Pergamon

European Journal of Cancer Vol. 33, No. 5, pp. 735–749, 1997

Printed in Great Britain
0959-8049/97 $17.00 + 0.00

PII: S0959-8049(97)00067-1

The DNA Structures at the Ends of Eukaryotic Chromosomes

R.J. Wellinger[1] and D. Sen[2]

[1]Faculté de Médecine, Department de Microbiologie et Infectiologie, Université de Sherbrooke, Sherbrooke QC J1H 5N4; and [2]Department of Chemistry and Institute of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, Canada BC V5A 1S6

The sequence organisation of the telomeric regions is extremely similar for all eukaryotes examined to date. Subtelomeric areas may contain large sequence arrays of middle repetitive, complex elements that sometimes have similarities to retrotransposons. In between and within these complex sequences are short, satellite-like repeats. These areas contain very few genes and are thought to be organised into a heterochromatin-like domain. The terminal regions almost invariably consist of short, direct repeats. These repeats usually contain clusters of 2–4 G residues and the strand that contains these clusters (the G strand) always forms the extreme 3′-end of the chromosome. Thus, most telomeric repeats are clearly related to each other which in turn suggests a common evolutionary origin. A number of different structures can be formed by single-stranded telomeric G strand repeats and, as has been suggested recently, by the C strand. Since the main mechanism for the maintenance of telomeric repeats predicts the occurrence of single-stranded extensions of the G strand, the propensity of G-rich DNA to fold into alternative DNA structures may have implications for telomere biology. © **1997 Elsevier Science Ltd. All rights reserved.**

Key words: telomeric repeats, G-tails, alternative DNA structures

Eur J Cancer, Vol. 33, No. 5, pp. 735–749, 1997

INTRODUCTION

THE ENDS of eukaryotic chromosomes have special properties when compared to ends created by random chromosomal breakage: native ends, or telomeres, are stable structures not prone to degradation, recombination or fusions with other chromosome ends [1–3]. This property of "capping" natural chromosome ends was the first essential function ascribed to telomeres [2, 4]. Since those seminal observations in the late 1930s, we have learned that DNA is at the base of the chromosome and with the advent of molecular biology, some remarkable features of the DNA sequences present at the telomeres have been uncovered. For instance, functional telomeres are not recognised by DNA damage recognition systems even though formally, a chromosomal end constitutes a double-stranded DNA break [3]. Furthermore, in the early 1970s, it was realised that due to the intrinsic requirements and the polarity of the normal DNA replication machinery, a DNA end cannot be fully replicated [5, 6]. Thus, telomeres must provide the basis for a second essential function, namely the complete replication of the DNA at chromosomal termini. However, cytologists have been studying the positioning and behaviour of telomeres for much longer. As early as 1885, Rabl described a peculiar clustering of the telomeres on one side of mitotic nuclei [7]. These observations of a chromosomal order with a non-random distribution of telomeres and centromeres have since been confirmed in many systems [8]. Moreover, the terminal regions of the chromosomes may be heterochromatic, repress transcription of nearby genes and determine the timing of replication initiation [9–18].

Being implicated in such a variety of biological phenomena, the definition of a eukaryotic telomere has become diffuse. The notion that the presence of a tract of telomeric repeat sequences (see below) ensures all functions of a telomere is widely used but remains far from being proven. Thus, as proposed by Biessmann and Mason [19], the telomere can be defined operationally as the complex at the ends of chromosomes that performs the functions studied in any given system. This definition has the advantage of including all *cis*-acting sequences and *trans*-acting factors binding to these sequences and involved in the functions ascribed to telomeric chromatin. Here we will concentrate on the description of the DNA sequences found at eukaryotic chromosomal termini and review our current knowledge of possible structures formed by them.

DNA ORGANISATION AT CHROMOSOMAL TERMINI

The overall sequence organisation of telomeric areas is remarkably conserved in all eukaryotes (Figure 1). The telo-

Correspondence to R.J. Wellinger.

meric DNA sequences have generally been subdivided into at least three distinct areas: (i) telomere associated sequences (TAS); (ii) double-stranded telomeric repeats; and (iii) the terminal DNA structure at the very ends of the chromosomal DNA.

Telomere associated sequences

In all organisms for which sequence information of subtelomeric areas is available, the region proximal to the terminus specific sequences (see below) contains a variety of repeated elements called TAS [19, 20]. Usually, the subtelomeric area with the TAS is described as the sequences centromere-distal to the last identified, functional, single-copy gene. The actual amounts of TAS vary between chromosomes of the same species and may make up considerable amounts of the total genome [19]. For instance, human autosomes have TAS comprising a few hundred kb [21, 22], whereas the sex chromosomes bear a subtelomeric pseudoautosomal region (PAR) of approximately 2600 kb that may contain mostly TAS [23]. In general, there are two different types of elements found in such subtelomeric areas. Relatively complex, middle-repetitive elements varying from a few hundred base pairs (bp) to a few thousand bp are known for plant [24, 25], fungal [26–28], invertebrate [29–31] and vertebrate species, including humans [32–35] (see [19, 36] for reviews). In every species for which enough sequence information is available, the distribution and/or copy number of individual repeats are variable. Human subtelomeric areas may contain telomere-specific repeats, but most repeats occur at least on a subset of the telomeres [37, 38]. However, only a small subset of all repeats may have been identified so far and it is therefore not clear whether there are TAS repeats present on all telomeres. In yeast, the 'core X' element seems to be present in all native subtelomeric areas [20], but the presence of this element is not essential for chromosome maintenance [39]. Interspersed in between and within these complex repeats are shorter, tandemly repeated satellite-like sequences. These repeats of 20–1000 bp seem to be recombinogenic, rearrange with high frequency and may constitute the most prominent reason for the observed high degree of polymorphism of subtelomeric chromosomal regions [20, 40–44]. In several organisms, blocks of the terminus-specific repeated sequences (see below), or variants of them, are also found imbedded in subtelomeric areas [26, 27, 37, 45, 46]. Furthermore, in yeast, copies of a number of open reading frames (ORFs) such as the *PAU, SUC, MAL* and *MEL* families of genes have been found in some, but not all subtelomeric areas [20, 47, 48]. In *Trypanosomes*, the expression of certain surface antigen genes is correlated with relocation of the genes into subtelomeric expression sites [10, 13], and in *Plasmodium*, the location of surface antigen genes in the polymorphic subtelomeric areas may be responsible for the antigenic polymorphism of these genes [44, 49, 50]. However, despite these examples, the subtelomeric areas in general are very low in gene density and transcriptional activity.

What are the possible functions of this sometimes vast mosaic of repeated sequence elements? It is important to note that TAS family members are relatively homologous within the families of each species, but that there is no evidence for sequence relationships between the repeats of various species [20]. In yeast and *Drosophilia*, genes located close to an actual chromosome end are subject to transcriptional repression [9, 11, 12]. Thus, it has been argued that the TAS provide a buffer zone between the actual terminal telomeric structure and more internal chromosomal domains [20, 36, 51]. Given the possibility of relatively rapid changes of the sequences within this area, it could also provide a region in which genes can undergo adaptive changes via amplification and recombination [36, 44]. Cytological and genetic studies have shown that telomeres interact with each other non-randomly during mitotic growth and during meiosis in a number of species [20, 52, 53]. Since the terminal repeats are identical for all telomeres in any given species, it could be the highly variable arrays of the different TAS that specify telomere to telomere interactions. Finally, TAS have been shown to provide alternative mechanisms for telomere maintenance via recombination or transposition in cases where the main telomere replication mechanism via telomerase is absent [54–57].

Double-stranded telomeric repeats

In striking contrast to the apparently unrelated repeat sequences of the TAS of various organisms, the sequences found at the very ends of the chromosomes of a variety of organisms are clearly related (see Table 1). The first terminal sequences to be identified were those of some ciliate species such as *Tetrahymena* and *Oxytricha* [58, 61]. The great abundance of telomeres in the macronuclei of the organisms was instrumental in the initial identification and characterisation of these sequences by direct chemical methods. Today, the terminal repeat sequences are known for a great deal of different species, and while common features are easily identified for a number of them, there are also notable differences and exceptions. For most species, the terminal repeats consist of short (6–8 bp), tandemly repeated sequences. For instance, the first telomeric repeat sequence identified, the one for *Tetrahymena*, is C_4A_2/T_2G_4

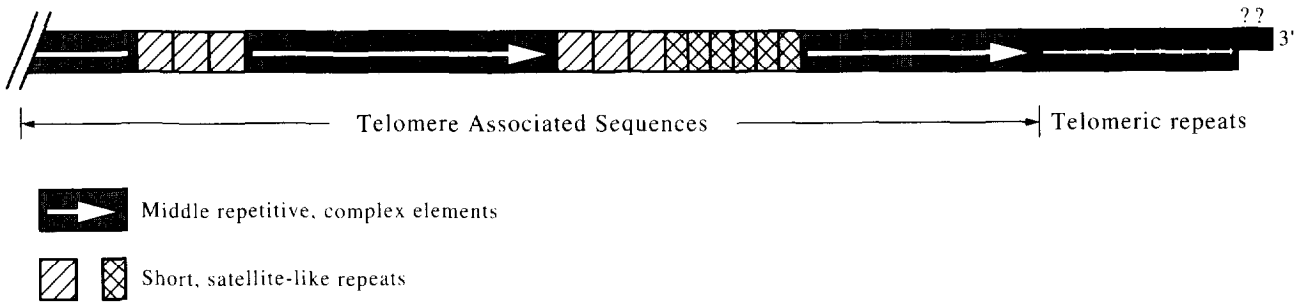

Figure 1. A generalised diagram of the sequence organisation at eukaryotic telomeres. See text for details.

