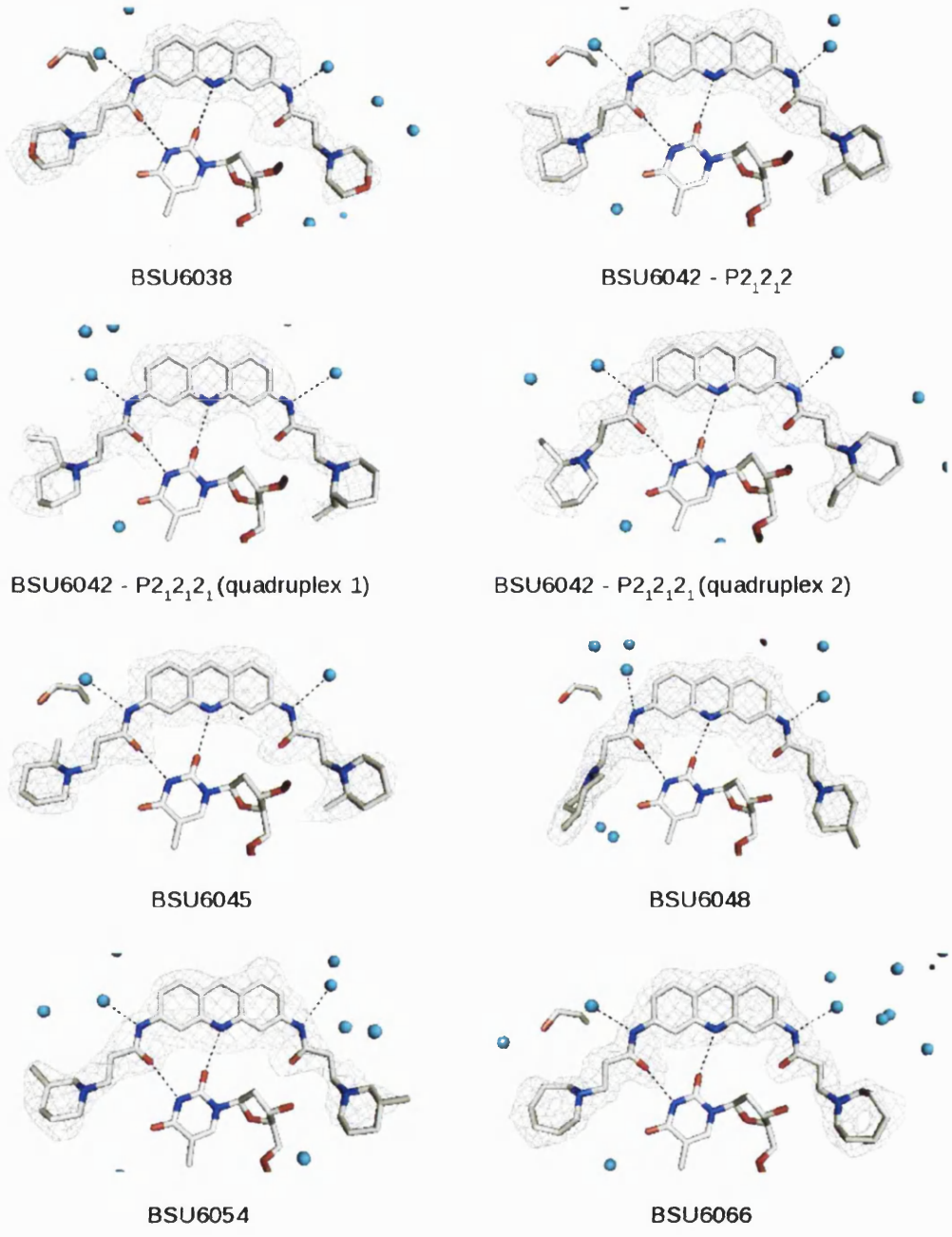


Figure 3.7: The ligands are shown fitted into the omit electron density map (at 1.5 σ). Immediate hydrogen bond contacts to thymine and water molecules are identical (shown in black dashes).

All of the atoms in the structures are fully occupied. In the P2$_1$2$_1$2$_1$ space group, the DNA is described by a total of 506 atoms, the ligand by 36 to 42 atoms depending on the ligand used, 4 potassium ions and 51 to 71 molecules of water depending on

structure. In the $P2_12_12$ space group, the DNA is described by a total of 1012 atoms, the two molecules of acridine ligand by 80 atoms, one molecule of spermine containing 14 atoms, 8 potassium ions and 162 molecules of water.

3.3.2 Crystal forms

All six ligands in the series formed complexes that crystallised in the same space group ($P2_12_12$) in accordance with previous observations for the complex (PDB ID 1L1H (Haider et al., 2003).

Additionally, ligand BSU6042 (compound 2 in Figure 3.4) crystallised in a novel space group ($P2_12_12_1$).

3.3.3 Overall structure description

Each asymmetric unit was formed of one quadruplex bound to one ligand present in one of the T_4 loops. The quadruplex was bimolecular, formed of a dimer of the sequence d[GGGGTTTTGGGG] and has anti-parallel conformation. The four thymine nucleotides formed a diagonal loop at either end of the quadruplex connecting two opposite guanine nucleotides in the G-quartet arrangement in a way so that it straddled the end G-quartet.

The complexes crystallised in two space groups $P2_12_12$ and $P2_12_12_1$.

In the $P2_12_12$ space group, the complex was formed of one ligand bound in the T_4 loop of the quadruplex. The other T_4 loop remained ligand-free (Figure 3.8). Both of the T_4 loops were diagonal. The asymmetric unit was formed of one complex with four asymmetric units in the unit cell.

In the $P2_12_12_1$ space group, the ligand BSU6042 was bound in an identical manner to the $P2_12_12$ space group. Here, two copies of the complex were found in the asymmetric unit. The unit cell contained four asymmetric units i.e. 8 complexes.

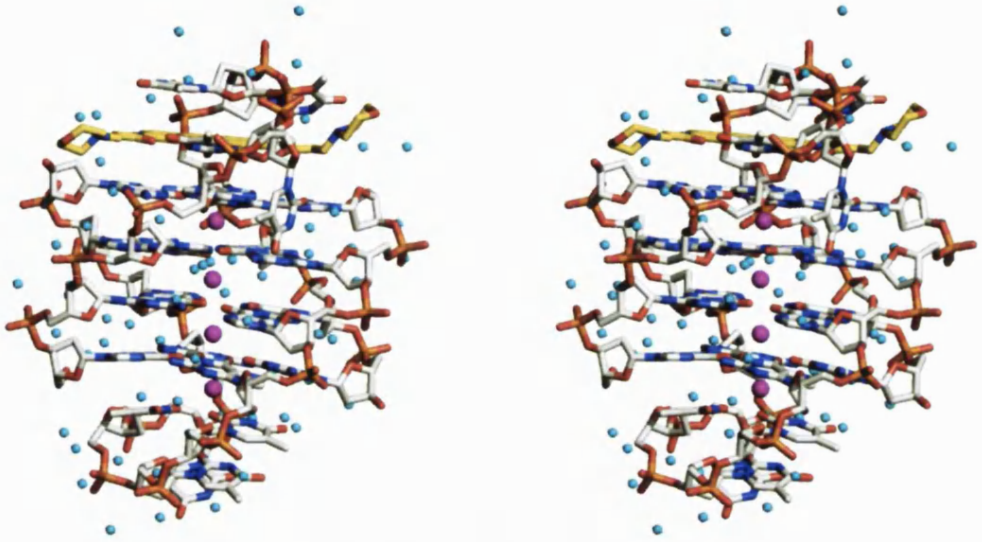


Figure 3.8: Stereo view of the complex formed between the Oxytricha nova telomeric quadruplex and the disubstituted ligand BSU6038. Space group is $P2_12_12$. Carbon atoms are shown in yellow in the ligand and in white in the DNA. Potassium ions and waters are shown in magenta and cyan spheres, respectively. Same colour scheme is used for the figures in this chapter unless stated otherwise.

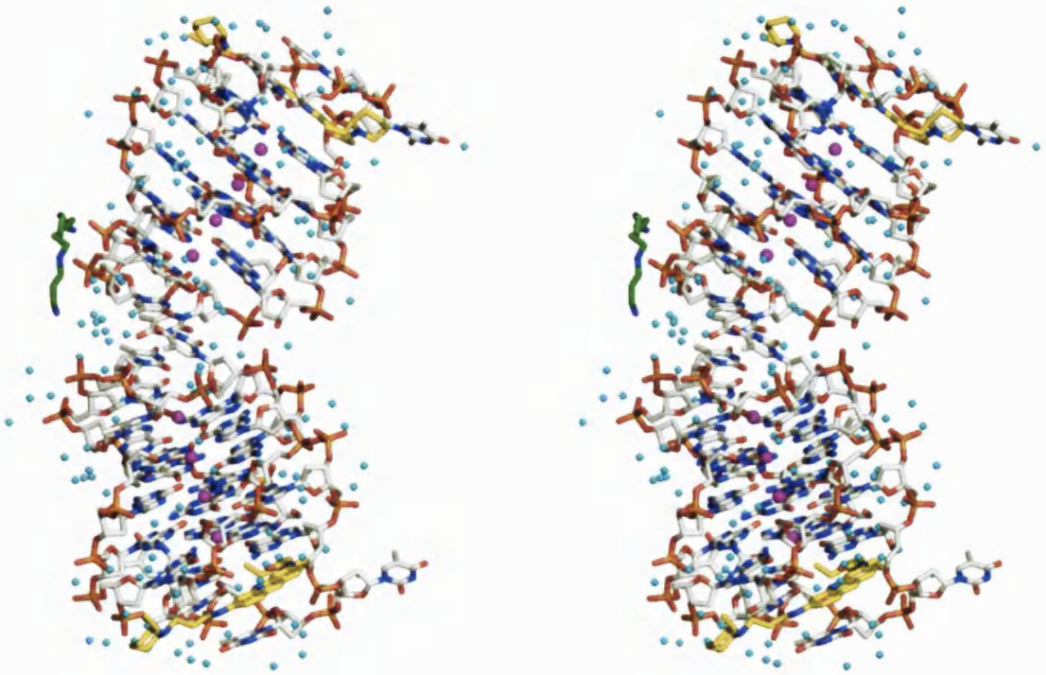


Figure 3.9: Stereo view of the complex formed between the bimolecular anti-parallel quadruplex formed by the telomeric sequence of Oxytricha nova and the disubstituted ligand BSU6042. Space group is $P2_12_12_1$. Carbon atoms in spermine are shown in green.

Additionally, the structure revealed one molecule of spermine - present in the crystallising solution - straddling the two complexes externally (Figure 3.9). This molecule of spermine was not seen in the other crystal form even though it was present in the crystallising conditions.

Crystal packing

$P2_12_12$

Three types of crystal packing interactions are observed in the $P2_12_12$ space group (Figure 3.10) and are in accordance with previous observations for the complex (PDB ID 1L1H (Haider et al., 2003)):

1. The first thymine in the ligand-bound loop flips out and away from the G-quartet so as it becomes perpendicular to it and forms π–π stacking interactions with a thymine from a symmetry related quadruplex (shown in (a) in Figure 3.10). However, the same thymine in the complex with BSU6066 forms two hydrogen bonds (3.0 and 3.1 Å) with the edges of a guanine base in another quadruplex (shown in (b) in Figure 3.10).

   The last thymine in the ligand-bound loop does not contact a symmetry related quadruplex.

2. The ligand-free loop (Figure 3.11): both of the second and fourth thymine nucleotides contribute to crystal packing arrangement by forming T·T base pairs with the fourth and second thymine nucleotides, respectively, belonging to the equivalent loop in a symmetry related quadruplex. Each base pair contains one hydrogen bond (2.7 Å shown in (c) in Figure 3.10) between the O4 atom of one thymine and the N3 atom of the other. This arrangement is observed in both of the native forms PDB ID 1JPQ and 1JRN (Haider et al., 2002), and in the ligand-free loop of the complexes' structures, including the complex with BSU6039 (PDB ID 1L1H (Haider et al., 2003)).

3. The ligand itself forms a hydrogen-bond (2.7 Å) with an oxygen atom

belonging to the DNA backbone of a symmetry related quadruplex (shown in (d) in Figure 3.10).

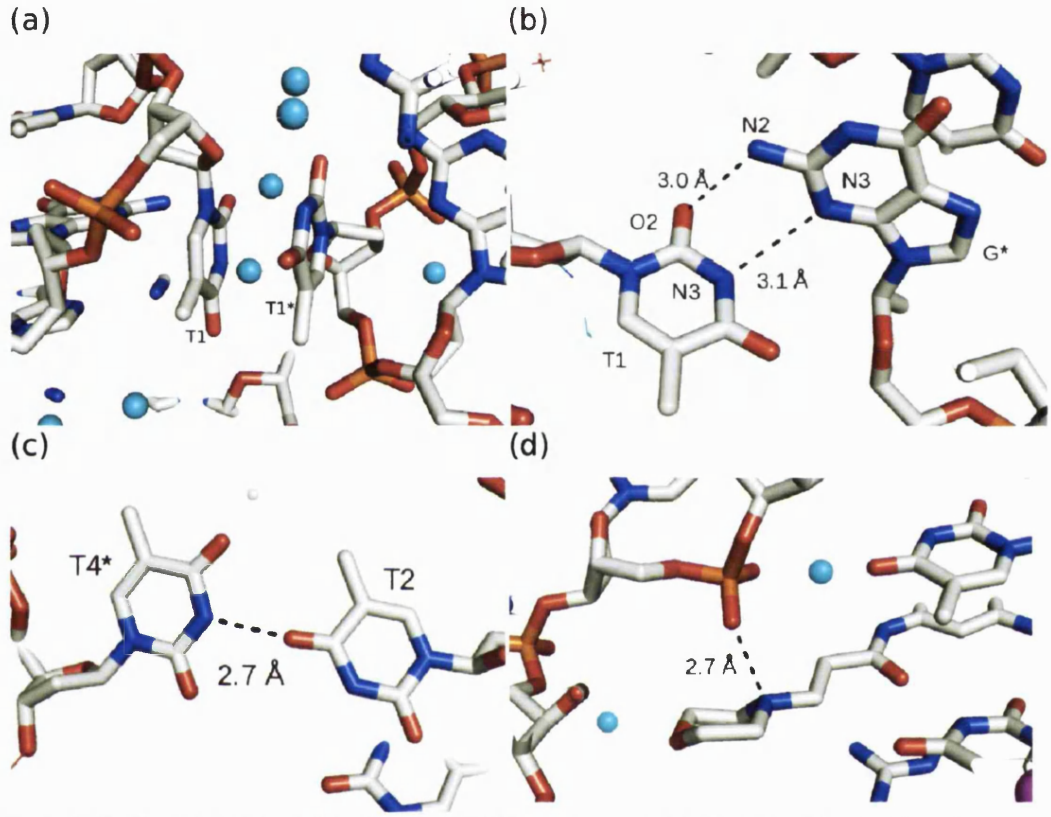


Figure 3.10: Crystal packing interactions in $P2_12_12$ space group. (a) Stacking of the first thymines (thymines belong to the ligand-bound loop). (b) In the complex formed with the ligand BSU6066, T1 (atoms N3 and O2) forms hydrogen bond interactions with the edges of a guanine base (N3 and N2, respectively) (c) T•T base pair. Hydrogen bond distances shown in dashed black lines. In (a+b+c), T1 and T2 belong to a quadruplex whereas T1, T4* and G* belong to a symmetry related quadruplex. (d) A hydrogen bond (2.7 Å) between the nitrogen in the end hetero ring of the flexible side chain with an oxygen belonging to the DNA backbone in a symmetry related quadruplex.*

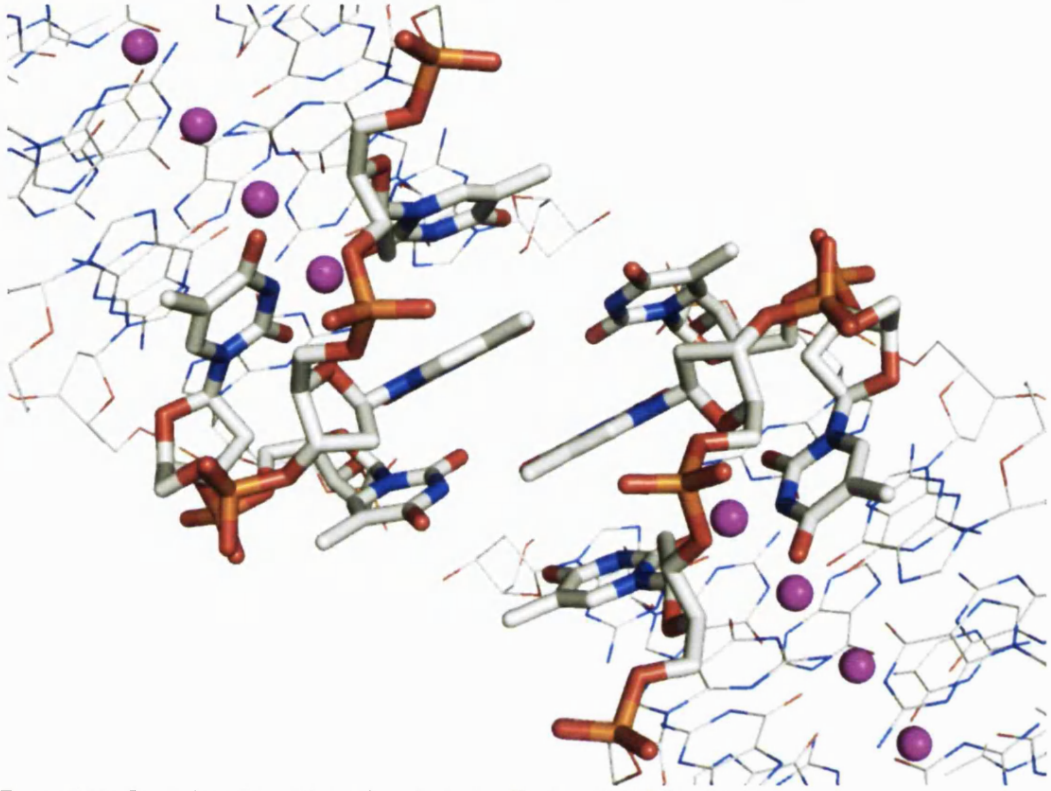


Figure 3.11: Crystal packing interactions between the two ligand-free loops.

$P2_12_12_1$

In this space group formed by the BSU6042-quadruplex complex:

1. The first thymine in the ligand-bound loop in both quadruplexes in the asymmetric unit has the same crystal packing interactions as in the $P2_12_12$ space group.

2. The last thymine in the ligand-bound loop in one of the quadruplexes in the asymmetric unit shows similar interactions to those observed for the first thymine in the complex with BSU6066 i.e. forms two hydrogen bond contacts with the edges of a guanine from a symmetry related quadruplex (shown in (b) in Figure 3.12). The equivalent thymine in the other quadruplex does not contact a symmetry related quadruplex reminiscent of the last thymine in the ligand-bound loop in space group $P2_12_12$. The conformation of these last thymines is compared to that of

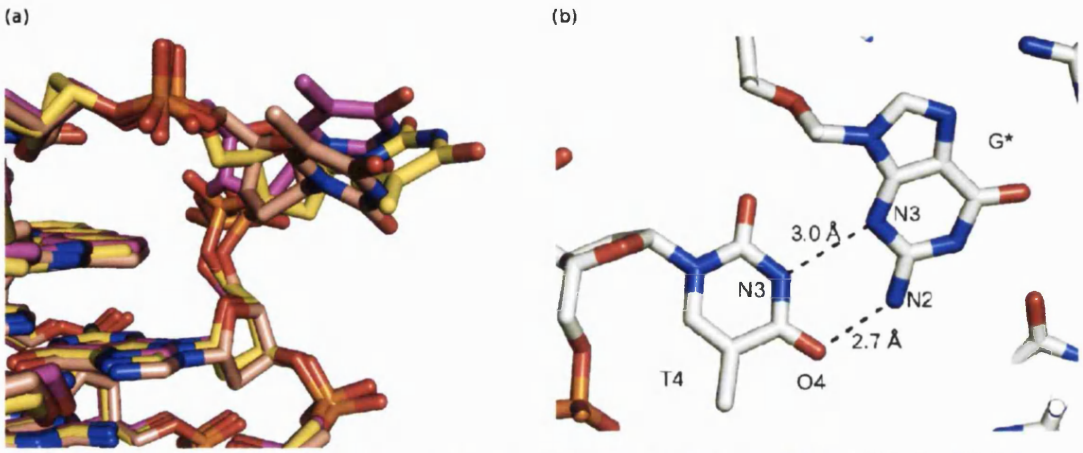


Figure 3.12: Crystal packing interactions in the complex formed with ligand BSU6042 in the $P2_12_12_1$ space group. (a) Alignment of the structures of the BSU6039 (carbon atoms in brown), quadruplex1 and quadruplex 2 (carbon atoms in magenta and yellow respectively) in the complex with ligand BSU6042 shows the conformational change undergone by the last thymine in the ligand-bound loop. (b) This thymine forms two hydrogen bonds (3.0 and 2.7 Å) with the edges of a guanine base from another quadruplex in a manner similar to the interactions formed by the first thymine in the complex with ligand BSU6066.

the same thymine in the complex with the ligand BSU6039 (PDB ID 1L1H (Haider et al., 2003)). The thymines adopt different conformations (shown in (a) in Figure 3.12) to adapt to the packing constraints in the local environmental.

3. The interactions formed by the ligand-free loop are identical to those seen previously in the $P2_12_12$ space group (Figure 3.11).

The guanine quartet

The conformation about the glycosidic bond for guanine bases is alternating *syn-anti* along each strand and *syn-syn-anti-anti* within each G-quartet. This in accordance with previous observations for the complex with the ligand BSU6039 (PDB ID 1L1H (Haider et al., 2003)).

Rise

The programme 3DNA (Lu & Olson, 2003) was used to calculate the rise between consecutive G-quartets. Rise is measured by the same strand C1' ··· C1' vectors

projected onto the normal vector of the "mean plane".

The rise values for the first, second and third step in the complexes presented in this chapter have mean values of 3.4±0.1, 3.4±0.1 and 3.3±0 Å respectively. This is in agreement with observations in native crystal structures;; in complex with a protein (PDB ID 1JB7 (Horvath & Schultz, 2001)), or in the crystal (PDB IDs 1JPQ and 1JRN (Haider et al., 2002)), as well as in the crystal of the complex ((Haider et al., 2003)) (mean values of 3.3±0.1, 3.4±0 and 3.3±0.1 Å, respectively).

Glycosidic torsional angle χ

The torsion angle χ was also calculated using the programme 3DNA (Lu & Olson, 2003). Torsion angles are defined in terms of four atoms: O4'-C1'-N9-C4 for purines, and O4'-C1'-N1-C2 for pyrimidines.

Along each of the DNA strand, the guanine nucleotide alternate *syn* and *anti* conformations whilst the thymine nucleotides in both of the loops; ligand-bound and ligand-free loops, have *anti* conformation. This is in accordance with the native (PDB ID 1JPQ (Haider et al., 2002)) and ligand-bound structures (PDB ID 1L1H (Haider et al., 2003)).

Glycosidic torsion angle values for the DNA backbone strand A (ligand-bound loop) show a significant change upon ligand binding for T6 (from -95±2 to -157±3), whereas those for the guanine nucleotides remain unchanged (Figure 3.13).

For strand B, all the guanine and thymine nucleotides remain unchanged (Figure 3.14).

Loops

Each quadruplex contains two T_4 loop, one at either end. Each T_4 loop is diagonal and extends across the G-quartet connecting two guanine bases on opposite corners of the G-quartet. Only one of the T_4 loops in the quadruplex contains a ligand which displaces the first and last thymine, forms a coplane with the second thymine and stacks onto the G-quartet. The fourth thymine stacks on the acridine core of the ligand on the other side. The other loop remains unchanged,

maintaining an identical arrangement to that observed in the native structure. All the thymine bases maintain their *anti* conformation with the regards to glycosidic torsion angles.

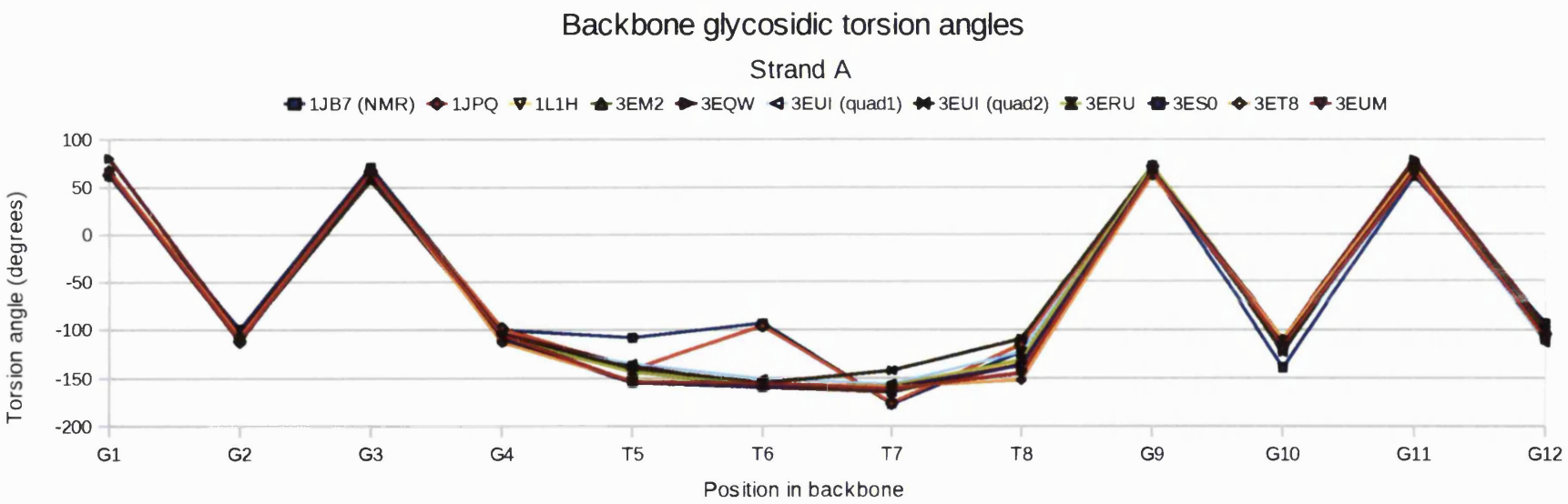


Figure 3.13: Glycosidic torsion angles for the DNA backbone strand A (ligand-bound loop) of the anti-parallel quadruplex formed by the Oxytricha nova telomeric sequence. The glycosidic angle values change significantly for T6 (from -95±2 to -157±3), whereas those for the guanine nucleotides remain unchanged. All nucleotides maintain their original syn and anti conformations as per previous observations.

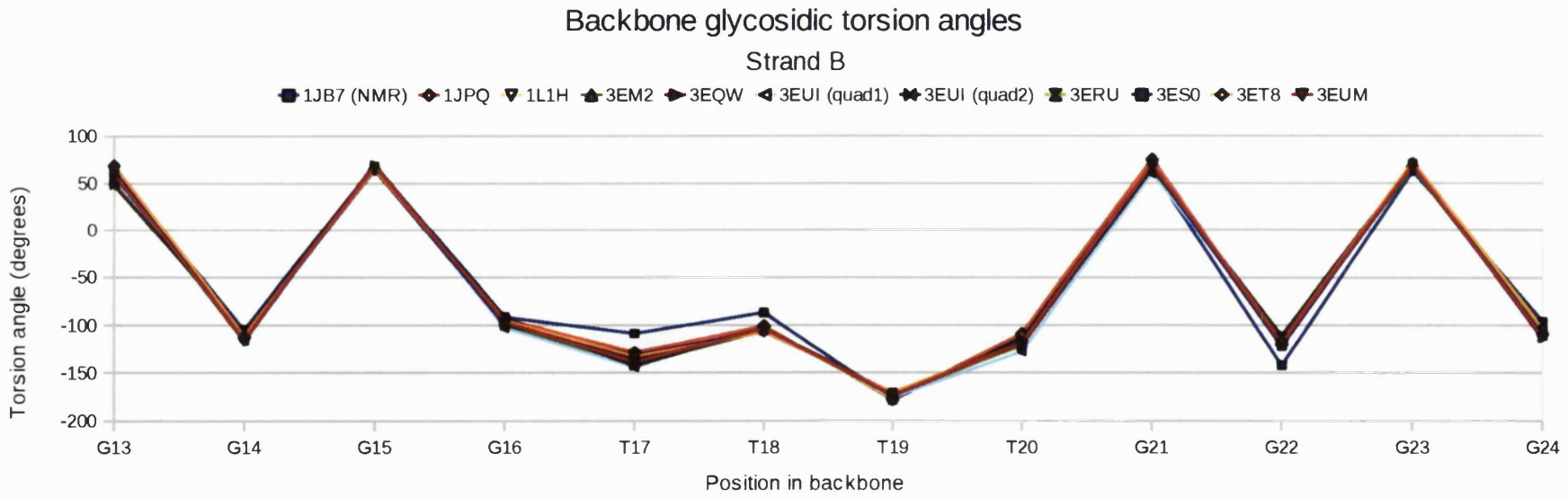


Figure 3.14: Glycosidic torsion angles for the DNA backbone strand B (ligand-free loop) in the Oxytricha nova telomeric quadruplex remain unchanged.

Groove types

Each quadruplex has four grooves. No changes occur in the structure of the grooves in of the complexes. The aligned DNA of all complexes is shown (Figure 3.15).

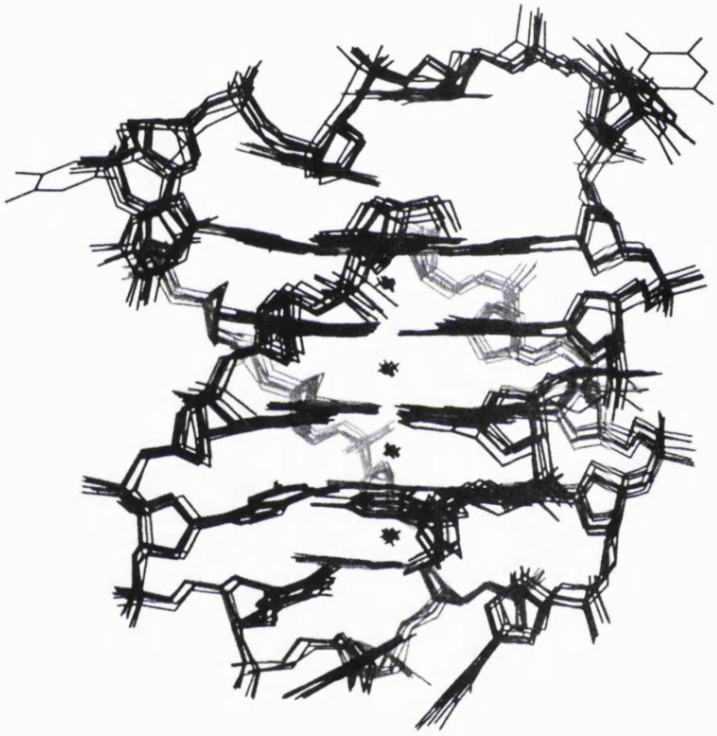


Figure 3.15: Aligned DNA quadruplexes. The DNA for the seven complexes reported here is shown in black lines. The ligands have been omitted for clarity.

The walls of each groove are formed of the negatively charged DNA backbone (phosphate OP1, OP2 and sugar O4´) and its floor is formed of the positively charged edges of the G-quartet-forming guanine bases (N2, N3 and N7). The anti-parallel topology and diagonal loops result in three types of grooves (Figure 3.16):

1. one wide groove: its walls are lined with the DNA backbone of the two anti-parallel 5' strands. This results in the oxygen atoms part of the deoxyribose sugar pointing towards the groove in both strands.

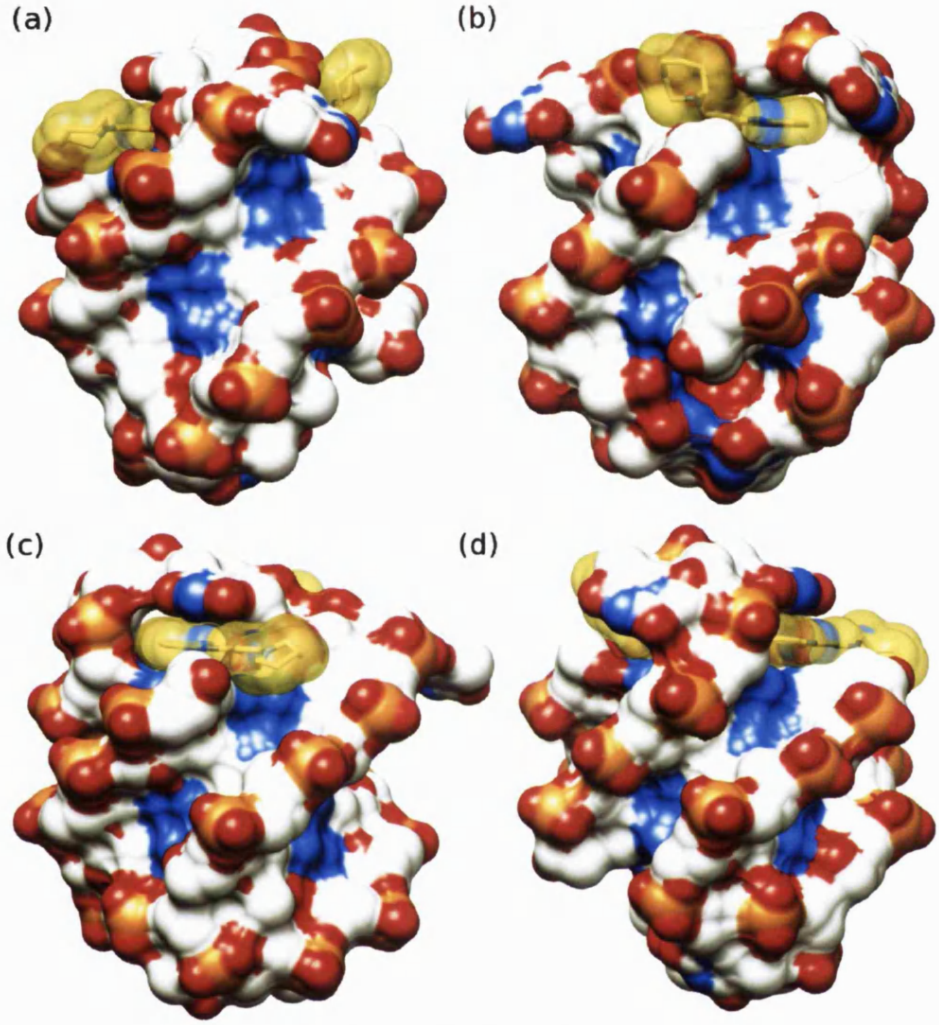


Figure 3.16: Grooves in the anti-parallel quadruplex formed by the telomeric Oxytricha nova. This quadruplex contains one wide groove (a), two medium grooves (b+c) and one narrow groove (d). Surface representation of the DNA (colours as per atom underneath) and ligand (transparent yellow) are shown.

2. two medium grooves: their walls are lined with the parallel DNA backbone of one 5' and one 3' strand. This results in the oxygen atoms part of the deoxyribose sugar in the 3' strand only pointing into the groove.
3. one narrow groove: its walls are lined with the anti-parallel DNA backbone of two 3' strands. This results in the oxygen atoms belonging to the deoxyribose moieties to point away from the groove and into the adjacent grooves on either side (each one of which is a medium groove).

The shape, volume and dimensions of the grooves are fixed due to the steric constraints on backbone and G-quartet movement.

All four grooves are highly hydrated. The water molecules tend to form networks that connect the phosphate backbone, the deoxyribose oxygen atoms and the edges of guanine bases.

The structures reveal that the 3- and 6-substituents side chains extend towards the grooves but not into them; one extends towards the wide groove and the other towards an adjacent medium groove. No interactions are observed , whether directly or bridged through water molecules, between the ligand and the DNA in the grooves.

