

Published in final edited form as:

Nat Rev Drug Discov. 2011 April ; 10(4): 261–275. doi:10.1038/nrd3428.

Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?

Shankar Balasubramanian,

Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK; Cancer Research UK, Cambridge Research Institute, Li Ka Shing Center, Cambridge, CB2 0RE, UK; and the School of Clinical Medicine, University of Cambridge, Cambridge, CB2 0SP, UK; sb10031@cam.ac.uk

Laurence H. Hurley, and

College of Pharmacy, University of Arizona, Tucson, Arizona 85721, USA; the BIO5 Institute, University of Arizona, Tucson, Arizona 85721, USA; and the Arizona Cancer Center, The University of Arizona, Tucson, Arizona 85724, USA; hurley@pharmacy.arizona.edu

Stephen Neidle

Cancer Research UK Biomolecular Structure Group, the School of Pharmacy, University of London, London, WC1N 1AX, UK; and the Centre for Cancer Medicines, The School of Pharmacy, University of London, London, WC1N 1AX, UK; stephen.neidle@pharmacy.ac.uk

Abstract

G-quadruplexes are four-stranded DNA structures that are over-represented in gene promoter regions and are viewed as emerging therapeutic targets in oncology, as transcriptional repression of oncogenes through stabilization of these structures could be a novel anticancer strategy. Many gene promoter G-quadruplexes have physicochemical properties and structural characteristics that might make them druggable, and their structural diversity suggests that a high degree of selectivity might be possible. Here, we describe the evidence for G-quadruplexes in gene promoters and discuss their potential as therapeutic targets, as well as progress in the development of strategies to harness this potential through intervention with small-molecule ligands.

DNA was the first defined target for anticancer drugs (for a historical perspective, see REF. ¹). However, with the advent of new molecular targets — such as kinases and cell surface receptors — that promised greater selectivity for cancer cells, interest in DNA-targeted drugs faded, even though they are still the mainstay of most treatment regimens².

The first glimpse of a new era for DNA-targeted therapeutics came through the realization that telomeres can form four-stranded DNA structures that are termed G-quadruplexes³⁻⁵ (BOX 1). Telomeres are DNA sequences that comprise large numbers of simple guanine-rich tandem repeats at the ends of eukaryotic chromosomes, which protect against degradation and genomic instability. Telomerase plays a major part in maintaining cellular immortalization by catalysing telomere extension, and the recognition in the mid-1990s that it was selectively expressed in the majority of human cancers stimulated interest in

© 2011 Macmillan Publishers Limited. All rights reserved

Competing interests statement

The authors declare competing financial interests: see Web version for details.

FURTHER INFORMATION

ClinicalTrials.gov: <http://clinicaltrials.gov>

RCSB Protein Data Bank (PDB): <http://www.rcsb.org/pdb/home/home.do>

inhibiting its activity as an anticancer strategy. One approach focused on targeting telomeric sequences that could potentially form G-quadruplexes⁶ (BOX 2).

As a result of research on telomeric G-quadruplexes and the cellular consequences of targeting them with small molecules that stabilize these structures (reviewed in REF. 7), their biological and therapeutic significance is well appreciated and continues to be an active field for drug discovery (BOX 2). Interest in the more general therapeutic significance of G-quadruplexes has expanded during the past decade to include G-quadruplexes in gene promoters as targets, which are the focus of this article.

This interest was sparked by a publication showing that transcriptional repression of the *MYC* proto-oncogene (herein, *MYC* refers to *c-MYC*), which is overexpressed in up to 80% of solid tumours, could be achieved by targeting a putative G-quadruplex in the nuclease hypersensitive element (NHE) in the promoter region of *MYC* with a small molecule⁸. Further credence has been lent to the strategy of targeting this class of nucleic acid structures by the more recent identification⁹ and small-molecule targeting¹⁰ of an RNA G-quadruplex in the 5'-untranslated region (UTR) of the small GTPase NRAS, and the identification of tight clusters of putative G-quadruplex-forming sequences immediately upstream and downstream of the transcription start site in proliferation-associated genes through bioinformatics studies (see below). Notably, G-quadruplexes are found in the gene promoters of a wide range of genes that are important in cell signalling, which have representatives from the six hallmarks of cancer¹¹. As well as transcription factors such as *MYC*, which is an undruggable target at the protein level owing to its short half-life and unstructured nature¹², other targets of particular interest include the small GTPase KRAS and the receptor tyrosine kinase KIT, which is a clinically validated drug target for treating gastrointestinal stromal tumours¹³.

Consequently, from a drug-development perspective, if robust strategies for therapeutically targeting gene promoter G-quadruplex structures could be established, their potential might be broad, perhaps even analogous to targeting current major drug target families, such as protein kinases. Moreover, the major advantage of targeting G-quadruplexes, compared with previously targeted families of DNA structures, is that they are discrete folded globular DNA structures, and thus small molecules could be designed to selectively recognize and stabilize these structures, which is analogous to drug targeting of folded protein molecules. Indeed, a first-in-class G-quadruplex-interactive compound, quarfloxin, progressed to Phase II clinical trials for cancer (discussed later), providing proof of principle for the potential therapeutic viability of targeting these structures. However, the compound was withdrawn from further trials owing to problems with bioavailability.

In this article, we first describe the evidence for the existence of G-quadruplexes in eukaryotic promoter regions, highlighting genes for which the presence of G-quadruplexes could have particular therapeutic potential. We then discuss progress in the discovery of small-molecule ligands that target these G-quadruplexes, highlighting the challenges for drug development and how they might be addressed.

Structure of promoter G-quadruplexes

Key structural characteristics of gene promoter G-quadruplexes

The basic elements of a G-quadruplex structure are shown in BOX 1. In principle, G-quadruplexes can be formed from four separate DNA strands (tetramolecular), two separate DNA strands (bimolecular), or a single DNA strand (unimolecular or intramolecular). Gene promoter sequences containing a continuous stretch of a G-quadruplex sequence with four or more G-tracts will normally fold into an intramolecular G-quadruplex, and all studies to

date on such sequences show this type of arrangement¹⁴. However, other possibilities, including bimolecular G-quadruplexes, are also evident in gene promoter regions¹⁵, and it is likely that other arrangements could exist. Indeed, many putative gene promoter G-quadruplex sequences contain more than four G-tracts, which suggests that they can form more than one discrete arrangement and that such arrangements may be in dynamic equilibrium with each other¹⁶.

Only a few detailed structures of gene promoter quadruplexes have been reported to date (BOX 3), and all are from nuclear magnetic resonance (NMR) studies^{17,23}, which contrasts with the success of crystallographic analyses on telomeric G-quadruplexes. Both NMR and crystallographic techniques require overcoming problems of conformational heterogeneity. It is usually necessary to screen several sequences in which ambiguity — resulting from G-tracts that contain differing numbers of guanines — is eliminated, and varying 5' and 3' ends are tried until a stable single conformer is found. Circular dichroism and chemical footprinting are useful techniques that can provide information on overall topology, although conformational heterogeneity can complicate their interpretation²⁴. Using these less precise and less definite techniques, the folding patterns of several promoter G-quadruplexes have been proposed. These include folding patterns for the genes encoding MYC²⁵, B cell lymphoma 2 (BCL-2)²⁶, vascular endothelial growth factor (VEGF)²⁷, hypoxia-inducible factor 1 α (HIF1 α)²⁸, the transcription factor MYB²⁹, platelet-derived growth factor α polypeptide (PDGFA)³⁰, PDGF receptor β polypeptide (PDGFR β)³¹, KRAS^{32,34}, retinoblastoma protein 1 (RB1)^{35,36} and human telomerase reverse transcriptase (TERT)³⁷. Although MYB was only discovered to have a G-quadruplex sequence downstream of the transcription start site in 2008, the unusual heptad–tetrad structure of the *MYB* G-quadruplex was originally solved by NMR in 2001 (REF. 20) and was the first naturally occurring parallel G-quadruplex to be characterized (BOX 3).

Unlike telomeric G-quadruplexes, which can be formed from the single-stranded DNA template at the 3' end of human telomeres, G-quadruplexes in gene promoter regions are constrained by the duplex nature of genomic DNA. Thus, an important question is how secondary structures such as G-quadruplexes can energetically compete with duplex DNA, particularly in view of the GC-rich nature of the putative sequences that form these structures, as at first sight they seem to have higher melting temperatures than AT-rich regions. Although conditions that mimic molecular crowding owing to high concentrations of macromolecules in the nucleus may stabilize these structures once they are formed, other dynamic forces that are derived from negative superhelicity generated by transcription are likely to be involved in the evolution of these structures from duplex DNA (BOX 4).

Evidence for G-quadruplexes in vivo

Most studies on G-quadruplexes have been carried out in *ex vivo* systems, although the occurrence of these structures in telomeres has been unequivocally demonstrated in ciliates³⁸. In the absence of direct evidence for G-quadruplexes in promoter regions of genes, the identification of proteins or enzymes that have important roles in facilitating either the formation of these structures or their unfolding (once they are formed) can provide strong circumstantial evidence of their presence in cells. This evidence would be strengthened if specificity could be demonstrated and the proteins are found to have the expected biological consequences in cells — for example, in silencing or activating gene transcription.

The best characterized system in this regard is the *MYC* G-quadruplex, in which both types of G-quadruplex-modulating proteins have been identified and their roles in cells have been demonstrated, as a consequence of their interactions with the *MYC* G-quadruplex^{39,40} (FIG. 1A). The first potential *MYC* transcriptional activation protein that was not a classical

transcription factor was identified in 1993 (REF. 41). This protein, NM23H2, which belongs to the non-metastasis 23 (NM23) family of proteins, was shown — via a reporter assay — to bind to the nuclease hypersensitive element III₁ (NHE III₁) sequence of the *MYC* promoter to stimulate promoter activity^{41, 42}. Before NM23H2 was identified as a transcription factor, it was known as nucleoside diphosphate (NDP) kinase, a housekeeping enzyme that was important for the maintenance of nucleotide pools⁴³. Mutational studies have shown that these two activities of the protein are in distinct domains⁴⁴. In 1995, NM23H2 was shown to be a footprint in the NHE III₁ of the translocated *MYC* allele in Burkitt's lymphoma cells⁴⁵. More recently, it has been shown that NM23H2 binds to the two single-stranded elements of the NHE III₁ sequence and that this binding affinity is lost following mutation of Arg88 to Ala88 in the active site, which binds to the phosphate of a key nucleotide in the 47-mer deoxynucleotide oligomer substrate³⁹. Significantly, NM23H2 did not bind to the *MYC* G-quadruplex, and in experiments in which small G-quadruplex-interactive molecules, such as TMPyP4, or ionic conditions, such as potassium chloride, stabilized the G-quadruplex (see later), NM23H2 was prevented from binding to this single-stranded template, thus suggesting that this enzyme could facilitate unwinding of the G-quadruplex and the i-motif (which is formed from the complementary C-rich strand)⁴⁶.

Using chromatin immunoprecipitation (ChIP) assays with an NM23H2-specific antibody, Chowdhury and colleagues⁴⁷ have shown that this protein binds to the *MYC* promoter. They also conclude, on the basis of luciferase assays, that NM23H2 induces *MYC* transcription and unfolds the G-quadruplex. This is proposed to occur in a stepwise manner in which the hexameric protein successively captures the single-stranded GC-rich strands in a weak binding complex that can be transferred to other duplex-binding proteins (for example, transcription factor Sp1) or single-stranded DNA-binding proteins (for example, CCHC-type zinc finger, cellular nucleic acid binding protein (CNBP) and heterogeneous nuclear ribonucleoprotein K (HNRNPK)) to activate transcription (FIG. 1). Although HNRNPK is a single-stranded DNA-binding protein, a recent publication has suggested that CNBP may also bind to the *MYC* G-quadruplex⁴⁸. In line with this role, cellular levels of NM23H2 correlate with *MYC* levels in colon cancer cell lines, and small interfering RNA (siRNA) knockdown experiments of NM23H2 demonstrate inhibition of *MYC* levels in these cell lines³⁹. The inhibition of *MYC* transcription by the small-molecule ligand TMPyP4 (see discussion below) can also be rationalized by the stabilization of the *MYC* G-quadruplex and prevention of NM23H2-induced unwinding of the G-quadruplex (FIG. 1B).