Table 1. Telomeric simple-sequence repeats

Organisms	Repeat sequence	Reference
Protozoa		
Tetrahymena	T_2G_4	[58]
Glaucoma	T_2G_4	[59]
Oxytricha	T_4G_4	[60, 61]
Euplotes	T_4G_4	[61]
Stylonychia	T_4G_4	[61]
Paramecium	$T_2[T/G]G_3$	[62, 63]
Trypanosoma	T_2AG_3	[64, 65]
Giardia	TAG_3	[66]
Plasmodium	$T_2[T/C]AG_3$	[67]
Theileria	$T_{3-4}AG_3$	[68]
Slime moulds		
Dictyostelium	AG_{1-8}	[69]
Physarum	T_2AG_3	[70]
Didymium	T_2AG_3	[70]
Fungi		
Neurospora	T_2AG_3	[71]
Podospora	T_2AG_3	[72]
Cladosporium	T_2AG_3	[73]
Histoplasma	T_2AG_3	[74]
Cryptococcus	T_2AG_{3-5}	[75]
Saccharomyces	$(TG)_{1-6}TG_{2-3}$	[76]
Kluyveromyces	$ACG_2AT_3GAT_2AG_2TATGTG_2TGT$	[77]
Candida albicans	$ACG_2ATGTCTA_2CT_2CT_2G_2TGT$	[78]
Candida tropicalis 4443	$A_2G_2ATGTCACGATCAT_2G_2TGT$	[77]
Candida guillermondii	$ACTG_2TGT$	[77]
Schizosaccharomyces	$T_{1-2}ACA_{0-1}C_{0-1}G_{1-6}$	[79]
Pneomocystis	T_2AG_3	[80]
Plants		
Chlamydomonas	T_4AG_3	[81]
Chlorella	T_3AG_3	[82]
Arabidopsis	T_3AG_3	[83]
Zea mays	T_3AG_3	[84]
Hordeum (Barley)	T_3AG_3	[85]
Nicotiana	T_3AG_3	[86]
Lycopersicum (Tomato)	$T_2(T/A)AG_3$	[87]
Invertebrates		
Ascaris	T_2AG_2C	[88]
Parascaris	T_2GCA	[89]
Caenorhabditis	T_2AG_2C	[90]
Bombyx (and other Lepidoptera)	T_2AG_2	[91]
Vertebrates		
Homo sapiens	T_2AG_3	[92]
Many other vertebrates	T_2AG_3	[93]

The sequence of the strand running 5′ to 3′ towards the physical end of the chromosomes is represented.

[58] and many protozan, fungal, plant, vertebrate and invertebrate repeats are clearly related to this repeat as their repeats can be summarised as $C_{2-4}T_{0-1}A_{2-4}/T_{2-4}A_{0-1}G_{2-4}$ (see Table 1). Some fungal species, however, bear a more heterogeneous repeat such as the one found for *S. cerevisiae* which can be abbreviated $C_{2-3}A(CA)_{1-6}/(TG)_{1-6}TG_{2-3}$ [76], or contain longer repeats of approximately 20–25 bp such as *Candida* or *Kluyveromyces* [77, 78] (Table 1). The actual number of repeats per telomere varies greatly between organisms as there may be as few as 2–3 repeats, making up some 20–30 bp (hypotrichous ciliates such as *Oxytricha* and *Euplotes* [61]), and there may be >10000 copies, making up >60–100 kb as in some mice [94, 95]. Moreover, the amount of telomeric repeats also varies between the chromosomes of the same cell and may, within a certain range, change gradually or stochastically during growth [33, 96–100]. It is thought that the actual lengths of the telomeric repeats are the result of a complex interplay between shortening and lengthening activities [19, 39]. The human telomeres for instance contain 500 to >2000 repeats, yielding 3 to approximately 12 kb of telomeric repeat sequences. However, in humans these repeat tracts are not stably maintained as they gradually diminish in cycling cells of most somatic lineages [33, 92, 101, 102]. These variable lengths of individual telomeres in any given species result in a very heterogeneous population of terminal restriction fragments and therefore in a "smeary" appearance of telomeric bands on Southern blots. This smeary appearance and the preferential sensitivity of the sequences in the smears to digestion with the DNA exonuclease BAL31 have been used widely as the main characteristics for *bona fide* telomeric DNA [36, 103]. As usual, there are exceptions to this generality: the telomeric DNA of the subchromosomal-length DNA fragments in the macronuclei of some hypotri-

chous ciliate species has a precise length and structure neither of which are subject to changes during vegetative outgrowth [61].

One of the most remarkable features of the telomeric repeat sequences is their polar arrangement. Due to the particular base composition of the repeats, one strand is usually rich in clusters of 2–8 G residues (Table 1). This strand, which in most cases completely lacks C residues, is also called the G strand, with the complementary strand being the C strand. In all cases where a G strand is identifiable, this strand is always oriented such that it forms the extreme 3′-end of the chromosome. In the very initial study of the terminal repeats of *Tetrahymena*, evidence for 1 base discontinuities, particularly in the C strand, were found, but their significance remains unclear [58]. Thus, the telomeric repeat sequences are highly conserved not only in a primary structure but also in sequence arrangement, a property that reflects the main mechanism of telomere repeat addition by telomerase (see article by G. Morin in this Special Issue, pages 750–760) and that may have other functional relevance (see below). It is important to note that telomeric repeats do also occur at subtelomeric and non-telomeric chromosomal loci. These repeats may be remnants of recombination or transposition events involving telomeric repeats [104, 105] or may indicate telomere fusion points of ancestral chromosomes [106].

Certain dipteran insect species do not seem to contain any simple repeat arrays at their telomeres. Instead, arrays of complex repeats form the ends of the chromosomes in *Chironomus* [31], an organisation reminiscent of the termini of some linear mitochondrial DNAs [107]. In *Drosophila*, the terminal areas contain elements termed HeT-A and TART that bear similarities to retrotransposons ([30, 56, 108], reviewed in [57, 109, 110]). Although the terminal sequence organisation in *Drosophila* shows similarities to the organisation of TAS in other organisms [109], it is in marked contrast to the organisation with short, direct repeats (see Table 1). However, it has been pointed out that this seemingly fundamental difference in terminal sequence organisation may have a common evolutionary origin, since in both cases the synthesis of telomeric DNA requires a RNA-dependent DNA polymerase [57, 109].

The very end

Although likely to be of primordial importance for the 'capping' function of the telomeres, suprisingly little is known about the actual physical structure of chromosomal DNA ends in plant, fungal or metazoan species. The ends of the macronuclear DNA of protozoan ciliates such as *Oxytricha* and *Euplotes* are formed by a 12–14 base overhang of the G strand (3′-end) over the C strand [61]. Furthermore, end-labelling studies using linear rDNA molecules isolated from *Tetrahymena* and *Didymium* suggest a similar end structure with a 12–16 base overhang of the G strand [111]. Since this end structure fits the predicted structure after telomere replication is completed [112, 113], it has been suggested that a short, 12–20 base overhang of the G strand is a conserved feature of eukaryotic chromosomal termini [36, 111]. It must be noted, however, that evidence for end structures of metazoan chromosomes are not available.

In yeast and *Euplotes*, the appearance of larger extensions of the G strand have been demonstrated to appear during telomere replication [18, 113–116] or during *de novo* telomere formation [117, 118]. These long, single-stranded tails are the predicted products of telomerase-mediated telomere repeat addition, although other mechanisms also contribute to their formation [112, 113, 116]. Due to their richness in G residues, these G-strand extensions may be implicated in alternative DNA conformations.

In several organisms, the very ends of the chromosomes are packaged into a non-nucleosomal end structure [119–122]. In organisms with $\geqslant$300 bp of telomeric repeats, this structure may comprise all the telomeric duplex repeats, such as in *Tetrahymena* or yeast [123, 124]. However, the very long arrays of telomeric repeat sequences of vertebrates appear to be nucleosomal, even though the nucleosomal repeat length is somewhat shorter for telomeric repeat DNA than for bulk DNA [125–127]. Evidence for a non-nucleosomal chromatin configuration in vertebrates has been obtained using human cells with particularly short telomeric repeat tracts [126]. Thus, a bipartite structure of the vertebrate telomeric repeat tracts with a nucleosomal organisation of the more proximal repeats and a non-nucleosomal end structure has been proposed [126, 128].

ALTERNATIVELY FOLDED STRUCTURES OF TELOMERIC DNA

The earliest report on an unorthodox DNA structure formed by telomeric DNA sequences was made by Oka and Thomas [129], who described unusual cohesive properties of purified macronuclear DNA molecules from the protozoan *Oxytricha*. Nuclease digestion studies established that telomeric DNA sequences were critical for this cohesiveness. A remarkable property of this cohesion was that it was stabilised significantly by potassium-containing buffers, as compared to sodium-containing buffers. Investigation of this particular cohesive property [130], as well as the separate but related investigation of G–G base-paired telomeric sequences, revealed that telomeric G-rich sequences were able to form a large and polymorphic class of higher-order structures such as G–G base pairs and four-stranded structures called G-quadruplexes, G-quartets or G4-DNA [131–133]. Although an *in vivo* role of any of these structures has not been demonstrated directly, there is considerable indirect evidence for their importance in telomere biology.