3.3.4 Binding site and and ligand binding mode

The binding site is formed of the G-quartet and the thymine nucleotides belonging to the T_4 loop. The acridine core of the ligand stacks onto two guanines part of the G-quartet. The second thymine in the T_4 loop forms a coplane with the ligand. The third thymine stacks onto the acridine core of the ligand. The first and fourth thymines flip out and away where they are available for crystal packing interactions

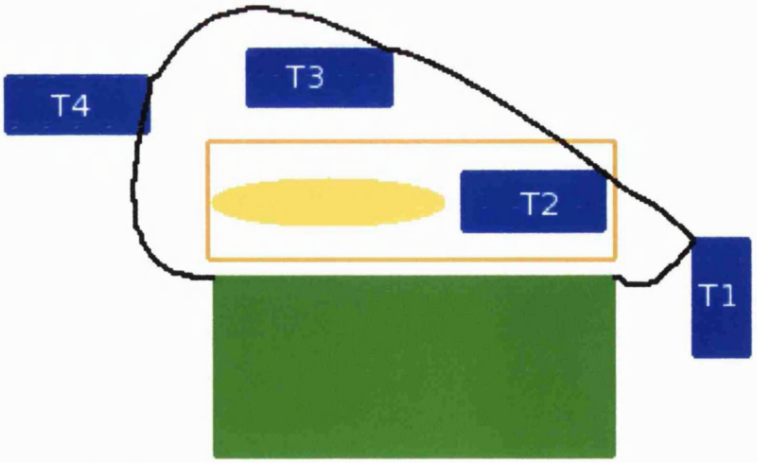


Figure 3.17: A schematic representation of the complex formed between a disubstituted acridine and the Oxytricha nova telomeric quadruplex. The ligand (shown in yellow oblong) forms a coplane (orange rectangle) with a thymine (T2=blue rectangle) which stacks onto the G-quartet belonging to the quadruplex (green rectangle). The third thymine (T3) stacks onto the ligand. The first and last thymines are displaced by the ligand (T1 and T4).

(Figure 3.17). The side chains of the ligand extend towards the grooves but no direct contacts are observed with the atoms within.

All ligands belonging to the series studied here adopt an idlentical binding mode to the ligand BSU6039 (compound 7: Figure 3.4) in its complex with the *Oxytricha nova* telomeric quadruplex (PDB ID 1L1H (Haider et al., 20l03)).

A summary of the pattern of hydrogen bond contacts as formed by the ligands in this series is listed below:

1. The central nitrogen of the acridine core forms one hydrogen bond with the O2 atom of the coplanar thymine (Figure 3.2).
2. The oxygen atom belonging to the carbonyl group of the amide linker of one of the side chain forms one hydrogen bond with the N3 atom of the coplanar thymine (Figure 3.2).
3. The nitrogen atom belonging to the same side chain as in (2) above forms one hydrogen bond to a water molecule that participates in local water networks ending in contacting the DNA.
4. The equivalent nitrogen atom belonging to the other side chain forms one hydrogen bond with a molecule of water that participates in local water networks that contact the DNA.
5. The protonated nitrogen in the end hetero ring forms a hydrogen bond with the OP1 oxygen of the DNA backbone of the loop (Figure 3.18).
6. The other end hetero ring contains an equivalent nitrogen which contacts an OP1 oxygen in a symmetry related quadruplex (Figure 3.18).

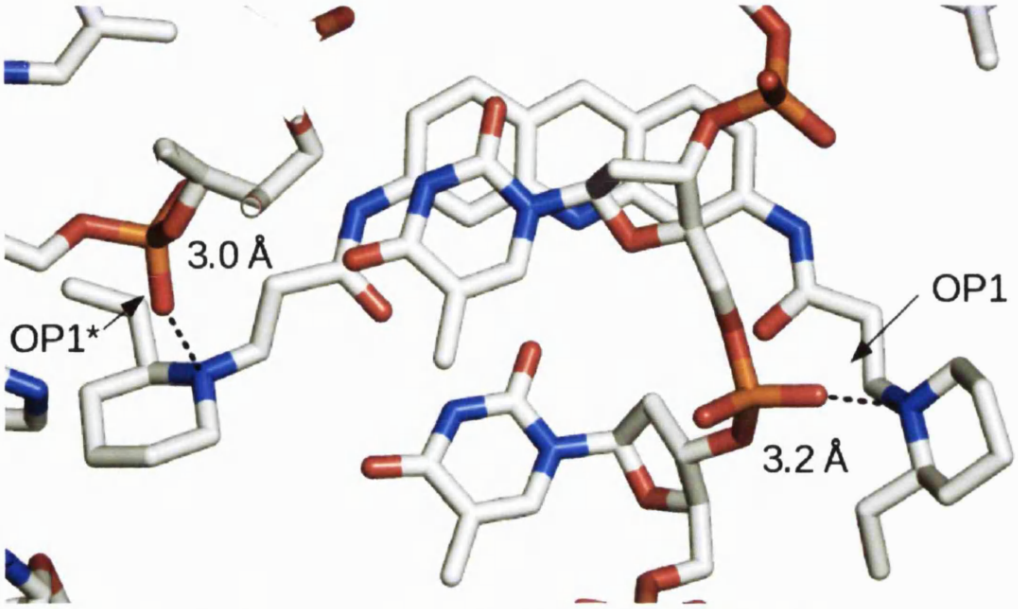


Figure 3.18: Hydrogen bond contacts between the DNA backbone and the protonated nitrogen atoms in the end hetero rings. One nitrogen atom contacts an OP1 oxygen in the backbone (3.2 Å) and the other nitrogen contacts an OP1 in another quadruplex (3.0 Å).*

Ligand binding mode

The acridine core and the amide linkers form identical patterns of hydrogen bonds with the coplanar thymine and immediate waters (Figure 3.7). However, the position of the alkyl substitution in the end hetero ring i.e. *ortho*, *meta* or *para*, affects the exact orientation of the substitution relative to the quadruplex (Figure 3.19). For example, the alkyl substitutions in ligands BSU6045 and BSU6042 (*ortho* methyl- and *ortho* ethyl-substituents, respectively) and ligand BSU6054 (*meta* methyl-substituent) have identical orientation. In contrast, BSU6048, which contains a *para* methyl substituent, has a significantly different orientation. It rotates away from the plane of the acridine core becoming perpendicular to it, as shown by the difference of ca. 60° in side chain torsion angle α_2 (Figure 3.20).

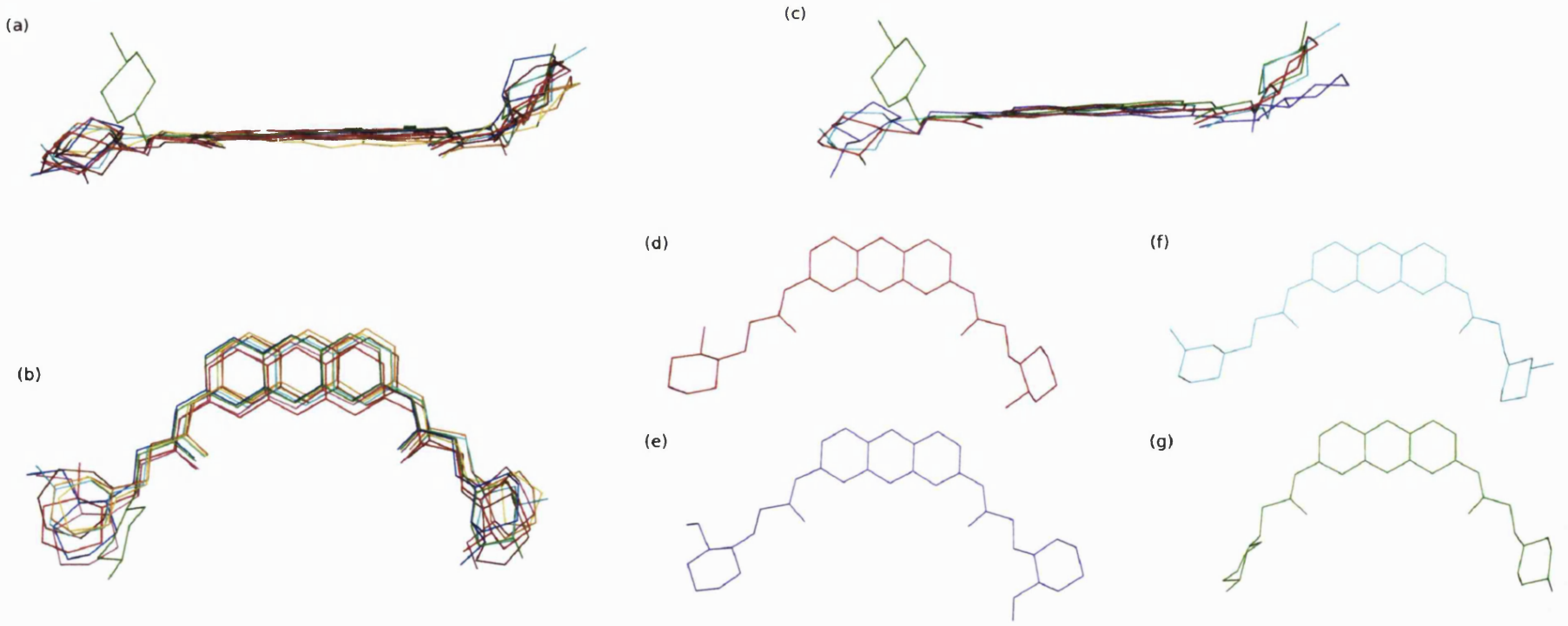
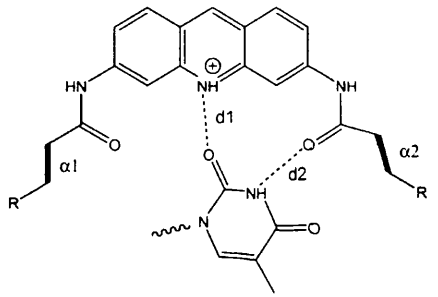


Figure 3.19: Ligand binding mode. (a) A side view of the ligands in the aligned complexes and a top view in (b). (c) A side view of ligands with ortho-, meta- and para- alkyl substituents in the aligned complexes and separately a top view in (d) BSU6045, (e) BSU6042, (f) BSU6054 and (g) BSU6048.



	1	2a	2b	3	4	5	6	7
PDB entry	3EM2	3EQW	3EUI	3ERU	3ES0	3ET8	3EUM	1L1H
d_1	3.1	3.1	3.1, 2.9	2.9	3.0	2.9	2.9	3.0
d_2	2.7	2.7	2.8, 2.8	2.8	2.7	2.7	2.7	2.7
α_1	238	230	225, 215	230	232	243	238	160
α_2	180	161	153, 145	175	112	187	180	212

Figure 3.20: Selected ligand-thymine intermolecular distances (Å) and side chain torsion angles (degrees). Structure with PDB ID 3EUI has two quadruplex-ligand complexes in the asymmetric unit.

In all instances, one hydrogen bond is maintained with the OP1 oxygen of the symmetry related quadruplex at a distance of 2.8 Å for BSU6048 and 2.7 Å for BSU6054 (Figure 3.21). Positioning the para substitution in this way allows for the methyl substituent to flip into a larger volume where it can be accommodated.

The size of the substituent also plays a role with regards to crystal packing. The ligand BSU6042 crystallises in a related space group $P2_12_12_1$. The extra bulk of the substitution could be destabilising the original arrangement resulting in a second crystal form.

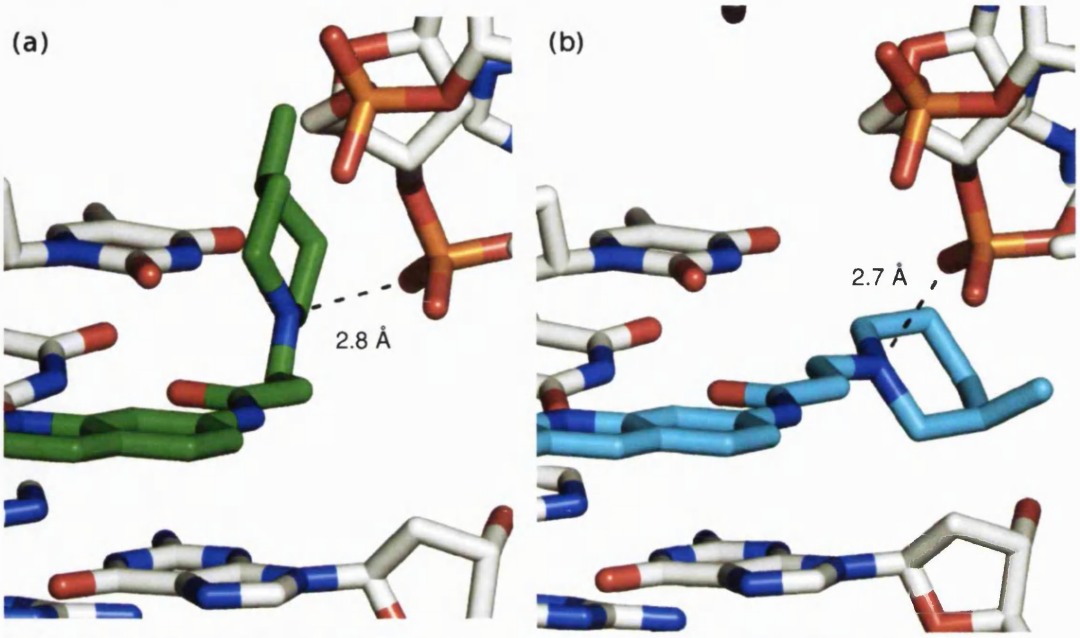


Figure 3.21: Effect of meta- and ortho- methyl substitutions in the end hetero ring on ligand binding mode. The end hetero ring (ortho-substituent) in the ligand BSU6048 (carbons shown in green) is rotated by approximately 60° (a). In contrast,the orientation of the end hetero ring in the ligand BSU6054 (carbons shown in cyan). Hydrogen bonds are shown in dashed black lines.

3.3.5 Water structure

The structures contain extensive water networks in accordance with previous observations for the native structure (PDB IDs 1JPQ and 1JRN (Haider et al., 2002)) and in complex with BSU6039 (PDB ID 1L1H (Haider et al., 2003)).

The water molecules form hydrogen bonds with N2 and N3 of the guanine bases and the O1P and O2P of the phosphate DNA backbone.

Water molecules vary between 51 to 71 in the $P2_12_12$ space group depending on structure (Table 3.1). There are 162 water molecules in the $P2_12_12_1$ space group structure.

However, with regards to the ligand, two water molecules play a significant role in fixing the ligand in place. Each of the water molecules contact an amido nitrogen in the side chains. These water molecules are seen in each of the complexes reported in this chapter (Figure 3.7) but are not seen in the previously solved structure for the complex with BSU6039. However, visualisation of the electron density maps

shows significant density for these water molecules adjacent to the nitrogens in the amide groups as expected. The density is spherical in shape with standard

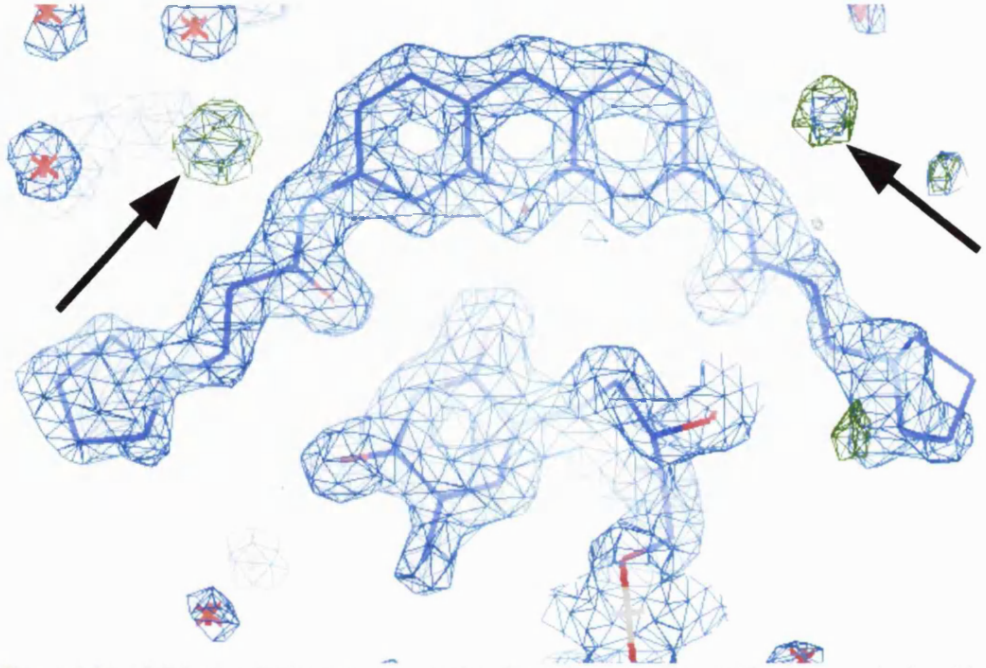


Figure 3.22: Visible density for two water molecules adjacent to the amide nitrogens in the side chains. $2F_o$-F_c maps are drawn in blue at 1.0 sigma and F_o-F_c maps are drawn in green at 3.00 sigma.

geometry (Figure 3.22).

3.4 Discussion

We have determined the crystal structures of six disubstituted acridine ligands in complex with the *Oxytricha nova* telomeric quadruplex and a series of corresponding disubstituted acridine ligands. The bimolecular quadruplex structures, formed by the sequence d[GGGGTTTTGGGG], are bimolecular and anti-parallel in accordance with previous studies in the crystal of the native form (Haider et al., 2002) and in a protein-quadruplex complex (Horvath & Schultz, 2001). It is also in agreement with the earlier ligand-bound crystal structure (Haider et al., 2003).

The structural changes between the complexes are minimal with an overall root-mean-square deviation (RMSD) of <0.2 Å. Amide and side chain conformations in

the acridine derivatives and first-shell water environments remained unchanged.

The ligand displaces the first and last thymines, forms a coplane with the second thymine, which in turns stacks onto the end G-quartet on one side and onto the third thymine on the other side. The coplane is formed through a hydrogen bond between O2 in the thymine and the central ring nitrogen atom of the acridine (with a mean distance of 3.0±1 Å). This suggests that this nitrogen atom is protonated. A similar arrangement is seen in the human telomeric-BRACO19 complex crystal structure (PDB ID 3CE5 (Campbell et al., 2008)). In the latter, a water molecule bridges the hydrogen bonds between the central ring nitrogen atom and the thymine ring (Figure 3.2).

The flexible side chains of the ligands extend towards adjacent grooves but does not contact them directly. Instead, the nitrogen atom belonging to the end hetero ring of one of the side chain form a hydrogen bond with the DNA backbone of the loop. The equivalent nitrogen in the other side chain forms a similar hydrogen bond but with another quadruplex. The morpholino end ring in compound 1 (Figure 3.4) is expected to be neutral at physiological pH, whereas the ring nitrogen atom in all the other compounds is expected to be protonated. This is reflected in the consistent findings of superior quadruplex binding and biological activity of acridine-based and many quadruplex-binding ligands that have cationic charged side-chains (Monchaud & Teulade-Fichou, 2008; Ou et al., 2008; Read et al., 1999; Read et al., 2001; Moore et al., 2006). However, as the structures reported in this chapter, including the complex formed with morpholino-containing compound 1, may suggest that these compounds exert their effect on ligand binding through long-range electrostatic interactions.

It is notable that there are no direct amine nitrogen-phosphate contacts with the groove in any of these structures or any indirect amine water-phosphate bridges, which have been observed in the human telomeric-BRACO19 complex crystal structure (PDB ID 3CE5 (Campbell et al., 2008)).

The conformations of the cyclic amine end groups and side chains of the acridine ligands all cluster together (a and b: Figure 3.19), although the alkyl groups

attached to the hetero rings are oriented in a variety of directions (d, e, f and g: Figure 3.19). The ligands are divided into two distinct groups with regards to the binding mode of the end hetero rings. This seems to be dependent on the position and size of the alkyl substitution in the hetero-cycle. The ligands which have an *ortho*- or *meta*-, methyl- or alkyl-substituent (or no substitution), bind in an identical binding mode, whereby, the ring that contains the nitrogen interacting with the backbone of the quadruplex adopts a conformation coming away from the plane of the acridine core and reaches up towards the loop. The other ring, the one that contains the nitrogen that interacts with the DNA backbone in another quadruplex, remains on the same plane as the acridine ring. This is in contrast to the position of the same ring in the complex formed with the ligand BSU6048 (compound 4: Figure 3.4). In this instance, the ring comes away from the plane of the acridine, whilst maintaining its hydrogen bond contact with the DNA backbone (a: Figure 3.21), in a manner similar to the ring at the other side chain (a and c: Figure 3.19). This new position enables the ligand to maintain its crystal packing interactions whilst avoiding steric clashes between the *para*-substituent on the ring and the nearby quadruplex. The *ortho*-ethyl substituents in BSU6042 (compound 2: Figure 3.4) are relatively close to backbone and sugar oxygen atoms, suggesting that a suitable ortho substituents such as a methylene hydroxyl group could form additional hydrogen bonds to O3' atoms or phosphate oxygen atoms.

Also, increasing the size of the substituent produced an effect on the crystal structures. The ligand BSU6042 (compound 2: Figure 3.4) with an *ortho*-ethyl substituent on the end hetero rings in the side chains, crystallised in two space groups. One form crystallised in the original space group $P2_12_12$ i.e. identical to the other ligands. The other form crystallised in the space group $P2_12_12_1$. A closely related ligand, BSU6045, contains an *ortho*-methyl substituent instead of an ethyl, and crystallised only in the original space group $P2_12_12$. The $P2_12_12_1$ space group contains new crystal packing interactions for the last thymine in the ligand-bound loop which are reminiscent of the those in the first thymine in the complex formed with the ligand BSU6066. BSU6066, similarly to BSU6042, contains a bulky hetero ring i.e. 7-membered ring. It seems that the bulky end hetero rings destabilise the

original crystal packing interactions with regards to the first and last thymines in the ligand-bound loop.

The structures reported in this work show that the end substituents in the side chains extend towards the grooves but do not contact them. On the contrary, the hetero rings extend towards the solvent and away from the quadruplex. This study shows that the ligands in this series, or at least the ones for which a crystal structure is reported here, do not show significant differences in ligand binding mode that explain the experimental observations with regards to biological assays of activity against telomerase (Harrison et al., 1999). At the same time, observations with regards to the new crystal packing interactions caused by the changes to the end hetero ring substituent with regards to position i.e. *ortho*-, *meta*- and *para*-substituent, or with regards to size i.e. *ortho*-methyl and *ortho*-ethyl substituents, suggest that the differences between these ligands have a significant effect in ligand-DNA interactions, at least with regards to crystal packing arrangement.

The similarities in the immediate environments for the disubstituted acridine ligands in complex with the *Oxytricha nova* anti-parallel telomeric quadruplex and for the trisubstituted acridine ligand BRACO19 in complex with the parallel human telomeric quadruplex are notable. BRACO19, containing an extra substitution at the 9th position, has superior quadruplex affinity, telomerase inhibition and cell cytotoxicity (discussed in chapter 2).

In both cases, the acridine core forms a coplane with a thymine through a hydrogen bond between the central nitrogen in the acridine ring and the O2 atom (disubstituted acridine) or N3 atom in the thymine base (BRACO19). The thymine in both cases originates from a loop thymine, but whereas the thymine base in the diagonal loop in the *Oxytricha nova* quadruplex is already in place over the binding site, the propeller loop in the human telomeric quadruplex would have to undergo more significant conformational adaptations to place the thymine in-plane. This coplanar thymine along with the original propeller loops and the stacking quadruplexes create a binding pocket that surrounds the end hetero ring substituent

i.e. pyrrolidino ring of BRACO19. This pocket is bounded by the DNA backbone of the loops (a: Figure 3.23). By contrast, in the *Oxytricha nova* - ligand complexes, the end hetero ring substituents in the ligand are pointing towards an open space and towards the solvent i.e. away from the DNA (b: Figure 3.23). This is due to the diagonal conformation of the loop in the *Oxytricha nova* quadruplex. The large open nature of the binding site in the *Oxytricha nova* quadruplex enables it to accommodate a piperidino ring with methyl or ethyl substituents (*ortho*-, *meta*- or *para*-substituted), as well as a seven-membered ring (compound 6: Figure 3.4). In contrast, replacing the smaller pyrrolidino ring with a larger piperidino or a seven-membered ring, results in a 7-fold lower binding constant to the human telomeric quadruplex (a reduction from 9.72×10^4 to 1.38×10^4 M^{-1}) (Read et al., 1999). However, a combination of a pyrrolidino ring - as 3- and 6-substituents - and a suitable 9-substituent such as the phenylamino present in BRACO19 (a: Figure 3.23) results in an increase in the binding constant by an order of magnitude to 4×10^5 M^{-1} (Read et al., 2001). The 9-substituent fits into a tight pocket bounded by DNA backbone atoms and a thymine base (discussed elsewhere).

Furthermore, the tight fit of the 9-substituent into a tight pocket lined by the DNA backbone and a thymine base, forces the acridine core to become rotated relative to the G-quartet it stacks on by approximately 45°. This further indicates that ligand binding to quadruplexes is greatly affected by loop constraints (Figure 3.24). Thus, in broad terms, a diagonal loop allows for a significant increase in the size of the 3- and 6-substituents by virtue of the loop's position, whereas, the propeller loop presents a more succinct binding site with severe constraints on substituent size.

These differences can be attributed to the distinct position of the loops in these two structures: the external placement of the propeller loop will constrain the size of the substituent size at the end of the side chains. In contrast, the diagonal loop is involved in π–π stacking with the aromatic core of the ligand. Structural studies suggest that diagonal loops are less prevalent among human telomeric quadruplexes (discussed elsewhere). Although a diagonal loop is present in the human *c-myc* promoter quadruplex in its complex with porphyrin as reported by

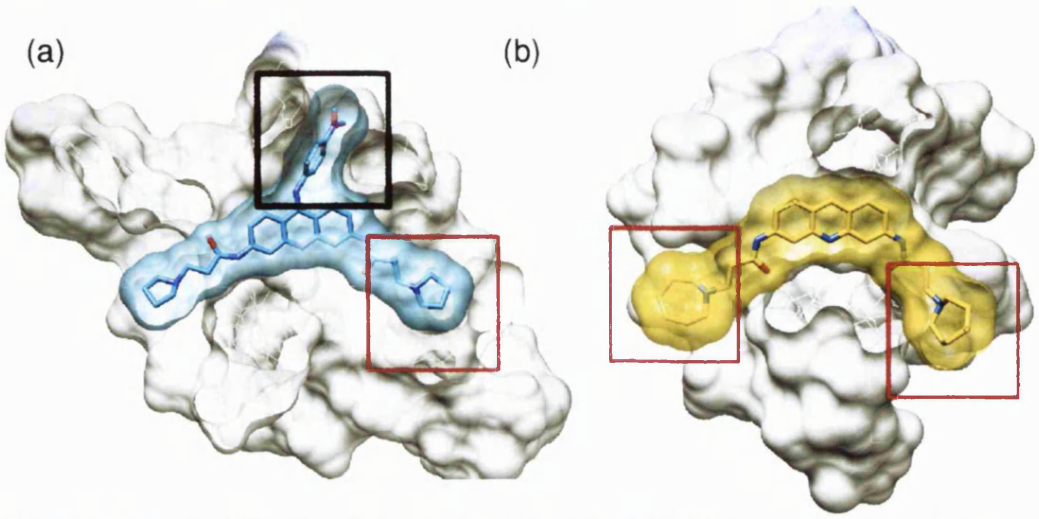


Figure 3.23: Position of the side chains. (a) The side chains in BRACO19 extend towards the loops and the DNA (pyrrolidino ring in the red box and 9-substituent in the black box). (b) The side chains in disubstituted acridine series extend towards solvent (red boxes).

NMR methods (Phan et al., 2005). It is notable that the porphyrin is bound in the loop environment where the bases originate from the propeller loops and not from the diagonal loop. Interestingly, the porphyrin substituents in this structure are exposed and do not form direct contacts with the DNA, analogous to the end groups in the structures presented here. Although loops in the human telomeric quadruplex and in the c-myc quadruplex are strictly of the propeller type i.e. maintain the parallel orientation of the DNA strands relative to each other, the actual loop structure in terms of base type and number in the *c-myc* structure is very different. The *c-myc* structure contains two single-base loops and one V-shaped 3-base loop.

We suggest that the principle of the loops constraining the shape and size of ligand substitutions should be useful in the future design of small molecules targeting individual quadruplexes, especially when the overall topology of the quadruplex and loop type is known.

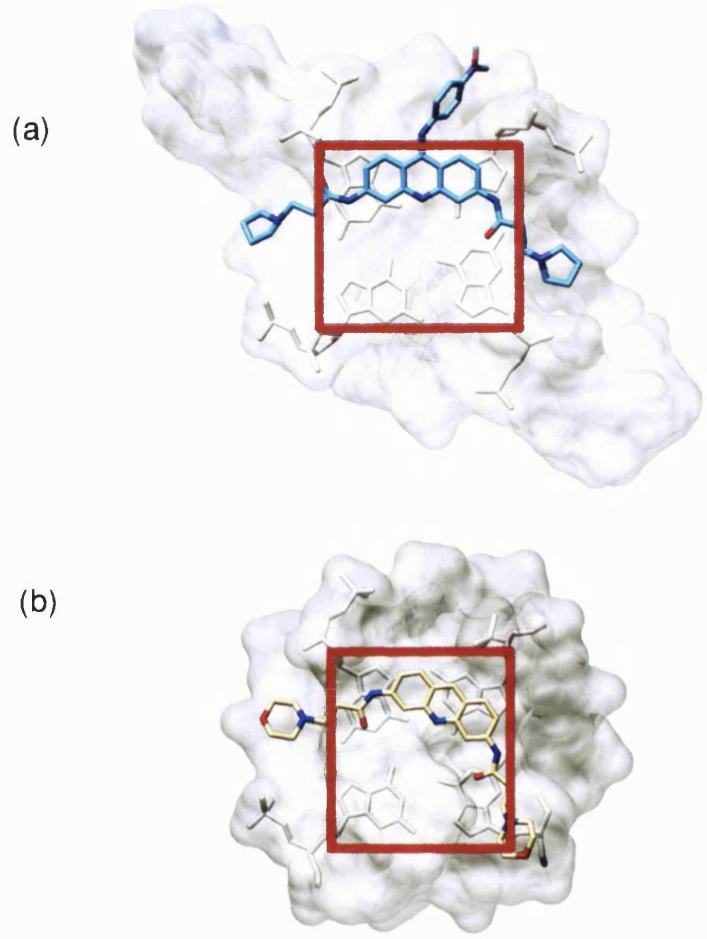


Figure 3.24: Ligand binding mode for the disubstituted and trisubstituted acridine family. The ligands belonging to the disubstituted acridine series stack onto the G-quartet (in grey sticks) with a twist of about 45º (b) relative to the position of the trisubstituted acridine BRACO19 (a). Surface representation for the DNA is shown in grey.

3.5 Conclusion

The crystal structures studied in this chapter revealed that all six ligands in the series bind to the *Oxytricha nova* bimolecular anti-parallel quadruplex in accordance with previously reported structures for the native and ligand-bound structures. The binding mode is identical for all ligands with regards to the triacridine core and the amido linkers. The size of the end hetero rings of the side chains, as well as, the size and position of the ring substituent affects the orientation of the ring in question. It also affects crystal packing interactions

resulting in one of the ligands BSU6042 with *ortho* ethyl ring substituents crystallising in the original space group $P2_12_12$ as well as in a novel space group $P2_12_12_1$. The structures also showed that the *Oxytricha nova* quadruplex, by virtue of its diagonal loop, can accommodate significant increments in the size of the hetero rings and ring substituents. This is due to the hetero rings in the side chains pointing towards an open space and towards solvent. This is in contrast to the human telomeric quadruplex where the propeller loops constrain the size of the end 3- and 6-substituents as shown by the decrease in quadruplex binding constants upon replacing the pyrrolidino moiety with a piperidino ring or with a seven-membered ring.

CHAPTER 4

COMPUTATIONAL STUDIES AND MOLECULAR MODELLING

4 COMPUTATIONAL STUDIES AND MOLECULAR MODELLING

4.1 General background

Molecular modelling involves studying and/or predicting the behaviour of molecules or molecular systems whereby a three-dimensional representation of molecules is built using a simplified description that is sufficient to mimic real behaviour.

Recent advances in computational power and in molecular mechanics methods (or empirical force field methods) have resulted in wider application of computational studies to large biological systems.

The application of molecular modelling can rationalise experimental results and provide missing pieces of information that cannot be obtained otherwise.

Computational investigations can provide a useful insight into the structure and dynamics of G-quadruplex nucleic acids, which is especially important in view of the small number of X-ray or NMR structures reported in the literature.

The detailed study of molecular modelling approaches is beyond the scope of this work and can be found in *Molecular Modelling: Principles and Applications* by Leach (Leach, 2001). However, an overview of empirical force field methods with direct impact on the system in this work is presented below. Also, a general discussion of molecular mechanics application to biological molecules and G-quadruplex structures in particular is presented.

4.1.1 Empirical force fields

By considering electrons, quantum mechanics enables the calculation of molecular attributes such as chemical reactions; for example the breaking and formation of chemical bonds. However, due to the large number of particles considered in the calculations, it becomes computationally expensive and time consuming, ultimately

resulting in limiting its application to small molecules. On the other hand, semi-empirical methods reduce the number of particles involved in the calculations by considering valence electrons, resulting in less demand for resources and thus offer an alternative to the former. Biomolecular systems are simply too large for *ab initio* computational studies over realistic and biologically-relevant time scales making it impractical and beyond the capability of most laboratories.

In contrast, molecular mechanics methods simplify matters further by ignoring electronic motion, thus calculating the energy of a system based on nuclear positions. In fact, both semi-empirical and empirical methods are possible due to the Born-Oppenheimer approximation which assumes that electronic and nuclear motion in molecules can be separated. In molecular mechanics, this simplification works in simulating biological processes making it an ideal method for studies conducted over biologically-relevant time scales. However, it trades the ability to describe atomic properties that are dependent on electronic motion for the capacity to extend these methods to study large systems.

Molecular mechanics methods are successful because studies of bimolecular processes, for example polypeptide folding or molecular complexation (such as protein-protein interactions, DNA- or protein-ligand interactions) are driven by weak non-bonded interatomic or intermolecular interactions at room temperature in the condensed phase and these are described well by force field methods (van Gunsteren et al., 2006).

Molecular mechanics methods focus on a simplified three-dimensional representation of molecules using simple potential energy terms that describe inter-atomic and intra-molecular interactions so as to reproduce molecular behaviour in biological processes adequately, thus ultimately mimicking the behaviour of molecules or molecular systems.

These terms describe the bonded (covalent) (such as bond-stretching, bond-angle bending, improper and proper dihedral angles) and non-bonded interactions between atoms in different molecules or between atoms in the same molecule but are separated by more than two or three bonds. Bonded terms include; atom types,

bond lengths, bond and rotational angles, and non-bonded terms include van der Waals (VDW) and electrostatic components.

Within the framework of classical physics, atoms are considered as balls and bonds as springs. For example, the potential energy of a bond varies with variations in the distance between the atoms it connects (as it would in a spring connecting two balls) and so, energy penalties are associated with deviation from optimal values.

Molecular mechanics force fields are divided into functions that pertain to each of the terms mentioned above. The resultant potential energy function or "force field" (the latter also used to refer to a set of parameters describing a system) hence provide a computational model of the system in question (Leach, 2001; van Gunsteren et al., 2006).

4.1.2 Force field approaches to date

An important attribute of molecular mechanics methods is transferability i.e. it should be possible to extend the applicability of a method to similar molecules and similar problems beyond its training sets with reliable results.

However, as force fields are empirically derived to reproduce different properties, the choice of a force field to study a particular problem in a particular system is essential.

Due to continuous improvements in parameterisation over the years, a set of reliable and rigorous force fields are available today that are suitable for the study of a wide range of problems.

4.1.3 Background for specific force fields

The AMBER force field

AMBER (Assisted Model Building with Energy Refinement) refers to a group of programs designed to work well together that are used to conduct and analyse molecular dynamics (MD) simulations in biomolecular systems containing proteins, nucleic acids or carbohydrates. "AMBER" is also used to describe the empirical

force fields implemented in the AMBER package (Case et al., 2005; Cornell et al., 1995).

The basic model as used in AMBER is described as follows (equation 4.1.1):

$$E_{total} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos(n\varphi - \gamma)\right] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right]$$

equation 4.1.1

The potential energy terms describes the following:

Bond and angle behaviour and energies This is described by the first two terms (equation 4.1.1) represented by a harmonic potential dependent on the interatomic distance r, and the bond angle θ, where K_r is the equilibrium bond length constant and K_θ is the equilibrium bond angle constant. r_{eq} and θ_{eq} are equilibrium bond distance and angle value and are derived from X-ray structural data. The energy of both terms increases with deviation away from equilibrium value.

Torsional energy A periodic function describes the torsional energy as a function of torsion angle φ. The incorporated parameters are often specified by the two central atoms in the dihedral angle considered. γ is the phase factor. V_n is a qualitative indicator of the relative barrier height to rotation. Improper torsions are also included to enforce planarity.