Affinity chromatography has also been used to identify nuclear proteins that bind to the *MYC* G-quadruplex⁴⁰. Nucleolin was identified as the major protein that binds to the *MYC* G-quadruplex, together with a number of other nucleolin-associated proteins. Importantly, nucleolin bound with higher affinity to the *MYC* G-quadruplex than its previously described RNA substrate (nucleolin recognition element) and had a higher selectivity for parallel G-quadruplexes over other naturally occurring G-quadruplexes that have different folded structures⁴⁰. Nucleolin not only stabilized the *MYC* G-quadruplex but also facilitated its formation from the single-stranded oligomer. Cellular experiments demonstrated that nucleolin repressed *MYC* transcription and significantly prevented Sp1-induced *MYC* transcription. Last, ChIP analysis showed that nucleolin bound to the NHE III₁ in the *MYC* promoter in cells. As nucleolin only binds to the G-quadruplex and not to other forms of DNA found in the NHE III₁, and ChIP analysis shows that nucleolin binds to this element in cells, this is the best evidence to date that the *MYC* G-quadruplex exists in a cellular context. The relationship among Sp1, CNBP, HNRNPK, NM23H2, nucleolin, the G-quadruplex/i-motif and *MYC* transcription is shown in FIG. 1A, which represents a general model for how secondary DNA structures can control transcription when the G-quadruplex/i-motif is the silencer element.

More recently, Sun *et al.*⁴⁹ provided strong evidence for the formation of a G-quadruplex in the promoter region of the human *VEGF* gene. This G-quadruplex is closely related to the one found in the *MYC* promoter, and consists of a parallel G-quadruplex that has loop ratios of 1/4/1. Parallel footprinting experiments with dimethyl sulphate (DMS) were carried out at three different levels on the same sequence: a single-stranded template, a duplex contained in a supercoiled plasmid and in cells. At the single-stranded and duplex DNA levels, DMS footprinting was used⁴⁹. Significantly, the DMS-protected G-tracts were almost the same at all three levels. Furthermore, deoxyribonuclease I footprinting and ChIP analysis have demonstrated that nucleolin is bound to the G-quadruplex-forming region in plasmids and in cells.

Genome-wide analysis of putative G-quadruplex-forming regions in gene promoters

The advancement in our understanding of the biophysical principles that link G-quadruplex formation *in vitro* with oligonucleotide sequences has provided a rational basis for constructing algorithms that predict sequences that are likely to fold into intramolecular G-quadruplexes⁵⁰⁻⁵². Various G-quadruplex search algorithms have been published, ranging from those that use more explicit definitions of nucleotide arrangements (for example, $GGG_{3+N_1-7}GGG_{3+N_1-7}GGG_{3+N_1-7}GGG_{3+N_1-7}$)^{52,53} to methods that identify windows of sequences that comprise sets of guanine clusters (for example, $GGGN_xGGGN_yGGGN_zGGG$, in which $12 + x + y + z < \text{size of window}$)^{54,55}. All such algorithms are imperfect and will generate some false positives and false negatives, but they have proven to be valuable for the elucidation of patterns of G-quadruplex motifs with respect to genome structure and function. A rigorous test for a predicted individual G-quadruplex motif will require biophysical analysis of an oligonucleotide of the corresponding sequence.

Early computational surveys of the human genome have revealed that G-quadruplex sequence motifs are prevalent in the genome, with a similar estimate of ~370,000 motifs observed in two independent studies using distinct algorithms^{50,52}. The initial genome-wide surveys have been followed up by computational studies that have concentrated on exploring the incidence of G-quadruplex motifs in regions of genes that are normally associated with function. A functionally important part of the genome that is under particular investigation is the sequence space that is proximal to the transcription start sites of genes, which are associated with core promoter elements. G-quadruplex motifs are enriched in regions that are proximal to the transcription start sites of protein-coding genes in genomes, including prokaryotic⁵⁶, human^{53,56,57}, chimpanzee⁵⁷, rat⁵⁷, mouse⁵⁷ and yeast⁵⁴ genomes. For example, FIG. 2 shows a probability density plot that exhibits a strong positional bias for G-quadruplex sequences, with a peak density of ~50 nucleotides upstream^{53,58}, which is analogous to the positional bias that is found for known transcription regulatory sites⁵⁹ (FIG. 2). Several helpful databases have now been compiled to facilitate the study of G-quadruplexes that are potentially involved in the regulation of genes^{60,61}.

A similar positional bias has been observed in the upstream promoter regions of mouse, rat⁵⁷ and yeast genomes⁵⁴ and in the upstream regions of genes in the *Escherichia coli* genome⁵⁶. A positional bias for G-quadruplexes has been observed in the first intron, which is downstream of gene promoters⁶². Several genomic computational studies have also revealed that G-quadruplex sequence motifs tend to be present outside nucleosome-bound regions in the genomes of *Saccharomyces cerevisiae*⁶³, *Caenorhabditis elegans* and humans⁶⁴. These cross-species observations might be explained in terms of a general regulatory role of these G-quadruplex sequence elements. A computational comparison has revealed that predicted transcription factor binding sites are either proximal to or overlap with G-quadruplex motifs in upstream elements^{62,65}, which would be consistent with a

molecular mechanism whereby G-quadruplex motifs and structure formation modulate transcription factor binding events.

These data demonstrate that, theoretically, G-quadruplex sequences in the human genome are statistically over-represented, but an important and as yet unanswered question is whether all of the sequence motifs that have been identified translate into the formation of stable G-quadruplex structures in a cellular context. Moreover, depending on the environment in which the sequence is located, different G-quadruplex structures can form. There have been a number of studies describing genome-wide experimental data that support a linkage between the formation of G-quadruplex structures in cells and gene expression. One study grouped genes according to the number of G-quadruplex motifs that are proximal to transcription start site regions and showed that the mean expression levels of genes that comprise downstream G-quadruplexes (transcription start site + 500 base pair region) were higher than that of genes that lack G-quadruplexes in this region⁶⁶. Furthermore, the mean expression level increased with the number of downstream G-quadruplex motifs. Other genome-wide expression studies have been reported, in which either a small molecule^{54,56} or protein⁶⁷ ligand that is specific for G-quadruplex DNA has been shown to perturb the expression of genes with a statistically significant bias toward genes with one or more G-quadruplex motifs proximal to the transcription start site.

Interestingly, in one genome-wide survey the prediction of G-quadruplex motifs correlated with the function of human genes — showing, for example, an over-representation in proto-oncogenes and an apparent depletion in tumour suppressor genes⁵⁵. This observation suggests that there may have been evolutionary selection for G-quadruplex structures based on function. It also suggests that G-quadruplexes, which are associated with proto-oncogenes, are a therapeutically relevant class of potential drug targets.

The following section considers selected examples of cancer-related genes for which there is substantial evidence for the existence of G-quadruplexes in gene promoter regions and strong support for the therapeutic potential of targeting the activity of the gene product.

Promoter G-quadruplexes as targets

In the following section, we discuss in more detail the therapeutic potential of targeting G-quadruplexes in *MYC*, *KIT* and *KRAS* to modulate the function of these genes. These genes were selected from the more comprehensive list detailed earlier, which also included *BCL-2*, *VEGF*, *HIF1A*, *MYB*, *PDGFA*, *PDGFRB*, *RBI* and *TERT*, because of their more in-depth evaluation; however, this is not meant to imply that the other genes are less attractive as drug targets.

MYC

Deregulation of the transcription factor MYC is almost always as a consequence of gene amplification, translocations, altered ploidy or enhanced transcription owing to upstream signalling abnormalities^{68,69}. Up to 80% of all solid tumours overexpress MYC, including gastrointestinal, ovarian and breast cancer tumours^{11,70,74}, as do many non-Hodgkin's lymphoma tumours¹¹, and so it has attracted considerable interest as a potential therapeutic target. Importantly, however, MYC expression is not always a negative prognostic factor. For example, in oestrogen-positive breast cancer, MYC expression is crucial for hormone therapy. Consequently, the genetic and therapeutic context of treatment must be considered before anti-MYC therapies are designed⁷⁵.

Finally, it is important to appreciate that because the control of MYC expression is extremely complex⁷⁴, there are other pathways and mechanisms that may be used as an

alternative to those involving the NHE III₁. A more complete description of the therapeutic implications of targeting MYC via the G-quadruplex is given in recent reviews^{11, 76}.

As mentioned above and described in BOX 4 and FIG. 1, the G-quadruplex-forming region in the MYC promoter is the most studied example of such a region. The 33-nucleotide sequence in NHE III₁(TGGGGAGGGTGGGGAGGGTGGGGAAGGTGGGGA) contains six G-tracts of unequal length (BOX 4), which unsurprisingly form a complex mixture of G-quadruplex structures in solution. Two separate NMR studies on the four central guanine tracts have shown that it has an intramolecular parallel topology, with one two-nucleotide and two single-nucleotide propeller loops, whereas another sequence containing the four central guanine tracts similarly forms a parallel quadruplex with a GGA triad stacked on one end^{17, 23, 77}.

It was first demonstrated in 2002 that the G-quadruplex in the MYC promoter element was a silencer element and that stabilization with the G-quadruplex interactive molecule TMPyP4 (FIG. 3) led to inhibition of MYC expression⁸. The ligand- or drug-mediated effect on MYC transcription by stabilization of the G-quadruplex has been independently demonstrated in subsequent work with quindolines and actinomycin D^{78, 79}. In situations in which an isolated G-quadruplex is present in the gene promoter, mutational studies that destabilize the G-quadruplex without preventing transcription factor binding can be used to demonstrate the role of G-quadruplexes in transcriptional activation. Thus, a specific G→T mutation in the MYC promoter that is known to destabilize the G-quadruplex resulted in an increase in transcription, which was partially alleviated by compounds that stabilize the G-quadruplex⁸. Finally, two Burkitt's lymphoma cell lines (CA46 and Ramos) have been used to directly demonstrate the differential effect on MYC gene expression of TMPyP4 and TMPyP2 (a positional isomer of TMPyP4 that is much less effective than TMPyP4 at stabilizing the MYC G-quadruplex). As only the Ramos cell line retains NHE III₁ under constitutive control of the heavy chain immunoglobulin region following translocation, it would be expected that the Ramos cell line would express significantly more MYC than the CA46. Furthermore, TMPyP4 but not TMPyP2 should selectively inhibit MYC expression uniquely in the Ramos cell line, and this is precisely what was observed⁸.

KIT

The *KIT* proto-oncogene encodes a receptor tyrosine kinase that, upon activation by its ligand, stimulates cell proliferation, differentiation and survival⁸⁰. Activating mutations or overexpression can result in aberrant function and oncogenic cellular transformation in melanocytes, mast cells, myeloid cells, germ cells and interstitial cells of Cajal lineages. Constitutive activation of *KIT* is considered to be the primary pathogenic event in gastrointestinal stromal tumours (GIST)⁸¹ and may have an important role in a number of other malignancies in which *KIT* expression is elevated. Consequently, the *KIT* protein has become a major molecular target of focus for GIST therapy, as exemplified by the multitargeted kinase inhibitors imatinib (Glivec/Gleevec; Novartis) and, more recently, sunitinib (Sutent; Pfizer)⁸². However, the clinical challenges posed by the emergence of drug resistance owing to mutations in the *KIT* protein suggest that alternative or indeed complementary approaches would be desirable, even with the availability of second-generation *KIT* inhibitors.