G–G base-paired duplexes

The first structural description of a class of non-canonically folded structures adopted by single-stranded, G-rich telomeric sequences was reported by Henderson and associates [131]. At temperatures under 40°C, the utilised oligomers seemed to be folding to putatively G–G or G–T base-paired, hairpin-like structures. These hairpin structures were indicated by aberrant gel mobilities of the oligomers as well as by the detection, by NMR (nuclear magnetic resonance), of the unusual *syn* glycosidic conformation for some of the guanosine residues. Guanines base-paired to each other by a Hoogsteen scheme were among the structural models proposed to explain the data. Since then, relatively little experimental work has been reported on G–G base-paired duplexes involving telomeric repeats. One NMR study on the sequence $d(G_4T_4G_4)$, comprising *Oxytricha* telomeric repeats suggests reverse Hoosgteen G–G base pairs in a hairpin loop [134]. However, studies of individual G–G mismatches within Watson–Crick duplexes indicate numerous ways in which two guanosines can hydrogen bond to

each other [135–137]. Recent work on eight consecutive G–G mismatches within a Watson–Crick duplex suggests that a number of the G–G mismatches must be Hoogsteen base-paired, since they are able to dimerise to G-quartets [138]. Such a dimerisation to form G-quartets would not be expected to be possible from reverse Hoogsteen base-paired guanines. However, G–G base pairs have been reported to occur in parallel or anti-parallel-stranded DNA duplexes formed by $d(GA)_n$ and $d(GCA)_n$ sequences [139–141] and it seems likely that telomeric DNA may adopt similar base-pairing configurations.

*C–G*G triplexes*

A note by Veselkov and associates [142] has reported methylation-protection data on a DNA model for the terminus of the *Tetrahymena* telomere. The molecules contain three T_2G_4 repeats in duplex form followed by a single-stranded overhang of two T_2G_4 repeats. The data are consistent with the G-rich overhang participating in triplex formation with the duplex component of the construct. In this putative structure, the bonding is expected to be of the canonical form of C–G*G, where C–G represents the Watson–Crick base-pair, and the two G strands run antiparallel to one another (reviewed in [143, 144]). However, in examining competing triplex–quadruplex equilibria involving guanine-rich oligonucleotides, it was found that at physiological concentrations of monovalent ions, particularly potassium, triplex formation is effectively inhibited and quadruplex formation is dominant [145].

Guanine quadruplexes, G-quartets and G4-DNA

The formation of guanine quadruplex complexes by telomeric G-strand sequences was first deduced from gel-mobility shift assays and chemical protection data ([130, 132, 133, 146–148], see also recent reviews [149, 150]) and these predictions were based on earlier work on the chemical behaviour of the guanine nucleotides and homopolymers [151, 152] (Figure 2). Two broad classes of structures, "parallel"- and "antiparallel"-stranded quadruplexes were found to be formed, often by the same telomeric oligomers. The identity of the monovalent cation present in solution (e.g. sodium, as opposed to potassium) seems to dictate which complex is formed [133, 153–157]. The unique interaction of G-quadruplexes with specific monovalent cations consists of the chelation of the ion, with varying degrees of affinity depending on the identity of the cation, between two adjacent guanine quartets (shown schematically in Figure 3). The actual interaction appears to depend on the balance between the energetics of coordination of the cation to G-quartets and the energetics of removal of water molecules ordinarily bound to the cation [149, 158]. Overall, the most stable G-quadruplexes are formed in the presence of potassium, with the order of stabilising ability by group Ia cations following the order K>Rb>Na>Li = Cs (and, for group IIa cations, Sr>Ba>Ca>Mg) [130, 132, 133, 154, 157, 159–162].

Subsequent to these early reports, a vast literature has addressed the structures and properties of telomeric G-quadruplexes. Three of the most striking properties of these quadruplexes are (i) their structural polymorphism; (ii) their unusual interactions with specific monovalent and divalent cations and (iii) their high thermodynamic stability and extraordinarily slow kinetics of formation and unravelling.

Structure. A large number of structures of different G-quadruplexes have been studied and solved by both X-ray crystallography and NMR. Other methodologies, such as protection from chemicals and nucleases, ultraviolet light-induced crosslinking as well as various electrophoretic techniques have also been utilised. The polymorphic features of the different quadruplex complexes can be placed into the following categories: (i) strand orientation; (ii) strand stoichiometry; (iii) glycosidic conformation; (iv) loop geometry and (iv) ion-binding geometries [149]. These categories are not entirely independent of one another. The classes of structures obtained to date are shown schematically in Figure 4. A distinctive structure is the purely parallel quadruplex ("G4-DNA") [146], which is composed of four DNA strands running in a parallel orientation (Figure 4a). The individual strands in this type of complex may contain

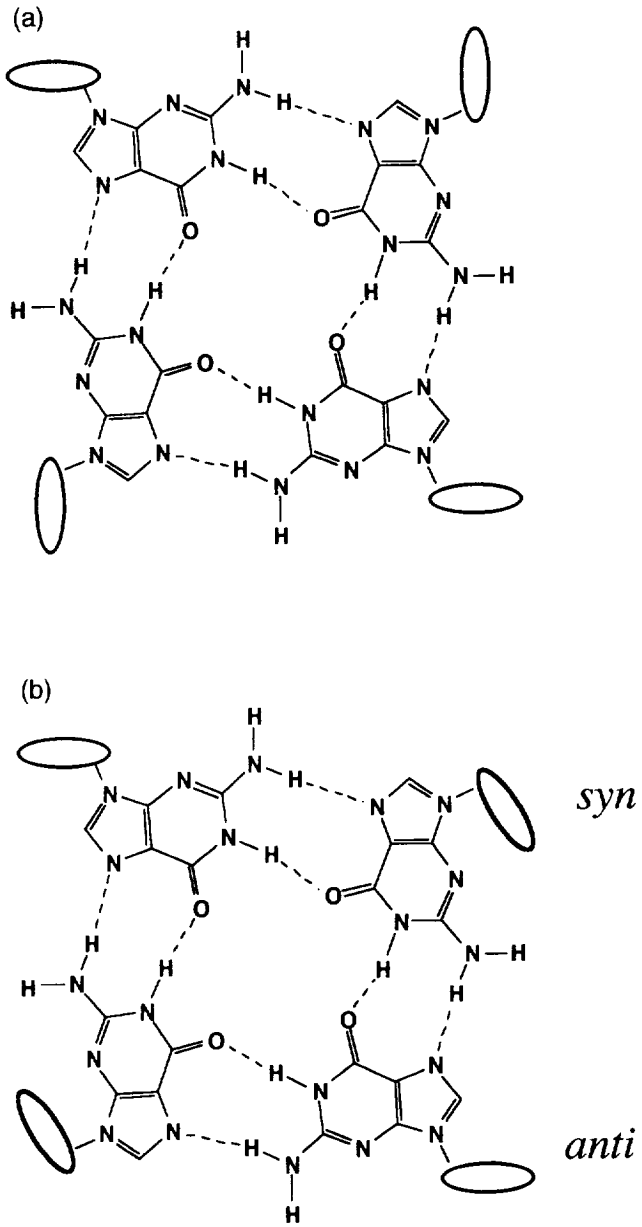

Figure 2. The hydrogen-bonding scheme in guanine-quartets. (a) and (b) show alternative glycosidic conformations found in different G-quadruplexes: (a) all '*anti*-' and (b) alternating '*syn*-' and '*anti*-'.

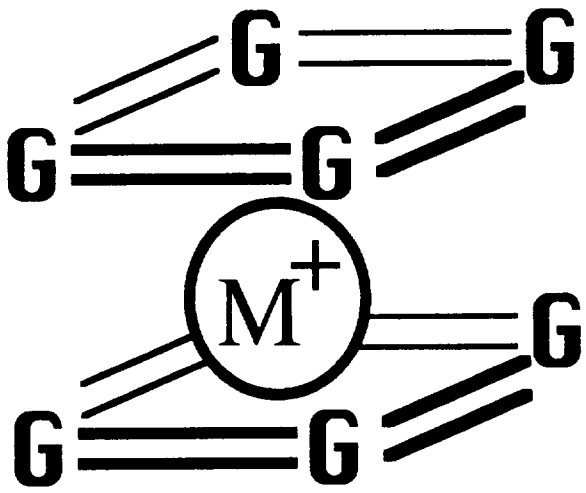

Figure 3. Schematic drawing showing the coordination site of a monovalent cation between successive guanine quartets.

a single or multiple G-rich repeats. A number of NMR and crystal structures, all in good agreement with one another, have been reported for this class [163–170]. Furthermore, telomere-like RNA sequences also form parallel quadruplexes [171]. The glycosidic conformation of the guanosines participating in quartet formation appeared uniformly to be *anti* (as in the B-form of double-helical DNA)(Figure 2a). The quadruple-stranded helices formed have four identical grooves, each resembling the minor groove of B-DNA. The recently reported crystal structure of $d(TG_4T)$ at 1.2 Å resolution has been particularly informative [170]. Crystallisation was carried out in the presence of sodium ions and an electron density corresponding to the sodium ions was localised between successive layers of the stacked G-quartets (see, for example, Figure 3).