Non-bonded interactions The fourth term broadly describes the non-bonded interactions and is calculated between atom pairs (i and j) in different molecules or atoms in the same molecule that are separated by three bonds or more. This term includes two components; VDW energies described by a Lennard-Jones potential and an electrostatic component represented by a Coulomb potential. In each atom pair, R represents the separation between the two atoms and q represents the

partial charge at the centres of the atoms involved. A is the repulsive component and B is the attractive component in the Lennard-Jones potential.

The AMBER force field and parm99SB parameter set (Cornell et al., 1995; Wang et al., 2000; Pérez et al., 2007) were designed for organic and biological molecules namely proteins and nucleic acids. This means that the parameters for common nucleotides and amino acids are already implemented.

The General AMBER Force Field (GAFF)

The General AMBER Force Field (GAFF) is a generic force field specific for small molecules supported by the AMBER suite of programs. GAFF is designed to be compatible with the AMBER force fields for proteins and nucleic acids (Wang et al., 2004). Initially, the lack of parameters for small molecules in AMBER prevented its wide use to study nucleic acid- or protein-ligand interactions and binding energies. In an environment where there is a demand for a quick robust method to study and/or rank the interactions and relative binding energies of libraries of ligands to a specific biomolecular target, the need for an efficient and compatible force field for small organic molecule i.e. binding ligands, to accompany bimolecular force fields becomes obvious.

Other methods are also available for examining ligand behaviour and binding when using the AMBER suite of programmes include using the force field MMFF94 (Halgren, 1996).

The default charge scheme in GAFF parameterisation is the restrained electrostatic potential (RESP) at HP/6-31G* level which is also the charge approach applied in protein and nucleic acid force fields in AMBER (Wang et al., 2004). However, RESP is expensive computationally as it requires running *ab initio* optimisation at the HF/6-31G* for each ligand which has prevented its wider application especially where the automated analysis of diverse libraries of ligands is concerned. Alternatively, high-quality atomic charges can be generated faster and cheaply using the AM1-BCC model (Jakalian et al., 2000). Using this approach, charges are obtained that are compatible with RESP charges. In fact, this approach is designed

to carry out a semiempirical AM1 calculation producing Mulliken charges that are then corrected to produce charges that match the electrostatic potential at the HP/6-31G* level. Similar to the AMBER macromolecular force fields, this force field also neglects the effects of polarisability in the ligand, thus affecting the fundamental limit to its accuracy.

AM1-BCC charges performed in calculations of hydration free energies as well as in instances where charges were derived from electrostatic potentials calculated at various levels of theory using computationally expensive and time consuming RESP charges (Mobley et al., 2007).

MM-PBSA and MM-GBSA methods for calculation of binding free energies

The method behind the calculation of relative binding energies for ligands in complex with a receptor presented in this chapter is presented briefly below.

Relatively fast calculations of binding energies is possible using molecular mechanics based methods such as the MM-GBSA (Tsui & Case, 2000) and the MM-PBSA (Kollman et al., 2000) methods (Molecular Mechanics-Generalised Born/Poisson Boltzmann Surface Area, respectively).

In MD simulations, the aim is to permit the system to evolve over time so as to allow it - provided adequate starting structures, parameters and simulation length are used - to sample a variety of conformations and energy states over the course of the simulation. From the data obtained at time-points over the course of the simulation, a trajectory representing the atomic motions of solute and solvent molecules is acquired and can be analysed. This provides information about molecular motions and the thermodynamic properties of the system.

The resultant conformations sampled by the receptor, ligand and complex at equilibrium are used to calculate their free energies.

In these methods, the solvent molecules are replaced with a continuum electrostatic model, thus omitting the need to calculate computationally expensive pair-wise non-bonded potentials for solvent molecules.

Free energy of binding ($\Delta G_{binding}$) of ligand A and receptor B is calculated according to the following scheme (Figure 4.1):

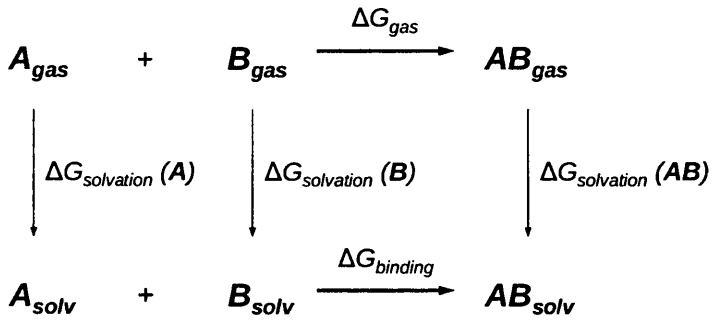


Figure 4.1: Energy cycle during calculations of binding energy of ligand A to receptor B to form the non-covalent complex AB.

The components of $\Delta G_{binding}$ are obtained by subtracting the sum of the average free energies of the reactants ($G_A + G_B$) from the average free energy of the complex (G_{AB}) according to the following equation (equation 4.1.2):

$$\Delta G_{binding} = G_{AB} - (G_A + G_B)$$

equation 4.1.2

Free energies can be broken down further into molecular mechanical energy (E_{MM}), the electrostatic and non-polar contributions to solvation free energy (G_{solv}), and the entropic (TS) energies of the system at temperature T. This is shown in the following scheme (equation 4.1.3):

$$G_{binding} = E_{MM} + G_{solv} - TS$$

equation 4.1.3

The molecular energies for the gas phase are calculated using the molecular mechanics force field.

The solvation free energy includes the electrostatic and non-polar solvation

energies as per the following scheme (equation 4.1.4):

$$G_{solvation} = G_{polar} + G_{non-polar}$$

equation 4.1.4

Both the Poisson-Boltzmann (PB) and Generalised Born (GB) methods calculate the electrostatic contribution from a set of charges placed in a defined volume of low dielectric constant surrounded by a dielectric continuum representing the solvent. However, they differ in the way they calculate G_{polar} in the above equation (equation 4.1.4). The PB method uses numerical solutions to the Poisson Boltzmann equation while the GB method uses the Generalised Born method to calculate the electrostatic contributions to solvation free energies.

Limitations for force field methods

Each force field is developed with a particular system in mind whilst aiming to create a balance in its ability to reflect atomic properties as well as intermolecular interactions; such as solute-ligand interactions.

Up until recently and due to limitations in computer power and time, force fields assigned static averaged properties to atoms. For example, fixed partial charges were assigned to the centres of atoms at the offset and were not revised as the study progresses to reflect the ensuing changes in the local environment. In real terms, the force field is limited in its ability to cope with modelling essential intermolecular electrostatic interactions; such as solute-ligand interactions.

However, current efforts are developing polarisable force fields which should be able to account for effects such as induced dipole (resulting when a strongly charged molecule approaches a neutral one) and to reflect variations in charge distribution brought on by the dielectric constant of the environment that the molecule is placed in. AMOEBA (Atomic Multipole Optimised Energetics for Biomolecular Applications) (Grossfield et al., 2003) and PFF (Polarisable Force Field) (Maple et al., 2005) are examples of recently developed polarisable force

fields. Furthermore, polarisable continuum electrostatics models such as PMPB (Polarisable Multipole Poisson-Boltzmann) are starting to emerge (Schnieders et al., 2007). The performance of polarisable force fields in reproducing biological processes, such as folding equilibria, has yet to be reported (van Gunsteren et al., 2006).

Furthermore, it is important to understand that computational methods such as dynamcs simulations deal with a single molecule with a specific starting geometry and its development at the picosecond (ps) time scale over the run time of the simulation (normally below 1 μs). Both the starting geometry and run time affect the ability to overcome free energy barriers. This is in contrast to experiments which deal with assemblies of molecules and are mostly at equilibrium.

MD simulations are closely related to X-ray crystal structures since the latter provides the best starting structures and can be used for validation of the methods applied. MD methods may not be able to predict a crystal structure but if the crystal structure is incorrect, then MD simulations are expected to be able to detect it.

Simulations of quadruplex nucleic acid structures

DNA and RNA structures are highly charged molecules which presents a challenge to MD simulations due to the effect this has on electrostatic properties and solvation.

However, the development of methods to treat long range electrostatics combined with increasing computer power, as well as the development of suitable force fields has resulted in a number of studies since the mid 1990s (Cheatham & Young, 2000; Pérez et al., 2007; Sponer & Spacková, 2007).

The availability of high quality X-ray crystallographic structural data for duplex nucleic acids has also enabled the development of MD force field methods.

MD methods were shown to reproduce the behaviour of duplex DNA in solution which showed that force field approaches can be used for this type of system. Significantly, MD methods demonstrated the transition of A-DNA to B-DNA in explicit solvent in conditions that favour the formation of the latter (Cheatham &

Kollman, 1996).

These methods turned out to work really well for other forms of DNA structures such as G-quadruplex structures, which were not considered in the original design. Thus, MD simulation studies enable one to study and understand the properties and ligand binding energies of four-stranded nucleic acids, which is extremely valuable, especially in view of the limited number of high quality structures in the literature (discussed in chapter 1).

MD simulations reported in the literature highlight the rigidity of the G-quartet stem, the flexibility of the loops and capture the overall stabilising effect of cations in the central channel of quadruplex structures (Sponer & Spacková, 2007).

Also, simulations were shown to be stable for the anti-parallel quadruplex structure formed by the *Oxytricha nova* telomeric sequence d($G_4T_4G_4$) on the nanosecond scale (Spačkova et al., 1999). However, the same study showed that the theoretical parallel quadruplex structure for the same sequence contained bifurcated hydrogen bonds for the two central G-quartets, which is not in agreement with the experimental data from the crystal structure.

The behaviour of the loops was also extensively studied using MD simulations. Hazel and co-workers showed that G-quartet stability as well as loop length and composition influence folding of G-quadruplex structures. Calculated free energies for multiple conformations were similar suggesting they are likely to coexist in solution (Hazel et al., 2004) even though a specific topology may be observed in the crystal structure (Hazel et al., 2006). A study by Fadrna and co-workers did not succeed in predicting the diagonal loop arrangement observed in the crystal structure (Fadrná et al., 2004).

The differences between crystal structure topologies and those produced by molecular dynamics simulation do not mean that the simulations are wrong. On the contrary, it suggests that more than one energetically favourable topology is possible and highlights the possibility that experimental factors may contribute to experimental observation.

More recent refinement of the AMBER force field for nucleic acids resulted in

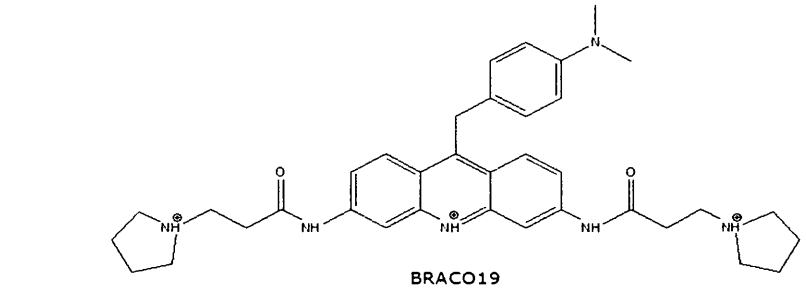
parmbsc0 (Pérez et al., 2007) which was shown by Perez and co-workers to produce stable trajectories and sample configurations close to the experiment for parallel and anti-parallel quadruplexes on a 10 - 200 nanosecond (ns) time scale. This force field requires further validation but is hoped to improve quadruplex, and nucleic acid simulations in general, and to enable, with growing computer power, state-of-the-art molecular dynamics simulations at the 1 microsecond (μs) time scale.

4.2 Scope of this work

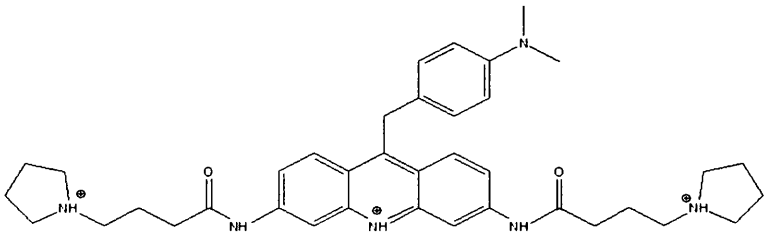
The aim of this work is the computational analysis of ligand-quadruplex interactions using molecular dynamics simulations and binding affinity calculations.

Structural analyses and calculations of binding affinities were conducted on a subset of four closely related trisubstituted acridine derivative compounds (Figure 4.2) in complex with the parallel human telomeric quadruplex. The existing crystal structure for the complex with BRACO19 (PDB ID 3CE5 (Campbell et al., 2008)) was used as the starting point in the first instance and then the ligand was replaced sequentially with compounds from the derivatives set.

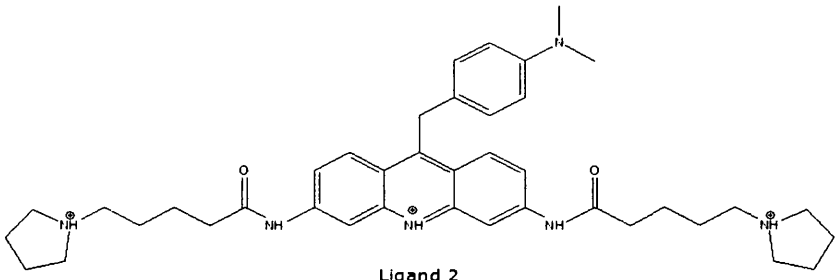
The four ligands cluster in two groups of affinities for the human telomeric quadruplex as calculated by experimental FRET data. BRACO19 and compound 1 change the melting temperature by a factor of approximately 4 relative to the same concentration of compounds 2 and 3. The change in melting temperature of the human telomeric quadruplex caused by 1 μM concentration (Moore et al., 2006) of each ligand is shown (Table 4.1). The binding affinities calculated using MM-PDBSA or MM-GBSA approach may then be compared with experimental FRET data for the compounds.



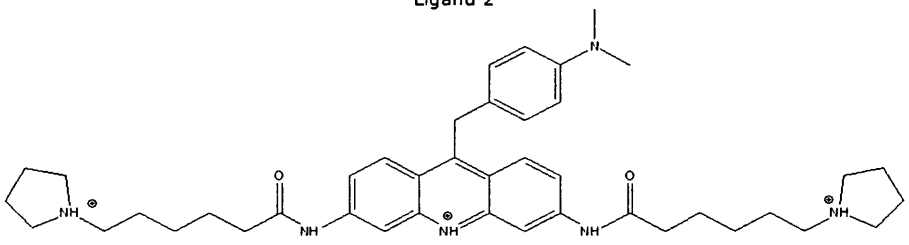
BRACO19



Ligand 1



Ligand 2



Ligand 3

Figure 4.2: Chemical structures of the ligands in this study. The ligands are derivatives of BRACO19 and differ by the length of the alkyl chain at the 3- and 6- position substituents.

The X-ray crystal structure used as the starting point in this work provides the first instance of 5' to 3' quadruplex stacking. However, the two quadruplexes are independent and not connected via a loop i.e. do not represent tandem repeats of quadruplexes. Though, with careful consideration it is possible to connect the two

quadruplexes via a linking DNA sequence formed of loop bases TTA. Indeed, molecular modelling has shown that a stable structure can be readily constructed for drug binding at the interface of tandem repeats of telomeric G-quadruplexes retaining most of the features from the crystal structure (Haider & Neidle, 2009).

This project aims to compare between closely related ligands in an environment where all the features from the crystal structure are retained. Thus, the DNA quadruplex part of the crystal structure remained intact and solely the ligand was sequentially replaced with the three different derivatives.

Table 4.1: The change in melting temperature of the human telomeric quadruplex caused by 1 μM ligand concentration (Moore et al., 2006).

Ligand	ΔT_m(1 μM) , (°C)
BRACO19	27.5
1	16.5
2	6.8
3	4.7

4.3 Materials and methods

4.3.1 Quadruplex preparation

The biological unit of the X-ray crystal structure of the parallel human telomeric bimolecular quadruplex in complex with the trisubstituted acridine ligand BRACO19 (PDB ID 3CE5 (Campbell et al., 2008)) was used as the starting point for the modelling study. The biological unit contains two quadruplexes stacked 5' to 3' sandwiching one ligand molecule (as part of a coplane containing a thymine nucleotide) at the interface (see Chapter 2). Coordinates were obtained from the Protein Data Bank. Preparation and manipulation of the structure and initial building of the ligand-DNA complexes prior to the molecular dynamics simulations was carried out using the SYBYL 7.3 molecular modelling suite (Tripos, Inc., St. Louis, USA) and the Chimera software version 1.3 (Pettersen et al., 2004). The ligand sandwiched at the binding site between two quadruplexes was left intact. The flanking nucleotides present at the 5' and 3' G-quartet faces which do not form part

of the binding side were omitted. Also, the second ligand positioned at the exposed 3' G-quartet and waters were omitted. The final structure contained one ligand bound at the interface of two 5' to 3' stacked quadruplexes and four potassium ions; two in the central channel of each quadruplex.

4.3.2 Ligand preparation

Ligands (Figure 4.2) were constructed in SYBYL, with charges and atom types applied using the MMFF94 method (Halgren, 1996; Halgren, 1996). Stepwise minimisations using the MMFF94s force field (Halgren, 1996) were subsequently carried out to a convergence of 0.02 kcal.mol^{-1}/Å over 5000 steps using the Powell method (Powell, 1977) with a cutoff of 11 Å.

4.3.3 Ligand-quadruplex complex preparation

Following output of the ligands from the SYBYL suite, the complex containing the BRACO19 ligand was ready for the next stage. For the other three ligands, quadruplex and corresponding ligand were superimposed in the Chimera software (Pettersen et al., 2004) in such a way that the acridine core, its 9-substituent (anilino group) and 3- and 6-side chains were superimposed. Each ligand positioned in this way was merged with the quadruplex forming the complex (DNA quadruplex plus ligand) and the output coordinates for each ligand relative to the quadruplex were identical. An output of the ligand-quadruplex complex coordinates for each compound-DNA pair would thus show the quadruplex in an identical spatial orientation.

Transfer from Chimera and complex preparation for MD

Preparation of the DNA-ligand complexes were performed using the XLEAP, ANTECHAMBER and SANDER modules of the AMBER9 package (Case et al., 2005; Cheatham & Young, 2000).

The individual complexes were initially exported from the superimposed Chimera database as files in pdb format. These pdb files were subsequently split manually

into new receptor and ligand pdb files, with residues renumbering and further manipulation of the data to prepare them for input into XLEAP. DNA and ligands were prepared and adjusted separately in this manner, with incorporation of the DNA into XLEAP accomplished with the necessary DNA residue templates while the ligands were individually defined as novel residues using the ANTECHAMBER and PARMCHK programs. ANTECHAMBER was used to calculate atom types and atomic point charges for ligands using the GAFF (General AMBER Force Field) (Wang et al., 2004) AM1-BCC charge model. Each ligand was assigned a total of three positive charges. Following merging of the receptor-ligand pdb files and the reading into XLEAP of the prep file defining the ligand parameters, the complex was imported into XLEAP and checked thoroughly for consistent connectivities, atom types and charges.

4.3.4 MD simulation parameters

All calculations were carried out in AMBER9 using the ff99 force field (Cheatham et al., 1999) and the parmbsc0 parameter set (Pérez et al., 2007). The TIP3P potential was used for waters (Jorgensen et al., 1983). GAFF atom types were used for the ligand (Wang et al., 2004). Ligand charges were derived using the AM1-BCC charge method (Jakalian et al., 2002) and defined in the ligand preparatory files as part of the new residue input.

Periodic boundary conditions were applied, with the particle-mesh Ewald (PME) method (Darden et al., 1993) used to treat long-range electrostatic interactions. The solute was first solvated in a waterbox, the dimensions of which extended to a distance at least 12 Å from any solute atom (total VDW box size approximately 63 x 72 x 64 $Å^3$). Potassium counterions were subsequently added using XLEAP to attain overall system neutrality, with standard potassium parameters as applied in the Cornell et al. Force field (VDW radius 2.658 Å). This process generally resulted in the addition of 7000 residues (ions and waters), depending on the exact box dimensions.

The minimisation procedure for the solute consisted of a two stage approach. In the first stage the solute was kept fixed and only the positions of the water and ions

were minimised. In the second stage, the entire system was minimised.

Positional restraints were used on each of the solute's atoms (DNA, ligand and central channel potassium ions) to keep them essentially fixed in the same position via using a force constant of 500 kcal.mol^{-1}/Å. This means that the water and counterions were free to move. Such restraints work by specifying a reference structure, in this case, the starting structure, and then restraining the selected atoms to conform to this structure. The solute was subjected to 5000 steps of minimisation with an initial 3000 steps of steepest descent minimisation followed by 2000 steps of conjugate gradient minimisation. Constant volume periodic boundaries and PME were used and a cutoff of 11 Å was employed.

The second stage of minimisation consisted of minimising the whole system. In this case, 5000 steps of minimisation were run without any restraints with an initial 3000 steps of steepest descent minimisation followed by 2000 steps of conjugate gradient minimisation.

The next step in the equilibration protocol involved allowing the system to heat up from 0 K to 300 K. In order to ensure this happened without any wild fluctuations in the solute, weak restraints of 10 kcal.mol^{-1}/Å were used on the solute similarlty to the first stage of the minimisation. The Langevin temperature equilibration scheme was used in maintaining and equalising the system temperature. Twenty picoseconds (ps) of MD with a timestep of 2 femtoseconds (fs) were run at constant volume with the SHAKE algorithm enabled for hydrogen atoms.

The final geometry was the the starting point for the full 2 nanoseconds (ns) production run, with energy information and averages printed out every 100 steps (0.2 ps intervals), coordinates captured every 1000 steps (2 ps intervals). The restraints were switched off and constant pressure was applied in the production run of 2 ns of at 300 K.

Preparation, minimisation and equilibrated MD were carried out for each complex in the series.

4.3.5 MD analyses

Trajectories were examined visually using the VMD software package (Humphrey et al., 1996). This allowed mapping of any large conformational changes to distinct time periods in the trajectory and the comparison of energy and RMSD values. Energies and other MD output values were extracted from the MD output files using a short PERL script.

Processing of the trajectories was performed using the PTRAJ tool in AMBER. Trajectories were first imaged to move coordinates from outside the periodic box back inside, and the centre of geometry of residues 1-47 (quadruplex, central potassium ions in the channel and the ligand) placed at the origin. Finally, the trajectories were stripped of all residues apart from the DNA (residues 1 to 42), the internal potassium ions (residues 43 to 46) and the ligand (residue 47). The stripped trajectories could then be used for structural analyses and relative binding affinity calculations.

4.3.6 Structural analyses

Preliminary analysis of the MD trajectories was carried out by examining the variations in the total energy (E_{TOT}) of the system (E_{TOT}=potential energy+kinetic energy) and by studying the structural and conformational changes in the structure over the course of the simulation.

System properties (such as E_{TOT}) were extracted from the data written to the molecular dynamics output files. Conformational and structural changes can be mapped out to specific time points in the simulation by calculating the RMSD values over the course of the production run. RMSD values were calculated with respect to the structure of the first frame after equilibration.

RMSD calculations provide a means of identifying major and sudden changes in the conformation of a system. RMSD analyses included calculating the RMSD for the heavy atoms of the structure as a whole, the G-quartets, the loop regions.

RMSD values were also calculated for the ligand as a whole, and separately, for the

acridine core, for each of the 3-, 6- and 9- substituents.

4.4 Results

4.4.1 Equilibration

An estimate of the equilibration status of the system can be assessed from the output data files of the MD run. The calculated total energy (E_{TOT}) includes the receptor, ligand and solvent molecules, reflecting the stability of the system over the course of the MD production run. In all the systems in this study, the variations in the total energy of the system and RMSD values over time indicated that an equilibrium state had been reached (Figure 4.3 and Figure 4.5). Equilibrium values were rapidly reached at around 500 ps.

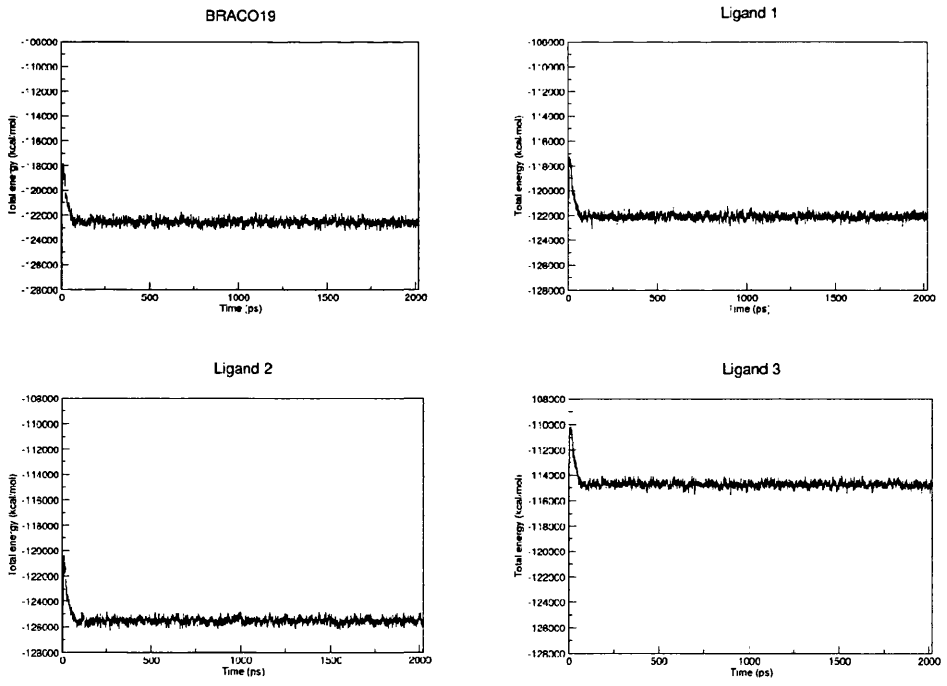


Figure 4.3: Total energy (E_{TOT}) for all four complexes over the course of the dynamics simulation.

4.4.2 Structural Analyses: visual and RMSD analysis

Visualisation of the trajectories and trace of RMSD values for the MD simulations illustrate stable trajectories in all four ligand complex systems. RMSD values do not represent the energetics of a system. However, they provide a means for identifying major changes in the conformation of the components of the system over the course of the MD run.

RMSD values were calculated for the heavy atoms belonging to the ligand as a whole, the acridine core and the atoms in each of the 3-, 6- and 9- substituents (Figure 4.4)

Simulations of all four compounds resulted in stable binding at the interface between the two quadruplexes. Visualisation of the trajectories showed that the stacking arrangement of the two quadruplexes remained intact with the ligand sandwiched in between, where the ligand remained in close contact with the 3' G-quartet and 5' TATA quartet surfaces. Little interaction was observed between quadruplex loops and the ligands.

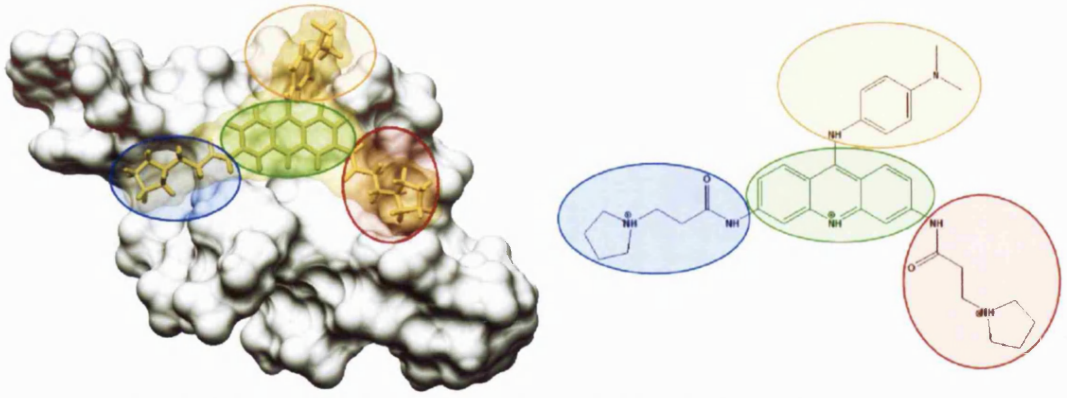


Figure 4.4: Defined Ligand components as used in the calculations of RMSD values. These components are shown as oblong coloured shapes superimposed onto the chemical structure of the ligand (right) and onto the bound ligand (left) (ligand is viewed laid onto the 3' face of the quadruplex). Green = acridine core, red = 3-substituent, blue = 6-substituent and orange = 9-substituent.

Trace of RMSD values for each complex as a whole with respect to the starting structure is shown over the course of the simulation run (Figure 4.5) and starting and finishing positions for each ligand are shown (Figure 4.6).

In general, all of the complexes were stable over the course of the simulation run and showed little variation from their respective starting positions.

The starting orientation of the carbonyl oxygen and the amide nitrogen in all the complexes was maintained throughout the course of the molecular dynamics simulations i.e. identical to the starting structure before the dynamics run and to the crystal structure (Figure 4.6).

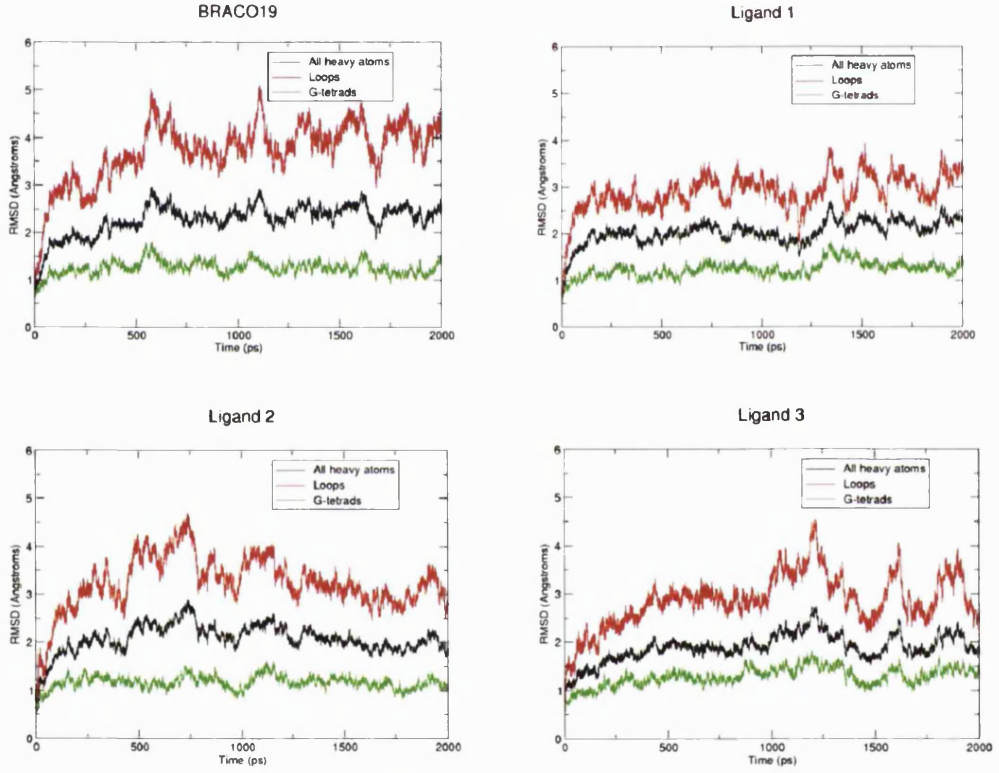


Figure 4.5: Trace of RMSD values for heavy atoms in each complex as a whole (black), for loops (red) and for G-quartets (green).

Over the course of the 2 ns simulation, the 9-substituent in all ligands remained in place relative to the starting position. It occupied a tight pocket formed by a thymine base and the DNA backbone (discussed in detail in Chapter 2).

The acridine core in BRACO19, ligand 1 and ligand 3 remained in place relative to the starting position indicating that no significant movement occurred over the course of the simulation. On the other hand, the acridine core in ligand 2 showed a

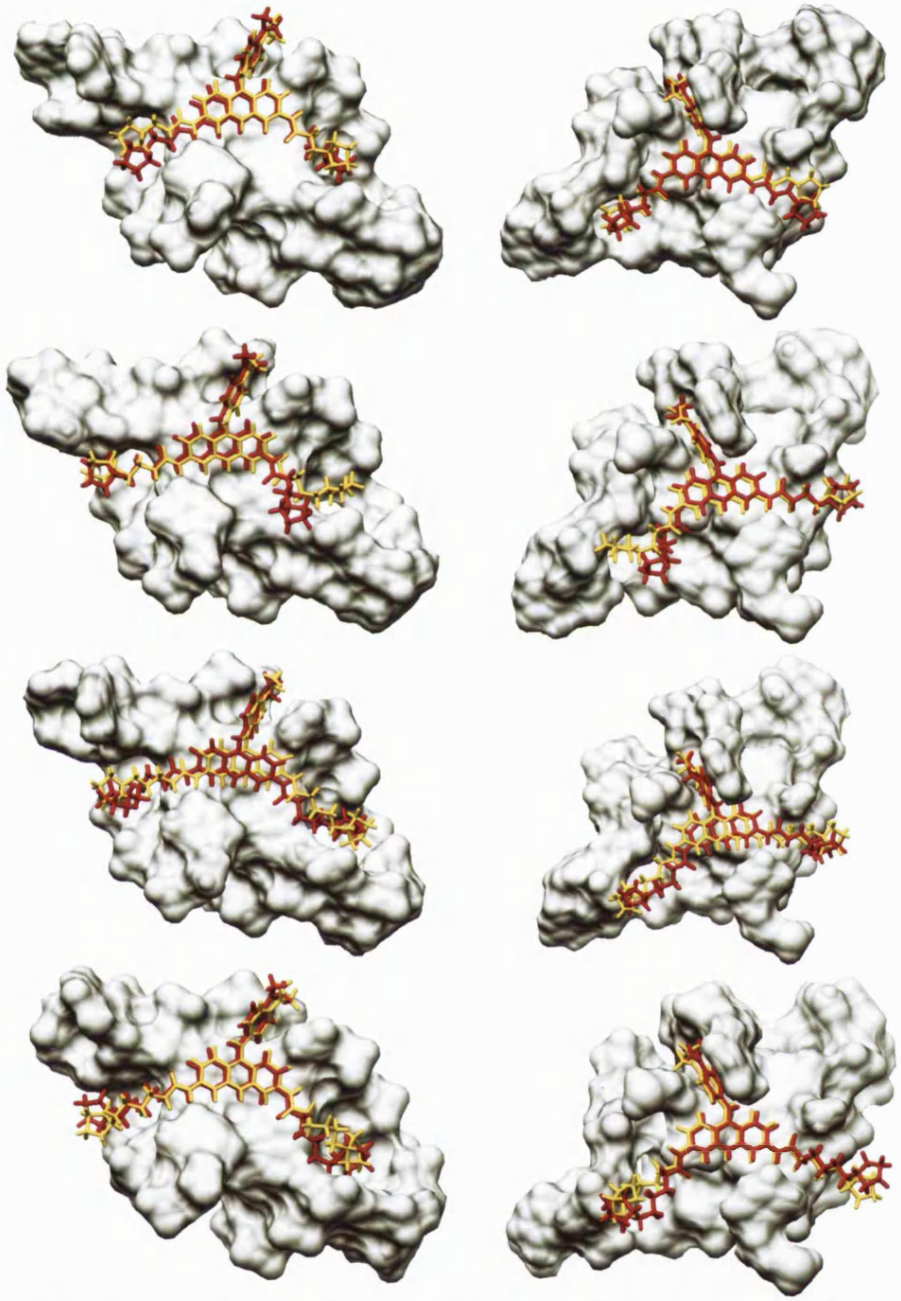


Figure 4.6: Starting and finishing positions for all four ligands. Starting positions are shown in yellow and finishing positions are shown in red. The starting position is after minimisation but before molecular dynamics simulations have started. The molecular surface (shown in grey) is generated from the starting quadruplex structure. Top row = BRACO19, second row = ligand 1, third row = ligand 2 and fourth row = ligand 3. Column on the left = ligand is shown overlaid onto the 3' face of the quadruplex. Column on the right = ligand is shown overlaid onto the 5' face (5' TATA quartet) of the other quadruplex.

small shift in its position along the G-quartet surface by about 1 Å. This motion

occurred as the side chains formed interactions with the groove regions. However, this shift remained very small as the 9-substituent anchored the ligand in place, thus resisting any large shifts in ligand position. Similarly, the 3' and 5' quadruplex faces restrict any rolling of the acridine core that may be caused, in the absence of such restrictions, by strong interactions between the side chains and the loops or grooves. In BRACO19, ligand 2 and ligand 3, the 3- and 6-substituents all point towards the grooves in the quadruplex at the 3' face over the course of the simulation. In contrast, the 3-substituent in ligand 1 point towards the groove in the quadruplex at the 5' face (Figure 4.6). In all four complexes, equilibration was reached early in the simulation (about 500 ps) as shown by the RMSD value (for the whole complex) remaining between 1.5 and 2.5 Å over the course of the simulation (Figure 4.5).