Two G-quadruplex-forming sequences, designated *KIT1* (REF. ⁸³) and *KIT2* (REF. ¹⁸), lie within a stretch that was previously shown to be part of the core promoter of *KIT*⁸⁴ (FIG. 4a). They occur between position -87 and -109 base pairs (*KIT1*) and between -140 and -160 base pairs relative to the transcription start site (*KIT2*) and flank a validated Sp1 binding site with close proximity^{18, 19, 21} (FIG. 4a). Both sequences have been shown to form stable G-quadruplexes *in vitro*, and the three-dimensional structures have been recently

solved for each sequence^{19,21} (BOX 3; FIG. 4b). More recently, Patel and colleagues⁸⁵ have shown that *KIT2* can also form a dimeric structure. The *KIT1* sequence is d(AGGGAGGGCGCTGGGAGGAGGG) with four apparently identical GGG tracts. Contrary to the initial expectation of a simple parallel topology, the core of three stacked G-quartets in the actual NMR structure is not built up from these four tracts alone, but instead involves the G in the CGCT sequence²¹ (FIG. 4a). Only two of the three 3' end guanines in the final G-tract participate in G-quartets and they do so by being at the end of the long pentanucleotide AGGAG lateral-type loop that turns back into the core of the quadruplex to put these two guanines in proximity. This large loop, which is stabilized by two A-G base pairs, forms the side and floor of a pronounced cleft, which is a potential ligand binding site (see later). Nineteen of the twenty-two nucleotides in the *KIT1* structure are conserved, and a search of the human genome database shows that the sequence has a unique occurrence in the human genome, which suggests that the structure itself is unique among gene promoter G-quadruplexes⁸⁶.

A more recently determined NMR structure of the *KIT2* quadruplex¹⁹, which was formed from the G21T mutant of the d(CGGGCGGGCGCGAGGGAGGGG) sequence, shows a more conventional parallel topology, although it is structurally more heterogeneous than the *KIT1* quadruplex. This heterogeneity is probably largely a consequence of the long flexible loop that is formed by the CGCGA sequence. The *KIT2* sequence is similarly of unique occurrence in the human genome and is highly conserved in eukaryotes. A second NMR analysis of this sequence has shown that a distinct *KIT2* quadruplex can also be formed by the dimerization of two monomers, which involves a novel parallel-stranded topology⁸⁵. This structure may well have implications for the folding of G-quadruplexes in G-rich strand-exchanged recombination regions of the genome, although it has little direct relevance to KIT function.

In the case of *KIT2*, through a biophysical study using single-molecule Förster resonance energy transfer (FRET) spectroscopy, it has been shown that the G-quadruplex-forming potential of the sequence was sufficient to promote non-duplex structure formation within the context of an extended DNA duplex⁸⁷. A 100-base-pair double helical DNA sequence, which was taken from the *KIT* promoter comprising one of the *KIT* G-quadruplexes, was flanked by donor and acceptor fluorophores as a FRET-based reporter of local DNA structure. Single-molecule analysis of this system demonstrated subpopulations in which the G-quadruplex region adopted either a duplex or a non-duplex structure. The non-duplex structure was absent when the G-quadruplex motif was mutated to prevent G-quadruplex formation. This property suggests that there is a relationship between the DNA sequence and a non-duplex structure, which may have temporal and/or functional relevance to the onset of *KIT* transcription. Chemical biology studies using small molecules have provided further evidence in support of *KIT* promoter G-quadruplexes as molecular targets (see later).

KRAS

Activating mutations in the gene coding for KRAS — which has a role in many signal transduction pathways — are of particular importance to the pathogenesis of pancreatic carcinomas. Although the KRAS protein has been actively pursued as a molecular target for therapeutics, it has proven to be challenging to find drugs that act at the protein level.

Parallel studies on the polypurine and polypyrimidine NHEs of the murine³³ and human³² *KRAS* promoters have been carried out, which are the focus of discussion here. Biophysical studies have shown that there are two different G-quadruplex structures that can form within the NHE of the human *KRAS* promoter³². Studies in pancreatic cancer cell lines that overexpress *KRAS* showed that *KRAS* mRNA levels were lowered in the presence of TMPyP4 (REF. 32). Furthermore, the *KRAS* DNA G-quadruplex-forming element has been

used as an affinity probe to isolate proteins such as 5'-deoxyribose-5-phosphate lyase Ku70, poly(ADP-ribose) polymerase 1 (PARP1) and heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1) from nuclear extracts³², and in a subsequent study, HNRNPA1 has been shown to unwind G-quadruplex DNA³⁴. In a different system, it is noteworthy that PARP1 has been shown to bind tightly to G-quadruplex DNA ($K_d = 65$ nM) and to be enzymatically activated following binding⁸⁸. Recent work on the murine *Kras* promoter has demonstrated that both PARP1 and the *MYC*-associated zinc finger can recognize the *KRAS* promoter quadruplex, which has led to a proposed transcription mechanism involving this zinc finger, PARP1 and the G-quadruplex structure⁸⁹.

Oligonucleotides that mimic one of the G-quadruplexes found in the human *KRAS* promoter have been shown to compete for protein binding with the G-quadruplexes in the natural promoter, in a pancreatic nuclear extract⁹⁰. In an interesting application of G-quadruplex-forming oligodeoxynucleotides, specific covalently linked polyaromatic stacking units were shown to both stabilize the G-quadruplexes and have a strong antiproliferative effect⁹⁰.

Implications for druggability

Although G-quadruplexes can take numerous conformations, there are common features among the G-quadruplex structures that may be harnessed to develop small molecules that bind to them. For example, a structural feature that is common to all G-quadruplexes — the large planar surface of a terminal G-quartet — has led to the development of many small-molecule families that are based on heteroaromatic systems that complement the G-quartet platform. However, this feature by itself is insufficient for high affinity and unlikely to confer selectivity.

Structural selectivity for a particular G-quadruplex might be achievable in several ways — for example, by restricting access to the G-quartet platform as a consequence of the steric hindrance imposed on them by loops, which will emphasize particular ligand shapes, or via sequence-dependent grooves and loops, which provide binding pockets for ligand side chains. There are as yet few molecular structures for promoter quadruplexes. Those that are currently known (*MYC* and *KIT*) are from NMR studies, and they do show considerable structural diversity such that selective small-molecule targeting could be feasible. This diversity is in contrast to the challenges posed in the kinase field by attempts to differentiate between ATP binding sites within a family of kinases. For example, a study by Kumar *et al.*⁹¹ that used a panel of 242 kinases highlighted the varying degrees of kinase promiscuity of the approved kinase inhibitors pazopanib (Votrient; GlaxoSmithKline), sorafenib (Nexavar; Onyx/Bayer) and sunitinib.

We suggest that the very diversity of G-quadruplex structures could be used to enhance selectivity. In particular, exploitation of the different grooves and loops that are present in G-quadruplex structures has not yet been undertaken in a systematic manner, not least because of the paucity of firm structural information. However, screening of focused libraries of small molecules has resulted in the identification of several high-affinity small molecules for the *KIT1* and *KIT2* G-quadruplexes (see later). In the case of the *KIT1* quadruplex structure, there is a pronounced cleft that may be suitable for ligand binding and high-throughput *in silico* screening. This cleft is distinct from the more usual G-quadruplex binding sites, and has so far not been found to be present in any other known quadruplex structure (FIG. 4b). A future goal would be to determine high-resolution crystal structures for promoter quadruplexes, as this will define groove and loop geometries.

Nevertheless, a major challenge is that it is not feasible, at least for the foreseeable future, to screen or design small molecules against the large number of gene promoters and other G-quadruplex structures that are likely to be prevalent in the human genome. One potential

solution is to focus on those gene promoter G-quadruplex targets that have unique sequences and possess structural features as well, such as the *KIT* quadruplexes, and to design specific molecules solely for them. One approach that could offer the potential of achieving the twin objectives of high affinity and specificity is gene targeting with a peptide nucleic acid probe, and this has been shown *in vitro* to be achievable with these probes⁹². However, if this concept is to be evaluated in cells, there will be significant hurdles to overcome, as short probe sequences may also hybridize to many duplex DNA sequences. Nevertheless, many small molecules are currently available that stabilize G-quadruplexes, and they could perhaps be used as a platform to develop more specific agents. These small molecules, and the most advanced G-quadruplex-interacting molecule developed to date, quarfloxin, are discussed in the next section.

Promoter G-quadruplex–ligand complexes

Small-molecule ligands as tools

The therapeutic potential of gene promoter G-quadruplexes has resulted in a rapidly increasing number of studies in which small-molecule ligands have been used to act as G-quadruplex stabilizers, in several instances with cell-based assays. The porphyrin TMPyP4 (FIG. 3) is a G-quadruplex-targeting ligand that has been used in a large number of studies, especially those concerned with elucidating the biological role of G-quadruplexes.

Quindoline and berberine (a related plant alkaloid) are classic G-quadruplex-binding agents that comprise planar aromatic chromophores with one or two aminoalkyl side chains^{93,94}. Derivatives of both these agents have been reported to show significant antiproliferative properties and to downregulate *MYC* expression in several cancer cell lines⁷⁹, as well as effectively binding to *MYC* *in vitro* and in PCR stop assays. However, their selectivity for one particular type of G-quadruplex type is modest, as shown by the quindoline derivative SYUIQ-5, which inhibits *MYC* promoter activity in a cell-free system and in two leukaemia cell lines⁹⁵, and acts as a telomerase inhibitor and inducer of telomere damage (possibly via stabilization of telomeric G-quadruplex DNA)⁹⁶. All of these ligands have significant affinity for duplex DNA, so their mode of action is complex; however, given the dominant role of *MYC* in many human cancers, their effects on *MYC* expression are likely to be major factors in their antiproliferative activity.

A recent report on the binding of some platinum complexes to *MYC* G-quadruplexes, with significant selectivity over duplex DNA, highlights the potential for ligand structural diversity⁹⁷, although the affinity of these compounds to other promoter quadruplexes remains to be explored.

A more selective agent for the *MYC* G-quadruplex is the telomestatin derivative S2T1-6OTD (FIG. 3), which has been shown to reduce the expression of *MYC* and *TERT* (which codes for the catalytic subunit of telomerase and is transcriptionally regulated by *MYC*) in childhood medulloblastoma and in atypical teratoid–rhabdoid tumour cells; this results in cell-cycle arrest, and has potent antiproliferative effects⁹⁸. Although S2T1-6OTD causes telomere shortening in these cells, this effect is largely *MYC*-dependent. Telomestatin (FIG. 3) has been shown to bind to the G-quadruplex in the *TERT* promoter, and this may mediate at least part of its effect on telomeres³⁷. Significantly, this ligand has no detectable effects on the stability of duplex DNA.

Trisubstituted isoalloxazines (FIG. 3) that bind to and stabilize the G-quadruplexes formed by *KIT1* and *KIT2* have been shown to reduce *KIT* transcript levels in cells that express *KIT*⁹⁹. Recent studies with patient-derived GIST882 cells that contain gain-of-function *KIT* mutations have shown that a naphthalene diimide G-quadruplex ligand (FIG. 3) can arrest

cell growth in a dose-dependent fashion¹⁰⁰. Mechanistic markers showed a suppression of *KIT* mRNA and KIT protein levels as well as effects on telomerase activity, suggesting a dual-mode of action. Recently, benzo[*a*]phenoxazines were identified as potentiators of KIT expression. Biophysical data suggest that these particular ligands preferentially bind to the G-quadruplex formed by *KIT2* (REF. ¹⁰¹).