The 'antiparallel' G-quadruplexes, unlike the 'parallel' ones, comprise a diverse family of structures. Two general types may be distinguished: one formed from the association of two distinct strands of DNA, each containing two G-clusters ("G′2-DNA"; Figure 4b–d), or a single strand of DNA containing four G-clusters ("G4′-DNA"; Figure 4e–g). Structural affinities between (b) and (e), and (c) and (f), can readily be identified. The polymorphism of these structures was highlighted when the first analysis of crystals [172] and the first NMR [173] of a quadruplex formed by the *Oxytricha* sequence, $d(G_4T_4G_4)$, reported different structures, the former resembling the structure in Figure 4d, and the latter the structure in Figure 4c. These two complexes differ in their loop geometries (Figure 4d has loops spanning two adjacent strands, whereas in Figure 4c, the loops run diagonally above the stack of quartets, connecting non-adjacent strands), as well as in the strand orientations (an alternating ABAB in the case of (d), but AABB in the case of (c)). The glycosidic conformations in either case alternate 'syn-anti-syn-anti' along a given strand (a 'syn-anti-syn-anti' alternation along a quartet is shown in Figure 2b). This arrangement is found in the structure in Figure 4d, whereas the structure in Figure 4c has 'syn-anti-syn-anti' in a given quartet. A major difference between the two studies was that the NMR was carried out in the presence of sodium, whereas the crystals were grown from a potassium solution. Assuming that the major stabilisation energy comes from the quartets themselves, and less so from the arrangements of backbones and loops, it appears that many of the individual structures within a particular class may be energetically quite similar. This is supported by a computational study in which a similarity in the energies of formation of such related structures was reported [174]. However, the category Figure 4c structures may be kinetically trapped intermediates in the formation of category Figure 4d structures [175].

Most of the structures reported for the intramolecular, G4′ type complexes are of the form shown in Figure 4e and f [166, 176–179]. An extremely anomalous structure for the *Tetrahymena* telomeric repeat sequence, $d(T_2G_4)_4$, has been reported recently [167]. This structure resembles the diagram shown in Figure 4g and has the notable feature that its strand orientations are of the form AAAB (three of the four strands are parallel to one another). In addition to having a flanking loop and a lateral loop like those described earlier, this structure also has a short TT 'loop' or 'bridge', that permits a reversal of the expected strand orientation.

Kinetics and thermodynamics of formation: salt effects. A number of investigators have reported thermodynamic data on various G-quadruplex complexes, using two fundamentally distinct approaches: (i) the prior assumption of a certain structural model for a complex (including its strand molecularity); followed by van't Hoff analysis; and (ii) structural model-independent calorimetric measurements. From the large number of studies, mutually consistent conclusions are not yet derivable because many different complexes have been studied under many different conditions (reviewed in [180]). However, some conclusions about the energetics of G-quadruplex complexes can be drawn. The formation of all quadruplexes appears to be enthalpy-driven, with the enthalpy per guanine quartet measured between −21 and −26 kcal/mol [163]. In addition, parallel structures were found to be more stable than antiparallel ones, and the potassium complex of any given structure was more stable than the corresponding sodium complex. An interesting study of the relative stabilities of parallel versus antiparallel G-quadruplex structures found that the antiparallel, G′2-type structure formed by $d(G_4T_4G_4)$ was approximately 5 kcal/mol less stable than the equivalent, but enforced parallel structure formed by $d(G_4T_2\text{-}5'\text{-}5'\text{-}T_2G_4)$ [181].

A major, and often defining characteristic of G-quadruplexes is that they show very slow kinetics of both formation and dissolution. The following are some of the formation constants that have been reported: a second-order constant of 5 $M^{-1}s^{-1}$ for the dimerisation of a G′2 complex by $d(T_4G_4T_4G_4)$ [182, 183], and between 0.72 $M^{-1}s^{-1}$ and 0.1 $M^{-1}s^{-1}$ for the same oligomer [184]. Formation of the parallel complexes is even slower but there is disagreement over the order of the reaction, ranging from two to four [133, 185].

Once formed, G-quadruplexes show an enormous kinetic stability [186, 187]. In NMR studies, guanine imino protons, which in double helices exchange within millisecond timescales, give stable resonances for weeks or months [164, 168, 171, 173]. Wang and Patel [164] report that the imino protons exchange over a timescale of minutes in complexes containing three guanine quartets, but of the order of months for a complex containing four quartets. These slow kinetics of formation and the enormous kinetic stability of

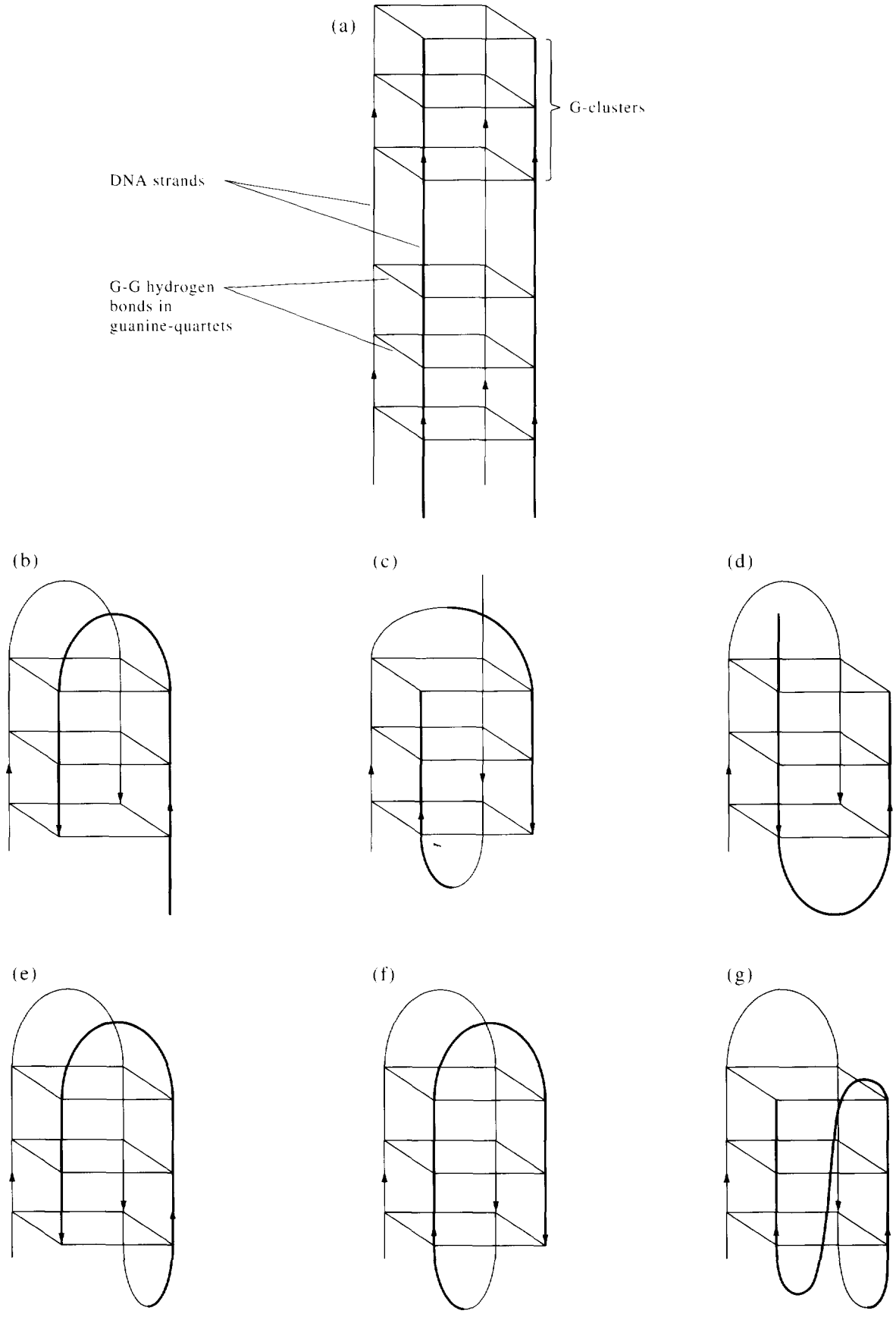

Figure 4. Different geometric arrangements for parallel (a) and antiparallel (b)–(g) G-quadruplexes. Note that the polarity of the DNA strands is indicated by arrowheads. See text for detailed description.

G-quadruplexes, once formed, raise important questions about how their formation or destruction might be regulated *in vivo*. The discovery of proteins that assist the formation of G-quadruplexes and of G-quadruplex specific nucleases is interesting in this respect.

The formation of different G-quadruplex complexes has also been studied in the presence of both sodium and potassium ions. At physiological concentrations, even small changes in the relative cation ratio can strongly affect the formation of a given class of G-quadruplexes, a result that recapitulates 'cationic switches' [133, 155, 157, 188]. Finally, a disproportionation of the B-DNA duplex, $d(G_4T_4G_4).d(C_4A_4C_4)$, to a parallel quadruplex occurs reversibly under conditions of relatively high sodium or potassium concentrations [189]. The transition region seems to occupy the cation concentration range of $[Na^+] + [K^+] = 150$ mM, which is close to the cumulative physiological concentrations for these two cations. These results imply that despite slow kinetics, the formation of G-quadruplexes from Watson–Crick source materials is thermodynamically reasonable.