As for the G-quartets, the calculated RMSD values which were maintained over the course of the simulation (after equilibration) at approximately 1 Å, demonstrated their characteristic stability. In contrast, those for the loops - ranging between 3 and 4 Å - demonstrated their flexibility. This is in accord with previous studies (Moore et al., 2006; Hazel et al., 2004).

The behaviour of each ligand during the course of the simulations is briefly outlined below.

BRACO19

On the whole, BRACO19 was the most stable ligand in the binding site. BRACO19 showed the least amount of conformational change as demonstrated by maintaining low RMSD values over the course of the simulation with very limited fluctuation remaining between 0.5 and 1 Å.

The RMSD values calculated over the course of the simulation (Figure 4.7) for the acridine core and 9- substituents also remained very low, between 0.1 and 0.3 Å.

The 6-substituent was also stable with RMSD values between 0.1 and 0.5 Å.

The 3-substituent showed a slightly higher variation with RMSD values ranging between 0.5 to 1 Å. This was due to the rotation of the end hetero ring whilst mostly

maintaining the conformation of the side chain itself.

The starting conformation for the alkyl chains, the carbonyl oxygens and the amide nitrogens was maintained throughout the simulation relative to the starting structure.

Ligand 1

Overall, ligand 1 showed the most movement, having RMSD values ranging between 1.5 and 2.5 Å.

Similar to the behaviour of BRACO19, the RMSD values calculated over the course of the simulation (Figure 4.7) for the acridine core and the 9- substituent remained very low between 0.1 and 0.3 Å.

Furthermore, the 6-substituent showed more mobility than BRACO19 but still maintained low RMSD values between 0.6 and 1.0 Å.

However, the 3-substituent showed a higher variation, with RMSD values ranging between 1.0 to 1.5 Å. This is due to a change in the conformation of the alkyl chain as it probed for interactions with the quadruplex. More specifically, as the positive charge on the nitrogen contained in the hetero ring at the end of alkyl chain probed for interactions with the quadruplex groove and negatively charged DNA backbone. This interaction was achieved very close to the start of the simulation (at approximately 250 ps) and was maintained throughout the length of the molecular dynamics run.

In contrast to all other complexes the 6-substituent in ligand 1 probed the loop and groove in the quadruplex belonging to the 5' face. This was achieved very early on in the simulation (Figure 4.7) and maintained throughout the course of the MD run where the RMSD value was maintained at approximately 1 Å.

Ligand 2

Ligand 2 maintained RMSD values between 1.0 and 1.5 Å over the course of the MD run (Figure 4.7).

Here, the calculated RMSD values (Figure 4.7) for the acridine core and 9-

substituents also remained low between 0.1 and 0.5 Å, but were higher than those for BRACO19, ligand 1 and ligand 3. This is due to a shift in the position of the acridine core along the G-quartet surface as the positively charged end hetero rings - enabled by the longer side chains - formed strong interactions with the negatively charged backbone in the loops and grooves.

The 6-substituent showed the highest variation in its conformation away from the starting structure over the course of the simulation with an RMSD value between 1.0 and 1.5 Å over the course of the simulation. This was due to a rotation in the end ring accompanied by a conformational change in the long alkyl chain.

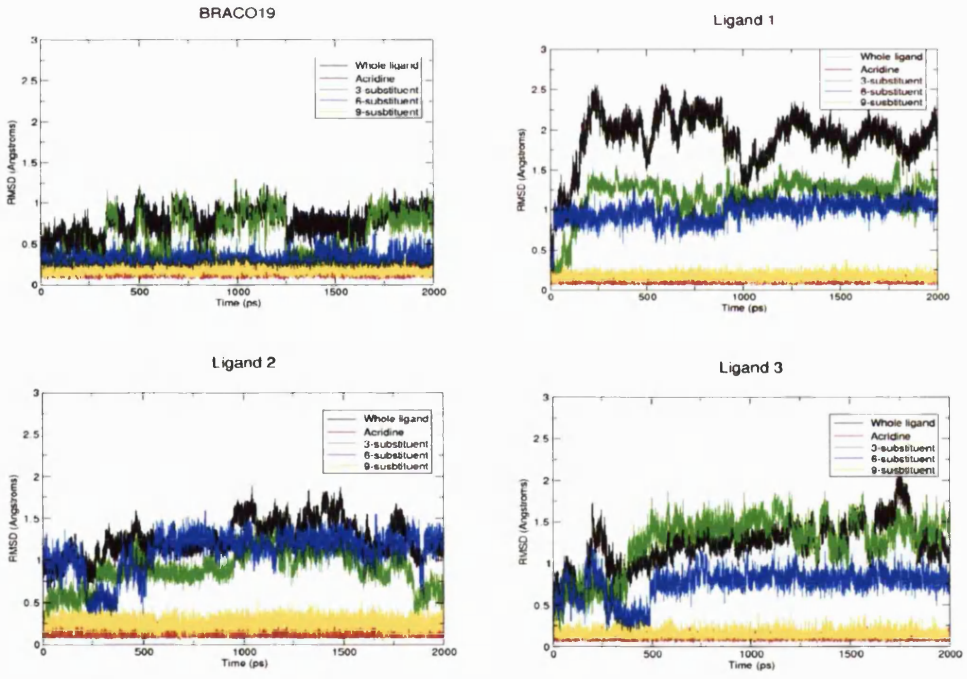


Figure 4.7: RMSD values for the ligand and its components (see text).

The 3-substituent showed a greater variation in RMSD values than BRACO19 but lower than ligand 1 and ligand 3 and was maintained between 0.7 and 1.3 Å. As the simulation progressed and the loops relaxed into an open position, the backbone of the loop belonging to the 5' quadruplex moved further away from the pyrrolidine ring, whereas the backbone of the loop belonging to the 3' quadruplex remained at a closer distance and thus physically available for interactions with the protonated

nitrogen. A change in conformation of the alkyl part of the side chain, and consequently a rotation in the orientation of the hydrogen in the protonated nitrogen by about 45° enabled successful interactions with the O2 of the DNA backbone.

Ligand 3

The calculated RMSD value for ligand 3 was maintained over the course of the simulation between 1.0 and 2.0 Å (Figure 4.7).

Here, the RMSD values calculated over the course of the simulation (Figure 4.7) for the acridine core and 9- substituents also remained low, between 0.1 and 0.3 Å, analogous to the behaviour of BRACO19, ligand 1 and ligand 2. No shift in the position of the acridine core was observed.

The 6-substituent showed variation in its conformation over the course of the simulation that was more significant than in BRACO19 but lower than ligand 1 and ligand 2, maintaining RMSD values between 0.5 and 1.0 Å. This was due to the long side chain extending deep into the quadruplex groove probing for interactions with the negatively charged DNA with its positively charged nitrogen atom in the end hetero ring.

The 3-substituent deviated from the starting structure and the calculated RMSD values were maintained at around 1.5 Å throughout the simulation. Unlike the relatively wide and open pocket available for the 6-substituent, the pocket containing the 3-substituent is much narrower and more shallow. This caused a strain on the longest side chain in this series to fit as it was constrained by the surrounding DNA atoms.

4.4.3 Calculation of binding free energy

Mean free binding energies for the four compounds studied here are presented including a breakdown of the calculated values (Table 4.2). The PB and GB methods for calculating the free energy of binding produced similar results for the four compounds in this work and are within the margin of standard error in the calculations of each other. The reason behind this is apparent from visual

inspection of the trajectories and comparison of the starting and final structures (Figure 4.6). The difference between these two positions is very small for all four complexes. This can be attributed to a combination of factors which led to a significant restriction to any deviation from starting structure, such as the 9-substituent which acts as an anchor thus holding the acridine core firmly in place and consequently the whole ligand. Also the presence of a G-quartet and a TATA-quartet on either side of the ligand restricts any rotational movement of the acridine core and any rotation of the amide bonds. This leaves the alkyl side chains. They have relative freedom to adapt their conformation in order to accommodate the probing of the groove regions by the positive charge in the end hetero rings.

The results of the free energy of binding calculated here are in contrast to the experimental results (Table 4.1) where BRACO19 and ligand 1 have significantly more favourable binding; BRACO19 produces the largest change in melting temperature with the human telomeric quadruplex, whereas Ligand 2 and ligand 3 give much less favourable stabilisation. The ligands are very closely related to each other from a structural point of view where only the number of carbon atoms in the 3- and 6- alkyl side chains differs; there are 2 in BRACO19, 3 in ligand 1, 4 in ligand 2 and 5 carbon atoms in ligand 3. It follows that they also have an equal net charge of +3 where the charges are at exactly the same positions in all ligands (one positive charge on the nitrogen atom in each of the end hetero rings in the 3- and 6-substituents and one positive charge on the nitrogen atom in the acridine core). This would account for the very similar electrostatic component of the calculated free energy of binding (Table 4.2).

It is notable that all the calculated values for binding free energy are negative (i.e. binding is favourable) which is in accordance with experimental binding data (Moore et al., 2006).

Table 4.2: Breakdown of free energy binding values (kcal.mol^{-1}) for all four complexes used in this study.

			BRACO19		LIGAND 1		LIGAND 2		LIGAND 3	
				σ		σ		σ		σ
ΔE_{ELEC}			-2196.67	15.34	-2224.92	9.98	-2239.54	14.49	-2186.70	11.49
ΔE_{MM}			-2291.97	15.34	-2319.90	10.13	-2335.59	14.99	-2293.75	12.34
	$\Delta G_{NONPOLAR}$	(PB)	-9.57	0.20	-9.44	0.11	-9.83	0.16	-11.35	0.10
	$\Delta G_{SOLVATION}$	(PB)	2207.17	16.50	2223.71	10.21	2232.03	13.64	2200.38	11.78
	$\Delta G_{MM+SOLVATION}$	(PB)	-84.80	4.70	-96.18	3.66	-103.56	4.44	-93.37	6.13
	$\Delta G_{NONPOLAR}$	(GB)	-9.57	0.20	-9.44	0.11	-9.83	0.16	-11.35	0.10
	$\Delta G_{SOLVATION}$	(GB)	2199.49	15.28	2222.56	9.77	2235.76	13.34	2198.49	12.61
	$\Delta G_{MM+SOLVATION}$	(GB)	-92.48	3.18	-97.33	3.54	-99.83	4.01	-95.26	3.29
$T\Delta S$			-35.52	6.85	-34.32	6.14	-36.59	7.72	-32.00	7.96
$\Delta G_{BINDING}$		(PB)	-49.28	8.30	-61.86	7.15	-66.97	8.91	-61.37	10.05
$\Delta G_{BINDING}$		(GB)	-56.96	7.55	-63.01	7.09	-63.24	8.70	-63.26	8.61

4.5 Discussion

Molecular dynamics simulations were conducted on four G-quadruplex ligand complexes; with BRACO19, ligand 1, ligand 2 and ligand 3 (Figure 4.2). The complexes were stable and reached equilibration rapidly (at approximately 500 ps) as demonstrated by the values of calculated RMSD ranging between 1.5 and 2.5 Å over the course of the simulation, and also the energetics of the system. The RMSD values for the G-quartets remained low throughout the course of the molecular dynamics simulation at approximately 1 Å, in accordance with previous reports on long-term quadruplex simulation (Spačkova et al., 1999; Spacková et al.,

2001; Chowdhury & Bansal, 2001). In contrast, the TTA loops showed a much larger variation reflecting their flexibility with RMSD values ranging between 3 to 4 Å, also in accord with previous MD studies (Moore et al., 2006; Hazel et al., 2004; Spačkova et al., 1999; Fadrná et al., 2004).

For all complexes studied here, the deviation from starting conformation was mainly limited to the 3- and 6- side chains. Notably, the bound ligand BRACO19 showed significant stability, over and above the other three ligands, demonstrated by maintaining the lowest range of RMSD values (0.5 to 1 Å for the ligand as a whole). Furthermore, in contrast to the other ligands, BRACO19 also demonstrated stability for its 3- and 6- side chains (RMSD values ranging between 0.1 to 1 Å). The reason for this becomes apparent upon visualisation of the binding site in the MD simulation and in view of the determined crystal structure for BRACO19 in a complex with the human telomeric quadruplex (Campbell et al., 2008). BRACO19, as in the other three ligands, fits tightly into a binding site where steric factors - a consequence of the G-quartet and the TATA-quartet above and below the ligand - restrict its movement. The combined restrictive effect of the planes of the quartets surrounding the acridine core and amide bonds, and the 9-substituent fitting into a tight pocket (discussed in detail in Chapter 2), effectively anchors the core of the ligand in place.

However, the 3- and 6- side chains point away from the quartet planes and towards pockets formed by the groove and loop regions. The ligands were originally designed with a positive charge at the end of long flexible side chains at the 3- and 6- positions so as to probe the pockets formed by the groove and loop regions for favourable electrostatic interactions with the negatively charged DNA backbone. It was hypothesized that lengthening the flexible alkyl component of the side chains would allow the positive charge in the pyrrolidine ring (or a positively charged group of similar dimensions) to probe deeper into these pockets.

Indeed, in 2002, Parkinson et al determined the crystal structure of the native human telomeric quadruplex sequence revealing an extensive network of waters in the groove and loop regions. This structure analysis supported the ongoing effort in designing tailored ligands, which in principle, are able to mimic the hydrogen

bonding pattern in the grooves and loops - as guided by the observed water networks - and thus have superior affinities. The crystal structure provided an excellent scaffold on which to base future structure-based ligand design, however, it did not explain some of the current experimental observations, for example, trends in binding affinities of compounds belonging to the extensively studied acridine family.

The crystal structure of the complex formed with BRACO19 (Campbell et al., 2008) provided a rational for several of the observed experimental observations and is further supported by the simulation work reported here. The crystal structure revealed that BRACO19 demonstrated optimum placement with regards to the positive charge in the end pyrrolidine ring so as to allow the nitrogen atom in the pyrrolidine moieties to contact the existing water network i.e. without disruption. It showed that lengthening the alkyl part of the side chains would cause the hydrophobic alkyl moiety to replace the pyrrolidine positive charge, thus destabilising the existing water network and interfering with its structural stability.

The molecular dynamics study shows that lengthening the side chains at the 3- and 6- position reduces the stability of the ligands as the increasingly hydrophobic longer side chains contact instead the well-defined polar pockets lined with the DNA backbone and nucleic acid bases belonging to the loop and groove regions.

This is in contrast to previous MD simulation studies in which binding between the same set of ligands used in this Chapter and the parallel human telomeric G-quadruplex was investigated (Moore et al., 2006). This work reported considerable deviation from starting structures over the course of a 1 ns simulation run; including shifts in the position of the acridine core, the 9-substituent and free rotation of the amide bond orientation relative to the acridine core. An orientation was preferred whereby the ligand forms a crescent shape allowing the oxygen to point towards to the centre of the G-quartet. The reason for the difference between the study conducted by Moore and co-workers and the work in this chapter becomes clear once the two types of structures are compared; the first is a single ligand bound onto the surface of a 3' G-quartet belonging to a singular G-quadruplex structure, whilst here the ligand is sandwiched between two quadruplexes. Moreover, each of

the ligands studied here formed a shared plane with a thymine base consistent with the crystal structure. These factors contribute to the rigidity seen in the ligands here, whereas in the study by Moore et al, the ligands were free to sample conformational space without significant steric hindrance. The ligands here faced greater restriction especially for the acridine core and the attached amide bonds.

Conformational changes seen here were restricted to rotation in the alkyl part of the side chains and in the end pyrrolidine ring. The side chains at the 6-position interacted with the groove belonging to the 3' quadruplex as the DNA backbone prevented interactions with the nearest grooves on the 5' face. On the other hand, the side chain at the 3-position could, in theory at least, interact with grooves belonging to both of the quadruplexes as shown by ligand 1. All three ligands maintained a closer contact with the groove belonging to the 3' quadruplex except for ligand 1 which interacted with the groove on the other side. The conformational changes observed in both of the 3- and 6- substituents are also affected by the conformational changes in the loops themselves as they relaxed and adopted a more open conformation over the course of the simulation, thus imposing steric forces onto the end of the side chains in the ligand.

The calculated values of free energy of binding for the complexes are closely similar within the likely error bars of the simulation and reflect the similarity in structural attributes of the four complexes making them energetically indistinguishable. This is especially significant in view of the similarity in the ligands with regards to their structures and charge as discussed earlier. We note that these calculations take no account of possible differences in water arrangements.

4.6 Conclusion

So far, G-quadruplex binding ligands were generally designed to have three points of contact with the loops or grooves of the quadruplex. In principle, the ligands contained flexible rotating elements that would probe the loop and groove regions for favourable interactions. This study showed the important role that the water networks - as observed in the crystal - play in the process of ligand design. The

significant consequences of ligand binding to tandem or stacking quadruplex multimers compared to singular quadruplexes are demonstrated. The former mimics molecular crowding conditions where the binding site is rich with features thus greatly affecting ligand binding and restricting its movement. The latter has a featureless surface where the ligand has freedom of movement in both the horizontal and vertical directions.

CHAPTER 5

CONCLUSION

5 CONCLUSIONS

G-rich regions capable of forming G-quadruplex structures are found in telomeric ends of chromosomes and in non-telomeric regions of the human genome such as promoter elements of oncogenes, RNA telomeric transcript and RNA 5' untranslated regions. Selective targeting of a specific G-rich region of the genome for a therapeutic effect depends on substantial knowledge of the structure and behaviour of the quadruplex target in question.

The crystal structure of the complex formed between BRACO19 and the human telomeric quadruplex presented here has shown that not only does ligand binding to tandem quadruplex multimers represent a real goal but also demonstrates that the binding site (formed at the interface of stacking quadruplexes) is rich with features that can be selectively targeted with suitable ligands.

BRACO19 was found to be sandwiched between two quadruplexes stacked 5' to 3' and anchored in place by flanking nucleotides (four thymine and two adenine nucleotides) which formed a significant part of the binding site. All six nucleotides flipped into the interface between the two quadruplexes forming multiple contacts with the ligand and leading to enhanced ligand-quadruplex "fit". In fact, the significant role played by the flexible flanking nucleotides in ligand binding was demonstrated in several places. The aromatic core of BRACO19, being of significantly smaller dimensions than the G-quartet surface and consequently less efficient at forming stacking interactions, was in effect extended with a thymine base to form a coplane that matched the dimensions of the quadruplex face more closely (i.e. by being sufficiently large). This led to greater stacking interactions and consequently more stable binding. The ligand-thymine coplane was a recurrent feature which was also observed in this work in the structures of the complexes formed between the *Oxytricha nova* telomeric quadruplex and a series of six disubstituted acridines in accord with previous studies. Another thymine base

flipped into the binding site almost perpendicular to the acridine core of BRACO19 and stacked onto the phenyl ring (part of the 9-substituent), thus anchoring the ligand in place. Moreover, the remaining thymine and adenine bases formed a TATA plane onto which the BRACO19-thymine coplane, mentioned earlier, stacked at the 5' quadruplex face.

All of the flanking nucleotides mentioned above can be accounted for by the nucleotides of two TTA loops (4 T + 2 A = 2 x TTA), which would need to be added to the singular stacking bimolecular quadruplexes in the crystal in order to create a tandem unimolecular quadruplex dimer. The two loops would be; the loop connecting two consecutive quadruplexes in the quadruplex dimer and the loop connecting the two DNA strands in the bimolecular quadruplex.

The nature of the loops in a quadruplex also has fundamental consequences with regards to ligand binding. A link between loop type and experimental observations became clear upon comparison of the propeller loops in the parallel BRACO19 complex to the diagonal loops in the disubstituted acridine series of complexes in the anti-parallel quadruplex, which emphasised the fundamental differences between the two (despite some common features being retained such as the ligand thymine coplane). The end hetero rings at the 3- and 6-substituent side chains point towards the solvent. The wider and open nature of the binding pocket formed by the diagonal loop enabled successful binding to increasingly bulky moieties (not only to five but also to six and seven membered rings). In the parallel structure, these moieties would have to be contained within a shallow pocket formed by the propeller loops. Based on this, a rationale for the experimental results became clear. Firstly, increasing the size of the end hetero rings would result in increasing steric constraints imposed by the shallow pocket of the parallel loop and consequently reduce affinity of binding (in contrast to the diagonal loop). This was in accord with experimental observations. Secondly, increasing the length of the alkyl part of the side chain would replace the positively charged nitrogen in the pyrrolidine ring (which is participating in local water networks in the crystal) with a hydrophobic alkyl part thus obliterating the established ligand-water-DNA structural pattern also in accord with the experiment. The modelling studies supported these

observations whereby increasing the length of the side chains resulted in increasingly mobile side chains compared to the more stable shorter side chains present in BRACO19. This was due to the combined effect of greater steric constraints imposed by the shallow pocket of the propeller loop on the longer (hence more bulky) side chains and the latter being predisposed to greater mobility due to a less stable (or absent) ligand-water-DNA network structure.

The role of favourable crystal packing interactions cannot be excluded as a contributing factor to the observed 5' to 3' stacking of the quadruplexes in the structure of the BRACO19 complex. However, the purpose of structure determination is not only to provide ready starting points for structure-based ligand design. It is also extremely useful in providing insight into the process of ligand binding itself, revealing information about conformational changes that lead to phenomena such as "induced-fit" between ligand and receptor.

The crystal structures presented here have had a significant impact on current perceptions of structure-based ligand design of telomeric G-quadruplex binding agents especially in view of the highly concentrated solute content in the crystals. The effect of molecular crowding conditions on the structure and stability of G-quadruplex structure (and biomolecules in general) is highly significant because this can illustrate how they behave under cell-mimicking conditions. Increasing the concentration of the solute, adding viscous agents such as polyethylene glycol (PEG) or adding ethanol are all methods commonly used to mimic molecular crowding conditions. Conditions in the crystal (more specifically the high solute concentration) are more closely related to molecular crowding conditions than dilute solutions used in other methods used in structural studies such as NMR.

However, NMR methods are extremely valuable in elucidating structural information, such as establishing that the specific fold adopted by the human telomeric quadruplex is sensitive to the nature of flanking nucleotides. Although it has not been shown yet whether ligands can affect the quadruplex fold in a similar way, G-quadruplex binding ligands can act as molecular chaperones for quadruplex formation.

Further, preferably high resolution, structural studies are necessary to clarify the structure and behaviour of tandem telomeric G-quadruplex multimers (native and in complex with ligands) in conditions mimicking cellular crowding conditions (such as in the crystal). This is essential if structure-based ligand design is to achieve increased affinity and decreased (or negligible) cross-binding with other quadruplexes; both characteristics are fundamental in advancing compounds from ligands to potential clinical anticancer agents.

BIBLIOGRAPHY

BIBLIOGRAPHY

Agbandje, M., Jenkins, T.C., McKenna, R., Reszka, A.P. & Neidle, S. (1992) **Anthracene-9,10-diones as potential anticancer agents. Synthesis, DNA-binding, and biologicai studies on a series of 2,6-disubstituted derivatives**. *Journal of Medicinal Chemistry*, **35**(8): 1418-1429.

Azzalin, C.M. & Lingner, J. (2008) **Telomeres: the silence is broken**. *Cell Cycle,* **7**(9): 1161-1165.

Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. (2007) **Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends**. *Science,* **318**(5851): 798-801.

Babushok, D.V., Ostertag, E.M. & Kazazian, H.H. (2007) **Current topics in genome evolution: molecular mechanisms of new gene formation**. *Cellular and Molecular Life Sciences*, **64**(5): 542-554.

Bendich, A.J. & Drlica, K. (2000) **Prokaryotic and eukaryotic chromosomes: what's the difference?** *BioEssays*, **22**(5): 481-486.

Blackburn, E.H. (2005) **Telomeres and telomerase: their mechanisms of action and the effects of altering their functions**. *FEBS Letters*, **579**(4): 859-862.

Blackburn, E.H. & Gall, J.G. (1978) **A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena**. *Journal of Molecular Biology*, **120**(1): 33-53.

Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S. & Wright, W.E. (1998) **Extension of life-span by introduction of telomerase into normal human cells**. *Science*,

279(5349): 349-352.

Bollmann, F.M. (2007) **Targeting ALT: the role of alternative lengthening of telomeres in pathogenesis and prevention of cancer**. *Cancer Treatment Reviews*, **33**(8): 704-709.

Brassart, B., Gomez, D., De Cian, A., Paterski, R., Montagnac, A., Qui, K-H., Temime-Smaali, N., Trentesaux, C., Mergny, J-L., Gueritte, F. & Riou, J-F. (2007) **A new steroid derivative stabilizes g-quadruplexes and induces telomere uncapping in human tumor cells**. *Molecular Pharmacology*, **72**(3): 631-640.

Brown, I.D. & McMahon, B. (2002) **CIF: the computer language of crystallography**. *Acta Crystallographica Section B*, **58**(3 Part 1): 317–324.

Bryan, T.M., Englezou, A., Gupta, J., Bacchetti, S. & Reddel, R.R. (1995) **Telomere elongation in immortal human cells without detectable telomerase activity**. *The EMBO Journal*, **14**(17): 4240-4248.

Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. & Neidle, S. (2006) **Quadruplex DNA: sequence, topology and structure**. *Nucleic Acids Research*, **34**(19): 5402-5415.

Burger, A.M., Dai, F., Schultes, C.M., Reszka, A.P., Moore, M.J., Double, J.A. & Neidle, S. (2005) **The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function**. *Cancer Research*, **65**(4): 1489-1496.

Cáceres, C., Wright, G., Gouyette, C., Parkinson, G. & Subirana, J.A. (2004) **A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tl^+/Na^+ ions**. *Nucleic Acids Research*, **32**(3): 1097-1102.

Cahoon, L.A. & Seifert, H.S. (2009) **An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae**. *Science (New*

York, N.Y.), **325**(5941): 764-767.

Cai, X., Gray, P.J. & Von Hoff, D.D. (2009) **DNA minor groove binders: Back in the groove**. *Cancer Treatment Reviews*. **35**(5): 437-450.

Cairney, C.J. & Keith, W.N. (2008) **Telomerase redefined: integrated regulation of hTR and hTERT for telomere maintenance and telomerase activity**. *Biochimie*, **90**(1): 13-23.

Campbell, N.H., Parkinson, G.N., Reszka, A.P. & Neidle, S. (2008) **Structural basis of DNA quadruplex recognition by an acridine drug**. *Journal of the American Chemical Society*, **130**(21): 6722-6724.

Campbell, N.H., Patel, M., Tofa, A.B., Ghosh, R., Parkinson, G.N. & Neidle, S. (2009) **Selectivity in Ligand Recognition of G-Quadruplex Loops (dagger)**. *Biochemistry*, **48**(8): 1675-1680.

Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B. & Woods, R.J. (2005) **The Amber biomolecular simulation programs**. *Journal of Computational Chemistry*, **26**(16): 1668-1688.

Cesare, A.J. & Reddel, R.R. (2008) **Telomere uncapping and alternative lengthening of telomeres**. *Mechanisms of Ageing and Development*, **129**(1-2): 99-108.

Chang, C., Kuo, I., Ling, I., Chen, C., Chen, H., Lou, P., Lin, J. & Chang, T. (2004) **Detection of quadruplex DNA structures in human telomeres by a fluorescent carbazole derivative**. *Analytical Chemistry*, **76**(15): 4490-4494.

Cheatham, T.E. & Kollman, P.A. (1996) **Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution**. *Journal of Molecular Biology*, **259**(3): 434-444.

Cheatham, T.E. & Young, M.A. (2000) **Molecular dynamics simulation of nucleic acids: successes, limitations, and promise**. *Biopolymers*, **56**(4): 232-256.

Cheatham, T.E., Cieplak, P. & Kollman, P.A. (1999) **A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat**. *Journal of Biomolecular Structure & Dynamics*, **16**(4): 845-862.

Chebotareva, N.A., Kurganov, B.I. & Livanova, N.B. (2004) **Biochemical effects of molecular crowding**. *Biochemistry (Moscow)*, **69**(11): 1239-1251.

Chowdhury, S. & Bansal, M. (2001) **G-Quadrupiex Structure Can Be Stable with Only Some Coordination Sites Being Occupied by Cations: A Six-Nanosecond Molecular Dynamics Study**. *The Journal of Physical Chemistry B*, **105**(31): 7572-7578.

Clark, G.R., Pytel, P.D., Squire, C.J. & Neidle, S. (2003) **Structure of the first parallel DNA quadruplex-drug complex**. *Journal of the American Chemical Society*, **125**(14): 4066-4067.

Collaborative Computational Project, N.4. (1994) **The CCP4 suite: programs for protein crystallography**. *Acta Crystallographica Section D*, **50**(5): 760–763.

Collier, D.A. & Neidle, S. (1988) **Synthesis, molecular modeling, DNA binding, and antitumor properties of some substituted amidoanthraquinones**. *Journal of Medicinal Chemistry*, **31**(4): 847-857.

Cookson, J.C., Dai, F., Smith, V., Heald, R.A., Laughton, C.A., Stevens, M.F.G. & Burger, A.M. (2005) **Pharmacodynamics of the G-quadrupiex-stabiiizing telomerase inhibitor 3,11-difluoro-6,8,13-trimethyl-8H-quino[4,3,2-kl]acridinium methosulfate (RHPS4) in vitro: activity in human tumor ceils correlates with telomere length and can be enhanced, or antagonized, with cytotoxic agents**. *Molecular Pharmacology*, **68**(6): 1551-1558.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M.,

Spellmeyer, D.C., Fox, T., Caldwell, J.W. & Kollman, P.A. (1995) **A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules**. *Journal of the American Chemical Society*, **117**(19): 5179-5197.

Cortez-Gonzalez, X. & Zanetti, M. (2007) **Telomerase immunity from bench to bedside: round one**. *Journal of Translational Medicine*, **5**: 12.

Cosme-Blanco, W. & Chang, S. (2008) **Dual roles of telomere dysfunction in initiation and suppression of tumorigenesis**. *Experimental Cell Research*, **314**(9): 1973-1979.

Counter, C.M., Avilion, A.A., LeFeuvre, C.E., Stewart, N.G., Greider, C.W., Harley, C.B. & Bacchetti, S. (1992) **Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity**. *The EMBO Journal*, **11**(5): 1921-1929.

Court, R., Chapman, L., Fairall, L. & Rhodes, D. (2005) **How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures**. *EMBO Reports*, **6**(1): 39–45.

Creze, C., Rinaldi, B., Haser, R., Bouvet, P. & Gouet, P. (2007) **Structure of a d(TGGGGT) quadruplex crystallized in the presence of Li^+ ions**. *Acta Crystallographica. Section D, Biological Crystallography*, **63**(Pt 6): 682-688.

Dai, J., Carver, M. & Yang, D. (2008) **Polymorphism of human telomeric quadruplex structures**. *Biochimie*, **90**(8): 1172-1183.

Dai, J., Carver, M., Punchihewa, C., Jones, R.A. & Yang, D. (2007) **Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K^+ solution: insights into structure polymorphism of the human telomeric sequence**. *Nucleic Acids Research*, **35**(15): 4927-4940.

Dai, J., Punchihewa, C., Ambrus, A., Chen, D., Jones, R.A. & Yang, D. (2007) **Structure of the intramolecular human telomeric G-quadruplex in**

potassium solution: a novel adenine triple formation. *Nucleic Acids Research*, **35**(7): 2440-2450.

Darden, T.A., York, D. & Pedersen, L.G. (1993) **Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems**. *Journal of Chemical Physics*, **98**: 10089-10092.

De Boeck, G., Forsyth, R.G., Praet, M. & Hogendoorn, P.C.W. (2009) **Telomere-associated proteins: cross-talk between telomere maintenance and telomere-lengthening mechanisms**. *The Journal of Pathology*, **217**(3): 327-344.

De Cian, A. & Mergny, J. (2007) **Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation**. *Nucleic Acids Research*, **35**(8): 2483-2493.

De Cian, A., Gauthier, L.R., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J. & Mergny, J. (2008) **Targeting telomeres and telomerase**. *Biochimie*, **90**(1): 131-155.

de Lange, T. (2002) **Protection of mammalian telomeres**. *Oncogene*, **21**(4): 532-540.

de Lange, T. (2005) **Shelterin: the protein complex that shapes and safeguards human telomeres**. *Genes & Development*, **19**(18): 2100-2110.

de Lange, T. & Jacks, T. (1999) **For Better or Worse? Telomerase Inhibition and Cancer**. *Cell*, **98**(3): 273-275.

de Lange, T., Shiue, L., Myers, R.M., Cox, D.R., Naylor, S.L., Killery, A.M. & Varmus, H.E. (1990) **Structure and variability of human chromosome ends**. *Molecular and Cellular Biology*, **10**(2): 518-527.

DeLano, W.L. (2008) **The PyMOL Molecular Graphics System. DeLano Scientific**

LLC, Palo Alto, CA, USA. http://www.pymol.org.

Deng, Y. & Chang, S. (2007) **Role of telomeres and telomerase in genomic instability, senescence and cancer**. *Laboratory Investigation*, **87**(11): 1071-1076.

Dionne, I. & Wellinger, R.J. (1996) **Cell cycle-regulated generation of single-stranded G-rich DNA in the absence of telomerase**. *Proceedings of the National Academy of Sciences of the United States of America*, **93**(24): 13902-13907.

Dodson, E. (2008) **The befores and afters of molecular replacement**. *Acta Crystallographica. Section D, Biological Crystallography*, **64**(Pt 1): 17-24.

Dong, C.K., Masutomi, K. & Hahn, W.C. (2005) **Telomerase: regulation, function and transformation**. *Critical Reviews in Oncology/Hematology*, **54**(2): 85-93.

Du, Z., Zhao, Y. & Li, N. (2008) **Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription**. *Genome Research*, **18**(2): 233-241.

Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F. & Maizels, N. (2004) **Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA**. *Genes & Development*, **18**(13): 1618-1629.

Eddy, J. & Maizels, N. (2006) **Gene function correlates with potential for G4 DNA formation in the human genome**. *Nucleic Acids Research*, **34**(14): 3887-3896.

Ellis, R.J. (2001) **Macromolecular crowding: an important but neglected aspect of the intracellular environment**. *Current Opinion in Structural Biology*, **11**(1): 114-119.

Emsley, P. & Cowtan, K. (2004) **Coot: model-building tools for molecular graphics**. *Acta Crystallographica. Section D, Biological Crystallography*, **60**(Pt 12 Pt 1): 2126-2132.

Evans, P. & McCoy, A. (2008) **An introduction to molecular replacement**. *Acta Crystallographica. Section D, Biological Crystallography*, **64**(Pt 1): 1-10.

Fadrná, E., Spacková, N., Stefl, R., Koca, J., Cheatham, T.E. & Sponer, J. (2004) **Molecular dynamics simulations of Guanine quadruplex loops: advances and force field limitations**. *Biophysical Journal*, **87**(1): 227-242.

Feldser, D.M. & Greider, C.W. (2007) **Short telomeres limit tumor progression in vivo by inducing senescence**. *Cancer Cell*, **11**(5): 461-469.

Feng, J., Funk, W.D., Wang, S.S., Weinrich, S.L., Avilion, A.A., Chiu, C.P., Adams, R.R., Chang, E., Allsopp, R.C. & Yu, J. (1995) **The RNA component of human telomerase**. *Science*, **269**(5228): 1236-1241.

Finkel, T., Serrano, M. & Blasco, M.A. (2007) **The common biology of cancer and ageing**. *Nature*, **448**(7155): 767-774.

Garman, E. (1999) **Cool data: quantity AND quality**. *Acta Crystallographica. Section D, Biological Crystallography*, **55**(Pt 10): 1641-1653.

Garman, E. (2003) **'Cool' crystals: macromolecular cryocrystallography and radiation damage**. *Current Opinion in Structural Biology*, **13**(5): 545-551.

Garman, E.F. & Owen, R.L. (2006) **Cryocooling and radiation damage in macromolecular crystallography**. *Acta Crystallographica. Section D, Biological Crystallography*, **62**(Pt 1): 32-47.

Garman, E.F. & Schneider, T.R. (1997) **Macromolecular Cryocrystallography**. *Journal of Applied Crystallography*, **30**: 211-237.

Gavathiotis, E., Heald, R.A., Stevens, M.F.G. & Searle, M.S. (2003) **Drug recognition and stabilisation of the parallel-stranded DNA quadruplex d(TTAGGGT)4 containing the human telomeric repeat**. *Journal of Molecular Biology*, **334**(1): 25-36.