The studies described in this section are examples of cases in which there is clear biophysical evidence of the associated small-molecule binding to the G-quadruplex DNA target *in vitro*, in some cases with selectivity for the target in question relative to duplex DNA or indeed other G-quadruplex sequences. Although the effects of such small molecules on specific cellular mechanisms and the associated growth inhibitory effects are promising in terms of their development into potential anticancer therapeutics, a number of future challenges must be addressed to confirm the robustness of the G-quadruplex targeting concept. These challenges include the need to definitively demonstrate that such molecules are interacting with the expected molecular targets in cellular DNA and to explore whether selectivity between various G-quadruplexes is attainable — or indeed required — in cells to achieve the desired selectivity windows for therapeutics. In this respect, the development of quarfloxin, which inhibits RNA polymerase (Pol I) transcription and displaces nucleolin from the nucleolus, is a highly interesting test case.

Quarfloxin, a first-in-class G-quadruplex-interacting drug

Quarfloxin (also known as CX-3543 or itarnafloxin), is a first-in-class G-quadruplex-interacting compound that had reached Phase II clinical trials for the treatment of neuroendocrine/carcinoid tumours (ClinicalTrials.gov identifier: NCT00780663)¹⁰².

Quarfloxin was originally derived from a group of fluoroquinolones that were shown to have dual topoisomerase II and G-quadruplex interactions¹⁰³ and was successfully optimized by Cylene Pharmaceuticals to eliminate topoisomerase II activity and maximize selectivity for G-quadruplexes (FIG. 5a). Quarfloxin has higher selectivity than TMPyP4 for all-parallel G-quadruplex structures, including those from *MYC*, *VEGF* and *T30695* promoters, in contrast to other alternatively folded structures¹⁰⁴.

Unexpectedly, quarfloxin was found to concentrate in the nucleolus and selectively inhibit Pol I transcription¹⁰². The inhibition of Pol I has been shown to be due to the disruption of the G-quadruplex–nucleolin complexes by quarfloxin, which in turn leads to cellular redistribution of nucleolin into the nucleoplasm. Examination of putative G-quadruplex-forming sequences in the non-template strand of recombinant DNA demonstrates a mixture of three parallel, two antiparallel and nine mixed (parallel and antiparallel) G-quadruplexes. Nucleolin and quarfloxin were both shown to bind to most of these G-quadruplexes, and quarfloxin was shown to disrupt all but one of these nucleolin–DNA complexes¹⁰². This effect was confirmed in cells by ChIP and immunofluorescence analyses, with the latter demonstrating that quarfloxin redistributed nucleolin within the cell (FIG. 5b). The immediate consequence of the disruption of nucleolin ribosomal RNA G-quadruplex interaction by quarfloxin is selective inhibition of elongation of Pol I-driven transcription. The relocalization of nucleolin to the nucleoplasm is a common response to cellular stress, which results in selective apoptosis through various different pathways^{105–111}.

Intriguingly, it has recently been shown that nucleolin has high selectivity for the *MYC* G-quadruplex, where it inhibits *MYC* transcription (see before)⁴⁰. Although quarfloxin did not inhibit *MYC* gene expression after a period of 2 hours in A549 cells¹⁰², it had previously been shown to inhibit *MYC* mRNA expression by 85% in HCT-116 tumours that were isolated from mice after treatment with quarfloxin¹⁰⁴. FIGURE 5b summarizes the current understanding of the mechanism of action of quarfloxin.

Future directions

Since the cellular proof-of-concept study that demonstrated small-molecule modulation of *MYC* expression in 2002, there has been considerable progress in our understanding of G-quadruplex structures in gene promoter regions as a potential new class of drug targets. However, from a drug discovery perspective, we are still in the very early days, and progress is severely hampered by the lack of structural information on small-molecule–G-quadruplex complexes in gene promoter elements.

Apart from the native structures listed in BOX 3, there is just one structure for a ligand–promoter G-quadruplex complex that is currently available in the Protein Data Bank, that of an *MYC* G-quadruplex–TMPyP4 complex, which is derived from a mutant sequence that forms a folding pattern with similarities to that of *KITI* (REF. 22). However, there are now several structures that are derived mostly from crystallographic analyses of telomeric G-quadruplex ligand complexes^{112,114}. It is hoped that the therapeutic potential shown by quarfloxin and other agents in earlier stages of development will encourage more chemists and biologists to undertake structural studies in this area. BOX 5 summarizes some of the advantages and challenges of gene promoter G-quadruplexes as therapeutic targets compared to enzymes, proteins or telomeric G-quadruplexes in oncology. The promise and potential is high: the challenges are considerable but surmountable.

Although the putative G-quadruplex-forming sequences present an opportunity for a range of drug targets, they also represent a potential problem of drug selectivity. Further progress on the development of small molecules with specificity and high affinity for a particular gene promoter G-quadruplex could come from either the screening of large compound libraries or the use of structure-based design methods, including *in silico* screening against virtual compound libraries.

The G-quadruplex field could thus be viewed as analogous to targeting a particular human kinase. Formerly, this was widely believed to be an unrealistic challenge in view of the similarity of ATP binding sites in many kinases; however, it is now clear that this challenge can be met. This has been illustrated by approved drugs such as gefitinib (Iressa; AstraZeneca), which targets a mutated tyrosine kinase domain in EGFR¹¹⁵, and by targeting the mutated serine-threonine BRAF kinase in patients with melanoma using the kinase inhibitor PLX4032 (REF. 116). Conversely, the clinical successes of some relatively non-selective kinase inhibitors, not least of which is imatinib itself, suggest that a measure of broad-spectrum activity against several oncogenic promoter G-quadruplexes may be therapeutically useful.

Studies with *MYC* suggest that it is the targeting of protein–G-quadruplex complexes (for example, the *NM23H2-MYC*–G-quadruplex) that leads to the biological consequences; therefore, identification of the specific proteins that bind to these sequences to modulate gene expression will be vital for understanding how the drugs exert their biological effects. This is analogous to the telomeric G-quadruplex studies, as well as DNA complexes with topoisomerases I and II.

Although quarfloxin did not proceed past Phase II trials because of bioavailability issues, the toxicity profile was very encouraging and suggests that medicinal chemistry optimization in the future could result in a compound with more favourable pharmacological properties. Quarfloxin was well tolerated by patients, and neither genotoxicity nor major organ toxicities were experienced. Furthermore, objective responses were observed during the Phase I clinical trials¹⁴³.

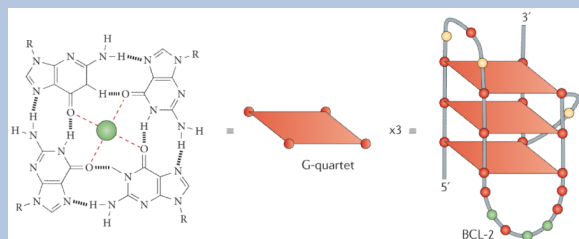
Last, the overwhelming emphasis has been on studies targeting G-quadruplexes in gene promoter regions. As the i-motif appears to exist under negative supercoiling conditions, this represents a second opportunity for involvement in the control of gene expression and drug targeting.

Box 1 | The key structural components of a G-quadruplex

Simple G-quadruplex structures are built from the following components: a core of two or more π - π stacked G-quartets (a very stable planar arrangement of four guanine bases that are hydrogen-bonded via Hoogsteen pairings); essential alkali metal ions (Na^+ or K^+) that coordinate to O6 atoms of the guanine bases and are organized in a linear array running through the centre of the core of G-quartets; and intervening variable-length sequences.

The intervening variable-length sequences hold together the G-quartets and form loops that are arranged on the exterior of the core. The loops are the major elements that define structural variability in G-quadruplexes, and are analogous to side-chains in amino acids. Even one short-length loop (<3 nucleotides) generally results in a G-quadruplex topology with at least two of the four backbones in a parallel orientation, and often with all four parallel to each other; this appears to be the dominant fold in most gene promoter G-quadruplexes examined to date.

The variability in loop sequences results in highly variable (and sometimes flexible) cavities on the exterior of G-quadruplexes that can form part of ligand binding sites. The example shown in the figure is for the dominant G-quadruplex structure that is found in the B cell lymphoma 2 (BCL-2) promoter^{26,117}, in which the loop lengths are 3, 7 and 1 nucleotides. In this case the small loop forms a parallel orientation and the two larger loops form antiparallel orientations. The stability of G-quadruplexes arises from the interplay between these various components. Thus, although other bases — such as thymine — can also form quartets, they are all significantly less stable than a G-quartet and, crucially, they are unable to effectively form the characteristic alkali metal ion channels.



Box 2 | Targeting telomeric G-quadruplexes in cancer

The telomerase enzyme complex, which is selectively expressed in most tumours, maintains telomere homeostasis and hence has a key role in cellular immortalization and transformation¹¹⁸. Telomerase maintains telomere length by virtue of its reverse transcriptase activity and, in humans, it catalyses the synthesis of TTAGGG telomeric DNA repeats. The concept that inhibition of telomerase may be an effective anticancer strategy has been supported by extensive preclinical data on the oligonucleotide GRN163L, which inhibits the catalytic function of telomerase by targeting its template domain (for example, see REF.¹¹⁹).

A distinct approach for telomerase inhibition via telomeric G-quadruplexes⁶ showed that appropriate small molecules can induce the 3' ends of human telomeric DNA (with the

terminal 150–200 nucleotides being single-stranded) to fold up into G-quadruplex-ligand structures, which render the ends inaccessible to telomerase template recognition and hence to telomere repeat addition. Subsequent work in the telomere field in the past decade has focused on the structural characterization of the human telomeric G-quadruplex structures^{120,123,124}, evidence for the occurrence of these structures in ciliates³⁸ and the cellular consequences of targeting these structures (reviewed in REF. 7). As a result of this body of work, the role of ligand-stabilized telomeric G-quadruplexes is well accepted by those involved in both chemical and biological research.

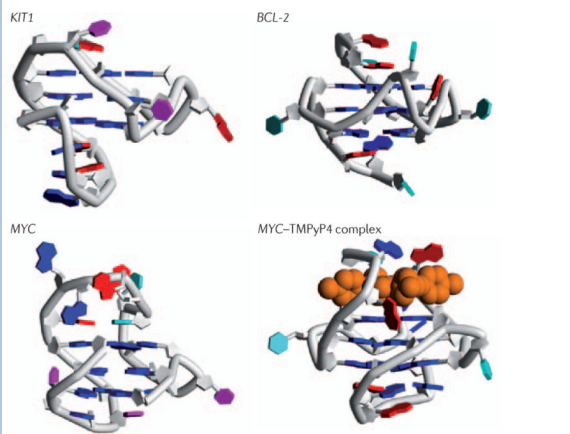
Several hundreds of small molecules that interact with telomeric G-quadruplexes have now been described in the literature. Their binding *in vitro* to human telomeric G-quadruplexes has been extensively explored^{125,126}, although cellular and *in vivo* data are only available for a small number of compounds⁷, in particular the cyclic polyoxazole natural product telomestatin¹²⁷ and the synthetic acridine compounds BRACO19 (REFS 128,129) and RHPS4 (REFS 130,131). There is compelling evidence that these compounds act not only by inhibiting the catalytic function of telomerase but also by uncapping telomerase from the 3' ends of telomeres. They also induce selective and potentially therapeutically important rapid DNA damage responses in cancer cells¹³² involving several pathways, notably the activation of poly(ADP-ribose) polymerase (PARP)¹³³ and ataxia telangiectasia mutated (ATM) signalling^{134,135}.

Evidence of antitumour activity in various xenograft models has been reported for telomestatin, BRACO19 and RHSP4, although no telomere-targeted G-quadruplex ligand has as yet progressed to clinical evaluation, probably because of the undrug-like characteristics of these and many other existing compounds. There have, however, been reports^{132,136} of non-polycyclic compounds with drug-like features and promising *in vitro* profiles.