The i-motif

In addition to the plethora of structures possible for telomeric G strands, the C strands may also form alternative structures. In an NMR study using the oligomer $d(TC_5)$, a new and unusual type of higher-order folding by cytosine-rich DNA sequences was reported [190]. This class of structures, referred to as the 'i-motif', consists of four-stranded complexes formed by the interdigitation of two parallel-stranded duplexes, which are themselves held together by hemi-protonated $C–C^+$ base pairs. The protonation of one of the two cytosines, promoted by pH 5–6 in solution [191], is necessary for the formation of three hydrogen bonds between each cytosine pair (for review, see [150]). In a manner analogous to G-quadruplexes, i-motifs may be formed by four distinct C-rich strands (a 'tetramer'), two strands containing two distinct C-rich motifs (a 'dimer'), and a single strand containing four distinct motifs (a 'monomer'). Such an intrastrand folding of telomeric C-strand repeats has been studied for the vertebrate repeat $d(C_3TA_2)_3C_3$ and the *Tetrahymena* repeat $d(C_4A_2)_3C_3$ [192, 193]. $C–C^+$ base-pairs, like G-quartets, have a slow opening rate of the order of hours, implying a considerable structural stability for the i-motif [194].

A number of the above investigators have proposed the existence of the i-motif *in vivo*. However, unlike the G-rich strands, which have been shown to exist as single-stranded overhangs at the telomeres [18, 61, 111, 115], the C-rich strand is not found naturally in a single-stranded form [115, 116, 195]. In addition, under solution conditions, there appears to be a requirement for a non-physiological, acidic pH for the formation and stability of i-motifs. Nonetheless, it has been proposed that double-stranded telomeric sequences might be induced, by protein factors or by superhelical stress, to unravel, giving rise to single-stranded stretches of both the C-rich and G-rich strands, which could then fold to form i-motifs and G-quadruplexes, respectively [192, 196]. In fact, a composite structure incorporating $C–C^+$ base-paired hairpins, a G-quadruplex, and a triplex region have been proposed for human telomeric repeats in a supercoiled plasmid under acidic conditions [197]. However, no proteins interacting with or promoting the formation of i-motifs have been described to date. It is also interesting that cytosine-rich RNA sequences are unable to form the i-motif [198], whereas G-quadruplexes can be formed by both DNA and RNA.

Interactions of proteins with G-quadruplex telomeric structures

A number of proteins that interact with telomeric DNA have been described. These proteins can be divided into three categories: (i) those that bind to double-stranded telometric repeats; (ii) those that bind primarily to the single-stranded G-rich strand; and (iii) those that bind to G-quadruplex structures. These categories are neither absolute nor exclusive, as a number of the proteins described fall into more than one category.

Important examples of a class (i) protein include 'TRFs', or 'T_2AG_3 repeat factors', defined for a variety of vertebrate sources and for yeast [199–201]. Such proteins are purely double-strand binders and do not seem to bind single-stranded telomeric sequences nor require a DNA terminus and, as such, they lie outside the scope of this review. RaP1p, an abundant yeast protein, binds the double-stranded yeast telomeric repeats and is the major protein component of the yeast telosome ([124, 202, 203], reviewed in [204]). RaP1p, however, is multifunctional and appears to interact with telomeric DNA in multiple ways, including binding both single-stranded G strands and quadruplex structures, as well as promoting the formation *in vitro* of the latter structures.

Proteins that bind to the G-rich single strand of telomeres fall into three classes. The first is exemplified by proteins that bind tenaciously to the G-rich single strand overhang and the association of which cannot be dissociated with high salt concentrations. These proteins are thought to be essential components of the telosome complexes of those organisms. Examples of this class of proteins include a heterodimeric protein (containing α and β subunits) from the protozoan *Oxytricha nova* [119, 120, 205, 206], as well as comparable proteins from *Euplotes* [207] and *Physarum* [208]. The β subunit of the *Oxytricha* proteins has been found to have a chaperone-like activity for the formation *in vitro* of G-quadruplex structures (see below). A G-strand binding protein similar to those from the protozoan sources has recently been identified in *Xenopus* [209]. Furthermore, a protein characterised from *Tetrahymena thermophila*, TEP, binds to the G-rich overhang but requires the presence of a duplex at the end of the overhang. The binding of TEP, unlike the other proteins of this class, appears to be salt-dependent [210].

In addition to the above, numerous mutually non-homologous proteins that bind to the G-rich telomeric single strands *in vitro* have been described [211–216]. The biological significance of these various proteins is unclear as a number of them are putative RNA-binding proteins which may or may not have a role in telomere biology.

The third class of proteins consists of those that, at least *in vitro*, bind G-quadruplex structures [217–221]. In addition, a vertebrate protein which was modelled to bind to G–G base pairs has been described [222], although binding to quadruplexes could not be ruled out. Chung and associates [223] described the binding of topoisomerase II to parallel-stranded G4-DNA, and a subsequent nuclease-like cutting at a site within the neighbouring single-stranded region. This cleaved sequence is also a substrate for topoisomerase II-cleaving when it is in a duplex context, but it is

not cleaved when it is purely single-stranded. However, the relationship of all of these proteins to telomere biology remains to be elucidated.

Recently, more direct evidence for the relevance of G-quadruplex-binding proteins has been obtained. First, the cloned and purified β subunit of the G-strand overhang binding protein from *Oxytricha* was found to catalyse the association of single-stranded *Oxytricha* telomeric DNA oligomers to a variety of parallel and antiparallel quadruplex structures [184, 224]. A similar activity has also been reported for the yeast Rap1p protein [225, 226]. Both proteins appear to promote G-quadruplex formation with nanomolar concentrations of single-stranded DNA and on a timescale of hours. As described above, uncatalysed G-quadruplex formation takes place with extraordinarily slow kinetics and requires DNA concentrations in the micromolar to millimolar range. Both the *Oxytricha* β subunit and Rap1p assist in the formation of the very quadruplex structures that are formed by the DNA alone (i.e. no 'new' or otherwise very rare DNA structures are promoted by the proteins). However, Rap1p appears to be more sequence-specific than the *Oxytricha* protein, utilising only its own cognate telomeric sequence. While the *Oxytricha* protein promotes the formation of all categories of quadruplexes that can be formed by the *Oxytricha* telomeric repeats, the structures promoted by Rap1p appear to be only parallel complexes (possibly owing to the relative instability of *Saccharomyces* repeats in antiparallel quadruplexes) [157].

A number of basic proteins, including histone H1 and polylysine, also accelerate the rate of G-quadruplex formation by telomeric sequences *in vitro*, although less efficiently than the *Oxytricha* β subunit with its cognate *Oxytricha* telomeric sequences [184, 224]. However, neither histone H1 nor polylysine appear to be able to promote quadruplex formation by yeast telomeric sequences. The possibility that the above telomeric proteins have quadruplex-accelerating activities because they are basic proteins is negated by the observation that point mutations in the β subunit of the *Oxytricha* protein generate mutants far less efficient at promoting G-quadruplex formation [224].

A third recent result of interest in this respect involves another yeast protein, called KEM1p (and also known as Sep1p, Dst1p, Xrn1p and Rar5p). This protein apparently has nucleolytic activity that is activated by recognition of a stack of guanine-quartets [227, 228]. The protein cleaves the DNA in single-stranded overhang regions 5′ to the guanine-quartets and the cutting sites appear to localise consistently 17–23 nucleotides from the edge of the guanine-quartets and at 2–4 nucleotides from the 5′ end. It has therefore been proposed that the KEM1 nuclease might remove the highly stable G-quadruplex structures at telomeres during meiosis, or that it is involved in self-recognition or recombination events mediated by the formation of guanine-quartets. However, a number of seemingly unrelated phenotypes are associated with deletion of *KEM1* (or *SEP1*). In addition, KEM1p was found to be a microtubule-associated protein, the role of which may be located exclusively in the cytoplasm [229]. Thus, the relationship of KEM1p with telomere biology is not quite clear.

Possible biological roles for G-quadruplexes

No conclusive role for either the G-quadruplexes or the i-motif structures has been confirmed. However, the evolutionary conservation and strandwise segregation of the telomeric G-rich and C-rich sequences by all but a few eukaryotic organisms suggests a conserved function for these sequences. Moreover, a study that examined the positioning of adenine residues relative to the guanine stretches of a number of telomeric repeats suggests that eukaryotes seem to have selected precisely those repeats for their telomeres, which are able to form quadruplex structures; conversely, those G-rich motifs that are unable to form quadruplexes have been avoided [230]. The case for the existence of G-quadruplex structures *in vitro* is further supported by the quadruplex-forming and removal activities of the groups of telomere-associated proteins discussed above. These proteins facilitate, under physiological conditions of salt concentration and temperature, the formation and removal of G-quadruplexes by the G-rich telomeric overhangs.

The most direct evidence for the existence of non-canonical telomeric structures based on guanine–guanine interactions is derived from studies on telomere replication in yeast. On telomeres of yeast chromosomes and linear plasmids, long (>30 nt) G-'tails' appear in a transient and cell-cycle-specific manner [18, 115, 116, 233]. When linear plasmids with these tails were isolated from the cells and analysed in two-dimensional agarose gels, a fraction of the plasmids behaved like a nicked circular DNA of the same size [115]. These and other data implied that the two G-rich tails on each plasmid telomere were interacting with one another by some non-Watson–Crick means [113, 115]. However, the existence of the associations *in vivo* could not be inferred unequivocally from these experiments: the association of the tails might have occurred during isolation of the DNA and was not necessarily present within the cell.

What functions, if any, might G-quadruplex structures have *in vivo*? The following proposals have been made.