Gellert, M., Lipsett, M.N. & Davies, D.R. (1962) **Helix formation by guanylic acid**. *Proceedings of the National Academy of Sciences of the United States of America*, **48**: 2013-2018.

Gomez, D., Wenner, T., Brassart, B., Douarre, C., O'Donohue, M-F., El Koury, V., Shin-Ya, K., Morjani, H., Trentesaux, C. & Riou, J-F. (2006) **Telomestatin-induced telomere uncapping is modulated by POT1 through G-overhang extension in HT1080 human tumor cells**. *The Journal of Biological Chemistry*, **281**(50): 38721-38729.

Gowan, S.M., Harrison, R.J., Patterson, L., Valenti, M., Read, M.A., Neidle, S. & Kelland, L.R. (2002) **A G-quadruplex-interactive potent small-molecule inhibitor of telomerase exhibiting in vitro and in vivo antitumor activity**. *Molecular Pharmacology*, **61**(5): 1154-1162.

Gowan, S.M., Heald, R., Stevens, M.F. & Kelland, L.R. (2001) **Potent inhibition of telomerase by small-molecule pentacyclic acridines capable of interacting with G-quadruplexes**. *Molecular Pharmacology*, **60**(5): 981-988.

Grandin, N. & Charbonneau, M. (2008) **Protection against chromosome degradation at the telomeres**. *Biochimie*, **90**(1): 41-59.

Granotier, C., Pennarun, G., Riou, L., Hoffschir, F., Gauthier, L.R., De Cian, A., Gomez, D., Mandine, E., Riou, J-F., Mergny, J-L., Mailliet, P., Dutrillaux, B. & Boussin, F.D. (2005) **Preferential binding of a G-quadruplex ligand to human chromosome ends**. *Nucleic Acids Research*, **33**(13): 4182-4190.

Greider, C.W. (1998) **Telomeres and senescence: the history, the experiment, the future**. *Current Biology*, **8**(5): R178-R181.

Greider, C.W. (1999) **Telomeres do D-loop-T-loop**. *Cell*, **97**(4): 419-422.

Greider, C.W. & Blackburn, E.H. (1985) **Identification of a specific telomere terminal transferase activity in Tetrahymena extracts**. *Cell*, **43**(2 Pt 1): 405-413.

Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H. & de Lange, T. (1999) **Mammalian telomeres end in a large duplex loop**. *Cell*, **97**(4): 503-514.

Grossfield, A., Ren, P. & Ponder, J.W. (2003) **Ion solvation thermodynamics from simulation with a polarizable force field**. *Journal of the American Chemical Society*, **125**(50): 15671-15682.

Gunaratnam, M., Greciano, O., Martins, C., Reszka, A.P., Schultes, C.M., Morjani, H., Riou, J. & Neidle, S. (2007) **Mechanism of acridine-based telomerase inhibition and telomere shortening**. *Biochemical Pharmacology*, **74**(5): 679-689.

Haider, S.M. & Neidle, S. (2009) **A molecular model for drug binding to tandem repeats of telomeric G-quadruplexes**. *Biochemical Society Transactions*, **37**(Pt 3): 583-588.

Haider, S.M., Parkinson, G.N. & Neidle, S. (2002) **Crystal structure of the potassium form of an Oxytricha nova G-quadruplex**. *Journal of Molecular Biology*, **320**(2): 189-200.

Haider, S.M., Parkinson, G.N. & Neidle, S. (2003) **Structure of a G-quadruplex-ligand complex**. *Journal of Molecular Biology*, **326**(1): 117-125.

Halgren, T.A. (1996) **Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94**. *Journal of Computational Chemistry*, **17**(5-6): 490-519.

Halgren, T.A. (1996) **Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions**. *Journal of Computational Chemistry*, **17**(5-6): 520-552.

Hanahan, D. & Weinberg, R.A. (2000) **The Hallmarks of Cancer**. *Cell*, **100**(1): 57-70.

Hanaoka, S., Nagadoi, A. & Nishimura, Y. (2005) **Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities**. *Protein Science*, **14**(1): 119-130.

Hara, E., Tsurui, H., Shinozaki, A., Nakada, S. & Oda, K. (1991) **Cooperative effect of antisense-Rb and antisense-p53 oligomers on the extension of life span in human diploid fibroblasts, TIG-1**. *Biochemical and Biophysical Research Communications*, **179**(1): 528-534.

Harley, C.B. (2008) **Telomerase and cancer therapeutics**. *Nature Reviews Cancer*, **8**(3): 167-179.

Harley, C.B., Futcher, A.B. & Greider, C.W. (1990) **Telomeres shorten during ageing of human fibroblasts**. *Nature*, **345**(6274): 458-460.

Harrington, L., Zhou, W., McPhail, T., Oulton, R., Yeung, D.S., Mar, V., Bass, M.B. & Robinson, M.O. (1997) **Human telomerase contains evolutionarily conserved catalytic and structural subunits**. *Genes & Development*, **11**(23): 3109-3115.

Harrison, R.J., Cuesta, J., Chessari, G., Read, M.A., Basra, S.K., Reszka, A.P., Morrell, J., Gowan, S.M., Incles, C.M., Tanious, F.A., Wilson, W.D., Kelland, L.R. & Neidle, S. (2003) **Trisubstituted acridine derivatives as potent and selective telomerase inhibitors**. *Journal of Medicinal Chemistry*, **46**(21): 4463-4476.

Harrison, R.J., Gowan, S.M., Kelland, L.R. & Neidle, S. (1999) **Human telomerase inhibition by substituted acridine derivatives**. *Bioorganic & Medicinal*

Chemistry Letters, **9**(17): 2463-2468.

Hashimoto, M., Kyo, S., Masutomi, K., Maida, Y., Sakaguchi, J., Mizumoto, Y., Nakamura, M., Hahn, W.C. & Inoue, M. (2005) **Analysis of telomeric single-strand overhang length in human endometrial cancers**. *FEBS Letters*, **579**(13): 2959-2964.

Hayflick, L. (1965) **The limited in vitro lifetime of human diploid cell strains*1, *2**. *Experimental Cell Research*, **37**(3): 614-636.

Hazel, P., Huppert, J., Balasubramanian, S. & Neidle, S. (2004) **Loop-length-dependent folding of G-quadruplexes**. *Journal of the American Chemical Society*, **126**(50): 16405-16415.

Hazel, P., Parkinson, G.N. & Neidle, S. (2006) **Predictive modelling of topology and loop variations in dimeric DNA quadruplex structures**. *Nucleic Acids Research*, **34**(7): 2117-2127.

He, H., Multani, A.S., Cosme-Blanco, W., Tahara, H., Ma, J., Pathak, S., Deng, Y. & Chang, S. (2006) **POT1b protects telomeres from end-to-end chromosomal fusions and aberrant homologous recombination**. *The EMBO Journal*, **25**(21): 5180-5190.

Hemann, M.T., Strong, M.A., Hao, L. & Greider, C.W. (2001) **The Shortest Telomere, Not Average Telomere Length, Is Critical for Cell Viability and Chromosome Stability**. *Cell*, **107**(1): 67-77.

Henderson, E.R. & Blackburn, E.H. (1989) **An overhanging 3' terminus is a conserved feature of telomeres**. *Molecular and Cellular Biology*, **9**(1): 345-348.

Hiyama, E. & Hiyama, K. (2003) **Telomerase as tumor marker**. *Cancer Letters*, **194**(2): 221-233.

Horvath, M.P. & Schultz, S.C. (2001) **DNA G-quartets in a 1.86 A resolution structure of an Oxytricha nova telomeric protein-DNA complex**. *Journal of Molecular Biology*, **310**(2): 367-377.

Huffman, K.E., Levene, S.D., Tesmer, V.M., Shay, J.W. & Wright, W.E. (2000) **Telomere shortening is proportional to the size of the G-rich telomeric 3'-overhang**. *The Journal of Biological Chemistry*, **275**(26): 19719-19722.

Humphrey, W., Dalke, A. & Schulten, K. (1996) **VMD: visual moiecular dynamics**. *Journal of Molecular Graphics*, **14**(1): 33-38, 27-28.

Huppert, J.L. & Balasubramanian, S. (2005) **Prevalence of quadruplexes in the human genome**. *Nucleic Acids Research*, **33**(9): 2908-2916.

Huppert, J.L. & Balasubramanian, S. (2007) **G-quadruplexes in promoters throughout the human genome**. *Nucleic Acids Research*, **35**(2): 406-413.

Huppert, J.L., Bugaut, A., Kumari, S. & Balasubramanian, S. (2008) **G-quadrupiexes: the beginning and end of UTRs**. *Nucleic Acids Research*, **36**(19): 6260-6268.

Incles, C.M., Schultes, C.M., Kempski, H., Koehler, H., Kelland, L.R. & Neidle, S. (2004) **A G-quadruplex telomere targeting agent produces p16-associated senescence and chromosomal fusions in human prostate cancer cells**. *Molecular Cancer Therapeutics*, **3**(10): 1201-1206.

Jakalian, A., Bush, B.L., Jack, D.B. & Bayly C.I. (2000) **Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method**. *Journal of Computational Chemistry*, **21**(2): 132-146.

Jakalian, A., Jack, D.B. & Bayly, C.I. (2002) **Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation**. *Journal of Computational Chemistry*, **23**(16): 1623-1641.

Johnson, K.A. (2008) **Role of induced fit in enzyme specificity: a molecular forward/reverse switch**. *The Journal of Biological Chemistry*, **283**(39): 26297-26301.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. (1983) **Comparison of simple potentiai functions for simulating liquid water**. *The Journal of Chemical Physics*, **79**(2): 926-935.

Kan, Z., Lin, Y., Wang, F., Zhuang, X., Zhao, Y., Pang, D., Hao, Y. & Tan, Z. (2007) **G-quadrupiex formation in human telomeric (TTAGGG)4 sequence with complementary strand in close vicinity under moiecuiarly crowded condition**. *Nucleic Acids Research*, **35**(11): 3646-3653.

Kan, Z., Yao, Y., Wang, P., Li, X., Hao, Y. & Tan, Z. (2006) **Molecuiar crowding induces telomere G-quadrupiex formation under salt-deficient conditions and enhances its competition with duplex formation**. *Angewandte Chemie (International Ed. in English)*, **45**(10): 1629-1632.

Karlseder, J., Smogorzewska, A. & de Lange, T. (2002) **Senescence induced by altered telomere state, not telomere loss**. *Science*, **295**(5564): 2446-2449.

Kelland, L.R. (2005) **Overcoming the immortality of tumour cells by telomere and telomerase based cancer therapeutics--current status and future prospects**. *European Journal of Cancer*, **41**(7): 971-979.

Keppler, M., Read, M., Perry, P., Trent, J., Jenkins, T., Reszka, A., Neidle, S. & Fox, K. (1999) **Stabilization of DNA triple helices by a series of mono- and disubstituted amidoanthraquinones**. *European Journal of Biochemistry*, **263**: 817-825.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L & Shay, J.W. (1994) **Specific association of human teiomerase activity with immortal cells and cancer**. *Science*, **266**(5193): 2011-2015.

Klobutcher, L.A., Swanton, M.T., Donini, P. & Prescott, D.M. (1981) **All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus**. *Proceedings of the National Academy of Sciences of the United States of America*, **78**(5): 3015-3019.

Kollman, P.A. et al. (2000) **Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models**. *Accounts of Chemical Research*, **33**(12): 889-897.

Kolquist, K.A., Ellisen, L.W., Counter, C.M., Meyerson, M., Tan, L.K., Weinberg, R.A., Haber, D.A. & Gerald, W.L. (1998) **Expression of TERT in early premalignant lesions and a subset of cells in normal tissues**. *Nature Genetics*, **19**(2): 182-186.

Kumari, S., Bugaut, A., Huppert, J.L. & Balasubramanian, S. (2007) **An RNA G-quadrupiex in the 5' UTR of the NRAS proto-oncogene modulates translation**. *Nature Chemical Biology*, **3**(4): 218-221.

Leach, A. (2001) *Molecular Modelling: Principles and application,* 2nd Ed., Pearson Education Ltd., Harlow.

Lee, M.E., Rha, S.Y., Jeung, H., Kim, T.S., Chung, H.C. & Oh, B. (2008) **Variation of the 3' telomeric overhang lengths in human celis**. *Cancer Letters*, **264**(1): 107-118.

Lei, M., Podell, E.R. & Cech, T.R. (2004) **Structure of human POT1 bound to teiomeric single-stranded DNA provides a model for chromosome end-protection**. *Nature Structural & Molecular Biology*, **11**(12): 1223-1229.

Lim, K.W., Amrane, S., Bouaziz, S., Xu, W., Mu, Y., Patel, D.J., Luu, K.N. & Phan, A.T. (2009) **Structure of the human telomere in K^+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers**. *Journal of the American Chemical Society*, **131**(12): 4301-4309.

Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. & Cech, T.R.

(1997) **Reverse transcriptase motifs in the catalytic subunit of telomerase**. *Science*, **276**(5312): 561-567.

Lipps, H.J. & Rhodes, D. (2009) **G-quadruplex structures: in vivo evidence and function**. *Trends in Cell Biology*, **19**(8): 414-422.

Loayza, D., Parsons, H., Donigian, J., Hoke, K. & de Lange, T. (2004) **DNA binding features of human POT1: a nonamer 5'-TAGGGTTAG-3' minimal binding site, sequence specificity, and internal binding to multimeric sites**. *The Journal of Biological Chemistry*, **279**(13): 13241-13248.

Londoño-Vallejo, J.A. (2008) **Telomere instability and cancer**. *Biochimie*, **90**(1): 73-82.

Lu, X. & Olson, W.K. (2003) **3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures**. *Nucleic Acids Research*, **31**(17): 5108-51021.

Luu, K.N., Phan, A.T., Kuryavyi, V., Lacroix, L. & Patel, D.J. (2006) **Structure of the human telomere in K^+ solution: an intramolecular (3 + 1) G-quadruplex scaffold**. *Journal of the American Chemical Society*, **128**(30): 9963-9970.

Maizels, N. (2006) **Dynamic roles for G4 DNA in the biology of eukaryotic cells**. *Nature Structural & Molecular Biology*, **13**(12): 1055-1059.

Maple, J.R., Cao, Y.X., Damm, W.G., Halgren, T.A., Kaminski, G.A., Zhang, L.Y. & Friesner, R.A. (2005) **A polarizable force field and continuum solvation methodology for modeling of protein-ligand interactions**. *JOURNAL OF CHEMICAL THEORY AND COMPUTATION*, **1**(4): 694-715.

Martadinata, H. & Phan, A.T. (2009) **Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K^+ solution**. *Journal of the American Chemical Society*, **131**(7): 2570-2578.

Martins, C., Gunaratnam, M., Stuart, J., Makwana, V., Greciano, O., Reszka, A.P., Kelland, L.R. & Neidle, S. (2007) **Structure-based design of benzylamino-acridine compounds as G-quadruplex DNA telomere targeting agents**. *Bioorganic & Medicinal Chemistry Letters*, **17**(8): 2293-2298.

Masutomi, K. & Hahn, W.C. (2003) **Telomerase and tumorigenesis**. *Cancer Letters*, **194**(2): 163-172.

Masutomi, K. et al. (2003) **Telomerase maintains telomere structure in normal human cells**. *Cell*, **114**(2): 241-253.

Matsugami, A., Xu, Y., Noguchi, Y., Sugiyama, H. & Katahira, M. (2007) **Structure of a human telomeric DNA sequence stabilized by 8-bromoguanosine substitutions, as determined by NMR in a K^+ solution**. *The FEBS Journal*, **274**(14): 3545-3556.

McClintock, B. (1939) **The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis**. *Proceedings of the National Academy of Sciences of the United States of America*, **25**(8): 405-416.

McClintock, B. (1941) **The Stability of Broken Ends of Chromosomes in Zea Mays**. *Genetics*, **26**(2): 234-282.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. & Read, R.J. (2007) **Phaser crystallographic software**. *Journal of Applied Crystallography*, **40**(4): 658–674.

Mimeault, M. & Batra, S.K. (2008) **Recent insights into the molecular mechanisms involved in aging and the malignant transformation of adult stem/progenitor cells and their therapeutic implications**. *Ageing Research Reviews*, **8**(2): 94-112.

Miyoshi, D. & Sugimoto, N. (2008) **Molecular crowding effects on structure and stability of DNA**. *Biochimie*, **90**(7): 1040-1051.

Mobley, D.L., Dumont, E., Chodera, J.D. & Dill, K.A. (2007) **Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent**. *The Journal of Physical Chemistry. B*, **111**(9): 2242-2254.

Monchaud, D. & Teulade-Fichou, M. (2008) **A hitchhiker's guide to G-quadruplex ligands**. *Organic & Biomolecular Chemistry*, **6**(4): 627-636.

Moore, M.J.B., Schultes, C.M., Cuesta, J., Cuenca, F., Gunaratnam, M., Tanious, F.A., Wilson, W.D. & Neidle, S. (2006) **Trisubstituted acridines as G-quadruplex telomere targeting agents. Effects of extensions of the 3,6- and 9-side chains on quadruplex binding, telomerase activity, and cell proliferation**. *Journal of Medicinal Chemistry*, **49**(2): 582-599.

Moustakas, D.T., Lang, P.T., Pegg, S., Pettersen, E., Kuntz, I.D., Brooijmans, N. & Rizzo, R.C. (2006) **Development and validation of a modular, extensible docking program: DOCK 5**. *Journal of Computer-Aided Molecular Design*, **20**(10-11): 601-619.

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. & Wu, J.R. (1988) **A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes**. *Proceedings of the National Academy of Sciences of the United States of America*, **85**(18): 6622-6626.

Muntoni, A. & Reddel, R.R. (2005) **The first molecular details of ALT in human tumor cells**. *Human Molecular Genetics*, **14 Spec No. 2**: R191-R196.

Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997) **Refinement of macromolecular structures by the maximum-likelihood method**. *Acta Crystallographica. Section D, Biological Crystallography*, **53**(Pt 3): 240-255.

Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B. & Cech, T.R. (1997) **Telomerase catalytic subunit homologs from fission yeast and human**. *Science*, **277**(5328): 955-959.

Neidle, S. & Parkinson, G.N. (2008) **Quadruplex DNA crystal structures and drug design**. *Biochimie*, **90**(8): 1184-1196.

Oganesian, L. & Bryan, T.M. (2007) **Physiological relevance of telomeric G-quadruplex formation: a potential drug target**. *BioEssays*, **29**(2): 155-165.

Olovnikov, A.M. (1973) **A theory of marginotomy : The incomplete copying of template margin in enzymic synthesis of polynucleotides and biooiogicai significance of the phenomenon**. *Journal of Theoretical Biology*, **41**(1): 181-190.

Ou, T., Lu, Y., Tan, J., Huang, Z., Wong, K. & Gu, L. (2008) **G-Quadruplexes: Targets in Anticancer Drug Design**. *ChemMedChem*, **3**(5): 690-713.

Paeschke, K., Juranek, S., Simonsson, T., Hempel, A., Rhodes, D. & Lipps, H.J. (2008) **Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates**. *Nature Structural & Molecular Biology*, **15**(6): 598-604.

Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H.J. (2005) **Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo**. *Nature Structural & Molecular Biology*, **12**(10): 847-854.

Pan, C., Xue, B., Ellis, T.M., Peace, D.J. & Díaz, M.O. (1997) **Changes in Telomerase Activity and Telomere Length during Human T Lymphocyte Senescence**. *Experimental Cell Research*, **231**(2): 346-353.

Parkinson, G.N., Cuenca, F. & Neidle, S. (2008) **Topology conservation and loop flexibility in quadruplex-drug recognition: crystal structures of inter- and intramolecular telomeric DNA quadruplex-drug complexes**. *Journal of Molecular Biology*, **381**(5): 1145-1156.

Parkinson, G.N., Ghosh, R. & Neidle, S. (2007) **Structural basis for binding of porphyrin to human telomeres**. *Biochemistry*, **46**(9): 2390-2397.

Parkinson, G.N., Lee, M.P.H. & Neidle, S. (2002) **Crystal structure of parallel quadruplexes from human telomeric DNA**. *Nature*, **417**(6891): 876-880.

Patel, D.J., Kozlowski, S.A., Nordheim, A. & Rich, A. (1982) **Right-handed and left-handed DNA: studies of B- and Z-DNA by using proton nuclear Overhauser effect and P NMR**. *Proceedings of the National Academy of Sciences of the United States of America*, **79**(5): 1413-1417.

Patel, D.J., Phan, A.T. & Kuryavyi, V. (2007) **Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics**. *Nucleic Acids Research*, **35**(22): 7429-7455.

Pedroso, I.M., Hayward, W. & Fletcher, T.M. (2009) **The effect of the TRF2 N-terminal and TRFH regions on telomeric G-quadruplex structures**. *Nucleic Acids Research*, **37**(5): 1541-1554.

Pennarun, G. et al. (2005) **Apoptosis related to telomere instability and cell cycle alterations in human glioma cells treated by new highly selective G-quadruplex ligands**. *Oncogene*, **24**(18): 2917-2928.

Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E., Laughton, C.A. & Orozco, M. (2007) **Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers**. *Biophysical Journal*, **92**(11): 3817-3829.

Perry, P.J. et al. (1998) **Human telomerase inhibition by regioisomeric disubstituted amidoanthracene-9,10-diones**. *Journal of Medicinal Chemistry*, **41**(24): 4873-4884.

Perry, P.J., Gowan, S.M., Reszka, A.P., Polucci, P., Jenkins, T.C., Kelland, L.R. & Neidle, S. (1998) **1,4- and 2,6-disubstituted amidoanthracene-9,10-dione derivatives as inhibitors of human telomerase**. *Journal of Medicinal Chemistry*, **41**(17): 3253-3260.

Perry, P.J., Read, M.A., Davies, R.T., Gowan, S.M., Reszka, A.P., Wood, A.A., Kelland, L.R. & Neidle, S. (1999) **2,7-Disubstituted amidofluorenone derivatives as inhibitors of human telomerase**. *Journal of Medicinal Chemistry*, **42**(14): 2679-2684.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. & Ferrin, T.E. (2004) **UCSF Chimera--a visualization system for exploratory research and analysis**. *Journal of Computational Chemistry*, **25**(13): 1605-1612.

Pflugrath, J.W. (1999) **The finer things in X-ray diffraction data collection**. *Acta Crystallographica. Section D, Biological Crystallography*, **55**(Pt 10): 1718-1725.

Phan, A.T., Kuryavyi, V. & Patel, D.J. (2006) **DNA architecture: from G to Z**. *Current Opinion in Structural Biology*, **16**(3): 288-298.

Phan, A.T., Kuryavyi, V., Burge, S., Neidle, S. & Patel, D.J. (2007) **Structure of an unprecedented G-quadrupiex scaffold in the human c-kit promoter**. *Journal of the American Chemical Society*, **129**(14): 4386-4392.

Phan, A.T., Kuryavyi, V., Gaw, H.Y. & Patel, D.J. (2005) **Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter**. *Nature Chemical Biology*, **1**(3): 167-173.

Phan, A.T., Kuryavyi, V., Luu, K.N. & Patel, D.J. (2007) **Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K^+ solution**. *Nucleic Acids Research*, **35**(19): 6517-6525.

Phatak, P. & Burger, A.M. (2007) **Telomerase and its potential for therapeutic intervention**. *British Journal of Pharmacology*, **152**(7): 1003-1011.

Phatak, P., Cookson, J.C., Dai, F., Smith, V., Gartenhaus, R.B., Stevens, M.F.G. & Burger, A.M. (2007) **Telomere uncapping by the G-quadruplex ligand RHPS4 inhibits clonogenic tumour cell growth in vitro and in vivo consistent with a cancer stem cell targeting mechanism**. *British Journal of Cancer*, **96**(8): 1223-1233.

Pisano, S., Galati, A. & Cacchione, S. (2008) **Telomeric nucleosomes: forgotten players at chromosome ends**. *Cellular and Molecular Life Sciences*, **65**(22): 3553-3563.

Powell, M. (1977) **Restart procedures for the conjugate gradient method**. *Mathematical Programming*, **12**(2): 241-254.

Qin, Y. & Hurley, L.H. (2008) **Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions**. *Biochimie*, **90**(8): 1149-1171.

Rawal, P., Kummarasetti, V.B.R., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S.K. & Chowdhury, S. (2006) **Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation**. *Genome Research*, **16**(5): 644-655.

Raynaud, C.M., Sabatier, L., Philipot, O., Olaussen, K.A. & Soria, J. (2008) **Telomere length, telomeric proteins and genomic instability during the muitistep carcinogenic process**. *Critical Reviews in Oncology/Hematology*, **66**(2): 99-117.

Read, M.A., Harrison, R.J., Romagnoli, B., Tanious, F.A., Gowan, S.H., Reszka, A.P., Wilson, W.D., Kelland, L.R. & Neidle, S. (2001) **Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors**. *Proceedings of the National Academy of Sciences of the United States of America*, **98**(9): 4844-4849.

Read, M.A., Wood, A.A., Harrison, R.J., Gowan, S.M., Kelland, L.R., Dosanjh, H.S. & Neidle, S. (1999) **Molecular modeling studies on G-quadruplex complexes of telomerase inhibitors: structure-activity relationships**. *Journal of Medicinal Chemistry*, **42**(22): 4538-4546.

Reed, J., Gunaratnam, M., Beltran, M., Reszka, A.P., Vilar, R. & Neidle, S. (2008) **TRAP-LIG, a modified telomere repeat amplification protocol assay to quantitate telomerase inhibition by small molecules**. *Analytical Biochemistry*, **380**(1): 99-105.

Renciuk, D., Kejnovská, I., Skoláková, P., Bednárová, K., Motlová, J. & Vorlícková, M. (2009) **Arrangements of human telomere DNA quadruplex in physiologically relevant K^+ solutions**. *Nucleic Acids Research*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19717545 [Accessed September 1, 2009].

Rigaku Corporation© 1997-2002 **CrystalClear: An Integrated Program for the Collection and Processing of Area Detector Data**.

Riou, J., Guittat, L., Mailliet, P., Laoui, A., Renou, E., Petitgenet, O., Mégnin-Chanet, F., Hélène, C. & Mergny, J. (2002) **Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands**. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(5): 2672-2677.

Rodriguez, R., Müller, S., Yeoman, J.A., Trentesaux, C., Riou, J. & Balasubramanian, S. (2008) **A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres**. *Journal of the American Chemical Society*, **130**(47): 15758-15759.

Saenger, W. (1983) *Principles of Nucleic Acid Structure*, Springer-Verlag.

Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J. & Plückthun, A. (2001) **In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Styionychia iemnae macronuclei**. *Proceedings of the National Academy of Sciences of the United States of America*, **98**(15): 8572-8577.

Schneider, B., Neidle, S. & Berman, H.M. (1997) **Conformations of the sugar-phosphate backbone in helical DNA crystal structures**. *Biopolymers*, **42**(1): 113-124.

Schnieders, M.J., Baker, N.A., Ren, P. & Ponder, J.W. (2007) **Polarizable atomic multipole solutes in a Poisson-Boltzmann continuum**. *The Journal of Chemical Physics*, **126**(12): 124114.

Schoeftner, S. & Blasco, M.A. (2008) **Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II**. *Nature Cell Biology*, **10**(2): 228-236.

Schuettelkopf, A.W. & van Aalten, D.M.F. (2004) **PRODRG - a tool for high-throughput crystallography of protein-ligand complexes**. *Acta Crystallographica Section D*, **60**(8): 1355-1363.

Schultes, C.M., Guyen, B., Cuesta, J. & Neidle, S. (2004) **Synthesis, biophysical and biological evaluation of 3,6-bis-amidoacridines with extended 9-anilino substituents as potent G-quadruplex-binding telomerase inhibitors**. *Bioorganic & Medicinal Chemistry Letters*, **14**(16): 4347-4351.

Shay, J.W. & Bacchetti, S. (1997) **A survey of telomerase activity in human cancer**. *European Journal of Cancer*, **33**(5): 787-791.

Shay, J.W. & Wright, W.E. (2005) **Senescence and immortalization: role of telomeres and telomerase**. *Carcinogenesis*, **26**(5): 867-874.

Shay, J.W., Pereira-Smith, O.M. & Rainey, W.E. (1991) **A role for both RB and p53 in the regulation of human cellular senescence**. *Experimental Cell Research*, **196**(1): 33-39.

Shay, J.W., Van Der Haegen, B.A., Ying, Y. & Wright, W.E. (1993) **The frequency of immortalization of human fibroblasts and mammary epithelial cells transfected with SV40 large T-antigen**. *Experimental Cell Research*, **209**(1): 45-52.

Shin-ya, K., Wierzba, K., Matsuo, K., Ohtani, T., Yamada, Y., Furihata, K., Hayakawa, Y. & Seto, H. (2001) **Telomestatin, a novel telomerase inhibitor from Streptomyces anulatus**. *Journal of the American Chemical Society*, **123**(6): 1262-1263.

Shiue, L., Myers, R.M., Cox, D.R., Naylor, S.L., Killery, A.M. & Varmus, H.E. (1990) **Structure and variability of human chromosome ends**. *Molecular and Cellular Biology*, **10**(2): 518-527.

Smogorzewska, A. & de Lange, T. (2002) **Different telomere damage signaling pathways in human and mouse cells**. *The EMBO Journal*, **21**(16): 4338-4348.

Smogorzewska, A., van Steensel, B., Bianchi, A., Oelmann, S., Schaefer, M.R., Schnapp, G. & de Lange, T. (2000) **Control of human telomere length by TRF1 and TRF2**. *Molecular and Cellular Biology*, **20**(5): 1659-1668.

Spačkova, N., Berger, I. & Sponer, J. (1999) **Nanosecond Molecular Dynamics Simulations of Parallel and Antiparallel Guanine Quadruplex DNA Molecules**. *Journal of the American Chemical Society*, **121**(23): 5519-5534.

Spacková, N., Berger, I. & Sponer, J. (2001) **Structural dynamics and cation interactions of DNA quadruplex molecules containing mixed guanine/cytosine quartets revealed by large-scale MD simulations**. *Journal of the American Chemical Society*, **123**(14): 3295-3307.

Sponer, J. & Spacková, N. (2007) **Molecular dynamics simulations and their application to four-stranded DNA**. *Methods (San Diego, Calif.)*, **43**(4): 278-290.

Starling, J.A., Maule, J., Hastie, N.D. & Allshire, R.C. (1990) **Extensive telomere repeat arrays in mouse are hypervariable**. *Nucleic Acids Research*, **18**(23): 6881-6888.

Sun, D. & Hurley, L.H. (2009) **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**. *Journal of Medicinal Chemistry*, **52**(9): 2863-2874.

Sun, D., Thompson, B., Cathers, B.E., Salazar, M., Kerwin, S.M., Trent, J.O., Jenkins, T.C., Neidle, S. & Hurley, L.H. (1997) **Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound**. *Journal of Medicinal Chemistry*, **40**(14): 2113-2116.

Sun, H., Karow, J.K., Hickson, I.D. & Maizels, N. (1998) **The Bloom's syndrome helicase unwinds G4 DNA**. *The Journal of Biological Chemistry*, **273**(42): 27587-27592.

Taetz, S., Baldes, C., Murdter, T.E., Kleideiter, E., Piotrowska, K., Bock, U., Haltner-Ukomadu, E., Mueller, J., Huwer, H., Schaefer, U.F., Klotz, U. & Lehr, C-M. (2006) **Biopharmaceutical characterization of the telomerase inhibitor BRACO19**. *Pharmaceutical Research*, **23**(5): 1031-1037.

Taetz, S., Murdter, T.E., Zapp, J., Boettcher, S., Baldes, C., Kleideiter, E., Piotrowska, K., Schaefer, U.F., Klotz, U., Lehr, C-M. (2008) **Decomposition of the telomere-targeting agent BRACO19 in physiological media results in products with decreased inhibitory potential**. *International Journal of Pharmaceutics*, **357**(1-2): 6-14.

Tan, J., Gu, L. & Wu, J. (2008) **Design of selective G-quadruplex ligands as potential anticancer agents**. *Mini Reviews in Medicinal Chemistry*, **8**(11): 1163-1178.

Taylor, R. & Kennard, O. (1982) **Molecular structures of nucleosides and nucleotides. 2. Orthogonal coordinates for standard nucleic acid base residues**. *Journal of the American Chemical Society*, **104**(11): 3209-3212.

Todd, A.K. & Neidle, S. (2008) **The relationship of potential G-quadruplex sequences in cis-upstream regions of the human genome to SP1-binding elements**. *Nucleic Acids Research*, **36**(8): 2700-2704.

Todd, A.K., Adams, A., Thorpe, J.H., Denny, W.A., Wakelin, L.P. & Cardin, C.J. (1999) **Major groove binding and 'DNA-induced' fit in the intercalation of a derivative of the mixed topoisomerase I/II poison N-(2-(dimethylamino)ethyl)acridine-4-carboxamide (DACA) into DNA: X-ray structure complexed to d(CG(5-BrU)ACG)2 at 1.3-A resolution**. *Journal of Medicinal Chemistry*, **42**(4): 536-540.

Todd, A.K., Johnston, M. & Neidle, S. (2005) **Highly prevalent putative quadruplex sequence motifs in human DNA**. *Nucleic Acids Research*, **33**(9): 2901-2907.

Tripos, Inc., St. Louis, USA.

Tsui, V. & Case, D.A. (2000) **Theory and applications of the generalized Born solvation model in macromolecular simulations**. *Biopolymers*, **56**(4): 275-291.

Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G.N. (2004) **REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use**. *Acta Crystallographica. Section D, Biological Crystallography*, **60**(Pt 12 Pt 1): 2184-2195.

Valente, A.P., Miyamoto, C.A. & Almeida, F.C.L. (2006) **Implications of protein conformational diversity for binding and development of new biological active compounds**. *Current Medicinal Chemistry*, **13**(30): 3697-3703.

van Gunsteren, W.F., Bakowies, D., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D.P., Glatti, A., Hunenberger, P.H., Kastenholz, M.A., Oostenbrink, C., Schenk, M., Trzeniak, D., van der Vegt, N.F.A., Yu, H.B. (2006) **Biomolecular modeling: Goals, problems, perspectives**. *Angewandte Chemie (International Ed. in English)*, **45**(25): 4064-4092.

van Mourik, T. & Dingley, A.J. (2005) **Characterization of the monovalent ion position and hydrogen-bond network in guanine quartets by DFT calculations of NMR parameters**. *Chemistry*, **11**(20): 6064-6079.

van Steensel, B., Smogorzewska, A. & de Lange, T. (1998) **TRF2 protects human telomeres from end-to-end fusions**. *Cell*, **92**(3): 401-413.

Verdun, R.E. & Karlseder, J. (2007) **Replication and protection of telomeres**. *Nature*, **447**(7147): 924-931.

Verma, A., Yadav, V.K., Basundra, R., Kumar, A. & Chowdhury, S. (2009) **Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells**. *Nucleic Acids Research*, **37**(13): 4194-4204.

Wang, J., Cieplak, P. & Kollman, P. (2000) **How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?** *Journal of Computational Chemistry*, **21**(12): 1049-1074.

Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. & Case, D.A. (2004) **Development and testing of a general amber force field**. *Journal of Computational Chemistry*, **25**(9): 1157-1174.

Wang, Y. & Patel, D.J. (1992) **Guanine residues in d(T2AG3) and d(T2G4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution**. *Biochemistry*, **31**(35): 8112-8119.

Wang, Y. & Patel, D.J. (1993) **Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex**. *Structure*, **1**(4): 263-282.

Wellinger, R.J., Ethier, K., Labrecque, P. & Zakian, V.A. (1996) **Evidence for a New Step in Telomere Maintenance**. *Cell*, **85**(3): 423-433.

Wheelhouse, R.T., Sun, D., Han, H., Han, F.X. & Hurley, L.H. (1998) **Cationic Porphyrins as Telomerase Inhibitors: the Interaction of Tetra-(N-methyl-4-pyridyl)porphine with Quadruplex DNA**. *Journal of the American Chemical Society*, **120**(13): 3261-3262.

Wright, W.E., Piatyszek, M.A., Rainey, W.E., Byrd, W. & Shay, J.W. (1996) **Telomerase activity in human germline and embryonic tissues and cells**. *Developmental Genetics*, **18**(2): 173-179.

Wright, W.E., Tesmer, V.M., Huffman, K.E., Levene, S.D. & Shay, J.W. (1997) **Normal human chromosomes have long G-rich telomeric overhangs at one end**. *Genes & Development*, **11**(21): 2801-2809.

Xu, Y., Kaminaga, K. & Komiyama, M. (2008) **G-quadruplex formation by human telomeric repeats-containing RNA in Na^+ solution**. *Journal of the American Chemical Society*, **130**(33): 11179-11184.