Box 3 | NMR structures of gene promoter G-quadruplexes

Only a few detailed structures of gene promoter G-quadruplexes have been reported to date, and all are from nuclear magnetic resonance studies^{17,23}, which contrasts with the relative success of crystallographic analyses on telomeric quadruplexes. Below are the sequences and structures of selected gene promoter G-quadruplexes.

Gene	Sequence	Topology	Loops	PDB ID	Refs
MYC	d[₁ TGAG ₁ TC ₁ TGAG ₃ TC ₃ A ₃ G ₃]	Parallel	3 ' P	2A5P	21
MYC	d[₁ TGAG ₁ TC ₁ TAG ₁ TC ₁ TA ₁]	Parallel	3 ' P	1XAV	17
MYC	d[₁ TGAG ₁ TC ₁ JGAG ₃ TC ₃ A ₃ G ₃ A ₁ +TMPyP4]	Parallel	3 ' P	2A5R	22
BCL-2	d[₁ G ₁ CGCG ₁ AG ₁ A ₁ T ₁ G ₁ CG ₁]	Mixed; parallel and antiparallel	2 ' L, 1 ' P	2F8U	117
KIT1	d[₁ AG ₁ AG ₁ CGCTG ₁ AG ₁ AG ₁]	Parallel	2 ' P	2O3M	21
KIT2	d[₁ CG ₁ CG ₁ CGCGAG ₁ AG ₁ T ₁]	Parallel	3 ' P	2KQG and 2KQH	19
KIT2	d[₁ CG ₁ CG ₁ CGCGAG ₁ AG ₁ T ₁] d[₁ CG ₁ CG ₁ CGCTAG ₁ AG ₁ T ₁]	Parallel	3 x P	2KYQ and 2KYP	85
MYB	d[₁ G ₁ AG ₁ AG ₁ AG ₁ A ₁]	Parallel	3 x P	1MYQ	20



BCL-2, B cell lymphoma 2; PDB, Protein Data Bank.

Figure is reproduced, with permission, from REF. ¹⁴ © (2009) Elsevier Science.

Box 4 | Negative superhelicity in G-quadruplex formation: MYC as an example

Telomeric G-quadruplexes are readily available in a single-stranded DNA template, whereas G-quadruplexes in gene promoter elements are constrained by the duplex nature of genomic DNA. An important question, therefore, is how secondary structures such as G-quadruplexes and i-motifs (which are formed from the complementary C-rich strand from hemiprotonated cytosine⁺–cytosine base pairs (bp)) can energetically compete with duplex DNA, particularly in view of the GC-rich nature of the putative sequences that form these structures, as they seem to have higher melting temperatures than AT-rich regions.

Conditions that mimic molecular crowding owing to high concentrations of macromolecules in the nucleus may stabilize these structures once they are formed; however, other dynamic forces derived from negative superhelicity generated by transcription are also highly likely to be involved in the evolution of these structures from duplex DNA. The energetics of duplex to G-quadruplex/i-motif structures has been discussed in more quantitative terms by Lane *et al.*¹³⁷ in terms of kinetic versus thermodynamic control. In addition to negative superhelicity and molecular crowding conditions, a key factor is stabilization of the resulting secondary DNA structures by proteins and ligands.

As a prerequisite to its formation from duplex DNA, the region containing the G-quadruplex/i-motif must first be nucleosome free. The observation that these putative G-quadruplex/i-motif-forming regions are clustered in close proximity upstream to the transcription start site, and that these regions are often associated with nuclease sensitive

sites provides indirect support for the premise that secondary DNA structures such as G-quadruplexes can be formed from duplex DNA by transcriptionally induced negative superhelicity. This is because the superhelical torque is highest in close proximity to the transcription start site and is likely to result in the melting of duplex DNA, which is detected by nucleases.

The promoter configuration for the *MYC* promoter is shown in the figure, part **a**, and shown in the inset is the 33-mer sequence of the purine-rich strand upstream of the P₁ promoter⁶⁸. Levens and colleagues¹³⁸ have elegantly demonstrated that negative superhelicity is generated from the forward movement of the DNA as it screws its way through the multiprotein transcriptional complex that is associated with RNA polymerase, and this superhelicity can be propagated at large distances from the transcription start site in the *MYC* promoter, even in the presence of topoisomerases. To mimic this situation in cells, the wild-type G-quadruplex-forming sequence in the *MYC* promoter and a mutant sequence containing a single guanine mutation in the centre of a critical run of three guanines required for G-quadruplex stability were cloned into a plasmid that inherently contained negative supercoiling⁴⁶. Footprinting experiments using chemical reagents (for example, DMS (dimethyl sulphate), KMnO₄ and Br₂) and enzymatic footprinting (for example, S1 nuclease and deoxyribonuclease I) demonstrated that both the wild-type sequence and a negatively supercoiled plasmid were necessary to facilitate the formation of the *MYC* G-quadruplex⁴⁶. Significantly, chemical footprinting signatures for both the G-quadruplex and the i-motif were present under negative supercoiling conditions in the wild-type sequence; however, these were offset such that the G-quadruplex was constrained in the 5' end of the G-rich strand and the i-motif was confined to the 5' end of the C-rich strand, producing overhangs of 5 and 14 bp at the end of the nuclease hypersensitive element III₁ (NHE III₁) (see the figure, part **b**). In previous studies that used circular dichroism, DMS footprinting and nuclear magnetic resonance techniques in a single-stranded DNA template, only the five 5' runs of three or more contiguous guanines were used^{8, 17, 22, 23, 25}. Under negative supercoiling conditions and in the context of duplex DNA, guanine runs 1–4 are used, and the i-motif uses a sixth run at the 5' end of the cytosine tract, in contrast to the results with single-stranded DNA GC-rich templates¹³⁹, which were not anticipated.

The remaining issue as to why this GC-rich region is the favoured sequence for unwinding and subsequent formation of the G-quadruplex/i-motif — rather than other adjacent AT-rich sequences, which would seem to have a lower energy barrier for formation of the intermediate single-stranded regions — can be addressed from two different perspectives. First, where there are successive runs of guanines that can form 'slipped mismatched structures', this may provide stable intermediates that can lead to these secondary structures¹⁴⁰. Consequently, once formed, the G-quadruplexes can be trapped out by proteins that specifically bind to and stabilize these structures (see main text). Second, under negative supercoiling conditions, the i-motif can form even in the absence of the G-quadruplex⁴⁶, and once one of these structures is formed, it is likely to favour the opposite strand to form the complementary structure. Using single-molecule fluorescence studies, the promotion of non-duplex structure in the *KIT* promoter DNA by a G-quadruplex motif has been demonstrated (see main text)⁸⁷.

Protein targets**Advantages**

- Straightforward if target has a well-defined active site
- Large specialized compound libraries are available (for example, for kinases)
- Structural data are available on many existing targets

Challenges

- Challenging if it involves protein–protein recognition
- Active site changes following ligand binding
- Undruggable if target is unstable or unfolded

Telomerases/telomeric G-quadruplex targeting**Advantages**

- Telomerase is expressed in most human cancers and not in normal somatic cells, so there is a possibility of broad clinical activity

Challenges

- Telomerase is not yet a clinically validated target
- High-throughput assays for telomerase inhibition by small molecules are not readily available and are under-developed

Acknowledgments

Research in the Hurley laboratory has been supported by grants from the US National Institutes of Health (CA95060, GM085585 and CA122952), the National Foundation for Cancer Research (VONHOFF0601) and the Leukemia & Lymphoma Society (6225-08). Research in the Balasubramanian laboratory has been supported by project grants from the Biotechnology and Biological Sciences Research Council of the UK and programme funding from Cancer Research UK. Research in the Neidle laboratory has been supported by programme funding from Cancer Research UK, and an FP6 grant from the European Union on molecular cancer medicine. We are grateful to T. Brooks for her careful reading of the manuscript and D. Bishop for preparing, proof-reading and editing the final version of the text and figures.

References

1. Kohn KW. Beyond DNA cross-linking: history and prospects of DNA-targeted cancer treatment — fifteenth Bruce F. Cain Memorial Award Lecture. *Cancer Res.* 1996; 56:5533–5546. [PubMed: 8971150]
2. Roche, VF. Foye's Principles of Medicinal Chemistry. Lemke, TL.; Williams, DA.; Roche, VF.; Zito, SW., editors. Lippincott Williams & Wilkins; Baltimore: 2008. p. 1147-1192.
3. Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature.* 1988; 334:364–366. [PubMed: 3393228]
4. Sundquist WI, Klug A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature.* 1989; 342:825–829. [PubMed: 2601741]
5. Zahler AM, Williamson JR, Cech TR, Prescott DM. Inhibition of telomerase by G-quartet DNA structures. *Nature.* 1991; 350:718–720. [PubMed: 2023635]
6. Sun D, et al. Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.* 1997; 40:2113–2116. [PubMed: 9216827]
7. Neidle S. Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.* 2010; 277:1118–1125. [PubMed: 19951354]

8. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress *c-MYC* transcription. *Proc. Natl Acad. Sci. USA.* 2002; 99:11593–11598. [PubMed: 12195017]
9. Kumari S, Bugaut A, Huppert JL, Balasubramanian S. An RNA G-quadruplex in the 5' UTR of the *NRAS* proto-oncogene modulates translation. *Nature Chem. Biol.* 2007; 3:218–221. [PubMed: 17322877]
10. Bugaut A, Rodriguez R, Kumari S, Hsu S-TD, Balasubramanian S. Small molecule-mediated inhibition of translation by targeting a native RNA G-quadruplex. *Org. Biomol. Chem.* 2010; 8:2771–2776. [PubMed: 20436976]
11. Brooks TA, Hurley LH. The role of supercoiling in transcriptional control of *MYC* and its importance in molecular therapeutics. *Nature Rev. Cancer.* 2009; 9:849–861. [PubMed: 19907434]
12. Gregory MA, Hann SR. c-Myc proteolysis by the ubiquitin-proteasome pathway: stabilization of c-Myc in Burkitt's lymphoma cells. *Mol. Cell. Biol.* 2000; 20:2423–2435. [PubMed: 10713166]
13. Heinrich MC, et al. Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J. Clin. Oncol.* 2003; 21:4342–4349. [PubMed: 14645423]
14. Neidle S. The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.* 2009; 19:239–250. [PubMed: 19487118]
15. Shklover J, Weisman-Shomer P, Yafe A, Fry M. Quadruplex structures of muscle gene promoter sequences enhance *in vivo* MyoD-dependent gene expression. *Nucleic Acids Res.* 2010; 38:2369–2377. [PubMed: 20053730]
16. Dailey MM, Miller MC, Bates PJ, Lane AN, Trent JO. Resolution and characterization of the structural polymorphism of a single quadruplex-forming sequence. *Nucleic Acids Res.* 2010; 38:4877–4888. [PubMed: 20348136]
17. Ambrus A, Chen D, Dai J, Jones RA, Yang DZ. Solution structure of the biologically relevant G-quadruplex element in the human *c-MYC* promoter. Implications for G-quadruplex stabilization. *Biochemistry.* 2005; 44:2048–2058. [PubMed: 15697230]
18. Fernando H, et al. A conserved quadruplex motif located in a transcription activation site of the human *c-kit* oncogene. *Biochemistry.* 2006; 45:7854–7860. [PubMed: 16784237]
19. Hsu ST, et al. A G-rich sequence within the *c-kit* oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. *J. Am. Chem. Soc.* 2009; 131:13399–13409. [PubMed: 19705869]
20. Matsugami A, et al. An intramolecular quadruplex of (GGA)₄ triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol.* 2001; 313:255–269. [PubMed: 11800555]
21. Phan AT, Kuryavvi V, Burge S, Neidle S, Patel DJ. Structure of an unprecedented G-quadruplex scaffold in the human *c-kit* promoter. *J. Am. Chem. Soc.* 2007; 129:4386–4392. [PubMed: 17362008]
22. Phan AT, Kuryavvi V, Gaw HY, Patel DJ. Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human *MYC* promoter. *Nature Chem. Biol.* 2005; 1:167–173. [PubMed: 16408022]
23. Phan AT, Modi YS, Patel DJ. Propeller-type parallel-stranded G-quadruplexes in the human *c-myc* promoter. *J. Am. Chem. Soc.* 2004; 126:8710–8716. [PubMed: 15250723]
24. Qin Y, Hurley LH. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie.* 2008; 90:1149–1171. [PubMed: 18355457]
25. Seenisamy J, et al. The dynamic character of the G-quadruplex element in the *c-MYC* promoter and modification by TMPyP4. *J. Am. Chem. Soc.* 2004; 126:8702–8709. [PubMed: 15250722]
26. Dexheimer TS, Sun D, Hurley LH. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the *bcl-2* P1 promoter. *J. Am. Chem. Soc.* 2006; 128:5404–5415. [PubMed: 16620112]
27. Sun D, Guo K, Rusche JJ, Hurley LH. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human *VEGF* gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* 2005; 33:6070–6080. [PubMed: 16239639]