Meiotic and mitotic pairing. The formation of guanine-quartets represents a different sort of 'recognition' process than the formation of Watson–Crick A–T and G–C base pairs: the G–G bonding can represent a recognition of 'self', whereas Watson–Crick base pairing involves the recognition of complementarity. Two or four *identical* single-stranded DNA molecules can recognise one another and bind specifically to one another via the formation of G-quartets [130, 146]. A self-recognition between the single-stranded, G-rich telomeric termini within a cell could take place at the onset of meiotic pairing of chromosomes [146]. Furthermore, it has been proposed that similar self-recognition mechanisms may also occur at internal chromosomal loci.

This concept of self-recognition has been used to propose a general mechanism for a DNA-mediated association of chromatids (whether mitotic or meiotic) in a 2-fold or 4-fold manner [130, 132, 146]. However, to date there is no direct evidence for such DNA–DNA interactions and whether any non-canonical base-pairings are involved in the above-mentioned processes *in vivo*.

Cell-cycle control related functions. Although formally being a double-stranded DNA break, a telomere is not recognised by the cellular control mechanisms that ensure DNA integrity [3]. The most prominent DNA damage checkpoint in yeast is mediated by the *RAD9* gene and arrests the cell cycle in late-S/G2-phase [232]. Within the precision of the experiments, this is exactly the phase in which yeast telomeres acquire long, single-stranded G-tails [18, 115]. It

is thus tempting to speculate that the formation of an 'alternative DNA structure', such as telomeric G-quadruplexes, at this particular stage in the cell cycle, might serve as signal for the completion of chromosome replication. In addition, such a structure could be a terminal DNA structure that is not recognised by the cell cycle checkpoint and therefore allows the cell cycle to progress normally [18, 233].

Recombination. Interstitial telomeric sequences in the Armenian hamster and in *Paramecium* exhibit high levels of recombination [234, 235]. An explanation offered for these recombinogenic properties of telomeric sequences, even when such sequences are interstitial, is that the sequence bias (and the presence of clusters of guanines) allow for the potential formation of strong, alternative, non-Watson–Crick base-pairing interactions by these sequences [146, 236–238]. For instance, telomeric sequences from *Tetrahymena* are highly susceptible to cleavage by S1 nuclease when present in a supercoiled plasmid [239]; and vertebrate telomeric sequences, under similar conditions, show evidence of being single-stranded and forming non-canonical structures [240]. In addition, a number of studies on illegitimate recombination have reported the predominance of telomere-like, repetitive, guanine-rich sequences at such sites [234–236, 241, 242]. Recently, it has been shown that G–G base paired regions within Watson–Crick duplexes, such as may transiently be formed by either terminal or interstitial telomeric sequences, are liable to form G-quadruplexes, which may in turn lead to illegitimate recombination events [138, 238].

Control of telomere replication and inhibition of telomerase. Shortly after the first description of the formation of G-quadruplex structures by telomeric sequences, Zahler and associates [243] reported that the *Oxytricha* telomeric sequence, while in the form of a G-quadruplex, was not a substrate for elongation by telomerase. This inhibition seemed to stem (i) from an inaccessibility of the telomeric primers to telomerase and (ii) from a limited extension of the primers by telomerase. It was therefore proposed that the formation of G-quadruplexes by telomeric sequences might serve to control telomere replication. However, it is important to note that *in vivo*, such a control function could only be effective once at least a few repeats have been added to the substrate. The short, 12-base overhang in the terminal complex of *Oxytricha* telomeres is recognised and elongated efficiently by telomerase [244].

CONCLUSION

The telomeres have a number of essential functions and a failure of any of them could lead to genetic instability, a possible prelude to cancerous cell transformation, or cell death. For these reasons, recently there has been heightened interest in telomere research and our knowledge of the actual sequence arrangements at the telomeres has increased dramatically. Somewhat surprisingly, however, despite these efforts, virtually nothing is known about the physical structure of the very ends of vertebrate chromosomes. Are mammalian chromosome ends also organised with an overhang of the G-strand? How long are those overhangs and what changes occur at the ends during the cell cycle, particularly at times when the telomeres are replicated? What roles, if any, have the proteins that bind those G-rich single strands? In addition, many telomeric repeat sequences have a number of very unique features, including an ability to form alternative base interactions that could lead to non-canonical DNA structures. Again however, while substantial progress to understanding the *in vitro* physico-chemical properties of these structures has been made, any occurrence or role of such structures *in vivo* still remain to be demonstrated conclusively. It is thus conceivable that the putative single-stranded tails, and the proteins that bind them, play a much more prominent role in telomere biology than was anticipated.

1. McClintock B. The behaviour in successive nuclear divisions of a chromosome broken at meiosis. *Proc Natl Acad Sci USA* 1939, **25**, 405–416.
2. McClintock B. The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 1941, **26**, 234–282.
3. Sandell LS, Zakian VA. Loss of a yeast telomere: arrest, recovery, and chromosome loss. *Cell* 1993, **75**, 729–739.
4. Muller HJ. The remaking of chromosomes. *The Collecting Net* 1938, **13**, 181–195, 198.
5. Olovnikov AM. A theory of marginotomy. *J Theor Biol* 1973, **41**, 181–190.
6. Watson JD. Origin of concatemeric DNA. *Nature* 1972, **239**, 197–201.
7. Rabl C. Ueber Zelltheilung. *Morphologisches Jahrbuch* 1885, **10**, 214–330.
8. Dernburg AF, Sedat JW, Cande WZ, Bass HW. Cytology of telomeres. In Blackburn EH, Greider CW, eds. *Telomeres*, 1st edition. Cold Spring Harbor Laboratory Press 1995, 295–338.
9. Levis R, Hazelrigg T, Rubin GM. Effects of genomic position on the expression of transduced copies of the *white* gene of *Drosophila*. *Science* 1985, **229**, 558–561.
10. Pays E, Steinert M. Control of antigen gene expression in African trypanosomes. *Annu Rev Genet* 1988, **22**, 107–126.
11. Gottschling DE, Aparicio OM, Billington BL, Zakian VA. Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* 1990, **63**, 751–762.
12. Nimmo ER, Cranston G, Allshire RC. Telomere-associated chromosome breakage in fission yeast results in variegated expression of adjacent genes. *EMBO J* 1994, **13**, 3801–3811.
13. Van der Werf A, Van Assel S, Aerts D, Steinert M, Pays E. Telomere interactions may condition the programming of antigen expression in *Trypanosoma brucei*. *EMBO J* 1990, **9**, 1035–1040.
14. Karpen GH, Spradling AC. Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome *Dp1187* by single P element insertional mutagenesis. *Genetics* 1992, **132**, 737–753.
15. Ferguson BM, Brewer BJ, Reynolds AE, Fangman WL. A yeast origin of replication is activated late in S phase. *Cell* 1991, **65**, 507–515.
16. Ferguson B, Fangman WL. A position effect on the time of replication origin activation in yeast. *Cell* 1992, **68**, 333–339.
17. Fangman WL, Brewer BJ. A question of time: replication origins of eukaryotic chromosomes. *Cell* 1992, **71**, 363–366.
18. Wellinger RJ, Wolf AJ, Zakian VA. Origin activation and formation of single strand TG_{1-3} tails occur sequentially in late S phase on a yeast linear plasmid. *Mol Cell Biol* 1993, **13**, 4057–4065.
19. Biessmann H, Mason JM. Genetics and molecular biology of telomeres. *Adv Genet* 1992, **30**, 185–249.
20. Louis EJ. The chromosome ends of *Saccharomyces cerevisiae*. *Yeast* 1995, **11**, 1553–1573.
21. Burmeister M, Kim S, Price ER, *et al.* A map of the distal region of the long arm of the human chromosome 21 constructed by radiation hybrid mapping an pulsed-field gel electrophoresis. *Genomics* 1991, **9**, 19–30.
22. Wilkie AO, Higgs DR, Rack KA, *et al.* Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* 1991, **64**, 595–606.