Xue, Y., Kan, Z., Wang, Q., Yao, Y., Liu, J., Hao, Y. & Tan, Z. (2007) **Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K^+ solution under molecular crowding condition**. *Journal of the American Chemical Society*, **129**(36): 11185-11191.

Yanez, G.H., Khan, S.J., Locovei, A.M., Pedroso, I.M. & Fletcher, T.M. (2005) **DNA structure-dependent recruitment of telomeric proteins to single-stranded/double-stranded DNA junctions**. *Biochemical and Biophysical Research Communications*, **328**(1): 49-56.

Yui, J., Chiu, C.P. & Lansdorp, P.M. (1998) **Telomerase activity in candidate stem cells from fetal liver and adult bone marrow**. *Blood*, **91**(9): 3255-3262.

Zahler, A.M. & Prescott, D.M. (1988) **Telomere terminal transferase activity in the hypotrichous ciliate Oxytricha nova and a model for replication of the ends of linear DNA molecules**. *Nucleic Acids Research*, **16**(14B): 6953-6972.

Zahler, A.M., Williamson, J.R., Cech, T.R. & Prescott, D.M. (1991) **Inhibition of telomerase by G-quartet DNA structures**. *Nature*, **350**(6320): 718-720.

Zaug, A.J., Podell, E.R. & Cech, T.R. (2005) **Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro**. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(31): 10864-10869.

Zhang, N., Phan, A.T. & Patel, D.J. (2005) **(3 + 1) Assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex**. *Journal of the American Chemical Society*, **127**(49): 17277-17285.

Zhou, J., Wei, C., Jia, G., Wang, X., Feng, Z. & Li, C. (2009) **Human telomeric G-quadruplex formed from duplex under near physiological conditions: spectroscopic evidence and kinetics**. *Biochimie*, **91**(9): 1104-1111.

Zimmermann, S. & Martens, U.M. (2007) **Telomeres and telomerase as targets for cancer therapy**. *Cellular and Molecular Life Sciences*, **64**(7-8): 906-921.

PUBLICATIONS (2007 - 2009)

Available online at www.sciencedirect.com
ScienceDirect
METHODS
Methods 43 (2007) 252–263
www.elsevier.com/locate/ymeth

Crystallographic studies of quadruplex nucleic acids

Nancy H. Campbell, Gary N. Parkinson *

Research UK Biomolecular Structure Group, The School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX, UK

Accepted 7 August 2007

ract

NA quadruplexes are formed from guanine-rich repeats that self-associate into higher order four-stranded structures. These G-rich sequences can be found in both telomeric regions as well as regions proximal to promoters of oncogenes. The compelling evidence abilizing these motifs by small molecule ligands can alter cell viability in certain cancer cell lines has led to identification of DNA uplex structures as therapeutic targets. Target-based design of selective ligands that target particular quadruplex topologies is heavant on the availability of high-resolution structural information of the intended target. X-ray crystallography can provide this level ail to atomic resolution. Recently drug discovery programs have refocused on the need for a fuller structural and molecular descripf the target molecule. This review describes a crystallographic route to the determination of quadruplex topology, and high-resoloop structures for target-based ligand design. The review also highlights the methods employed in the design of appropriate DNA nces and crystallization techniques to solve these unusual DNA structures.

rds: Quadruplex; Crystallization; Structure solution; Structure determination; X-ray crystallography; Nucleic acids

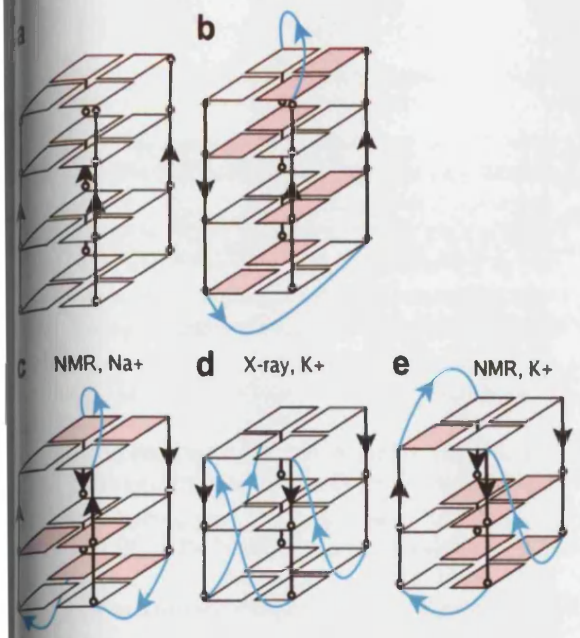
roduction

uctural studies of quadruplex nucleic acid were initiwith the discovery that the synthetic poly G oligonude forms a fourfold right-handed helix, using X-ray diffraction methods to define helical geometry. Finetural detail, including unambiguous determination of rrangement in the G-tetrad motif, requires the use of -crystal crystallography. Structural analyses applied ort-length oligonucleotides have subsequently pro- detailed information on quadruplexes. Crystallotic techniques have played a pivotal role, in tandem NMR methods, in advancing our knowledge of the range of topologies and detailed geometries available ese molecules. Such studies are assuming increasing rtance and attracting more general interest as the role uadruplex nucleic acids in biology becomes more ent and as their potential as therapeutic targets es realised. There is a natural tendency to attempt to establish common structural principles and folding rules; however, the current small size of the quadruplex structural data suggests that this is premature, not least because the topological variability that is a consequence of the diversity of quadruplex sequence (and sequence context), has barely been explored as yet. It is also apparent that more crystal structures need to be determined particularly as the human telomeric sequence has revealed several topological isoforms in solution that are different to the topology observed in the crystalline state (Fig. 1).

This chapter describes the methodology for obtaining crystal structures of quadruplexes and their ligand complexes. Crystallization protocols are discussed in detail since obtaining viable crystals remains the rate-determining step in all structural projects. These necessarily focus on the conditions that have to date been successful, and emphasise their variety; every quadruplex presents distinct surface characteristics to its environment, and it is therefore essential to explore a wide range of conditions if success is to be achieved. The route chosen for the collection of crystallographic data depends on the quality of the crystals, and on the methods to be used for structure

esponding author. Fax: +44 (0) 20 7753 5970.
ail address: gary.parkinson@pharmacy.ac.uk (G.N. Parkinson).

023/S - see front matter
1016/j.ymeth.2007.08.005



1. A schematic diagram showing the loop direction and polarity sting in both solution and solid state for (a) tetramolecular TGGGGT) and (b) bimolecular *Oxytricha nova* d(GGGGTTTTG GG), and of different forms of the unimolecular human telomere ence in solution (c) + (e) and in the crystal (d). Loops in blue. Base entation; white is *anti* and pink is *syn*. (For interpretation of the rences to colour in this figure legend, the reader is referred to the web sion of this article.)

lutions. The wide-spread availability of synchrotron am lines with multiple-wavelength anomalous diffraction (AD) capability has long been of a great advantage for cleic acid crystallographers as anomalously scattering mines substituted at the 5′ methyl of thymine provide fficient substitution sites and anomalous data for struc- re solution. However, the increasing availability of stable enium derivatives, which have a larger anomalous dis- rsion signal, enables effective phasing to be achieved with house data collection facilities. We will discuss here the ethods used to crystallize and solve these novel DNA uctures, bringing together the experience from our own oratory and information from other research groups.

Crystallization of native G-quadruplexes

Sample preparation

Starting DNA samples for both crystallographic and MR techniques should be as pure as possible. The sample usually dialysed against double distilled water or an propriate buffer for cartridge purified oligonucleotides, ile HPLC purified oligonucleotides are suitable for use thout additional purification. The sample should then dried to a powder for storage. A standard protocol material preparation commonly utilized in our labora- y is provided below. DNA samples are first dissolved o ultra-pure (type 1) water to a concentration of no more than 5 mM, verified by UV spectrographic methods. Appropriate buffer and salts are then added bringing the DNA to a concentration of typically 3 mM for annealing. The annealing of guanine-rich DNA at high concentrations can result in the formation of gel like solutions due to the random association of the strands via the hydrogen bonding between the abundant guanine nucleotides. Annealing buffer solutions typically comprise of a cacodylate salt, buffered to pH 6.5 at around 20 mM with the addition of monovalent cations such as sodium or potassium ions at concentrations between 30 and 100 mM to allow quadruplex formation. Before annealing the DNA is typically heated to 90 °C for 5–15 min to ensure the disassociation of all the strands. The samples are then allowed to cool overnight to room temperature in a covered heating block, allowing the gradual ordered re-association of the strands. Folding of the DNA into a single species can also be achieved by dialysis and the use of appropriate buffering conditions. This method has the advantage of incorporating an additional purification step. Typically, during dialysis the DNA is equilibrated over several days to form the quadruplex motif. The samples can then be diluted to a final concentration of around 2 mM prior to crystallization trials or NMR analysis.

2.2. Buffers, precipitating agents and additives

The formulations used for crystallization of biological samples are usually divided into three classes, buffering agents, precipitating agents and additives; this distinction can also applied to quadruplex crystallization. The choice of components used in each of these classes are discussed in detail with specific relevance to the stabilization of quadruplexes.

Buffering agents and pH are not usually critical factors in DNA crystallization, being ideally at neutral to slightly acidic pH 6.5–7.0. However the selection of buffering agent may sometimes have an influence on sample crystallization. Typically we use cacodylic acid, at about 20 mM and pH 6.5 either as the sodium or potassium salt. Cacodylic acid also allows for long-term storage of the DNA, pre-annealed and ready for crystallization. Phosphate buffers are usually avoided due to the tendency for phosphate salts to crystallize from the components used in the crystallization screens.

A limited range of precipitating agents has so far been employed in the growth of quadruplex crystals. Those agents that have led to successful X-ray structure determination are; MPD (62% of those quadruplexes crystallized), polyethylene glycols (PEG) with molecular weights between 400 and 4000 (29%), and ammonium sulphate (9%). This limited set is in contrast to the wider range of conditions utilized for duplex DNA crystallizations, which include 2-propanol, polyethylene glycol monomethyl ether (PEG MME), and 1,6-hexanediol.

As additives, cations will always be included in the crystallization mix as they are required for the formation and

ization of the quadruplexes. Their role is critical as can modulate the final folded topology as observed man telomeric sequences [1,2]. Monovalent metal cat- such as sodium (45%), potassium (50%) are normally ded in the annealing step and supplemented in the llization buffer to maintain the initial annealing con- ation during crystallization. Other monovalent metal such as lithium (9%), and divalent metal ions such gnesium (32%) and calcium (14%), are common addi- n crystallization buffers for quadruplexes. The choice ion is important as it can both stabilize and destabi- he folded quadruplex. The affinity of monovalent izing cations in the central channel is $NH_4^+ > Rb^+ > Na^+ > Cs^+ > Li^+$. Therefore the e of ammonium sulphate as a precipitating agent modify a quadruplex annealed in Na^+. Several diva- ations actively destabilize quadruplexes as shown in $Zn^{2+} > Co^{2+} > Mn^{2+} > Mg^{2+} > Ca^{2+}$ with the last ommonly used in quadruplex DNA crystallography. divalent cations play an important role in linking the ed phosphate backbones together in the lattice envi- ent. Cobalt is commonly used in DNA duplex crystal- phy but should most likely be avoided in quadruplex llization solutions. The most common non-salt addi- s the polyamine spermine (59%), successfully utilized th duplex and quadruplex crystallization.

Crystallization strategies: crystal screening and isation

ere are several strategies that the crystallographer can from when attempting to crystallize nucleic acids. commercially available crystallization kits are on some specifically designed for this purpose. However, orth bearing in mind that these kits were designed, in irst instance for crystallization of duplex DNA or , both of which demonstrate resilience to certain addi- even at relatively high concentration. On the other G-quadruplexes react completely differently in the conditions; by crashing out of the solution and form- dark heavy precipitate—even at low concentrations. ent cations such as magnesium, calcium, zinc and , invariably cause G-quadruplexes to come out of on. Magnesium, especially, is present in many of ecifically designed commercially available kits at rel- y high concentrations making these kits unsuitable itial screening purposes. However, one can still chose mponents of these kits that are more suitable i.e. the that contain only monovalent cations such as sodium otassium or perhaps mixed with very low concentra- of divalent cations, in the 1–5 mM range.

verse screening [3] is another strategy whereby repre- tives of the various families of crystallizing agents, ols, glycols, diols, monovalent and divalent salts, vestigated against a reasonable concentration of the . Initial screening to find a suitable range of concen- n is carried out as described in T. Bergfors book "Protein Crystallization: Techniques, Strategies and Tips" [4]. A 5 μl drop of the DNA at 0.5–2 mM is pipetted onto a siliconized coverslip. Then 0.1–0.2 μl of high concentration crystallizing agent is added, the drop is covered and the result recorded after 20–30 s. If no precipitate is formed, keep adding the crystallizing agent as described until a change occurs. The concentration of the DNA and that of the crystallizing agent is calculated at the point where the drop last remained clear and the reverse screening experiments are conducted around the calculated values of concentration. Some complexes of quadruplexes are very sensitive and initial screening should be carried out at a low range of DNA concentration i.e. 0.5–1.0 mM. Screening around the precipitation point is sensible since the crystals are formed near this point. Precipitation can be achieved either by increasing the concentration of the DNA or by increasing the concentration of the crystallizing agent. However, the higher the concentration of DNA at super-saturation, the more likely nucleation will take place; a necessary prerequisite for crystals to form. When setting up screening trays, crystallizing agents of various families should and can be mixed to induce crystal formation.

Dedicated, focused in-house screens provided us with a reliable method for rapidly identifying potential lead conditions for subsequent crystallization trials. In designing an in-house sparse matrix crystallization screen it should contain a selected set of precipitating agents that contain a range of concentration to allow for saturation, super-saturation and precipitation of the DNA to occur during the period of drop equilibration. Equilibrium of the drops can take between 2 weeks and 5 weeks depending upon the precipitating agent used. An example of such a screen used in our laboratory in shown in Table 1. Quadruplex crystals grown in these screens can appear over several days to several months. It is important to note that when using these screens a range of annealing conditions should be employed to ensure that the crystallization trials are not biased by the annealing buffer. Annealing should then be carried out in several different salts to ensure diversity in the sparse matrix screening. All screening trials would ideally incorporate at least two crystallization temperatures. We have found that 22 and 10 °C provide sufficient diversity. In addition high salts concentrations of up to 300 mM seem advantageous (Table 2).

The two most commonly exploited and popular crystallization techniques, sitting and hanging drops utilizing vapour diffusion, have been successfully used for the vast majority of the quadruplex DNA crystallizations. Batch method crystallizations using under-oil techniques have also yielded in a few cases suitable crystals for X-ray diffraction studies. From the limited set of crystallizations performed for quadruplexes, shown in Table 3 we find that typical DNA in-drop concentrations range from 0.5 to 2 mM. Interestingly the concentration used appears to be independent of the number of guanine repeats used to form either inter- or intra- molecular quadruplexes. Crystallization drop volumes used are between 1 and 2 μl, which are typical

ample of a sparse matrix quadruplex crystallization screen

	Salt	Buffer	Precipitant
	0.1 M KCl	0.05 M K Caco pH 6.5	20% v/v MPD
	0.3 M KCl	0.05 M K Caco pH 6.5	10% v/v MPD
	0.1 M KF, 0.025 $MgSO_4$	0.05 M K Caco pH 6.5	25% v/v MPD
	0.2 M KI	0.05 M K Caco pH 6.5	25% v/v MPD
	0.1 M KF, 0.02 M $(NH_4)_2SO_4$	0.05 M Na Caco pH 6.5	15% v/v MPD
	0.3 M NaCl	0.05 M Na Caco pH 6.5	30% v/v MPD
	0.1 M NaI	0.05 M Na Caco pH 6.5	35% v/v MPD
	0.3 M KF, 0.02 M $MgSO_4$	0.05 M K Caco pH 6.5	35% v/v MPD
	0.15 M NaF, 0.15 M KF	0.05 M Na Caco pH 6.5	40% v/v MPD
	0.1 M KCl, 0.1 M $(NH_4)_2SO_4$	0.05 M K Caco pH 6.5	20% v/v MPD
	0.1 M NaI, 0.1 M NaF, 0.1 M NaCl	0.05 M Na Caco pH 6.5	30% v/v MPD
	0.2 M KCl, 0.05 M KI	0.05 M K Caco pH 6.5	25% v/v PEG400
	0.1 M KCl, 0.05 M $MgCl_2$	0.05 M K Caco pH 6.5	10% v/v PEG400
	0.1 M Li_2SO_4	0.05 M K Caco pH 6.5	20% v/v PEG400
	0.3 M KI	0.05 M K Caco pH 6.5	15% v/v PEG400
	0.3 M NaI	0.05 M Na Caco pH 6.5	15% v/v PEG400
	0.1 M KCl, 0.03 M $CaCl_2$	0.05 M K Caco pH 6.5	1.3 M ammonium sulphate
	0.2 M KCl, 0.05 M Li_2SO_4	0.05 M K Caco pH 6.5	1.8 M ammonium sulphate
	0.05 M Li_2SO_4, 0.05 M $MgSO_4$	0.05 M K Caco pH 6.5	20% v/v *iso*-propanol
	0.2 M NaCl	0.05 M Na Caco pH 6.5	20% JEFF 2000
	0.2 M NaCl, 0.1 M NaF	0.05 M Na Caco pH 6.5	10% JEFF 2000
	0.05 M KI, 0.05 M $MgCl_2$	0.05 M K Caco pH 6.5	15% JEFF 2000
	0.1 M NaCl, 0.05 M $MgCl_2$	0.05 M Na Caco pH 6.5	1 M ammonium sulphate
	0.05 M Li_2SO_4	0.05 M Na Caco pH 6.5	1 M ammonium sulphate

macromolecular crystallizations. Highly automated otic systems are now becoming more widely employed ncing the time and material required for crystal screen- The reduced drop volumes, typically a 100 nl drop size, nbined with a selected range of commercial screens allow large scale screening to identify potential crystallization ditions.

Recording and collating experimental crystallization a are a major task and identifying trends become critical btaining diffraction quality crystals. With enough sam- even sparse matrix sampling provides sufficient data correlating the effects of various additives and salts.

Sequence modification

Choosing the most appropriate oligonucleotide ence, while still preserving the topology and structure interest is one of the most challenging aspects of druplex crystallization. In protein crystallization exper- nts a common technique employed is to vary the N- and erminal regions while retaining the core region of inter- This approach can also be successfully employed in A crystallization but must be applied with caution.

It is well understood that short G-rich sequences associ- g together as intermolecular quadruplexes provide the st stable and crystallizable platform for the study of etrads, however these structures only provide informa- limited to the G-quadruplex core motif. It is the link- nucleotides between the G-runs that are essential in viding the correct topological form and specific ligand ding surfaces. The length and nucleotide sequence of se connecting loops, define strand alignment, topology, and importantly ligand selectivity between quadruplexes. It is therefore critical that modifications to intervening loop sequences do not affect topology or ligand binding surfaces by creating new nucleotide associations.

In the human telomeric sequences the conformational heterogeneity observed is not only linked to ionic conditions and hydration, as observed in duplex DNA, but is also dependent upon the modification of the terminal 5′ and 3′ sequences positioned external to the central G-quartet core. This can be seen with the addition of adenine nucleotides where A-triads can form and stabilize new loop arrangements on top of the guanine quartets. Even subtle modification within the loops regions such as modifying thymines to uracils, or the removal of a single methyl group has been shown to have profound effects on the quadruplex topology in solution [5] resulting in the switching from parallel to anti-parallel conformations of the bimolecular quadruplexes. Thus care needs to be taken when adapting the nucleotide sequence as certain modifications can result in the stabilization of one structural topology over other folded topologies in solution and during crystallization.

With care, modifications such as the simple addition of nucleotides at the 5′ and 3′ ends to the central G-tetrad core motif can be used to enhance the association between quadruplexes and in doing so stabilize packing interactions. Modification of the native sequence can also be used to enhance stability of the quadruplex during crystallization. NMR is a valuable tool in those cases particularly where clean and identifiable spectra can be recorded. Mutations can be followed by rapid 1D NMR data collection. Spectra in the imino regions can then compared to the original 1D NMR and if identical it is a strong indication

ID				(Å)			
1D59	GGGGTTTTGGGG	Crystal structure of four-stranded *Oxytricha* telomeric DNA	P 2_1 2_1 2_1: $a = 27.73$, $b = 49.57$, $c = 97.27$	2.300	n/a	0.1950	[26]
244D	TGGGGT	The high-resolution crystal structure of a parallel-stranded guanine tetraplex	P 1: $a = 28.75$, $b = 35.47$, $c = 56.77$: $\alpha = 74.39$, $\beta = 77.64$, $\gamma = 89.73$: $\alpha = 74.29$, $\beta = 77.68$, $\gamma = 89.73$	1.200	0.1760	0.1240	[27]
352D	TGGGGT	The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution	P 1: $a = 28.28$, $b = 34.78$, $c = 56.23$: $\alpha = 74.31$, $\beta = 77.68$, $\gamma = 89.81$	0.950	n/a	0.1520	[28]
1OTC	GGGGTTTTGGGG	Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA	P 6_1 2 2: $a = b = 94.50$, $c = 426.2$	2.800	0.2920	0.2510	[29]
1J8G	UGGGGU	X-ray analysis of an RNA tetraplex (UGGGGU)(4) with divalent Sr (2^+) ions at subatomic resolution (0.61)	P 4 2_1 2: $a = b = 36.15$	0.610	0.112	0.103	[21]
1JB7	GGGGTTTTGGGG	DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein–DNA complex	P 6_1 2 2: $a = b = 93.10$, $c = 421.8$	1.860	0.2460	0.2320	[8]
1JPQ	GGGGUTTTGGGG	Crystal structure of the potassium form of an *Oxytricha nova* G-quadruplex	P 3_2 2 1: $a = b = 27.54$, $c = 145.82$	1.600	0.2759	0.2293	[9]
1JRN	GGGGTTTTGGGG	Crystal structure of the potassium form of an *Oxytricha nova* G-quadruplex	P 2_1 2_1 2_1: $a = 26.51$, $b = 47.44$, $c = 96.46$	2.000	0.2759	0.2624	[9]
1K8P	UAGGGUTAGGGT	Crystal structure of parallel quadruplexes from human telomeric DNA	P 3_1 2 1: $a = 56.61$, $b = 56.607$, $c = 40.546$	2.400	0.2800	0.1930	[2]
1KF1	AGGGTTAGGGTTA GGGTTAGGG	Crystal structure of parallel quadruplexes from human telomeric DNA	P 6: $a = 56.68$, $b = 56.682$, $c = 42.106$	2.100	0.2630	0.2310	[2]
1L1H	GGGGTTTTGGGG	Structure of a G-quadruplex–ligand complex	P 2_1 2_1 2: $a = 55.45$, $b = 42.736$, $c = 26.926$	1.750	0.2187	0.1438	[10]
1O0K	TGGGGT	Structure of the first parallel DNA quadruplex–drug complex	C 2 (C 1 2 1): $a = 53.00$, $b = 47.073$, $c = 31.865$: $\beta = 119.80$	1.170	Not reported	Not reported	[20]
1V3N	G(Br5C)GAGAGC	Crystal Structures of a DNA octaplex with I-motif of G-quartets and its splitting into two quadruplexes suggest a folding mechanism of eight tandem repeats	I 2 2 2: $a = 33.81$, $b = 43.47$, $c = 64.2$	1.800	0.2490	0.2250	[30]
1V3O	G(I5C)GAGAGC	Crystal Structures of a DNA octaplex with I-motif of G-quartets and its splitting into two quadruplexes suggest a folding mechanism of eight tandem repeats	I 2 2 2: $a = 34.65$, $b = 42.49$, $c = 64.08$	1.700	0.2960	0.2620	[30]
1V3P	G(I5C)GAGAGC	Crystal Structures of a DNA octaplex with I-motif of G-quartets and its splitting into two quadruplexes suggest a folding mechanism of eight tandem repeats	I 2 2 2: $a = 36.18$, $b = 36.38$, $c = 63.12$	2.300	0.2650	0.2340	[30]
1S45	TGGGGT	A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tl^{+1}/Na^{+1} ions	P 1: $a = 28.26$, $b = 35.41$, $c = 32.057$: $\alpha = 83.74$, $\beta = 61.78$, $\gamma = 76.68$	2.200	0.2390	0.1730	[24]
1S47	TGGGGT	A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tl^{+1}/Na^{+1} ions	P 2_1 (P 1 2_1 1): $a = 28.32$, $b = 56.341$, $c = 51.746$: $\beta = 99.17$	2.500	0.3110	0.2240	[24]
2AVH	GGGGTTTGGGG	Topology variation and loop structural homology in crystal and simulated structures of a bimolecular DNA quadruplex	C 2 2 2: $a = 29.93$, $b = 37.11$, $c = 43.65$	1.500	0.2080	0.1990	[31]
2AVJ	GGGGUTTGGGG	Topology variation and loop structural homology in crystal and simulated structures of a bimolecular DNA quadruplex	P 2_1 (P 1 2_1 1): $a = 31.68$, $b = 33.309$, $c = 79.234$, $\beta = 91.49$	2.390	0.2420	0.1720	[31]
2GRB	UGIGGU	Crystal structure of an RNA quadruplex containing inosine tetrad: implications for the roles of NH_2 group in purine tetrads	P 2_1 (P 1 2_1 1): $a = 29.06$, $b = 52.738$, $c = 37.344$, $\beta = 102.98$	1.400	0.2500	0.1978	[32]
2HBN	GGGGTTTTGGGG	Crystallization and characterization of the thallium form of the *Oxytricha nova* G-quadruplex	P 2_1 2_1 2_1: $a = 27.38$, $b = 48.21$, $c = 96.198$	1.550	0.2480	0.2260	[33]
2HRI	TAGGGTTAGGG	Structural basis for binding of porphyrin to human telomeres	C 2 2 2_1: $a = 37.29$, $b = 61.99$, $c = 61.40$	2.090	0.2570	0.2080	[34]
2O4F	TGGGGT	Structure of a d(TGGGGT) quadruplex crystallized in the presence of Li[$^+$] ions	P 2_1 (P 1 2_1 1)	1.5	0.257	0.206	[35]

[a] Protein data bank.

Initial in-drop and well crystallization conditions

PDB ID	Initial in-drop conditions					Well conditions	pH	K	Crystallization	Ref.
	DNA	Salt	Additives	Precipitant	Buffer					
1D59	1 mM	10 mM $MgCl_2$ 40 mM KCl	6 mM spermine	5% MPD	20 mM K cacodylate (pH 7.0)	40%MPD	7.0	Not reported	Vapour diffusion/ hanging drop	[26]
244D	1 mM	12 mM $CaCl_2$ 130–180 mM NaCl	6 mM spermine	5% v/v MPD	20 mM Na cacodylate (pH 6.0)	120 mM Na cacodylate (pH 6.6), 70 mM $CaCl_2$,700 mM to 1 M NaCl, 26–32% MPD	6.6	277.00	Vapour diffusion/ hanging drop	[27]
352D	1 mM	12 mM $CaCl_2$ 130–180 mM NaCl	6 mM spermine	5% v/v MPD	20 mM Na cacodylate (pH 6.6)	120 mM Na cacodylate (pH 6.6),70 mM $CaCl_2$, 700 mM to 1 M NaCl, 26–32% MPD	6.6	278.00	Vapour diffusion/ hanging drop	[28]
1OTC	100–150 μM	25 mM NaCl	0.05 mM EDTA, 0.02% NaN_3, 2 mM DTT	15% ethylene glycol 1.5–4% PEG 4000	2.5 mM Tris (pH 7.5), 20 mM MES (pH 5.5–6.5)	30% ethylene glycol, 3–8% PEG 4000, 40 mM MES (pH 5.5–6.5), 0.02% NaN_3, 2 mM DTT, 50 mM NaCl	6.0	293.00–298.00	Vapour diffusion/ hanging drop	[29]
1J8G	Not reported	20 mM $MgCl_2$, 80 mM LiCl, 40 mM SrCl, 20 mM $CaCl_2$	12 mM spermine	10% v/v MPD	40 mM Na cacodylate (pH 7.0)	35% v/v MPD	7.0	293.00	Vapour diffusion/ hanging drop	[21]
1JB7	100–150 μM	25 mM NaCl	0.05 mM EDTA, 0.02% NaN_3, 2 mM DTT	15% ethylene glycol 1.5–4% PEG 4000	2.5 mM Tris (pH 7.5), 20 mM MES (pH 5.5–6.5)	30% ethylene glycol, 3–8% PEG 4000, 40 mM MES (pH 5.5–6.5), 0.02% NaN_3, 2 mM DTT, 50 mM NaCl	6.0	282.00	Vapour diffusion/ hanging drop	[8]
1JPQ	1 mM	10 mM $MgCl_2$ 40 mM KCl	3.3 mM spermine	5.0% v/v MPD	K cacodylate used in annealing the oligo	25% MPD (v/v)	7.0	285.00	Vapour diffusion/ hanging drop	[9]
1JRN	1 mM	10 mM $MgCl_2$ 40 mM KCl	4.1 mM spermine	5.0% v/v MPD	K cacodylate used in annealing the oligo	35.0% MPD (v/v)	7.0	285.00	Vapour diffusion/ hanging drop	[9]
1K8P	0.5 mM	50 mM NaCl,50 mM KCl, 50 mM Li_2SO_4	—	500 mM $(NH_4)_2SO_4$	50 mM K cacodylate (pH 6.5)	1M $(NH_4)_2SO_4$	7.0	286.00	Vapour diffusion/ hanging drop	[2]
1KF1	1.7 mM	300 mM KI	2 mM BRACO19[a]	15% v/v PEG400	50 mM K cacodylate (pH 6.5)	45% v/v PEG400	6.5	285.00	Vapour diffusion/ hanging drop	[2]
1L1H	0.5 mM	10 mM $MgCl_2$ 40 mM KCl	1 mM BSU 6039[a], 1.6 mM spermine	5% w/v MPD	20 mM K cacodylate (pH 7.0)	30% MPD in 20 mM KCl	7.0	285.00	Vapour diffusion/ hanging drop	[10]
1O0K	Not reported	$CaCl_2$	Spermine	MPD	Cacodylate buffer	Not reported	7.0	278.00	Vapour diffusion/ sitting drop	[20]
1V3N	0.75 mM	6 mM NaCl 40 mM KCl	6 mM spermine	5% v/v MPD	20 mM Na cacodylate (pH 7.0)	40 mM Na cacodylate (pH 7.0), 12 mM spermine, 12 mM NaCl, 80 mM KCl, 10% (v/v) MPD	7.0	277.00	Vapour diffusion/ hanging drop	[30]
1V3O	0.75 mM	6 mM NaCl 40 mM KCl	6 mM spermine	5% MPD	20 mM Na cacodylate (pH 7.0)	41 mM Na cacodylate (pH 7.0), 12 mM spermine, 12 mM NaCl, 80mMKCl, 10% (v/v) MPD	7.0	227.00	Vapour diffusion/ hanging drop	[30]
1V3P	1.75 mM	6 mM NaCl 40 mM KCl	6 mM spermine	5% MPD	20 mM K cacodylate (pH 7.0)	42 mM Na cacodylate (pH 7.0), 12 mM spermine, 12 mM NaCl, 80 mM KCl, 10% (v/v) MPD	7.0	277.00	Vapour diffusion/ hanging drop	[30]

(*continued on next page*)

	DNA	Salt	Additives	Precipitant	Buffer					
1S45	0.5 mM	30 mM $MgCl_2$ 0.25 mM Thallium acetate	1 mM spermine, 20 μM lysine–anthraquinone	—	20 mM Na cacodylate (pH 6.5)	60 mM $MgCl_2$, 10% (v/v) MPD	6.5	277.00	Vapour diffusion/ hanging drop	[24]
1S47	0.5 mM	30 mM $MgCl_2$ 0.25 mM Thallium acetate	1 mM spermine, 20 μM lysine–anthraquinone	2.5% v/v PEG400	20 mM Na cacodylate (pH 6.5)	60 mM $MgCl_2$ 25% (v/v) PEG 400	6.5	277.00	Vapour diffusion/ hanging drop	[24]
2AVH	0.5 mM	150 mM NaI	—	7.5% PEG400	250 mM Na cacodylate (pH 6.5)	0.3 M NaI, 0.05 M Na cacodylate (pH 6.5), 15% PEG400	6.5	284.15	Vapour diffusion/ hanging drop	[31]
2AVJ	0.5 mM	150 mM NaI	—	7.5% PEG400	250 mM Na cacodylate (pH 6.5)	0.3 M NaI, 0.05 M Na cacodylate (pH 6.5), 15% PEG400	6.5	293.15	Under-oil evaporation	[31]
2GRB	1 mM	25 mM KCl 30 mM $SrCl_2$	2 mM spermine	2.5% v/v MPD	40 mM Na cacodylate (pH 7.0)	25% (v/v) MPD	7.0	298.00	Vapour diffusion/ hanging drop	[32]
2HBN	1.5 mM	10 mM Mg acetate 40 mM K acetate	3.5 mM spermine	5% v/v MPD	50 mM K cacodylate (pH 6.5)	35.0% (v/v) MPD	6.5	291.00	Vapour diffusion/ hanging drop	[33]
2HRI	1 mM	80 mM Li_2SO_4, 80 mM NaCl, 80 mM KCl	1–5 mM TMPyP4[a]	500 mM $(NH_4)_2SO_4$	20 mM K cacodylate (pH 6.5)	1.4 M $(NH_4)_2SO_4$	6.5	285.15	Vapour diffusion/ hanging drop	[34]
2O4F	0.65 mM	5 mM $MgSO_4$		0.9 M Li_2SO_4	25 mM Na cacodylate (pH 6.0)	1.8 M Li_2SO_4	7.4	277	Vapour diffusion/ hanging drop	[35]

[a] Ligand.

that the mutant sequence forms the same topology. Crystallization experiments may then follow.

3. Crystallization of G-quadruplexes/ligand complexes

When forming complexes between DNA and small molecule ligands the ratio used for crystallization trials can be a critical factor in whether or not crystals of the complex would be formed. For example, the native *Oxytricha* sequence d[GGGGTTTTGGG] folds into a symmetric, bimolecular quadruplex, as shown in both solution and crystalline forms [6–10]. The quadruplex motif has two equivalent planar binding sites external to the stacked guanine-tetrads. Crystallization trials of a DNA/ligand complex used a series of di-substituted acridines to reveal that a molar ratio of one ligand to one bimolecular quadruplex was required for crystals to form. The final refined structure revealed that only one of the two equivalent G-tetrad sites was occupied by the ligand. This family of ligands is typical of the majority of quadruplex stabilizing ligands that interact by stacking onto the external G-tetrads, whereas in duplex DNA, minor groove binding are very common. It might be then possible to target the four grooves characteristic of G-quadruplex DNA. In parallel-stranded quadruplexes the grooves are equivalent, presenting four equally accessible binding surfaces. When one of the backbone strands runs anti-parallel the groove widths are altered, generating a range of narrow to wide groove widths. The situation can be more complicated as seen for the human telomeric sequences that can fold into several topologies. Here one isoform maybe stabilized over another by either cations, or even particular stabilizing ligands, as observed for the ligand telomestatin [11]. In the case of the co-crystallizations of the porphyrin molecule TMPyP4 and the human telomeric sequence, the topology remained the same but the external propeller TTA loops were remodelled by the ligand to generate two new binding surfaces. A range of ligand concentrations need to be used when performing co-crystallizations. If binding surfaces are available in a crystalline environment then ligand soaking experiments can be used. This technique provides for rapid screening and phasing, generating high quality structural information. G-quartet targeting ligands tend to be excluded in this technique as the binding surfaces are usually utilized in packing interactions.

4. Structural determination

The determination of a validated three-dimensional model is a prerequisite for rational structure-based ligand design. Several biophysical techniques can be employed to determine quadruplex DNA topologies providing low resolution models, however NMR and crystallographic techniques provide the structural detail required for rational drug design. In solution, NMR determinations provide important structural information that is indepen-

t of the constraints of the crystalline lattice environ-nt; however the advantage of crystallographic tech-ues is the ability to determine structures de-novo and provide high-resolution coordinate data for both the ll-defined quadruplex core motif but also for the variable ctures adopted by the nucleotides seen in the connect-loops. The determination of these loop structures is ical in the design of selective ligands. In fact the close king of the quadruplexes in a lattice environment pro-es additional information regarding molecular associa-n, either through the planar tetrads or through loop–p interactions. The remodelling of these loops, either ough packing or ligand association can be readily deter-ed using crystallographic methods.