28. De Armond R, Wood S, Sun D, Hurley LH, Ebbinghaus SW. Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1 α promoter. *Biochemistry*. 2005; 44:16341–16350. [PubMed: 16331995]
29. Palumbo SL, et al. A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. *Nucleic Acids Res*. 2008; 36:1755–1769. [PubMed: 18252774]
30. Qin Y, Rezler EM, Gokhale V, Sun D, Hurley LH. Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Res*. 2007; 35:7698–7713. [PubMed: 17984069]
31. Qin Y, et al. Molecular cloning of the human platelet-derived growth factor receptor β (PDGFR- β) promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR- β expression. *Biochemistry*. 2010; 49:4208–4219. [PubMed: 20377208]
32. Cogoi S, Paramasivam M, Spolaore B, Xodo LE. Structural polymorphism within a regulatory element of the human *KRAS* promoter: formation of G4-DNA recognized by nuclear proteins. *Nucleic Acids Res*. 2008; 36:3765–3780. [PubMed: 18490377]
33. Cogoi S, Xodo LE. G-quadruplex formation within the promoter of the *KRAS* proto-oncogene and its effect on transcription. *Nucleic Acids Res*. 2006; 34:2536–2549. [PubMed: 16687659]
34. Paramasivam M, et al. Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human *KRAS* promoter: implications for transcription. *Nucleic Acids Res*. 2009; 37:2841–2853. [PubMed: 19282454]
35. Xu Y, Sugiyama H. Structural and functional characterizations of the G-quartet and i-motif elements in retinoblastoma susceptibility genes (*Rb*). *Nucleic Acids Symp. Ser. (Oxf.)*. 2005; 49:177–178.
36. Xu Y, Sugiyama H. Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (*Rb*). *Nucleic Acids Res*. 2006; 34:949–954. [PubMed: 16464825]
37. Palumbo SL, Ebbinghaus SW, Hurley LH. Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *J. Am. Chem. Soc*. 2009; 131:10878–10891. [PubMed: 19601575]
38. Schaffitzel C, et al. *In vitro* generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylyonchia lemnae* macronuclei. *Proc. Natl Acad. Sci. USA*. 2001; 98:8572–8577. [PubMed: 11438689]
39. Dexheimer TS, et al. NM23-H2 may play an indirect role in transcriptional activation of *c-myc* gene expression but does not cleave the nuclease hypersensitive element III. *Mol. Cancer Ther*. 2009; 8:1363–1377. [PubMed: 19435876]
40. González V, Guo K, Hurley LH, Sun D. Identification and characterization of nucleolin as a *c-myc* G-quadruplex-binding protein. *J. Biol. Chem*. 2009; 284:23622–23635. [PubMed: 19581307]
41. Postel EH, Berberich SJ, Flint SJ, Ferrone CA. Human *c-myc* transcription factor PuF identified as nm23-H2 nucleoside diphosphate kinase, a candidate suppressor of tumor metastasis. *Science*. 1993; 261:478–480. [PubMed: 8392752]
42. Berberich SJ, Postel EH. PuF/NM23-H2/NDPK-B transactivates a human *c-myc* promoter-CAT gene via a functional nuclease hypersensitive element. *Oncogene*. 1995; 10:2343–2347. [PubMed: 7784082]
43. Lascu L. The nucleoside diphosphate kinases 1973–2000. *J. Bioenerg. Biomembr*. 2000; 32:213–214. [PubMed: 11768304]
44. Postel EH, Ferrone CA. Nucleoside diphosphate kinase enzyme activity of NM23-H2/PuF is not required for its DNA binding and *in vitro* transcriptional functions. *J. Biol. Chem*. 1994; 269:8627–8630. [PubMed: 8132589]
45. Ji L, Arcinas M, Boxer LM. The transcription factor, Nm23H2, binds to and activates the translocated *c-myc* allele in Burkitt's lymphoma. *J. Biol. Chem*. 1995; 270:13392–13398. [PubMed: 7768941]
46. Sun D, Hurley LH. The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the *c-Myc* promoter: implications for drug targeting and control of gene expression. *J. Med. Chem*. 2009; 52:2863–2874. [PubMed: 19385599]

47. Thakur RK, et al. Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within *c-MYC* promoter nuclease hypersensitive element induces *c-MYC* expression. *Nucleic Acids Res.* 2009; 37:172–183. [PubMed: 19033359]
48. Borgognone M, Armas P, Calcaterra NB. Cellular nucleic-acid-binding protein, a transcriptional enhancer of c-Myc, promotes the formation of parallel G-quadruplexes. *Biochem. J.* 2010; 428:491–498. [PubMed: 20394585]
49. Sun D, Guo K, Shin Y-J. Evidence of the formation of G-quadruplex structures in the promoter region of the human vascular endothelial growth factor gene. *Nucleic Acids Res.* 2011; 39:1256–1265. [PubMed: 20959293]
50. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* 2005; 33:2908–2916. [PubMed: 15914667]
51. Kikin O, Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 2006; 34:W676–W682. [PubMed: 16845096]
52. Todd AK, Johnston M, Neidle S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* 2005; 33:2901–2907. [PubMed: 15914666]
53. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* 2007; 35:406–413. [PubMed: 17169996]
54. Hershman SG, et al. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 2008; 36:144–156. [PubMed: 17999996]
55. Eddy J, Maizels N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* 2006; 34:3887–3896. [PubMed: 16914419]
56. Rawal P, et al. Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res.* 2006; 16:644–655. [PubMed: 16651665]
57. Verma A, et al. Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved *cis*-regulatory elements in human and related species. *J. Med. Chem.* 2008; 51:5641–5649. [PubMed: 18767830]
58. Huppert JL, Bugaut A, Kumari S, Balasubramanian S. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.* 2008; 36:6260–6268. [PubMed: 18832370]
59. Xie X, et al. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature.* 2005; 434:338–345. [PubMed: 15735639]
60. Yadav VK, Abraham JK, Mani P, Kulshrestha R, Chowdhury S. QuadBase: genome-wide database of G4 DNA — occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic Acids Res.* 2008; 36:D381–D385. [PubMed: 17962308]
61. Zhang R, Lin Y, Zhang CT. GregList: a database listing potential G-quadruplex regulated genes. *Nucleic Acids Res.* 2008; 36:D372–D376. [PubMed: 17916572]
62. Eddy J, Maizels N. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.* 2008; 36:1321–1333. [PubMed: 18187510]
63. Halder K, Halder R, Chowdhury S. Genome-wide analysis predicts DNA structural motifs as nucleosome exclusion signals. *Mol. Biosyst.* 2009; 5:1703–1712. [PubMed: 19587895]
64. Wong HM, Huppert JL. Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol. Biosyst.* 2009; 5:1713–1719. [PubMed: 19585004]
65. Todd AK, Neidle S. The relationship of potential G-quadruplex sequences in *cis*-upstream regions of the human genome to SP1-binding elements. *Nucleic Acids Res.* 2008; 36:2700–2704. [PubMed: 18353860]
66. Du Z, Zhao Y, Li N. Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res.* 2008; 18:233–241. [PubMed: 18096746]
67. Fernando H, et al. Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression. *Nucleic Acids Res.* 2009; 37:6716–6722. [PubMed: 19745055]
68. Marcu KB, Bossone SA, Patel AJ. *myc* Function and regulation. *Annu. Rev. Biochem.* 1992; 61:809–860. [PubMed: 1497324]

69. Spencer CA, Groudine M. Control of *c-myc* regulation in normal and neoplastic cells. *Adv. Cancer Res.* 1991; 56:1–48. [PubMed: 2028839]
70. Eilers M, Eisenman RN. Myc's broad reach. *Genes Dev.* 2008; 22:2755–2766. [PubMed: 18923074]
71. Lutz W, Leon J, Eilers M. Contributions of Myc to tumorigenesis. *Biochim. Biophys. Acta.* 2002; 1602:61–71. [PubMed: 11960695]
72. Meyer N, Penn LZ. Reflecting on 25 years with MYC. *Nature Rev. Cancer.* 2008; 8:976–990. [PubMed: 19029958]
73. Pelengaris S, Khan M, Evan G. c-MYC: more than just a matter of life and death. *Nature Rev. Cancer.* 2002; 2:764–776. [PubMed: 12360279]
74. Wierstra I, Alves J. The *c-myc* promoter: still *Myster Y* and Challenge. *Adv. Cancer Res.* 2008; 99:113–333. [PubMed: 18037408]
75. Musgrove EA, et al. Identification of functional networks of estrogen- and *c-Myc*-responsive genes and their relationship to response to tamoxifen therapy in breast cancer. *PLoS ONE.* 2008; 3:e2987. [PubMed: 18714337]
76. Brooks TA, Hurley LH. Targeting MYC expression through G-quadruplexes. *Genes Cancer.* 2010; 1:641–649. [PubMed: 21113409]
77. Kettani A, et al. A dimeric DNA interface stabilized by stacked A.(G.G.G.G.)A hexads and coordinated monovalent cations. *J. Mol. Biol.* 2000; 31:627–644. [PubMed: 10731417]
78. Kang HJ, Park HJ. Novel molecular mechanism for actinomycin D activity as an oncogenic promoter G-quadruplex binder. *Biochemistry.* 2009; 48:7392–7398. [PubMed: 19496619]
79. Ou TM, et al. Stabilization of G-quadruplex DNA and down-regulation of oncogene *c-myc* by quindoline derivatives. *J. Med. Chem.* 2007; 50:1465–1474. [PubMed: 17346034]
80. Yarden Y, et al. Human proto-oncogene *c-kit*: a new cell surface receptor tyrosine kinase for an unidentified ligand. *EMBO J.* 1987; 6:3341–3351. [PubMed: 2448137]
81. Sakurai S, Fukasawa T, Chong JM, Tanaka A, Fukayama M. *C-kit* gene abnormalities in gastrointestinal stromal tumors (tumors of interstitial cells of Cajal). *Jpn J. Cancer Res.* 1999; 90:1321–1328. [PubMed: 10665649]
82. Tuveson DA, et al. STI571 inactivation of the gastrointestinal stromal tumor c-KIT oncoprotein: biological and clinical implications. *Oncogene.* 2001; 20:5054–5058. [PubMed: 11526490]
83. Rankin S, et al. Putative DNA quadruplex formation within the human *c-kit* oncogene. *J. Am. Chem. Soc.* 2005; 127:10584–10589. [PubMed: 16045346]
84. Park GH, Plummer HK, Krystal GW. Selective Sp1 binding is critical for maximal activity of the human *c-kit* promoter. *Blood.* 1998; 92:4138–4149. [PubMed: 9834219]
85. Kuryavyy V, Phan AT, Patel DJ. Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human *c-kit2* promoter. *Nucleic Acids Res.* 2010; 38:6757–6773. [PubMed: 20566478]
86. Todd AK, Haider SM, Parkinson GN, Neidle S. Sequence occurrence and structural uniqueness of a G-quadruplex in the human *c-kit* promoter. *Nucleic Acids Res.* 2007; 35:5799–5808. [PubMed: 17720713]
87. Shirude PS, Okumus B, Ying L, Ha T, Balasubramanian S. Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene *c-kit*. *J. Am. Chem. Soc.* 2007; 129:7484–7485. [PubMed: 17523641]
88. Soldatenkov VA, Vetcher AA, Duka T, Ladame S. First evidence of a functional interaction between DNA quadruplexes and poly(ADP-ribose) polymerase-1. *ACS Chem. Biol.* 2008; 3:214–219. [PubMed: 18338862]
89. Cogoi S, Paramasivam M, Membrino A, Yokoyama KK, Xodo LE. The *KRAS* promoter responds to MYC-associated zinc finger and poly [ADP-ribose]polymerase 1 proteins which recognize a critical quadruplex-forming *GA*-element. *J. Biol. Chem.* 2010; 285:22003–22016. [PubMed: 20457603]
90. Cogoi S, et al. Identification of a new G-quadruplex motif in the *KRAS* promoter and design of pyrene-modified G4-decoys with antiproliferative activity in pancreatic cancer cells. *J. Med. Chem.* 2009; 52:564–568. [PubMed: 19099510]