23. Ellis NA, Goodfellow PJ, Pym B, *et al.* The pseudoautosomal boundary in man is defined by an Alu repeat sequence inserted on the Y chromosome. *Nature* 1989, **337**, 81–84.
24. Bedbrook JR, O'Dell M, Flawell RB. Amplification of rearranged repeated DNA sequences in cereal plants. *Nature* 1980, **288**, 133–137.
25. Richards EJ, Chao S, Vongs A, Yang J. Characterization of *Arabidopsis thaliana* telomeres isolated in yeast. *Nucleic Acids Res* 1992, **20**, 4039–4046.
26. Chan CSM, Tye B-K. A family of *Saccharomyces cerevisiae* repetitive autonomously replication sequences that have very similar genomic environments. *J Mol Biol* 1983, **168**, 505–523.
27. Chan CSM, Tye B-K. Organization of DNA sequences and replication origins at yeast telomeres. *Cell* 1983, **33**, 563–573.
28. Louis EJ, Naumova ES, Lee A, Naumov G, Haber JE. The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. *Genetics* 1994, **136**, 789–802.
29. Rubin GM. Isolation of a telomeric DNA sequence from Drosophila melanogaster. *Cold Spring Harbor Symp Quant Biol* 1977, **42**, 1041–1046.
30. Young BS, Pession A, Traverse KL, French C, Pardue ML. Telomere regions in Drosophila share complex DNA sequences with pericentric heterochromatin. *Cell* 1983, **34**, 85–94.
31. Lòpez CC, Nielsen L, Edström J-E. Terminal long tandem repeats in chromosomes from *Chironomus pallivittatus*. *Mol Cell Biol* 1996, **16**, 3285–3290.
32. Cross SH, Allshire RC, McKay SJ, McGill NI, Cooke JH. Cloning of human telomeres by complementation in yeast. *Nature* 1989, **338**, 771–774.
33. de Lange T, Shiue L, Myers RM, *et al.* Structure and variability of human chromosome ends. *Mol Cell Biol* 1990, **10**, 518–527.
34. Rouyer F, de la Chapelle M, Andersson M, Weissenbach J. An interspersed repeated sequence specific for human subtelomeric regions. *EMBO J* 1990, **9**, 505–514.
35. Cheng JF, Smith CL, Cantor CR. Structural and transcriptional analysis of a human subtelomeric repeat. *Nucleic Acids Res* 1991, **19**, 149–154.
36. Henderson E. Telomere DNA structure. In Blackburn EH, Greider CW, eds. *Telomeres,* 1st edition. Cold Spring Harbor Laboratory Press, 1995, 11–34.
37. Brown WR, MacKinnon PJ, Villasante A, Spurr N, Buckle VJ, Dobson MJ. Structure and polymorphism of human telomere-associated DNA. *Cell* 1990, **63**, 119–132.
38. Weber BL, Allen L, Magenis RE, Hayden MR. A low-copy repeat located in subtelomeric regions of 14 different human chromosomal termini. *Cytogenet Cell Genet* 1991, **57**, 179–183.
39. Zakian VA. Telomeres: beginning to understand the end. *Science* 1995, **270**, 1601–1607.
40. Simmler MC, Rouyer F, Vergnaud G, *et al.* Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature* 1985, **317**, 692–697.
41. Cooke HJ, Smith BA. Variability at the telomeres of the human X/Y pseudoautosomal region. *Cold Spring Harbor Symp Quant Biol* 1986, **51**, 213–219.
42. Foote S, Kemp DJ. Chromosomes of malaria parasites. *Trends Genet* 1989, **5**, 337–342.
43. Walter MF, Jang C, Kasravi B, *et al.* DNA organization and polymorphism of a wild-type *Drosophila* telomere region. *Chromosoma* 1995, **104**, 229–241.
44. de Bruin D, Lanzer M, Ravetch JV. The polymorphic subtelomeric regions of *Plasmodium falciparum* chromosomes contain arrays of repetitive sequence elements. *Proc Natl Acad Sci USA* 1994, **91**, 619–623.
45. Allshire RC, Dempster M, Hastie ND. Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Res* 1989, **17**, 4611–4627.
46. Dore E, Pace T, Ponzi M, Picci L, Fontali C. Organization of subtelomeric repeats in *Plasmodium berghei*. *Mol Cell Biol* 1990, **10**, 2423–2427.
47. Carlson M, Celenza JL, Eng FJ. Evolution of the dispersed SUC gene family of *Saccharomyces* by rearrangements of chromosome telomeres. *Mol Cell Biol* 1985, **5**, 2894–2902.
48. Michels CA, Read E, NAt K, Charron MJ. The telomere-associated MAL3 locus of *Saccharomyces* is a tandem array of repeated genes. *Yeast* 1992, **8**, 655–665.
49. Pologe LG, Ravetch JV. A chromosomal rearrangement in a *P. falciparum* histidine-rich protein gene is associated with the knobless phenotype. *Nature* 1986, **322**, 474–477.
50. Corcoran LM, Thompson JK, Walliker D, Kemp DJ. Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P. falciparum*. *Cell* 1988, **53**, 807–813.
51. Wright WE, Shay JW. Telomere positional effects and the regulation of cellular senescence. *Trends Genet* 1992, **8**, 193–197.
52. Lima-de-Faria A. *Molecular Evolution and Organization of the Chromosome*. Amsterdam, Elsevier, 1983.
53. Gilson E, Laroche T, Gasser SM. Telomeres and the functional architecture of the nucleus. *Trends Cell Biol* 1993, **3**, 128–134.
54. Lundblad V, Blackburn EH. An alternative pathway for yeast telomere maintenance rescues *estl*$^-$ senescence. *Cell* 1993, **73**, 347–360.
55. Biessmann H, Champion LE, O'Hair M, Ikenaga K, Kasravi B, Mason JM. Frequent transposition of *Drosophila melanogaster* HeT-A retroposons to receding chromosome ends. *EMBO J* 1992, **11**, 4459–4469.
56. Levis RW, Ganesan R, Houtchens K, Tolar LA, Sheen F-M. Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* 1993, **75**, 1083–1093.
57. Pardue ML, Danilevskaya ON, Lowenhaupt K, Slot F, Traverse KL. *Drosophila* telomeres: new views on chromosome evolution. *Trends Genet* 1996, **12**, 48–52.
58. Blackburn EH, Gall JG. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. *J Mol Biol* 1978, **120**, 33–53.
59. Katzen AL, Cann GM, Blackburn EH. Sequence-specific fragmentation of macronuclear DNA in a Holotrichous ciliate. *Cell* 1981, **24**, 313–320.
60. Oka Y, Shiota S, Nakai S, Nishida Y, Okubo S. Inverted terminal repeat sequence in the macronuclear DNA of *Stylonychia pustulata*. *Gene* 1980, **10**, 301–306.
61. Klobutcher LA, Swanton MT, Donini P, Prescott DM. All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unsual 3′ terminus. *Proc Natl Acad Sci USA* 1981, **78**, 3015–3019.
62. Baroin A, Prat A, Caron F. Telomeric site position heterogeneity in macronuclear DNA of *Paramecium primaurelia*. *Nucleic Acids Res* 1987, **15**, 1717–1728.
63. Forney JD, Blackburn EH. Developmentally controlled telomere addition in wild type and mutant paramecia. *Mol Cell Biol* 1988, **8**, 251–258.
64. Blackburn EH, Challoner PB. Identification of a telomeric DNA sequence in *Trypanosoma brucei*. *Cell* 1984, **36**, 447–457.
65. Vander Ploeg LHT, Liu AUC, Borst P. Structure of growing telomeres of trypanosomes. *Cell* 1984, **36**, 459–468.
66. Le Blancq SM, Kase RS, Van der Ploeg LH. Analysis of *Giardia lamblia* rDNA encoding telomere with $(TAGGG)_n$ as the telomere repeat. *Nucleic Acids Res* 1991, **19**, 5790.
67. Ponzi M, Pace T, Dore E, Frontali C. Identification of a telomeric DNA sequence in *Plasmodium berghei*. *EMBO J* 1985, **4**, 2991–2995.
68. Sohanpal BK, Morzaria SP, Gobright EI, Bishop RP. Characterization of the telomeres at opposite ends of a 3 Mb *Theileria parva* chromosome. *Nucleic Acids Res* 1995, **23**, 1942–1947.
69. Emery HS, Weiner AM. An irregular satellite sequence is found at the termini of the linear extrachromosomal rDNA in *Dictyostelium discoideum*. *Cell* 1981, **26**, 411–419.
70. Forney J, Henderson ER, Blackburn EH. Identification of the telomeric sequence of the acellular slime molds *Didymium iridis* and *Physarum polycephalum*. *Nucleic Acids Res* 1987, **14**, 9143–9151.
71. Schechtman MG. Characterization of telomere DNA from *Neurospora crassa*. *Gene* 1990, **88**, 159–165.
72. Javerzat JP, Bhattarcherjee V, Barreau C. Isolation of telomeric DNA from the filamentous fungus *Podospora anserina* and construction of a self-replicating linear plasmid showing