. Crystallography

Structure solution in X-ray crystallography relies on the ermination of the phases associated with each measured cture factor. Once determined the phases and measured cture factor amplitudes can be combined to generate an ctron density map that will represent the electron distri-ion within the crystalline lattice. Into this map an mic resolution model can be built and the resulting cture determined as a set of atomic coordinates for h atom type. There are four main routes available to stallographers to determine the phase and so structural cidation; these are: direct methods, molecular replace-nt, isomorphous replacement and multiple-wavelength omalous dispersion. The specific merits of each method l be discussed as pathways to quadruplex structure ermination.

4.2. Molecular replacement

Molecular replacement is the most utilized technique for phase determination of quadruplex crystal structures (74%, Table 4), relying on the structural features commonly associated in these DNA motifs. The technique can be applied to low resolution data, less than 4 Å, which can be collected in-house without the need of specialist synchrotron radiation. The strength of this structure solution technique comes from the ability to separate the rotational component of the search from the translational component. This then reduces a six-dimensional search of a model into the crystal lattice into two three-dimensional searches. The initial randomly oriented search model can then be rotated and positioned into a packed lattice environment. The closer the search model is to the final structure the easier it is to determine the correct structural solution. This method then requires a reliable three-dimensional coordinate model as the search molecule. Fortunately the core of the quadruplex, the guanine-tetrad, is well defined, having a typical diameter of 21–23 Å and a fixed hydrogen bonding pattern. When stacked these tetrads have a rise of 3.13 Å and a twist of 30° between each layer. The regularity of this core G-tetrad motif can then be employed to define a search model even if an exact match in sequence or length is not available from the PDB or NDB databases. This predefined arrangement of the stacked guanine-tetrads, typical of quadruplex DNA can then be used to aid in the building of a reliable structural model.

Although molecular replacement as a method of structure solution works well, particularly for protein structure determination, it is less successful for the solution of

ble 4
cture solution method and software used in the solution and refinement processes

ID	Structure solution method	Structure solution software	Structure refinement software	Ref.
	Molecular replacement	X-PLOR [12]	X-PLOR	[26]
	Molecular replacement	AMoRE [13]	SHELX-93	[27]
	Molecular replacement	n/a	SHELX-93	[28]
	MIR	MLPHARE [14]	CNS 0.5	[29]
	MAD	MLPHARE	SHELXL-97	[21]
	Molecular replacement	n/a	CNS 1.0	[8]
	Molecular replacement	X-PLOR	SHELX-97 [18]	[9]
	Molecular replacement	X-PLOR	SHELX-97	[9]
	MAD	CNS [15]	SHELX-97	[2]
	Molecular replacement	CNS	SHELX-97	[2]
	Molecular replacement	EPMR [16]	SHELX-97	[10]
	Molecular replacement	AMoRE	SHELX-97	[20]
	MAD	SOLVE [17]	CNS 1.0	[30]
	Molecular replacement	AMoRE	CNS 1.0	[30]
	Molecular replacement	AMoRE	CNS 1.0	[30]
	Molecular replacement	AMoRE	CNS 1.1	[24]
	Molecular replacement	EPMR	CNS 1.1	[24]
	Molecular replacement	EPMR	REFMAC 5.2.	[31]
	SAD	SHELXD	REFMAC 5.2.	[31]
	MAD	SOLVE	REFMAC 5.2.	[32]
	Molecular replacement	REFMAC 5.2. [36]	REFMAC 5.2.	[33]
	Molecular replacement	PHASER [19]	REFMAC 5.2	[34]
	Molecular replacement	AMoRE	CNS 1.1	[35]

tures that contain highly repetitive units, particularly repeating units are stacked which is typical for A. In addition, the building of a suitable DNA quadruearch model for molecular replacement is compliby the variation in groove width associated with anine orientation about the glycosidic bonds. Unlike ex DNA with an anti-parallel arrangement for its two ds, the four phosphate backbone strands in quadrus can potentially have any 5′ to 3′ arrangement. In ruplexes the G-tetrads retain a fixed geometry, while lycosidic angle, the link from the guanine base to ibose sugar, depends on the backbone 5′ to 3′ orienta- Changes in this glycosidic angle alter the relative posiof the sugar ribose and so the groove widths (Fig. 2). all-parallel structures all the glycosidic angles are alent and so the four grooves widths are also equiva- This equivalence changes when at least one phosphate bone runs in an anti-parallel arrangement to the other ds. The variation in glycosidic angle reflects changes e groove widths, generating narrow, medium and wide es. Each variation in groove width alters the positions e strongly scattering phosphate groups which can then sely affect the validity of the search model chosen. program packages used successfully for molecular cement are X-PLOR [12], AMoRE [13], MLPHARE CNS [15], EPMR [16], SOLVE [17], SHELX [18], PHASER [19] (Table 4).

lthough prior structural information is required to the structure, and a level of bias in the resulting struc- may be introduced, this does not have to limit its use in g novel structures. Phase information derived from partial models is sufficient to calculate electron density for unknown regions of the unit cell. The missing ponents; the ligands and/or the nucleotides forming nnecting loops, can be built into this electron density, ding an unbiased structural description of the com- tructure.

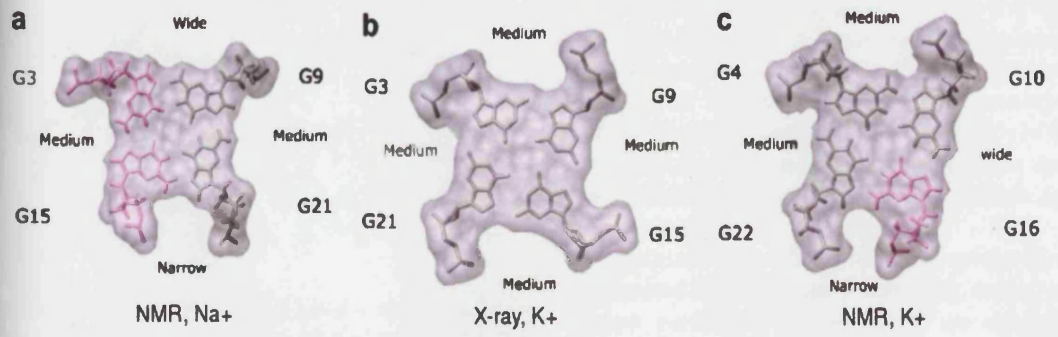
Direct methods

is method has not yet been utilized as a path to phase mination for quadruplex structures. Only two datasets with sufficient resolution for direct methods to succeed have been collected, 1O0K [20] at 1.17 and 1J8G at 0.61 Å [21] however they were solved by MR and MAD methods, respectively. In general good quality diffraction data needs to be collected above 1.2 Å resolution before this technique for structure solution can be applied with any degree of success. The great advantage of this method is that it does not require prior structural knowledge to generate a complete structural model.

4.4. Multiple-wavelength anomalous diffraction (MAD) and single anomalous diffraction (SAD)

The intensity contributions of anomalously scattering electrons can be used to estimate the phases and so calculate electron density maps. By selecting an appropriate wavelength of X-ray radiation the contribution of the anomalous scattering can be enhanced. Scattering atoms with a high anomalous contribution can be soaked into preformed crystals or be engineered into the samples prior to crystallization. Synchrotron X-ray sources provide an ideal source as they are tunable, thus enabling collection on a single crystal at several wavelengths, allowing the collection of both maximum dispersive and anomalous contributions.

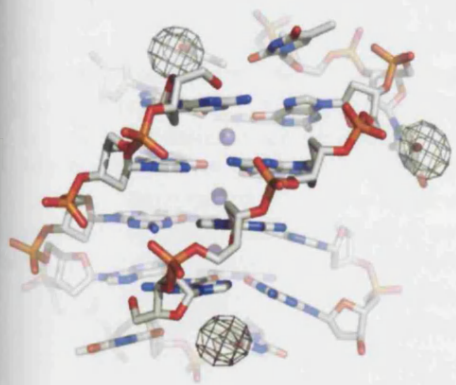
The advantage of sequential solid phase synthesis of DNA is that it allows for the insertion of modified nucleotide bases at specific locations in the sequence enabling the incorporation of specific heavy atoms for anomalous diffraction and structure solution. The two most attractive modifications for DNA are 5-bromo-deoxyuracil and 5-iodo-deoxyuracil, both thymidine analogues. Being close to the selenium K-edge, the bromo-U K-edge at 0.920 Å is ideal for automated data collection systems setup at synchrotron sources idealized for protein structure solution. The 5′ methyl group substitution places the bromine group into the major groove of duplex DNA. In quadruplex DNA the thymines are normally 3′ and 5′ of the stacked G-quartets, and are either part of the loop structure or stacking externally on the G-tetrads. This introduces a complicating factor in that the modified bases may not be structurally ordered. Ordering of the atoms in a lattice is



Glycosidic bond and variation in groove width in the human telomere. (a) + (c) in solution and (b) in the crystal. Shown for the middle tetrad. *Anti* n grey and *syn* in pink. Surfaces in purple. (For interpretation of the references to colour in this figure legend, the reader is referred to the web of this article.)

uired to retain occupancy resulting in sufficient levels of omalous scattering at a particular location. In addition thermal and static disorder, halogen derivatives have disadvantage of being labile, making them light sensi- during crystallization, and sensitive to X-ray exposure ring collection. This results in a reduction of occupancy ring prolonged X-ray exposure, which can be very signif- nt for iodo-substituted derivatives. MAD and SAD hniques have been successfully applied in six cases of ucture solution, 1K8P UAGGGUTAGGGT, 2AVJ GGGUTTGGGG, 2GRB UGIGGU. In the case of VJ the loss of the anomalous scattering contribution ring collection required the use of one wavelength for ucture solution, utilizing a single wavelength collected the anomalous peak (SAD), (Fig. 3). Fortunately one omine substitution provides sufficient anomalous scatter- for the phasing of about thirty nucleotides. It is a com- on practice to introduce several substitutions into one uence thereby increasing the probability of detection at least one anomalous site and subsequent phasing of structure. An example of a SAD phased map is shown Fig. 4, highlighting how the strongly scattering atoms ist in the placement of a quadruplex model, with the osphate groups defining the backbone geometry and central potassium ions aligning the G-quartets.

Following on from the successes of protein structure ution using selenium as an anomalous scatterers, archers have recently employed the use of a Se–CH3 up inserted onto the C2′ of the ribose sugar ring [22] solve both A-form DNA and RNA structures. The mod- ation places a selenium atom into the minor groove of plex DNA. Structural analyses have shown a minimal tortion to the A-form. This technique has yet to be plied to solving quadruplex structures but may have advantage of increasing stability, both during crystalli- on and data collection, and that it can be readily syn- sized. The stabilization of glycosidic torsion angles, over *syn*, might then be utilized using this ribose type dification, ensuring a greater stability of one quadruplex



3. Anomalous difference Fourier map revealing the location of the mine atoms on the Br-uracil nucleotides, substituting for the 5-methyl p of thymines. The refined bimolecular quadruplex structure of ence d(GGGG^{Br}UTTGGGG)$_2$ is shown overlaid. Anomalous data cted at 0.92 Å, the maximum anomalous contribution for bromine.

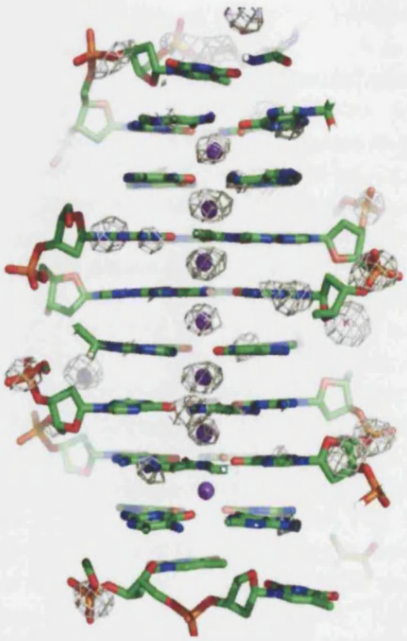
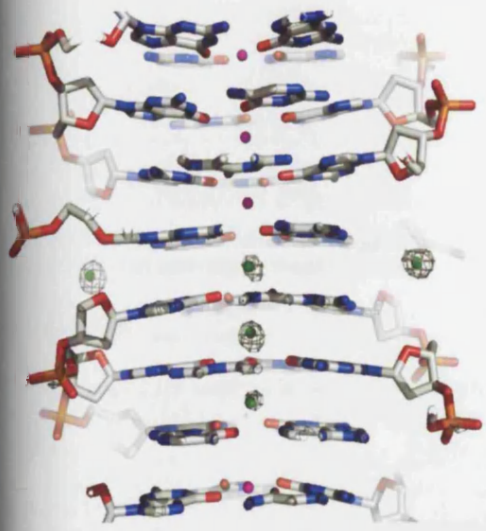


Fig. 4. Two bimolecular quadruplexes of sequence d(GGGGTTTG GGG)$_2$ with lateral connecting loops are shown with electron density drawn at 4.0σ calculated from SAD derived phases. The phosphates (yellow) and potassium cations (purple) in the central channel can be used to orient and build the quadruplex structures. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

topology over another [23]. Topology is defined by back- bone strand polarity requiring the appropriate glycosidic torsion angle of the guanines, modifications favouring one glycosidic torsion angle over another would selectively enhance the stability of one topology over another.

Other anomalous scattering elements have been employed that rely on their preferential uptake and order- ing within the central core of a G-quadruplex. Monovalent metal cations are an integral component of quadruplex DNA, and position themselves between the negatively charged carbonyl O6 oxygens of the guanines. Thallium can be substituted for sodium and potassium monovalent metal cations in the central quadruplex core. With an L3 peak anomalous signal at 0.9772 Å, $\Delta f'' = 10e$ and at 1.54 Å, $\Delta f'' = 8e$ for Cu-Kα radiation, thallium could potentially be used for structure solution [24] with data col- lections taking place in-house using Cu sources. Although, in this case the d(TGGGGT)$_4$ parallel-stranded structure annealed in sodium, was solved using conventional MR methods, possibly due to the low occupancy of the thallium atoms, in the difference anomalous electron density maps the anomalous scattering from the thallium atoms could be clearly seen.

Ions used for stabilizing the quadruplexes can usually be determined by their relative scattering, radius, and coordi- nation, particularly in the central ion channel, although some ions are more difficult to differentiate, especially in the grooves such as sodium and oxygen of water molecules as their scattering is almost equivalent. However, we were



5. Stacked head-to-head are two d(TGGGGT)$_4$ quadruplex mole- with Na ions shown as purple and calcium ions as green spheres ating the central channel. Overlain is drawn the anomalous ence maps at 2.5σ, identifying the anomalously scattering atoms. nterpretation of the references to colour in this figure legend, the is referred to the web version of this article.)

to utilize the small amount of anomalous scattering of ium, using Cu-Kα radiation at 1.54178 Å, with a $\Delta f''$ of 6 electrons to identify the metal ions and occupancy in parallel-stranded G-quadruplex of sequence GGGGT) (PDB ID 2GW0: Fig. 5). Potassium ions also be identified by their anomalous scattering after able phases have been derived. These ions have a weak malous scattering contribution, $\Delta f'' = 1e$ for Cu-Kα ation, but significantly higher $\Delta f'' = 3e$ for Cr-Kα radi- (2.29 Å) [25].

Isomorphous replacement

his technique relies on the difference of one structure or data set over another usually by the substitution ntroduction of heavy atoms. These substitutions are ally introduced by heavy atom soaks into the pre- ed crystals or by modification of the nucleotides in DNA sequence. The substitution of sodium ions with ium ions in the central channel can provide an oppor- ty for isomorphous replacement, although the use of malous scattering on these ions appears the most reli- path to phase determination. Currently there is only example of isomorphous replacement as a path to druplex structure determination [25]. In this case the druplex is embedded in a protein/DNA complex, and ed using mercury derivatives. The mercury atoms are d to the protein component of the complex.

ncluding remarks

he importance of telomerase and its inhibition as a er therapeutic target has re-directed attention onto quadruplexes and their role in disrupting telomere function. Structural studies are currently focusing on characterizing relationships between DNA sequence and topology in order to enhance selectivity through specific structural motifs, such as lateral, propeller and diagonal loops. Any well-defined binding motif can be exploited by engineering small molecule ligands to selectively target one quadruplex over other related structures and to ensure a reduction in overall non-specific duplex DNA binding.

Structural data from crystallographic determinations also need to be placed into the context of their relevance in cellular environments. Associating the relevant folded topology for quadruplexes formed from repeats of the human telomeric sequence d(TTAGGG) is currently an issue due to variation in observed topology, which appears to be related to the environment during structure determination, sequence used and ionic conditions. Fig. 1 shows schematically three quite different topologies determined by NMR and crystallographic methods. NMR analysis of the four-repeat 22 mer sequence d[AGGG(TTAGGG)$_3$] in sodium solution showed formation of an anti-parallel topology with a diagonal TTA loop at one end of a G-tetrad stack, and two lateral loops at the other end. Biophysical studies suggest that there is only a single species in sodium solution, and the reasonable assumption is that this corresponds to the NMR structure. In striking contrast, biophysical studies (CD, single molecule) in potassium ion solution suggest the presence of more than one species for this sequence, although there is no general agreement as to their identity, or even the number of species. Even more controversial have been the quadruplexes formed by repeats of the human telomeric sequence folded in potassium, at concentrations close to a cellular environment. The crystal structure of d[AGGG(TTAGGG)$_3$] determined by crystallographic means from potassium-containing solution, revealed an unexpected all-parallel topology, with TTA propeller loops. Subsequent NMR studies on sequences related to this (with additional short 5′ and 3′ flanking sequences added), have determined in K^+ solutions that the major species is yet another topology, the (3 + 1) fold, which has both propeller and lateral loops, Fig. 1e. The dominant form of d[AGGG(TTAGGG)$_3$] itself in solution, in the absence of stabilizing sequences, has yet to be determined, but the implication of energy calculations (Burge and Neidle, unpublished observations), is that the propeller form is likely to dominate over others. Further structural studies are certainly required to clarify the dominant structural form in a cellular environment but it is interesting to note that a crystallographic determination of a human telomeric sequence bound to a porphyrin ligand revealed the all-propeller topology [34], consistent with previous structural determinations.

Whether the guanine repeat sequences are from telomeric regions or the promoter regions of oncogenes the self-association of these sequences into quadruplexes provides a unique platform for ligand interaction and the design of selective therapeutic agents. The crystallization and

gh-resolution structural determination of these motifs in mplex with ligands is an important goal that is finally ng realised. It is hoped that the crystallographic techques discussed above will provide a basis for the determiion of new and interesting quadruplex motifs.

knowledgment

We are grateful for support from Cancer Research UK.

ferences

Y. Wang, D.J. Patel, Structure 1 (1993) 263–282.
G.N. Parkinson, M.P. Lee, S. Neidle, Nature 417 (2002) 876–880.
E.A. Stura, in: T.Bergfors (Ed.), Protein Crystallisation: Techniques, Strategies and Tips, 1999, pp. 113–124.
T.Bergfors, in T.Bergfors (Ed.), Protein Crystallisation: Techniques, Strategies and Tips , 1999, pp. 41–50.
A.T. Phan, D.J. Patel, J. Am. Chem. Soc. 125 (2003) 15021–15027.
F.W. Smith, J. Feigon, Nature 356 (1992) 164–168.
P. Schultze, N.V. Hud, F.W. Smith, J. Feigon, Nucleic Acids Res. 27 (1999) 3018–3028.
M.P. Horvath, S.C. Schultz, J. Mol. Biol. 310 (2001) 367–377.
S. Haider, G.N. Parkinson, S. Neidle, J. Mol. Biol. 320 (2002) 189–200.
S.M. Haider, G.N. Parkinson, S. Neidle, J. Mol. Biol. 326 (2003) 117–125.
E.M. Rezler, J. Seenisamy, S. Bashyam, M.Y. Kim, E. White, W.D. Wilson, L.H. Hurley, J. Am. Chem. Soc. 127 (2005) 9439–9447.
A.T. Brunger, J. Kuriyan, M. Karplus, Science 235 (1987) 458–460.
J. Navaza, P. Saludjian, in: W.C. Charles (Ed.), Methods in Enzymology Macromolecular Crystallography Part A, Academic Press, 1997, pp. 581–594.
Z. Otwinowski, in: W. Wolf, P.R. Evans, A.G.W. Leslie (Eds.), Isomorphous Replacement and Anomalous Scattering, 1991, pp. 80–86.
A.T. Brunger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Acta Crystallogr. D. Biol. Crystallogr. 54 (1998) 905–921.
[16] C.R. Kissinger, D.K. Gehlhaar, D.B. Fogel, Acta Crystallogr. D. Biol. Crystallogr. 55 (1999) 484–491.
[17] T.C. Terwilliger, J. Berendzen, Acta Crystallogr. D. Biol. Crystallogr. 55 (1999) 849–861.
[18] G.M. Sheldrick, T.R. Schneider, in: W.C. Charles Jr. (Ed.), Methods in Enzymology Macromolecular Crystallography Part B, Academic Press, 1997, pp. 319–343.
[19] A.J. McCoy, R.W. Grosse-Kunstleve, L.C. Storoni, R.J. Read, Acta Crystallogr. D. Biol. Crystallogr. 61 (2005) 458–464.
[20] G.R. Clark, P.D. Pytel, C.J. Squire, S. Neidle, J. Am. Chem. Soc. 125 (2003) 4066–4067.
[21] J. Deng, Y. Xiong, M. Sundaralingam, Proc. Natl. Acad. Sci. USA 98 (2001) 13665–13670.
[22] J. Jiang, J. Sheng, N. Carrasco, Z. Huang, Nucleic Acids Res. 35 (2007) 477–485.
[23] C.F. Tang, R.H. Shafer, J. Am. Chem. Soc. 128 (2006) 5966–5973.
[24] C. Caceres, G. Wright, C. Gouyette, G. Parkinson, J.A. Subirana, Nucleic Acids Res. 32 (2004) 1097–1102.
[25] N. Watanabe, Acta Crystallogr. D. Biol. Crystallogr. 62 (2006) 891–896.
[26] C. Kang, X. Zhang, R. Ratliff, R. Moyzis, A. Rich, Nature 356 (1992) 126–131.
[27] G. Laughlan, A.I. Murchie, D.G. Norman, M.H. Moore, P.C. Moody, D.M. Lilley, B. Luisi, Science 265 (1994) 520–524.
[28] K. Phillips, Z. Dauter, A.I. Murchie, D.M. Lilley, B. Luisi, J. Mol. Biol. 273 (1997) 171–182.
[29] M.P. Horvath, V.L. Schweiker, J.M. Bevilacqua, J.A. Ruggles, S.C. Schultz, Cell 95 (1998) 963–974.
[30] J. Kondo, W. Adachi, S. Umeda, T. Sunami, A. Takenaka, Nucleic Acids Res. 32 (2004) 2541–2549.
[31] P. Hazel, G.N. Parkinson, S. Neidle, J. Am. Chem. Soc. 128 (2006) 5480–5487.
[32] B. Pan, K. Shi, M. Sundaralingam, J. Mol. Biol. 363 (2006) 451–459.
[33] M.L. Gill, S.A. Strobel, J.P. Loria, Nucleic Acids Res. 34 (2006) 4506–4514.
[34] G.N. Parkinson, R. Ghosh, S. Neidle, Biochemistry 46 (2007) 2390–2397.
[35] C. Creze, B. Rinaldi, R. Haser, P. Bouvet, P. Gouet, Acta Crystallogr. D. Biol. Crystallogr. 63 (2007) 682–688.
[36] G.N. Murshudov, A.A. Vagin, E.J. Dodson, Acta Crystallogr. D. Biol. Crystallogr. 53 (1997) 240–255.

J|A|C|S
COMMUNICATIONS

Published on Web 05/06/2008

Structural Basis of DNA Quadruplex Recognition by an Acridine Drug

Nancy H. Campbell, Gary N. Parkinson, Anthony P. Reszka, and Stephen Neidle*

The Cancer Research UK Biomolecular Structure Group, The School of Pharmacy, University of London, 29−39 Brunswick Square, London WC1N 1AX, United Kingdom

Received March 6, 2008; E-mail: stephen.neidle@pharmacy.ac.uk

Cancer cells have evolved mechanisms to maintain telomere length, in most instances by the synthesis of further telomeric DNA repeats using the telomerase enzyme.[1] The consequence of indefinite telomere length maintenance is cellular immortalization, a key step along the pathway to tumorigenesis. The telomerase complex, which maintains telomere length in some 80−85% of cancer cells[2] by virtue of its reverse transcriptase activity, is thus a potential target for therapeutic intervention.[3] The 3′-end of human telomeres, which comprises repeats of the sequence d(TTAGGG), is single-stranded for the final 100−200 nucleotides of its length, and can be induced to form both intra- and intermolecular four-stranded G-quadruplex structures[4] by quadruplex-binding ligands. These can inhibit telomerase activity since quadruplex DNA is not recognized by the single-stranded template of the RNA component of the telomerase complex.[5] Many such ligands have been reported,[6] and empirical SARs have been defined for a number of them, although structural information to date on quadruplex-ligand complexes is limited. The biological consequences of telomere maintenance disruption have been reported for a few ligands and include telomere uncapping and consequent induction of a DNA damage response.[7]

The 3,6,9-trisubstituted acridine ligand BRACO-19 (Figure 1) was designed by qualitative molecular modeling, on the assumption that the three substituents would each occupy a groove in a quadruplex.[8a] BRACO-19 inhibits telomerase enzymatic activity, resulting in telomere shortening, and also produces end-to-end chromosomal fusions in cancer cells[8b] as a consequence of quadruplex disruption of uncapping of proteins associated with the single-strand overhang. It shows significant *in vivo* anticancer activity in tumor xenografts, which is associated with telomere uncapping.[9]

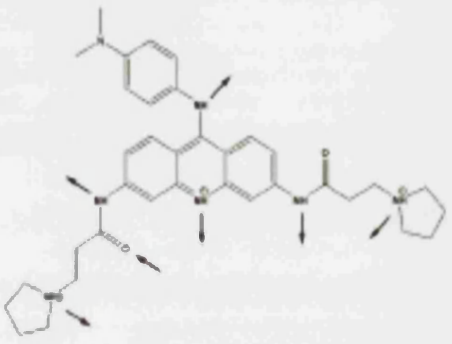


Figure 1. Structure of BRACO-19. Arrows indicate directionality of hydrogen bonding found in the crystal structure reported here.

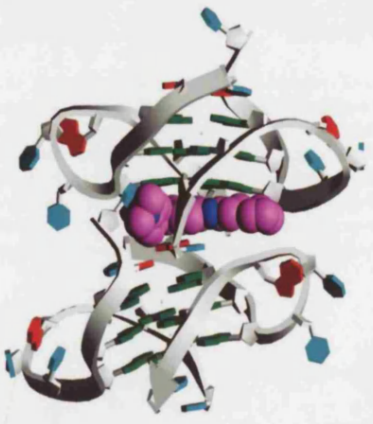


Figure 2. The biological unit in the crystal (PDB id 3CE5). A BRACO-19 molecule (mauve) is shown at the interface of the two quadruplexes in the unit, stacked between a G-quartet (top) and a TATA tetrad (bottom).

We report here the 2.5 Å crystal structure (R factor = 18.3% and R_{free} = 21.3%) of a complex between a bimolecular human telomeric G-quadruplex of sequence d(TAGGGTTAGGGT) and BRACO-19. The crystals grew in the presence of physiological levels of K^+ ions and revealed a parallel-stranded quadruplex arrangement, with the biological unit being two 5′ to 3′ stacked quadruplexes and a well-resolved BRACO19 molecule (Figure 2); the unit packs into infinite columns in the crystal. This topology has been found in the crystal structures of native unimolecular and bimolecular human telomeric quadruplexes,[10] in a complex with the porphyrin TMPyP4,[11] indicating that it is a preferred topology for quadruplex-ligand complexes. Antiparallel structures have been assigned by solution NMR studies[12,13] for human telomeric DNA quadruplexes, with one parallel and two lateral loops. These may not be compatible with drug binding since a G-tetrad platform has to be unrestricted by lateral or diagonal loops for effective drug stacking onto the platform.

Each bimolecular quadruplex in the present structure contains three planar stacked G-tetrads with a BRACO-19 molecule stacking directly onto the 3′ end G-tetrad face. The TTA sequences, which intersperse the G-tracts, form propeller loops that are similar to those observed in the native crystal structures.[10] BRACO-19 fits into a site bounded on one side by the 3′-end G-tetrad and on the other by a 5′-end TATA tetrad face formed from adenine and thymine bases at the interface between the two 5′ to 3′ stacked bimolecular quadruplexes in the biological unit. The drug is asymmetrically stacked on the G-tetrad at the end of one quadruplex, and there is $\pi-\pi$ overlap with just two guanine bases (Figure 4). The cationic ring nitrogen atom of the acridine ring is in-line with the K^+ ion channel that runs through the quadruplex. (Ions were unequivocally assigned as K^+ in the crystallographic analysis). The other side of the acridine chromophore surface is stacked not onto a second G-tetrad but onto a reverse Watson−Crick A•T base pair that forms part of a TATA quartet formed at the dimer interface. This TATA platform is constructed from the 5′-TA ends of two strands of one quadruplex, and the 3′-thymine from a third strand on the other. The G- and the TATA tetrads are off-set with respect to each other, and inclined by ca. 30° in two directions (Figure 4). The other 3′ end thymine base is flipped into the binding site so

10.1021/ja8016973 CCC: $40.75

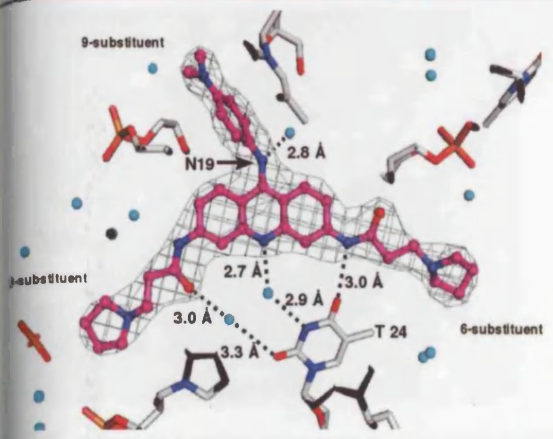


Figure 3. View onto the face of the BRACO-19 molecule in an omit map, showing the hydrogen bonds to water molecules and the flipped-in thymine base T24, with electron density for the drug contoured at 1.0 σ.

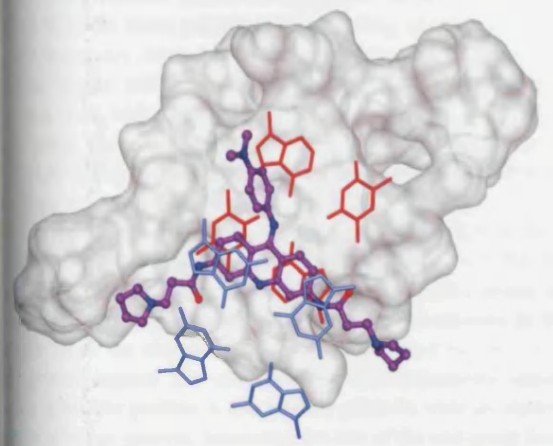


Figure 4. View on the face of the BRACO-19 molecule (mauve), showing the G-tetrad (in blue) above the plane and the TATA quartet (in red) stacked below the drug. The solvent-accessible surface of the lower quadruplex is also shown. This and other figures have been produced by the UCSF CHIMERA package (*J. Comput. Chem.* **2004**, *25*, 1605−1612).

that it is able to interact with the BRACO-19 molecule (Figure 3), in arrangement reminiscent of the *Oxytricha nova* antiparallel bimolecular quadruplex crystal structure complexed with a disubstituted acridine ligand.[14] This thymine thus plays a significant role in ligand binding, with a direct H-bond to a side-chain amide nitrogen atom, and two water-mediated interactions, one with the carbonyl atom of the other side-chain and the other to the central ring nitrogen atom of the acridine (Figure 3).

The 3- and 6-position substituents of BRACO19, comprising flexible side chains with cationic termini each extending into a wide groove, are located on opposite sides of the G-tetrad face, while the 9-position anilino substituent fits into a narrow pocket. These wide shallow grooves, whose walls are formed by the negatively charged phosphate backbones and floors by the edges of the G-tetrads, are filled with water molecules that contact phosphate oxygen atoms and base edges (Figure 5a,b). There are no direct H-bonds from the cationic pyrrolidino ring nitrogen atoms at the ends of the side chains to any of the negatively charged phosphate backbones. Instead, both nitrogen atoms participate in the networks of water molecules in the grooves, and these make contact with phosphate groups (Figure 6). Water molecules also contact the amide groups of the side chains, on the opposite side of the H-bonded thymine base. The phenyl ring of the 9-substituent fits tightly into a narrow hydrophobic pocket at the dimer interface,

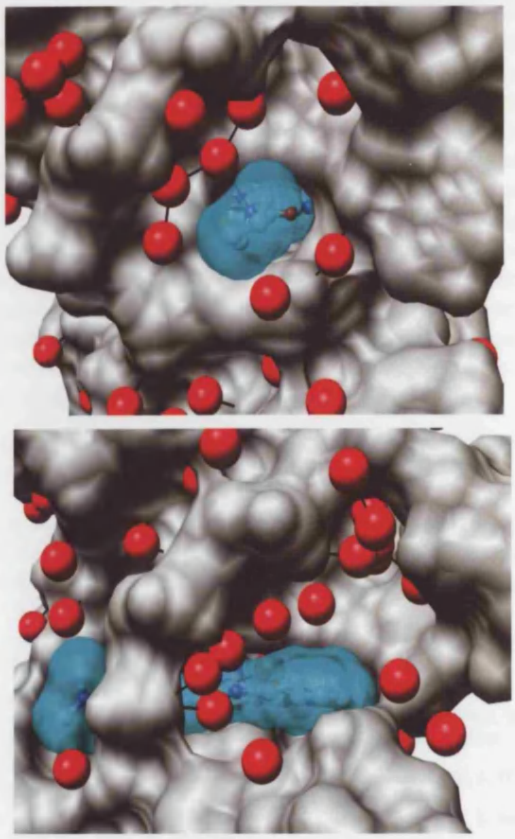


Figure 5. (a,b) Views of the biological unit. The BRACO-19 molecule is shown in a transparent cyan-colored solvent-accessible surface representation with individual atoms shown in ball-and stick form. The solvent-accessible surface of the quadruplex is colored grey. The figures each show one of the two 3,6-substituents protruding into the wide shallow grooves and the arrangement of waters (red spheres) in the grooves. The pyrrolidino nitrogen atoms of each side chain are each H-bonded to a water molecule, which are within water networks. H bonds are shown by black lines.

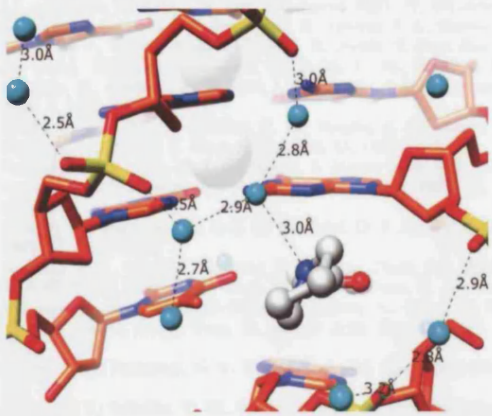


Figure 6. Detailed view of the hydrogen-bonding arrangement involving the water molecule network and the pyrrolidino nitrogen atom at the end of one of the BRACO-19 side chains. This nitrogen is directly hydrogen-bonded to one water molecule, which in turn hydrogen bonds to two others, each of which contacts a phosphate group.

which is formed by a terminal 5′ thymine base from one quadruplex in the biological dimer and the C5′ from the other quadruplex (Figure 7). This pocket is extended into a narrow groove that is filled with water molecules, reminiscent of a DNA duplex minor groove. There is also a H-bond between a water molecule and the nitrogen atom N19 bridging between the acridine core and the phenyl group, although other water molecules contacting this one

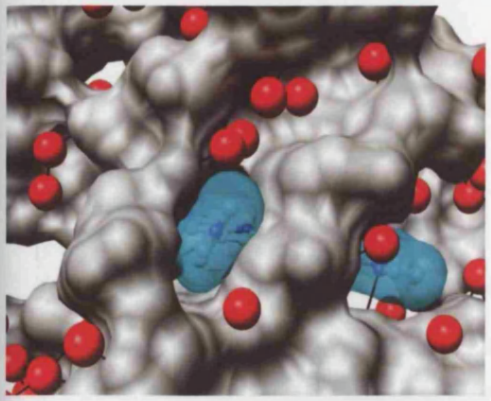


Figure 7. View of the 9-position aniline group (in cyan) tightly bound in one end of a narrow channel, with the upper part of the channel occupied by a number of water molecules. Colors are as in Figure 5.

are not well resolved. Thus of the eight donor−acceptor substituents in BRACO-19, seven participate in H-bonding, of which six are to water molecules rather than directly to quadruplex substituents.