91. Kumar R, et al. Myelosuppression and kinase selectivity of multikinase angiogenesis inhibitors. *Br. J. Cancer*. 2009; 101:1717–1723. [PubMed: 19844230]
92. Lusvardi S, et al. Loop and backbone modifications of peptide nucleic acid improve G-quadruplex binding selectivity. *J. Am. Chem. Soc.* 2009; 131:18415–18424. [PubMed: 19947597]
93. Caprio V, et al. A novel inhibitor of human telomerase derived from 10*H*-indolo[3,2-*b*]quinoline. *Bioorg. Med. Chem. Lett.* 2000; 10:2063–2066. [PubMed: 10999471]
94. Guyen B, Schultes CM, Hazel P, Mann J, Neidle S. Synthesis and evaluation of analogues of 10*H*-indolo[3,2-*b*]quinoline as G-quadruplex stabilising ligands and potential inhibitors of the enzyme telomerase. *Org. Biomol. Chem.* 2004; 2:981–988. [PubMed: 15034620]
95. Liu J-N, et al. Inhibition of myc promoter and telomerase activity and induction of delayed apoptosis by SYUIQ-5, a novel G-quadruplex interactive agent in leukemia cells. *Leukemia*. 2007; 21:1300–1302. [PubMed: 17392822]
96. Zhou W-J, et al. G-quadruplex ligand SYUIQ-5 induces autophagy by telomere damage and TRF2 delocalization in cancer cells. *Mol. Cancer Ther.* 2009; 8:3203–3213. [PubMed: 19996277]
97. Wang P, Leung C-H, Ma D-L, Yan S-C, Che C-M. Structure-based design of platinum(II) complexes as *c-myc* oncogene down-regulators and luminescent probes for G-quadruplex DNA. *Chem. Eur. J.* 2010; 16:6900–6911. [PubMed: 20437426]
98. Shalaby T, et al. Disabling *c-Myc* in childhood medulloblastoma and atypical teratoid/rhabdoid tumor cells by the potent G-quadruplex interactive agent S2T1-6OTD. *Mol. Cancer Ther.* 2010; 9:167–179. [PubMed: 20053783]
99. Bejugam M, et al. Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: small molecule regulation of *c-kit* oncogene expression. *J. Am. Chem. Soc.* 2007; 129:12926–12927. [PubMed: 17918848]
100. Gunaratnam M, et al. Targeting human gastrointestinal stromal tumor cells with a quadruplex-binding small molecule. *J. Med. Chem.* 2009; 52:3774–3783. [PubMed: 19469547]
101. McLuckie KI, et al. G-quadruplex-binding benzo[*a*] phenoxazines down-regulate *c-KIT* expression in human gastric carcinoma cells. *J. Am. Chem. Soc.* 2011; 133:2658–2663. [PubMed: 21294544]
102. Drygin D, et al. Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res.* 2009; 69:7653–7661. [PubMed: 19738048]
103. Duan W, et al. Design and synthesis of fluoroquinophenoxazines that interact with human telomeric G-quadruplexes and their biological effects. *Mol. Cancer Ther.* 2001; 1:103–120. [PubMed: 12467228]
104. Jin, CH., et al. *In vivo* efficacy of CX-3543, a novel c-Myc oncogene inhibitor; 95th Annual Meeting, Orlando, Florida. *Proc. Am. Assoc. Cancer Res.*; 2004; 2004. Abstr. LB-243
105. Daniely Y, Borowiec JA. Formation of a complex between nucleolin and replication protein A after cell stress prevents initiation of DNA replication. *J. Cell Biol.* 2000; 149:799–810. [PubMed: 10811822]
106. Daniely Y, Dimitrova DD, Borowiec JA. Stress-dependent nucleolin mobilization mediated by p53-nucleolin complex formation. *Mol. Cell. Biol.* 2002; 22:6014–6022. [PubMed: 12138209]
107. Kim K, et al. Novel checkpoint response to genotoxic stress mediated by nucleolin-replication protein A complex formation. *Mol. Cell. Biol.* 2005; 25:2463–2474. [PubMed: 15743838]
108. Kito S, Morimoto Y, Tanaka T, Haneji T, Ohba T. Cleavage of nucleolin and AgNOR proteins during apoptosis induced by anticancer drugs in human salivary gland cells. *J. Oral Pathol. Med.* 2005; 34:478–485. [PubMed: 16091115]
109. Saxena A, Rorie CJ, Dimitrova D, Daniely Y, Borowiec JA. Nucleolin inhibits Hdm2 by multiple pathways leading to p53 stabilization. *Oncogene*. 2006; 25:7274–7288. [PubMed: 16751805]
110. Storck S, Shukla M, Dimitrov S, Bouvet P. Functions of the histone chaperone nucleolin in diseases. *Subcell. Biochem.* 2007; 41:125–144. [PubMed: 17484127]
111. Wang Y, et al. Regulation of DNA replication after heat shock by replication protein A-nucleolin interactions. *J. Biol. Chem.* 2001; 276:20579–20588. [PubMed: 11389152]
112. Campbell NH, Parkinson GN, Reszka AP, Neidle S. Structural basis of DNA quadruplex recognition by an acridine drug. *J. Am. Chem. Soc.* 2008; 130:6722–6724. [PubMed: 18457389]

113. Parkinson GN, Cuenca F, Neidle S. Topology conservation and loop flexibility in quadruplex–drug recognition: crystal structures of inter- and intramolecular telomeric DNA quadruplex–drug complexes. *J. Mol. Biol.* 2008; 381:1145–1156. [PubMed: 18619463]
114. Collie GW, et al. Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures. *J. Am. Chem. Soc.* 2010; 132:9328–9334. [PubMed: 20565109]
115. Lynch TJ, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *New Engl. J. Med.* 2004; 350:2129–2139. [PubMed: 15118073]
116. Flaherty KT, et al. Inhibition of mutated, activated BRAF in metastatic melanoma. *New Engl. J. Med.* 2010; 363:809–819. [PubMed: 20818844]
117. Dai JX, Chen D, Jones RA, Hurley LH, Yang DZ. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* 2006; 34:5133–5144. [PubMed: 16998187]
118. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell.* 2000; 100:57–70. [PubMed: 10647931]
119. Marian CO, et al. The telomerase antagonist, imetelstat, efficiently targets glioblastoma tumor-initiating cells leading to decreased proliferation and tumor growth. *Clin. Cancer Res.* 2010; 16:154–163. [PubMed: 20048334]
120. Ambrus A, et al. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* 2006; 34:2723–2735. [PubMed: 16714449]
121. Parkinson GN, Lee MP, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature.* 2002; 417:876–880. [PubMed: 12050675]
122. Wang Y, Patel DJ. Solution structure of the human telomeric repeat d[AG₃(T₂AG₃)₃] G-tetraplex. *Structure.* 1993; 1:263–282. [PubMed: 8081740]
123. Xu Y, Noguchi Y, Sugiyama H. The new models of the human telomere d[AGGG(TTAGGG)₃] in K⁺ solution. *Bioorg. Med. Chem.* 2006; 14:5584–5591. [PubMed: 16682210]
124. Luu KN, et al. Structure of the human telomere in K⁺ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.* 2006; 128:9963–9970. [PubMed: 16866556]
125. Monchaud D, Teulade-Fichou MP. A hitchhiker’s guide to G-quadruplex ligands. *Org. Biomol. Chem.* 2008; 6:627–636. [PubMed: 18264563]
126. Yang D, Okamoto K. Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med. Chem.* 2010; 2:619–646. [PubMed: 20563318]
127. Tauchi T, et al. Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: *in vitro* and *in vivo* studies in acute leukemia. *Oncogene.* 2006; 25:5719–5725. [PubMed: 16652154]
128. Burger AM, et al. The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res.* 2005; 65:1489–1496. [PubMed: 15735037]
129. Gunaratnam M, et al. Mechanism of acridine-based telomerase inhibition and telomere shortening. *Biochem. Pharmacol.* 2007; 74:679–689. [PubMed: 17631279]
130. Cookson JC, et al. Pharmacodynamics of the G-quadruplex-stabilizing telomerase inhibitor 3,11-difluoro-6,8,13-trimethyl-8*H*-quino[4,3,2-*kl*] acridinium methosulfate (RHPS4) *in vitro*: activity in human tumor cells correlates with telomere length and can be enhanced, or antagonized, with cytotoxic agents. *Mol. Pharmacol.* 2005; 68:1551–1558. [PubMed: 16150933]
131. Salvati E, et al. Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J. Clin. Invest.* 2007; 117:3236–3247. [PubMed: 17932567]
132. Rodriguez R, et al. A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.* 2008; 130:15758–15759. [PubMed: 18975896]
133. Salvati E, et al. PARP1 is activated at telomeres upon G4 stabilization: possible target for telomere-based therapy. *Oncogene.* 2010; 29:6280–6293. [PubMed: 20802516]

134. Pennarun G, et al. Role of ATM in the telomere response to the G-quadruplex ligand 360A. *Nucleic Acids Res.* 2008; 36:1741–1754. [PubMed: 18263609]
135. Rizzo A, et al. Stabilization of quadruplex DNA perturbs telomere replication leading to the activation of an ATR-dependent ATM signaling pathway. *Nucleic Acids Res.* 2009; 37:5353–5364. [PubMed: 19596811]
136. Rahman KM, et al. Biaryl polyamides as a new class of DNA quadruplex-binding ligands. *Chem. Commun.* 2009; 2009:4097–4099.
137. Lane AN, Chaires JB, Gray RD, Trent JO. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* 2008; 36:5482–5515. [PubMed: 18718931]
138. Kouzine F, Sanford S, Elisha-Feil Z, Levens D. The functional response of upstream DNA to dynamic supercoiling *in vivo*. *Nature Struct. Mol. Biol.* 2008; 15:146–154. [PubMed: 18193062]
139. Simonsson T, Pribylova M, Vorlickova M. A nuclease hypersensitive element in the human c-myc promoter adopts several distinct i-tetraplex structures. *Biochem. Biophys. Res. Commun.* 2000; 278:158–166. [PubMed: 11071868]
140. Sinden, RR. *DNA Structure and Function*. Academic Press; San Diego: 1994. p. 259-286.
141. Kendrick S, Hurley LH. Asserting the role of G-quadruplex/i-motif secondary structures as cis-acting regulatory elements. *Pure Appl. Chem.* 2010; 82:1609–1621. [PubMed: 21796223]
142. González V, Hurley LH. The *c-MYC*NHE III: function and regulation. *Annu. Rev. Pharmacol. Toxicol.* 2010; 50:111–129. [PubMed: 19922264]
143. Lim, JK., et al. Quarfloxin phase I clinical data and scientific findings supporting the selection of carcinoid/neuroendocrine tumors as the phase II indication; 100th AACR Annual Meeting 2009 Proceedings; Denver, USA. 2009; p. 868-869. Abstr. 3599