This crystal structure provides a rationale for the quadruplex binding and biological behavior of a number of BRACO-19 analogues. Extending the length of the 3- and 6- side chains by $-(CH_2)-$ groups results in a progressive decrease in quadruplex binding; the binding constant for BRACO-19 itself is 3.1×10^7 M^{-1}, whereas it is 6×10^5 M^{-1} for the analogue with three more $-(CH_2)-$ units in each side chain.[15] The structure shows that the increasingly hydrophobic side chains when extended would no longer be able to participate in a water H-bond network in the groove since the charged end-group would extend too far out of the groove region. The alternative, of the hydrophobic chains packing into the grooves, is also unlikely given the wide yet shallow nature of these grooves. Increasing the size of the end-group from the five-membered pyrrolidino ring to a six-membered piperidino or similar ring does not have a deleterious effect on binding. The crystal structure also shows that the side chains are both in narrow regions at the entrance to the grooves, which would exclude the binding of larger ring systems, in agreement with experiment.[15−17]

The crystal structure explains the key role of the 9-position aniline substituent, which increases quadruplex affinity 10-fold. The narrow cross-section of the binding pocket accommodates a planar aromatic ring, but would also fit other hydrophobic groups, such as a linear $-(CH_2)_n-$ chain, in accord with experiment. A substituent attached to one of the anilino amine nitrogen atoms can confer additional stability.[15−17] For example the analogue with a $-(C=O)-(CH_2)_3-$pyrrolidino has a binding constant of 7.7×10^7 M^{-1}. Qualitative modeling of this substituent into the crystal structure shows that it lies snugly in the groove and suggests that substituents longer than this would not fit.

The development of many G-quadruplex ligands has to date assumed that stacking to a G-tetrad is an adequate description of the binding. This crystal structure shows that this is not the case, with a structurally complex ligand such as BRACO-19 inducing the formation of a quadruplex binding site with a high degree of three-dimensional complexity. Although the details of the many stabilizing interactions involving water molecules are probably specific to acridines and BRACO-19 analogues, the active involvement of loop and other nucleotides in helping to define and shape the binding site are likely to be general features of ligand binding to telomeric and other quadruplexes. This has been observed in our recently determined structures of two quadruplex complexes with a naphthalene diimide ligand (PBD nos. 3CDM, 3CC0), which have the TTA loops altering their conformation to form part of the binding site platforms for this molecule. The BRACO-19 complex, with its all-parallel topology and 3′ to 5′ packing of repeating units in the crystal structure, is a direct model for the drug-bound single-strand overhang and suggests that the drug may promote the formation of a highly compacted secondary structural arrangement that would effectively compete with telomerase uncapping and with other single-strand binding proteins such as hPOT1.

Acknowledgment. This work is supported by Cancer Research UK (Programme Grant No. C129/A4489 to S.N.).

Supporting Information Available: Crystallographic detail. This material is available free of charge via the Internet at http://pubs.acs.org. Coordinates and structure factors are available from the PDB as entry no. 3CE5.

References

(1) Xu, L.; Blackburn, E. H. *Mol. Cell* **2007**, *28*, 315–327.
(2) Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L. C.; Coviello, G. M.; Wright, W. E.; Weinrich, R.; Shay, J. W. *Science* **1994**, *266*, 2011–2015.
(3) (a) Shay, J. W.; Wright, W. E. *Nat. Rev. Drug Discovery* **2006**, *5*, 577. (b) Oganesian, L.; Bryan, T. M. *Bioessays* **2007**, *29*, 155–165.
(4) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402–5415.
(5) (a) Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. *Nature* **1991**, *350*, 718–720. (b) Sun, D.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H. *J. Med. Chem.* **1997**, *40*, 2113–2116.
(6) For a recent survey see: De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J.-F.; Mergny, J.-L. *Biochimie* **2008**, *90*, 131–155.
(7) For example: (a) Gunaratnam, M.; Greciano, O.; Martins, C.; Reszka, A. P.; Schultes, C. M.; Morjani, H.; Riou, J.-F.; Neidle, S. *Biochem. Pharmacol.* **2007**, *74*, 679–689. (b) Brassart, B.; Gomez, D.; De Cian, A.; Paterski, R.; Montagnac, A.; Qui, K. H.; Temime-Smaali, N.; Trentesaux, C.; Mergny, J.-L.; Gueritte, F.; Riou, J.-F. *Mol. Pharmacol.* **2007**, *72*, 631–640.
(8) (a) Read, M.; Harrison, R. J; Romagnoli, B.; Tanious, F. A.; Gowan, S. H.; Reszka, A. P.; Wilson, W. D.; Kelland, L. R.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4844–4849. (b) Incles, C. M.; Schultes, C. M.; Kempski, H.; Koehler, H.; Kelland, L. R.; Neidle, S *Mol. Cancer Ther.* **2004**, *3*, 1201–1206.
(9) Burger, A. M.; Dai, F.; Schultes, C. M.; Reszka, A. P.; Moore, M. J.; Double, J. A.; Neidle, S. *Cancer Res.* **2005**, *65*, 1489–1496.
(10) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876–880.
(11) Parkinson, G. N.; Ghosh, R.; Neidle, S. *Biochemistry* **2007**, *46*, 2390–2397.
(12) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2007**, *35*, 6517–6525.
(13) (a) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 8710–8716. (b) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. *Biochemistry* **2005**, *44*, 2048–2058. (c) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723–2735.
(14) Haider, S. M.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2003**, *326*, 117–125.
(15) Moore, M. J.; Schultes, C. M.; Cuesta, J.; Cuenca, F.; Gunaratnam, M.; Tanious, F. A.; Wilson, W. D.; Neidle, S. *J. Med. Chem.* **2006**, *49*, 582–599.
(16) Schultes, C. M.; Guyen, B.; Cuesta, J.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4347–4351.
(17) Harrison, R. J.; Cuesta, J.; Chessari, G.; Read, M. A.; Basra, S. K.; Reszka, A. P.; Morrell, J.; Gowan, S. M.; Incles, C. M.; Tanious, F. A.; Wilson, W. D.; Kelland, L. R.; Neidle, S. *J. Med. Chem.* **2003**, *46*, 4463–4476.

JA8016973

Accelerated Publications

Selectivity in Ligand Recognition of G-Quadruplex Loops†

Nancy H. Campbell, Manisha Patel, Amina B. Tofa, Ragina Ghosh, Gary N. Parkinson, and Stephen Neidle*

Cancer Research UK Biomolecular Structure Group, The School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX, U.K.

Received December 7, 2008; Revised Manuscript Received January 15, 2009

ABSTRACT: A series of disubstituted acridine ligands have been cocrystallized with a bimolecular DNA G-quadruplex. The ligands have a range of cyclic amino end groups of varying size. The crystal structures show that the diagonal loop in this quadruplex results in a large cavity for these groups, in contrast to the steric constraints imposed by propeller loops in human telomeric quadruplexes. We conclude that the nature of the loop has a significant influence on ligand selectivity for particular quadruplex folds.

Nucleic acid sequences containing short tracts of guanine residues can form higher-order quadruplex structures (*1, 2*). Quadruplexes can be intramolecular, with four or more G-tracts, or intermolecular, with one or two G-tracts. They have been extensively investigated by structural and biophysical methods, and a number of distinct topologies have been found. They depend in large part on the sequence and size of the intervening loops. Chain-reversal (propeller) loops are positioned on the exterior of quadruplex structures, whereas diagonal and edge (lateral) loops are stacked onto the terminal G-quartets of a quadruplex structure.

Two principal categories of quadruplex nucleic acids have received attention as potential therapeutic targets: (i) those capable of being formed at the single-stranded overhang of telomeric DNA, for which appropriate small-molecule ligands can induce quadruplex formation, inhibit the activity of the telomerase enzyme complex, and selectively target telomere maintenance in cancer cells (*3, 4*), and (ii) those within genomes (*5, 6*) and especially within promoter sequences of genes involved in cellular proliferation (*7*) such as vascular endothelial growth factor (*8, 9*), and oncogenes such as *c-myc* (*10*−*14*) and *c-kit* (*15*−*18*). It has been suggested (*9, 10*) that ligands selective for promoter quadruplexes could in principle downregulate the expression of these genes and thus have an anticancer effect.

A wide range of small molecules have been investigated as quadruplex-binding and -stabilizing ligands (see recent reviews in refs *19* and *20*). The majority share the common structural features of (i) a planar heteroaromatic chromophore, which stacks by $\pi-\pi$ interactions onto the G-quartet motif at the terminus of a quadruplex, and (ii) short alkyl chain substituents normally terminating in an amino group that is fully cationic at physiological pH. The precise nature of these substituents has been found to influence quadruplex affinity and selectivity. For example, the size of the terminal amino substituent has a major influence on the affinity and energetics of substituted acridine binding to a human intramolecular telomeric DNA quadruplex (*21*). The nature of the substituent together with side chain length has been rationalized for series of trisubstituted acridine ligands (having 3-pyrrolopropionamido chains at the 3- and 6-positions) (*22, 23*) on the basis of a cocrystal structure with a human telomeric DNA quadruplex (*24*). The propeller loops on the exterior of this quadruplex change their conformation from those found in the native structure (*25*) to form discrete binding pockets for the substituents. By contrast, the bimolecular quadruplex from the organism *Oxytricha nova* has a diagonal loop of sequence d(TTTT) stacked on top of the quadruplex (*26*). A cocrystal structure with an acridine derivative disubstituted with the same 3-pyrrolopropionamido group (Figure 1, compound **7**) shows this ligand bound between the loop and a terminal G-quartet (*27*). Structural

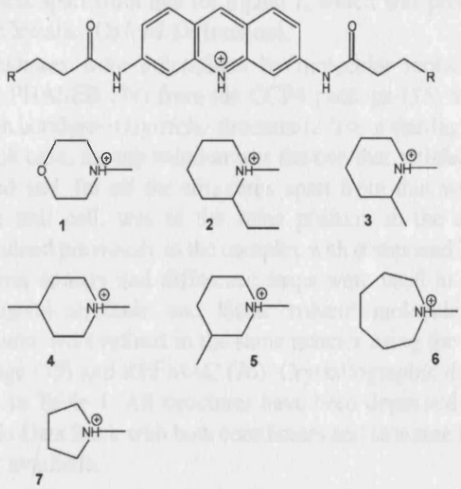


FIGURE 1: Structures of 3,6-disubstituted acridines in this study.

† This work was supported by CRUK and the EU (FP6 Project on Molecular Cancer Medicine).

* To whom correspondence should be addressed. Phone: +44 (0)20 7753 5971. Fax: +44 (0)20 7753 5970. E-mail: stephen.neidle@pharmacy.ac.uk.

10.1021/bi802233v CCC: $40.75
Published on Web 01/27/2009

Table 1: Crystallographic Details of the Quadruplex−Acridine Complexes That Are Reported in This Study[a]

	1	2a	2b	3	4	5	6
PDB entry	3EM2	3EQW	3EUI	3ERU	3ES0	3ET8	3EUM
Cell dimensions [*a*, *b*, *c* (Å)]	55.23, 42.71, 26.84	55.28, 42.73, 26.62	55.36, 42.54, 48.64	55.22, 42.63, 26.85	55.53, 42.46, 27.30	55.40, 42.59, 27.14	55.62, 42.31, 27.08
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12_1$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Resolution (Å)	2.30	2.20	2.20	2.00	2.20	2.45	1.78
No. of observed reflections	3000	3454	5575	4422	3549	2602	5372
% observed reflections	97.0	99.6	94.5	95.7	99.7	99.7	81.7
R_{merge}	0.20	22.6	23.4	20.1	19.7	21.1	22.4
R_{free}	0.29	31.4	28.5	24.9	24.5	27.7	27.0
Average B ($Å^2$)	0.057	0.092	0.087	0.059	0.041	0.084	0.079
No. of observed water molecules	17.0	12.9	20.7	17.4	11.1	16.6	16.5
	64	66	159	71	56	51	52

[a] Further information is available in the individual PDB entry.

Information about quadruplex−ligand complexes is limited to a small number of NMR-derived (*13, 28, 29*) and crystal structures (*24, 27, 30, 31*), and as yet, it has not been possible to define the structural features that differentiate small molecule binding to one quadruplex compared to another. Evidence is accumulating that it is possible to target particular quadruplex types, as shown by the binding of diarylethynyl imides to a parallel topology formed by a quadruplex formed by a sequence within the *c-kit* promoter (*32*).

This study addresses the issue of ligand selectivity by examining how ligand binding is constrained by loop geometry in two distinct quadruplex architectures. The *O. nova* bimolecular quadruplex has been used as an exemplar for one particular type of loop, the diagonal type, formed in this instance by four thymine nucleotides. We report here six cocrystal structures of disubstituted acridines with this quadruplex (Figure 1), to examine how the T_4 diagonal loop accommodates differences in substituent size in a range of closely related compounds (ligands **1**−**7**). The results are compared to the geometry observed with the trisubstituted acridine compound BRACO-19 bound to the human bimolecular quadruplex that has purely propeller loops (*24*).

MATERIALS AND METHODS

General. The synthesis of the disubstituted acridines has been previously described (*33*). All compounds were analytically pure and were used as the hydrochloride salts, apart from compound **1**, which was used as the free base. The oligonucleotide d(GGGGTTTTGGGG) was purchased from Eurogentec and HPLC-purified by them. It was annealed at 90 °C prior to use.

Crystallography. All crystals were obtained at 12 °C by the hanging-drop vapor-diffusion method. Typical conditions were 2 μL drops containing 5% (v/v) methyl-2,4-pentanediol (MPD), 0.50 mM oligonucleotide, 0.25 mM ligand, 40 mM KCl, 5 mM $MgCl_2$, and 4.1 mM spermine equilibrated against 35% (v/v) MPD, at pH 7.0.

Crystallographic data for all except one structure were collected on an in-house Rigaku R-AXIS-IV image plate system with a rotating-anode source using Osmic mirror optics to produce monochromatic Cu Kα radiation at a wavelength of 1.54178 Å. Data for Protein Data Bank (PDB) entry 3EUM (ligand **6**) were collected on an Oxford Diffraction Nova system equipped with a ccd detector and sealed tube/microfocus system. All crystals were flash-frozen using glycerol cryoprotectant and were maintained at 105 K during data collection. Unit cell parameters and data collection information are given in Table 1. Processing and data reduction were carried out using the DTREK part of the Crystal Clear software package (Rigaku Inc.) for all data sets apart from that for ligand **7**, which was processed with Crysalis (Oxford Diffraction).

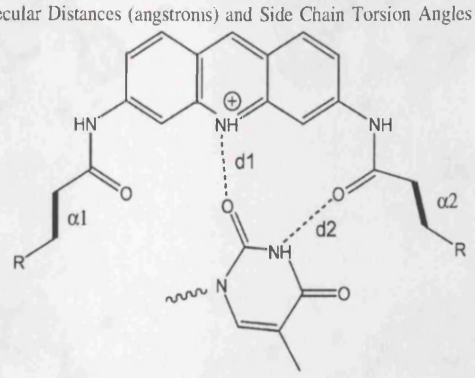
Structures were determined by molecular replacement using PHASER (*34*) from the CCP4 package (*35*) and the known acridine−*Oxytricha* structure (*27*) as a starting point. In each case, the top solution was the one that satisfactorily refined and, for all the structures apart from that with the larger unit cell, was in the same position in the cell as determined previously in the complex with compound **7** (*27*). Electron density and difference maps were used to define the ligand structure and locate solvent molecules. All structures were refined in the same manner using the CCP4 package (*35*) and REFMAC (*36*). Crystallographic data are given in Table 1. All structures have been deposited in the Protein Data Bank with both coordinates and structure factors being available.

RESULTS AND DISCUSSION

A total of six ligands were cocrystallized with the *Oxytricha* telomeric DNA sequence $d(G_4T_4G_4)$ and form bimolecular quadruplex crystal structures isomorphous with the native structure (PDB entries 1JPQ and 1JRN) (*26*) and the previously reported complex with compound **7**, in space group $P2_12_12$ (PDB entry 1L1H) (*27*). Crystals of a second form with compound **2** were also obtained, having two independent quadruplex complexes in the asymmetric unit. These two structures are closely similar to the other complexes.

All structures (Table 2) have the acridine compounds bound in a manner identical to that previously found (*27*) for compound **7** (Figure 2), with one ligand bound per bimolecular quadruplex, and stacked between a terminal G-quartet of the quadruplex and thymine bases from the T_4 loop. The acridine chromophore forms a pair of hydrogen bonds in all the structures with a loop thymine base that is in the same plane as the acridine (Table 2 and Figure 2). The formation of a hydrogen bond between thymine O2 and the central ring nitrogen atom of the acridine [with a mean distance of 3.0(1) Å in all the complexes] suggests that this nitrogen atom is protonated. A similar arrangement has been

Table 2: Selected Ligand–Thymine Intermolecular Distances (angstroms) and Side Chain Torsion Angles (deg)[a]



	1	2a	2b	3	4	5	6	7	BRACO-19
PDB entry	3EM2	3EQW	3EUI	3ERU	3ES0	3ET8	3EUM	1L1H	3CE5
d_1	3.1	3.1	3.1, 2.9	2.9	3.0	2.9	2.9	3.0	NA
d_2	2.7	2.7	2.8, 2.8	2.8	2.7	2.7	2.7	2.7	NA
α_1	238	230	225, 215	230	232	243	238	160	217
α_2	180	161	153, 145	175	112	187	180	212	172

[a] Structure 3EUI has two quadruplex–ligand complexes in the asymmetric unit. Structure 3CE5 has water-mediated hydrogen bonds between the thymine ring and the acridine group, so distances d_1 and d_2 are not directly comparable.

served in the telomeric BRACO-19 acridine quadruplex mplex (*24*), but with a water molecule bridging by drogen bonds the central ring of the acridine and the ymine ring. Structural changes in the *Oxytricha* quadruplex lf are minimal between all the ligand–complex crystal uctures and the native structure, and the overall root-mean-are (rsmd) is <0.2 Å for all structures. Amide and side ain conformations in the acridine derivatives and first-ll water environments are also unchanged in all the uctures.

The morpholino groups in compound **1** are expected to neutral at physiological pH, whereas the ring nitrogen om in all the other compounds should be protonated. This fference is reflected in the consistent findings of superior adruplex binding and biological responses of acridine-sed and many quadruplex-binding ligands that have ionic charged side chains (*19*–*23*). It is notable that there e no direct amine nitrogen–phosphate contacts in any of ese structures or any indirect amine–water–phosphate dges, which have been observed in the BRACO-19 mplex crystal structure (*24*). We cannot exclude the ssibility that some water molecules in the vicinity of the ine groups in these diacridine structures have not been served due to increased water mobility or the resolution its of the data. Otherwise, it is suggested that the charged ine groups in compounds **1**–**7** exert their effect on ligand nding through long-range electrostatic interactions. Alter-ive end groups with an amine nitrogen atom next to the de chain would bring the charged nitrogen closer to a sphate group; such analogues do not appear to have been vestigated to date.

The conformations of the cyclic amine end groups and de chains of the acridine ligands all cluster together (Figure although the methyl groups attached to the piperidino ngs are oriented in a variety of directions (Table 2). This nfirms what can be observed visually, that there is a gnificant volume of space that is accessible to all the rings d their substituents. The para methyl compound **4** has its bstituted piperidino ring flipped into a region with a larger olume that can accommodate its para substitution, as shown by the difference of ca. 60° in side chain torsion angle α_2. Both ortho ethyl substituents in compound **2** are relatively close to backbone and sugar oxygen atoms, suggesting that a suitable ortho substituent such as a methylene hydroxyl group could form additional hydrogen bonds to O3′ or phosphate oxygen atoms.

The regions in the telomeric quadruplex acridine complex (*24*) around the BRACO-19 ligand have some resemblance to those around the disubstituted acridines. In both cases, the in-plane thymine originates from a loop, but whereas the thymines require only minimal conformational change compared with the native state to achieve this since the diagonal loop is already in place over the binding site, the parallel loop has to undergo a more profound change to place the thymine in-plane. This and changes in the other parallel loop in the telomeric structure result in the formation of discrete binding pockets that are bounded by phosphate groups from the loops and accommodate the pyrrolidino rings at the end of the BRACO-19 side chains (Figure 4a). Acridine compounds with larger ring substituents have reduced quadruplex affinity; a piperidino ring results in a 7-fold lower K_a, and an eight-membered ring gives a further 3-fold reduction (*21*). By contrast, the results presented here show that the binding site beneath the diagonal loop in the *Oxytricha* quadruplex is able to accommodate, without any distortion, a piperidino ring, even when it has methyl or ethyl substituents, as well as a seven-membered ring. An eight-membered ring has been modeled into an *Oxytricha* binding site (Figure 4b), and this similarly does not produce any structural perturbations; it is too bulky to fit into the telomeric quadruplex pocket.

We ascribe these differences to the distinct disposition of the loops in these structures: the external placement of propeller loops will affect substituent groups attached to a chromophore, whereas the bases in four-nucleotide diagonal loops are involved in π–π stacking with the chromophore itself. Thus, in broad terms, a diagonal loop allows considerable latitude in the length and size of substituents, whereas the presence of a propeller loop would impose severe constraints on substituents. The overall volume of the

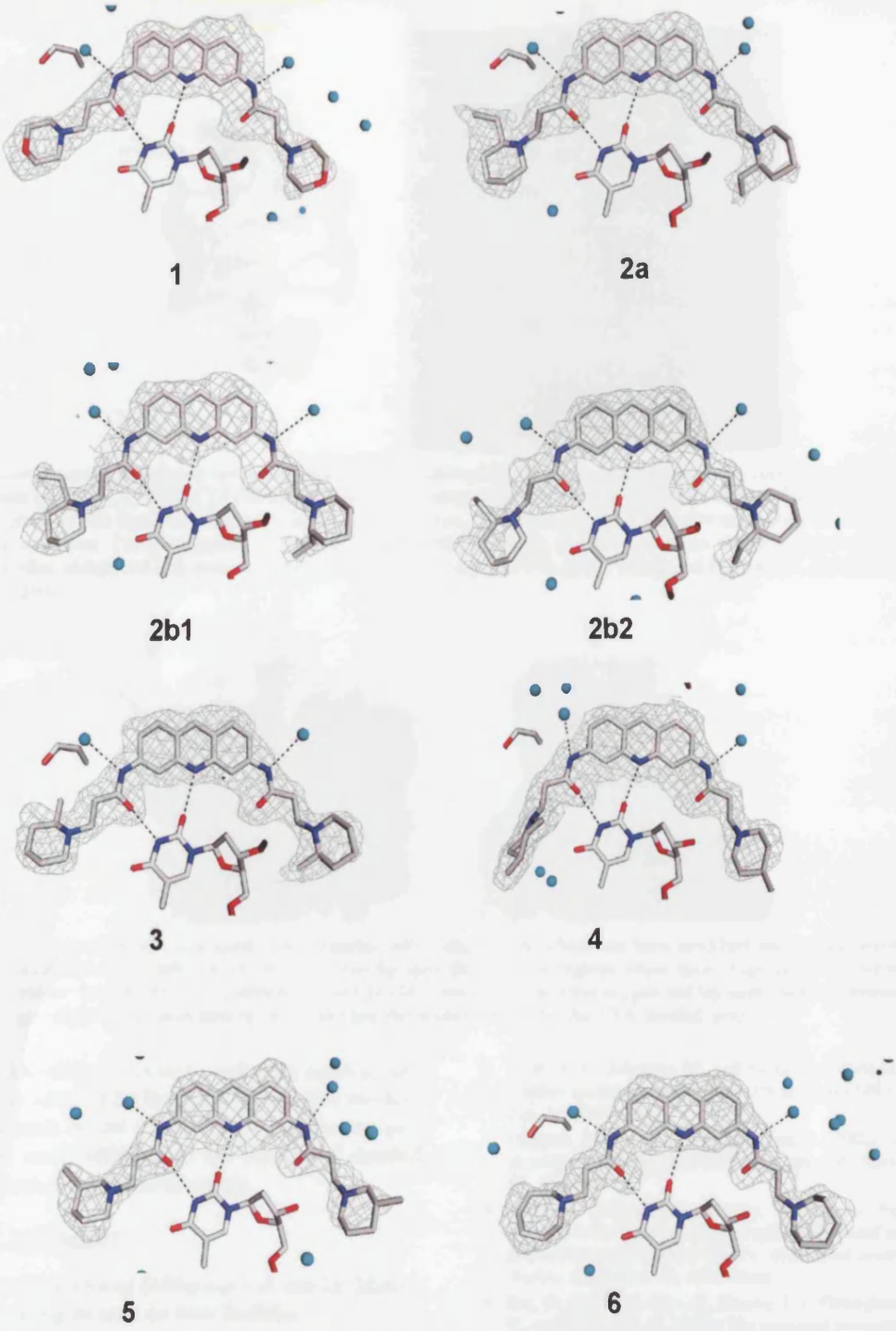


FIGURE 2: Omit map views of the crystal structures, looking onto the acridine plane. In each case, the thymine hydrogen bonding to the acridine chromophore is shown. Electron density is drawn at a 1.5σ level.

overlaid end group rings, taken together, is 66 Å^3 greater than that of the piperidino rings of BRACO-19 in its complex, and the surface area is 54 Å^2 greater. Examination of the *Oxytricha* structures indicates that there is no steric barrier to an individual end group being larger than suggested by this total overlaid volume.

The exact length and sequence of a propeller loop will define the size and surface charge characteristics of the pocket that it can form. Since these loops are formed by short ≤3-nucleotide sequences (*37*−*39*), it is likely that there will be similar steric constraints on the size of ligand substituents for binding to all quadruplexes that contain propeller loops (for example, refs *8*, *9*, *12*, *17*, *25*, and *40*−*45*). Structural studies have suggested that diagonal loops may be rather less prevalent among human genomic quadruplexes, although one is present in the quadruplex formed by the five guanine tracts in the *c-myc* promoter (*13*), with the porphyrin TMPyP4 ligand bound in the loop environment. The porphyrin substituent groups in this structure are fully exposed, analogous to the end groups in the structures presented here.

We suggest that the principle of loops constraining side chain substitution should be useful in the future design of small molecules targeting individual quadruplexes, when

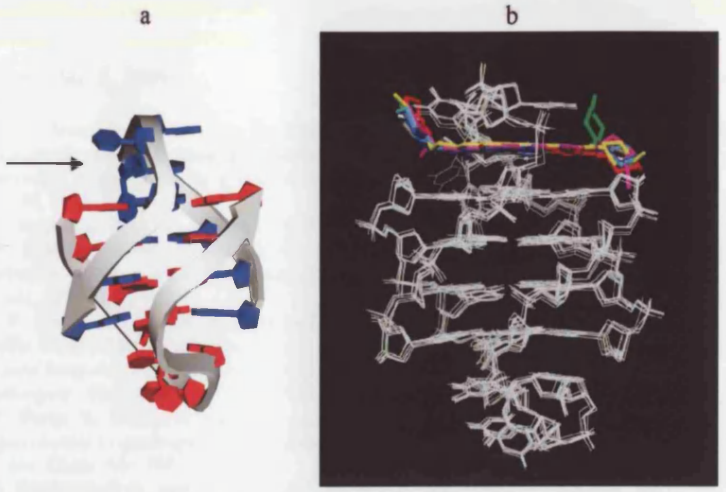


FIGURE 3: (a) Cartoon representation of the bimolecular *O. nova* quadruplex, drawn using crystallographic coordinates of structur 33EUM. The bases of one individual d(GGGGTTTTGGGG) molecule are colored red, and the second is colored blue. The view of the molecule is the same as in panel b. The ligand binding site is indicated by an arrow. (b) Overlay of the quadruplex complexes in all eight disubstituted acridine crystal structures. The quadruplexes are shown in white wireframe form, and the ligands are shown in color stick representation: compound **1** in blue, compound **2** in mauve, compound **3** in cyan, compound **4** in green, compound **5** in yellow, compound **6** in red, and compound **7** in gray.

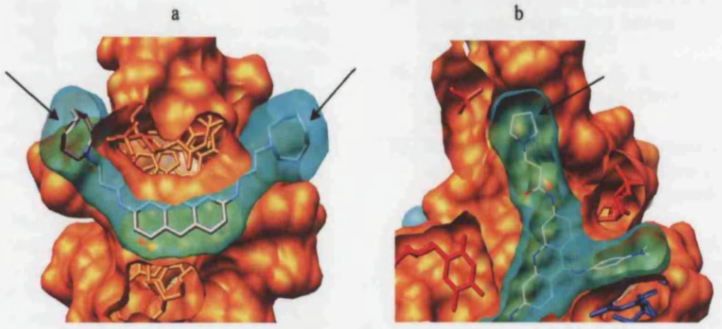


FIGURE 4: (a) Structure of the *O. nova* quadruplex complex with compound **6**, which has been modified and energy-minimized to have terminal eight-membered rings (indicated by arrows). Note the open nature of the regions where these rings reside. (b) Solvent-accessible surface of the acridine−human telomeric quadruplex structure (*24*), showing the acridine in cyan and the close packing between the surfaces of the terminal pyrrolidino ring (indicated by an arrow) and the pocket formed by the TTA parallel loop.

there is prior knowledge of overall quadruplex topology and therefore of the nature of the loops. Further insights into how ligand substituents impart selectivity for a particular quadruplex must await additions to the number of detailed quadruplex three-dimensional structures.

ACKNOWLEDGMENT

We are grateful to Oxford Diffraction Ltd. and Dr. Marcus Winter for granting us time on their facilities.

REFERENCES

1. Burge, S. E., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006) Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res. 34*, 5402–5415.
2. Patel, D. J., Phan, A. T., and Kuryavyi, V. (2007) Human telomere, oncogenic promoter and 5′-UTR G-quadruplexes: Diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res. 35*, 7429–7455.
3. Oganesian, L., and Bryan, T. M. (2007) Physiological relevance of telomeric G-quadruplex formation: A potential drug target. *BioEssays 29*, 155–165.
4. de Cian, A., Lacroix, L., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J.-F., and Mergny, J.-L. (2008) Targeting telomeres and telomerase. *Biochimie 90*, 131–155.
5. Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res. 33*, 2908–2916.
6. Todd, A. K., Johnston, M., and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res. 33*, 2901–2907.
7. Huppert, J. L., and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res. 35*, 406–413.
8. Guo, K., Gokhale, V., Hurley, L. H., and Sun, D. (2008) Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic Acids Res. 36*, 4598–4608.
9. Sun, D., Liu, W. J., Guo, K., Rusche, J. J., Ebbinghaus, S., Gokhale, V., and Hurley, L. H. (2008) The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents. *Mol. Cancer Ther. 7*, 880–889.
10. 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.
11. Hurley, L. H., Von Hoff, D. D., Siddiqui-Jain, A., and Yang, D. (2006) Drug targeting of the c-MYC promoter to repress gene expression via a G-quadruplex silencer element. *Semin. Oncol. 33*, 498–512.
12. Phan, A. T., Modi, Y. S., and Patel, D. J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc. 126*, 8710–8716.
13. Phan, A. T., Kuryavyi, V., Gaw, H. Y., and Patel, D. J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol. 1*, 167–173.

14. Ambrus, A., Chen, D., Dai, J., Jones, R. A., and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry 44*, 2048–2058.
15. Rankin, S., Reszka, A. P., Huppert, J., Zloh, M., Parkinson, G. N., Todd, A. K., Ladame, S., Balasubramanian, S., and Neidle, S. (2005) Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc. 127*, 10584–10589.
16. Fernando, H., Reszka, A. P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A. R., Neidle, S., and Balasubramanian, S. (2006) A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry 45*, 7854–7860.
17. Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D. J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc. 129*, 4386–4392.
18. Todd, A. K., Haider, S. M., Parkinson, G. N., and Neidle, S. (2007) Sequence occurrence and structural uniqueness of a G-quadruplex in the human c-kit promoter. *Nucleic Acids Res. 35*, 5799–5808.
19. Monchaud, D., and Teulade-Fichou, M. P. (2008) A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem. 6*, 627–636.
20. Ou, T., Lu, Y., Tan, J., Huang, Z., Wong, K., and Gu, L. (2008) G-Quadruplexes: Targets in anticancer drug design. *ChemMedChem 3*, 690–713.
21. Read, M. A., Wood, A. A., Harrison, J. R., Gowan, S. M., Kelland, L. R., Dosanjh, H. S., and Neidle, S. (1999) Molecular modeling studies on G-quadruplex complexes of telomerase inhibitors: Structure-activity relationships. *J. Med. Chem. 42*, 4538–4546.
22. Read, M., Harrison, R. J., Romagnoli, B., Tanious, F. A., Gowan, S. H., Reszka, A. P., Wilson, W. D., Kelland, L. R., and Neidle, S. (2001) Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proc. Natl. Acad. Sci. U.S.A. 98*, 4844–4849.
23. Moore, M. J., Schultes, C. M., Cuesta, J., Cuenca, F., Gunaratnam, M., Tanious, F. A., Wilson, W. D., and Neidle, S. (2006) Trisubstituted acridines as G-quadruplex telomere targeting agents. Effects of extensions of the 3,6- and 9-side chains on quadruplex binding, telomerase activity, and cell proliferation. *J. Med. Chem. 49*, 582–599.
24. Campbell, N. H., Parkinson, G. N., Reszka, A. P., and Neidle, S. (2008) Structural basis of DNA quadruplex recognition by an acridine drug. *J. Am. Chem. Soc. 130*, 6722–6724.
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. Haider, S., Parkinson, G. N., and Neidle, S. (2002) Crystal structure of the potassium form of an *Oxytricha nova* G-quadruplex. *J. Mol. Biol. 320*, 189–200.
27. Haider, S. M., Parkinson, G. N., and Neidle, S. (2003) Structure of a G-quadruplex-ligand complex. *J. Mol. Biol. 326*, 117–125.
28. Fedoroff, O. Y., Salazar, M., Han, H., Chemeris, V. V., Kerwin, S. M., and Hurley, L. H. (1998) NMR-based model of a telomerase-inhibiting compound bound to G-quadruplex DNA. *Biochemistry 37*, 12367–12374.
29. Gavathiotis, E., Heald, R. A., Stevens, M. F., and Searle, M. S. (2003) Drug recognition and stabilisation of the parallel-stranded DNA quadruplex d(TTAGGGT)$_4$ containing the human telomeric repeat. *J. Mol. Biol. 334*, 25–36.
30. Parkinson, G. N., Ghosh, R., and Neidle, S. (2007) Structural basis for binding of porphyrin to human telomeres. *Biochemistry 46*, 2390–2397.
31. Parkinson, G. N., Cuenca, F., and Neidle, S. (2008) Topology conservation and loop flexibility in quadruplex-drug recognition: Crystal structures of inter- and intramolecular telomeric DNA quadruplex-drug complexes. *J. Mol. Biol. 381*, 1145–1156.
32. Dash, J., Shirude, P. S., Hsu, S. T., and Balasubramanian, S. (2008) Diarylethynyl amides that recognize the parallel conformation of genomic promoter DNA G-quadruplexes. *J. Am. Chem. Soc. 130*, 15950–15956.
33. Harrison, R. J., Gowan, S. M., Kelland, L. R., and Neidle, S. (1999) Human telomerase inhibition by substituted acridine derivatives. *Bioorg. Med. Chem. Lett. 9*, 2463–2468.
34. Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D60*, 432–438.
35. Collaborative Computational Project No. 4 (1995) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D50*, 760–763.
36. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D53*, 240–255.
37. Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc. 126*, 16405–16415.
38. Risitano, A., and Fox, K. R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res. 32*, 2598–2606.
39. Bugaut, A., and Balasubramanian, S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry 47*, 689–697.
40. Phan, A. T., Luu, K. N., and Patel, D. J. (2006) Different loop arrangements of intramolecular human telomeric (3 + 1) G-quadruplexes in K^+ solution. *Nucleic Acids Res. 34*, 5715–5719.
41. Phan, A. T., Kuryavyi, V., Luu, K. N., and Patel, D. J. (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K^+ solution. *Nucleic Acids Res. 35*, 6517–6525.
42. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R. A., and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res. 34*, 2723–2735.
43. Dai, J., Carver, M., Punchihewa, C., Jones, R. A., and Yang, D. (2007) Structure of the hybrid-2 type intramolecular human telomeric G-quadruplex in K^+ solution: Insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res. 35*, 4927–4740.
44. Cogoi, S., and Xodo, L. E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res. 34*, 2536–2549.
45. Qin, Y., and Hurley, L. H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie 90*, 1149–1171.

BI802233V