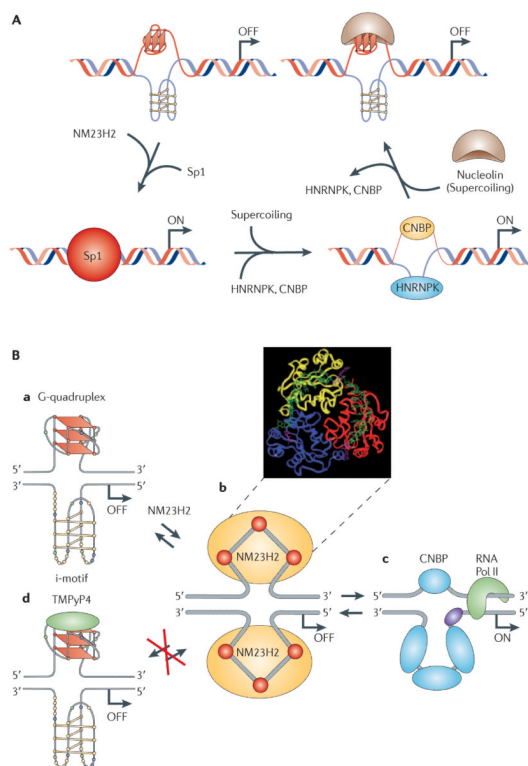


Figure 1. G-quadruplexes in promoter regions: *MYC* as an example

A | A proposed model of *MYC* transcriptional regulation that involves the resolution of the G-quadruplex by NM23H2 (from the non-metastasis 23 (NM23) protein family) for duplex DNA formation and subsequent transcriptional activation by transcription factor Sp1. The binding of heterogeneous nuclear ribonucleoprotein K (HNRNPK) and cellular nucleic acid binding protein (CNBP) to single-stranded DNA induced by negative supercoiling also leads to activation of *MYC* transcription. The stabilization of the G-quadruplex by nucleolin results in negative regulation of *MYC* transcription. **B** | This diagram shows the involvement of NM23H2 and a G-quadruplex-interactive compound in modulating the activation and silencing of the nuclease hypersensitive element III₁ (NHE III₁) in the *MYC* promoter. **Ba** shows the G-quadruplex/i-motif form of the NHE III₁, which is the silencer element. **Bb** shows a molecular model of the purine-rich strand unwound from the G-quadruplex form, which is superimposed on an NM23H2 trimer. **Ba** to **Bc** via **Bb** illustrates the remodelling of the G-quadruplex/i-motif complex by NM23H2, in which a stepwise unfolding of the secondary DNA structure is proposed to take place. Part **A** reproduced, with permission, from REF. ¹⁴¹ © (2010) International Union of Pure and Applied Chemistry. Part **B** is modified, with permission, from REF. ¹⁴² © (2010) Annual Reviews. RNA Pol II, RNA polymerase II.

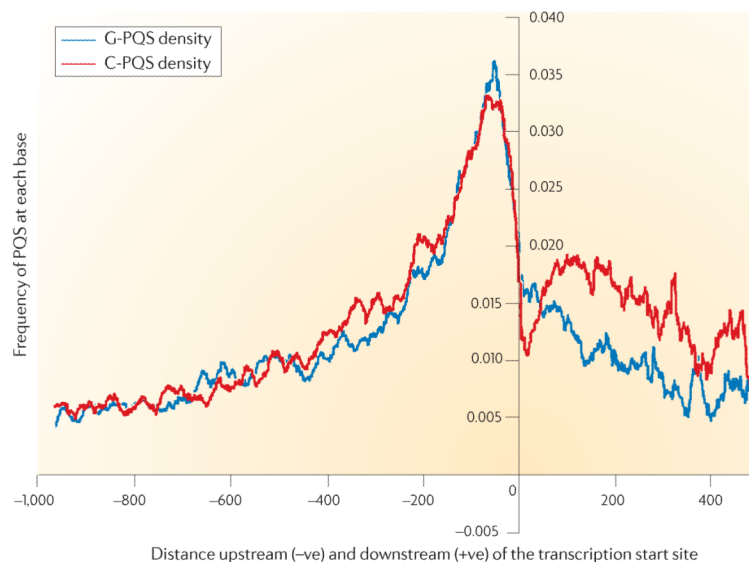


Figure 2. Identification of putative G-quadruplex-forming regions in gene promoters using computational analysis

The density plot shows the frequency (that is, the probability) of each nucleotide upstream (–ve) or downstream (+ve) of the transcription start site being part of a putative G-quadruplex-forming sequence (PQS), based on the Quadparser algorithm⁵⁰. The data has been averaged over all human protein coding genes in the genome. The blue and red plots reflect the presence of a G-quadruplex motif (G-PQS) or the C-rich complement of a G-quadruplex motif (C-PQS), respectively. The downstream data are related to the spliced sequence (that is, exons), and there is clearly some evidence of strand bias, in contrast with the symmetry between the blue and red plots in the upstream sequence. Figure is modified, with permission, from REF. ⁵⁸ © (2008) Oxford University Press.

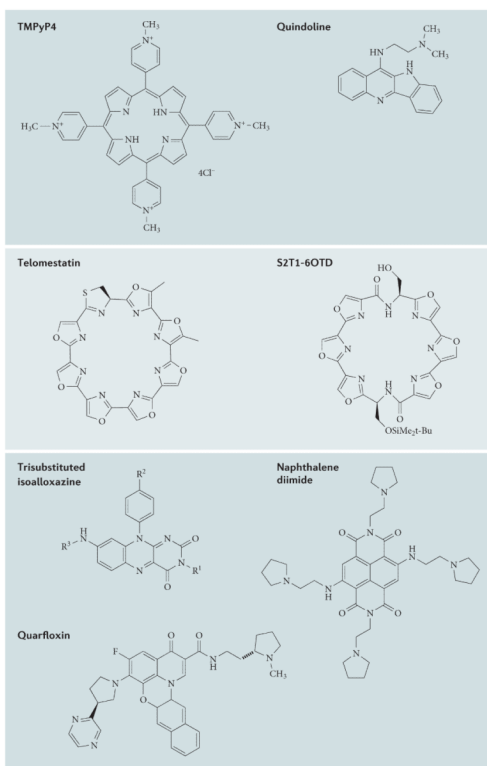


Figure 3. Structures of selected G-quadruplex ligands

Each of these small molecules has shown a G-quadruplex-associated biological activity in one or more studies.

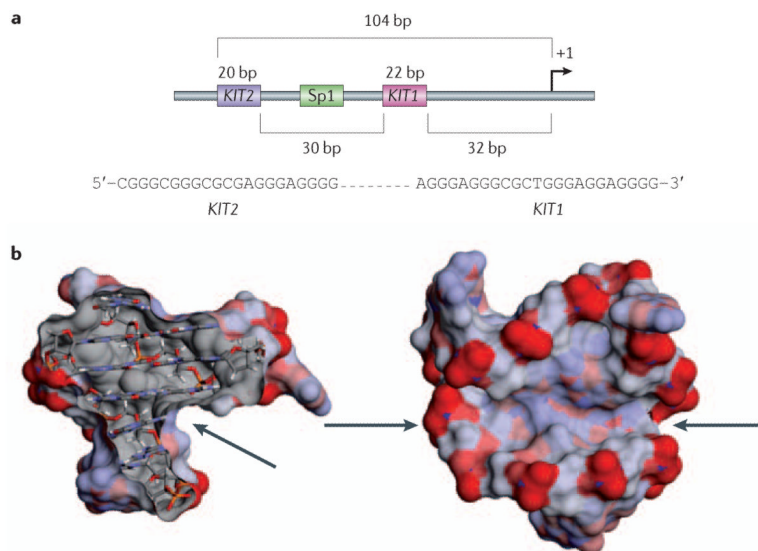


Figure 4. G-quadruplexes in the promoter of the gene coding for KIT

a | The minimal promoter structure of the *KIT* gene showing the spatial relationship between G-quadruplexes *KIT1* and *KIT2* and a transcription factor Sp1 binding site. Shown in the inset are the sequences of *KIT1* and *KIT2*. **b** | The two orthogonal views of the averaged nuclear magnetic resonance structure of the *KIT1* quadruplex are shown, with the solvent-accessible surface coloured by charge. The arrows indicate the large irregular cleft in the structure, which has not been found in any other G-quadruplex structure determined to date. bp, base pair.

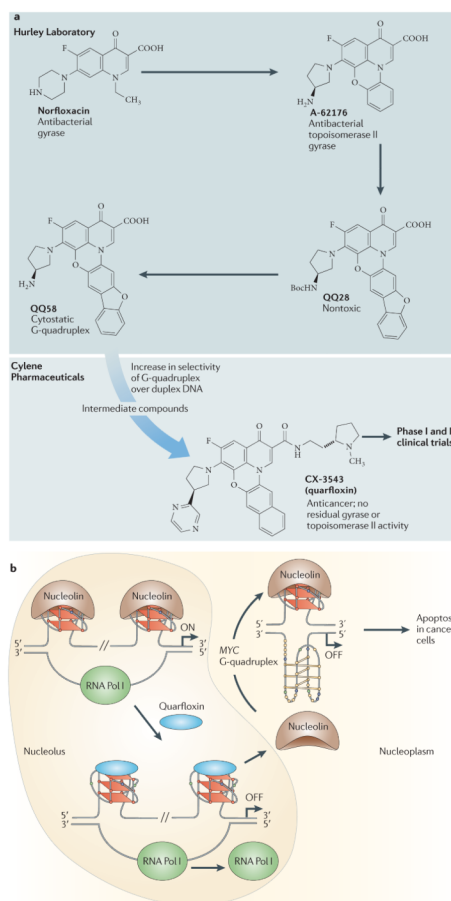


Figure 5. Properties of quarfloxin

a | Origin and lead optimization of quarfloxin (CX-3543). The design of the fluoroquinolone QQ58 (REF. ¹⁰³) was based on the principle that nor-floxacin and A-62176 inhibited gyrase and topoisomerase II by interacting with the enzyme–duplex DNA complex, and this scaffold was amenable to G-quadruplex interaction by extending the tetracyclic nucleus to a larger and approximately planar system that would interact more favourably with the tetrad in the G-quadruplex than with duplex DNA. Cylene Pharmaceuticals then optimized the planar, presumably G-tetrad-interacting moiety and identified groove-binding arms that would both prevent topoisomerase poisoning and increase selectivity between different G-quadruplexes. **b** | Diagram showing the proposed mechanism of action of quarfloxin specifically involving the *MYC* G-quadruplex^{40,102}. The other target proteins of nucleolin are mentioned in the main text. Quarfloxin is preferentially concentrated in the nucleolus. Binding of quarfloxin to the non-template G-quadruplexes in the recombinant DNA displaces nucleolin, which is relocated to the nucleoplasm, where it has been shown to bind to the *MYC* G-quadruplex to inhibit *MYC* expression⁷⁶. RNA Pol I, RNA polymerase I.