high transformation frequency. *Nucleic Acids Res* 1993, **21**, 497–504.
73. Coleman MJ, McHale MT, Arnau J, Watson A, Oliver RP. Cloning and characterization of telomeric DNA from *Cladosporium fulvum*. *Gene* 1993, **132**, 67–73.
74. Woods JP, Goldman WE. In vivo generation of linear plasmids with addition of telomeric sequences by *Histoplasma capsulatum*. *Mol Microbiol* 1992, **6**, 3603–3610.
75. Edman JC. Isolation of telomerelike sequences from *Cryptococcus neoformans* and their use in high-efficiency transformation. *Mol Cell Biol* 1992, **12**, 2777–2783.
76. Shampay J, Szostak JW, Blackburn EH. DNA sequences of telomeres maintained in yeast. *Nature* 1984, **310**, 154–157.
77. McEachern MJ, Blackburn EH. A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts. *Proc Natl Acad Sci USA* 1994, **91**, 3453–3457.
78. McEachern MJ, Hicks JB. Unusually large telomeric repeats in the yeast *Candida albicans*. *Mol Cell Biol* 1993, **13**, 551–560.
79. Sugawara N, Szostak JW. Telomeres of *Schizosaccharomyces pombe*. *Yeast* (Suppl.) 1986, **1**, 373.
80. Underwood AP, Louis EJ, Borts RH, Stringer JR, Wakefield AE. *Pneumocystis carinii* telomere repeats are composed of TTAGGG and the subtelomeric sequence contains a gene encoding the major surface glycoprotein. *Mol Microbiol* 1996, **19**, 273–281.
81. Petracek ME, Lefebvre PA, Silflow CD, Berman J. *Chlamydomonas* telomere sequences are A + T-rich but contain three consecutive GC base pairs. *Proc Natl Acad Sci USA* 1990, **87**, 8222–8226.
82. Higashiyama T, Maki S, Yamada T. Molecular organization of *Chlorella vulgaris* chromosome I: presence of telomeric repeats that are conserved in higher plants. *Mol Gen Genet* 1995, **246**, 29–34.
83. Richards EJ, Ausubel FM. Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* 1988, **53**, 127–136.
84. Burr B, Burr FA, Matz EC, Romero-Severson J. Pinning down lose ends: mapping telomeres and factors affecting their length. *Plant Cell* 1992, **4**, 953–960.
85. Kilian A, Kleinhofs A. Cloning and mapping of telomere-associated sequences from *Hordeum vulgare*. *Mol Gen Genet* 1992, **235**, 153–156.
86. Fajkus J, Kovarik A, Kralovics R, Bezdek M. Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. *Mol Gen Genet* 1995, **247**, 633–638.
87. Ganal MW, Lapitan NL, Tanksley SD. Macrostructure of the tomato telomeres. *Plant Cell* 1991, **3**, 87–94.
88. Müller F, Wicky C, Spicher A, Tobler H. New telomere formation after developmentally regulated chromosomal breakage during the process of chromatin diminuition in *Ascaris lumbricoides*. *Cell* 1991, **67**, 815–822.
89. Teschke C, Solleder G, Moritz KB. The highly variable pentameric repeats of the AT-rich germline limited DNA in *Parascaris universalens* are the telomeric repeats of somatic chromosomes. *Nucleic Acids Res* 1991, **19**, 2677–2684.
90. Cangiano G, La Volpe A. Repetitive DNA sequences located at the terminal portion of the *Caenorhabditis elegans* chromosomes. *Nucleic Acids Res* 1993, **21**, 1133–1139.
91. Okazaki S, Tsuchida K, Maekawa H, Ishikawa H, Fujiwara H. Identification of a pentanucleotide telomeric sequence, $(TTAGG)_n$ in the silkworm *Bombyx mori* and in other insects. *Mol Cell Biol* 1993, **13**, 1424–1432.
92. Moyzis RK, Buckingham JM, Cram LS, *et al.* A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, present at the telomeres of human chromosomes. *Proc Natl Acad Sci USA* 1988, **85**, 6622–6626.
93. Meyne J, Ratliff RL, Moyzis RK. Conservation of the human telomere sequence $(TTAGGG)_n$ among vertebrates. *Proc Natl Acad Sci USA* 1989, **86**, 7049–7053.
94. Kipling D, Cooke HJ. Hypervariable ultra-long telomeres in mice. *Nature* 1990, **347**, 400–402.
95. Starling JA, Maule J, Hastie ND, Allshire RC. Extensive telomere repeat arrays in mouse are hypervariable. *Nucleic Acids Res* 1990, **18**, 6881–6888.
96. Murnane JP, Sabatier L, Marder BA, Morgan WF. Telomere dynamics in an immortal human cell line. *EMBO J* 1994, **13**, 4953–4962.
97. Shampay J, Blackburn EH. Generation of telomere-length heterogeneity in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 1988, **85**, 534–538.
98. Li B, Lustig AJ. A novel mechanism for telomere size control in *Saccharomyces cerevisiae*. *Genes Dev* 1996, **10**, 1310–1326.
99. Larson DD, Spangler EA, Blackburn EH. Dynamics of telomere length variation in *Tetrahymena thermophila*. *Cell* 1987, **50**, 477–483.
100. Bernards A, Michels PAM, Lincke CR, Borst P. Growth of chromosome ends in multiplying trypanosomes. *Nature* 1983, **303**, 592–597.
101. Harley CB, Futcher AB, Greider CW. Telomeres shorten during ageing of human fibroblasts. *Nature* 1990, **345**, 458–460.
102. Hastie ND, Dempster M, Dunlop MG, Thompson AM, Green DK, Allshire RC. Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 1990, **346**, 866–868.
103. Yao M-C, Yao C-H. Repeated hexanucleotide C-C-C-C-A-A is present near free ends of macronuclear DNA of *tetrahymena*. *Proc Natl Acad Sci USA* 1981, **78**, 7436–7439.
104. Cherry JM, Blackburn EH. The internally located telomeric sequences in the germ-line chromosomes of *Tetrahymena* are at the ends of transposon-like elements. *Cell* 1985, **43**, 747–753.
105. Rossi E, Floridia M, Casali M, *et al.* Stability and phenotypic consequences of chromosome rearrangements leading to interstitial telomeric sequences. *J Med Genet* 1993, **30**, 926–931.
106. IJido JW, Baldini A, Ward DC, Reeders ST, Wells RA. Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc Natl Acad Sci USA* 1991, **88**, 9051–9055.
107. Morin GB, Cech TR. Mitrochondrial telomeres: surprising diversity of repeated telomeric DNA sequences among six species of *Tetrahymena*. *Cell* 1988, **52**, 367–374.
108. Biessmann H, Carter SB, Mason JM. Chromosome ends in *Drosophila* without telomeric DNA sequences. *Proc Natl Acad Sci USA* 1990, **87**, 1758–1761.
109. Mason JM, Biessmann H. The unusual telomeres of *Drosophila*. *Trends Genet* 1995, **11**, 58–62.
110. Pardue M-L. *Drosophila* telomeres: another way to end it all. In Blackburn E, Greider CW, eds. *Telomeres*, 1st edition, Monograph 29. Cold Spring Harbor Laboratory Press, 1995, 339–370.
111. Henderson ER, Blackburn EH. An overhanging 3′ terminus is a conserved feature of telomeres. *Mol Cell Biol* 1989, **9**, 345–348.
112. Zahler AM, Prescott DM. DNA primase and the replication of the telomeres in Oxytricha nova. *Nucleic Acids Res* 1989, **17**, 6299–6317.
113. Wellinger RJ, Ethier K, Labrecque P, Zakian VA. Evidence for a new step in telomere maintenance. *Cell* 1996, **85**, 423–433.
114. Wellinger RJ, Wolf AJ, Zakian VA. Use of non-denaturing Southern hybridization and two dimensional agarose gels to detect putative intermediates in telomere replication in *Saccharomyces cerevisiae*. *Chromosoma* 1992, **102**, S150–S156.
115. Wellinger RJ, Wolf AJ, Zakian VA. *Saccharomyces* telomeres acquire single-strand TG_{1-3} tails late in S phase. *Cell* 1993, **72**, 51–60.
116. Dionne I, Wellinger RJ. Cell cycle regulated generation of single-stranded G-rich DNA in the absence of telomerase. *Proc Natl Acad Sci* 1996, **93**, 13902–13907.
117. Roth M, Prescott DM. DNA intermediates and telomere addition during genome reorganization in *Euplotes crassus*. *Cell* 1985, **41**, 411–417.
118. Blackburn EH, Greider CW. *Telomeres*. Cold Spring Harbor Laboratory Press, 1995.
119. Gottschling DE, Cech TR. Chromatin structure of the molecular ends of *Oxytricha* macronuclear DNA: phased nucleosomes and a telomeric complex. *Cell* 1984, **38**, 501–510.
120. Gottschling DE, Zakian VA. Telomere proteins: specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. *Cell* 1986, **47**, 195–205.
121. Budarf ML, Blackburn EH. Chromatin structure of the telomeric region and 3′-nontranscribed spacer of *Tetrahymena* ribosomal RNA genes. *J Biol Chem* 1986, **261**, 363–369.

122. Price CM. Telomere structure in Euplotes crassus: characterization of DNA-protein interactions and isolation of a telomere-binding protein. *Mol Cell Biol* 1990, **10**, 3421–3431.
123. Blackburn EH, Chiou S-S. Non-nucleosomal packaging of a tandemly repeated DNA sequence at termini of extra-chromosomal DNA coding for rDNA in Tetrahymena. *Proc Natl Acad Sci USA* 1981, **78**, 2263–2267.
124. Wright JH, Gottschling DE, Zakian VA. *Saccharomyces* telomeres assume a non-nucleosomal chromatin structure. *Genes Devel* 1992, **6**, 197–210.
125. Makarov VL, Lejnine S, Bedoyan J, Langmore JP. Nucleosomal organization of telomere-specific chromatin in rat. *Cell* 1993, **73**, 775–787.
126. Tommerup H, Dousmanis A, de Lange T. Unusual chromatin in human telomeres. *Mol Cell Biol* 1994, **14**, 5777–5785.
127. Lejnine S, Makarov V, Langmore JP. Conserved nucleoprotein structure at the ends of vertebrate and invertebrate chromosomes. *Proc Natl Acad Sci USA* 1995, **92**, 2393–2397.
128. de Lange T. In search of vertebrate telomeric proteins. *Cell Dev Biol* 1996, 7, 23–29.
129. Oka Y, Thomas CA Jr. The cohering telomeres of *Oxytricha*. *Nucleic Acids Res* 1987, **15**, 8877–8898.
130. Sundquist WI, Klug A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* 1989, **342**, 825–829.
131. Henderson E, Hardin CC, Walk SK, Tinoco JI, Blackburn EH. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell* 1987, **51**, 899–908.
132. Williamson JR, Raghuraman MK, Cech TR. Monovalent cation-induced structure of telomeric DNA: the G-Quartet model. *Cell* 1989, **59**, 871–880.
133. Sen D, Gilbert W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* 1990, **344**, 410–414.
134. Choi KH, Choi BS. Formation of a hairpin structure by telomere 3′ overhang. *Biochim Biophys Acta* 1994, **217**, 341–344.
135. Cognet JAH, Gabarro-Arpa J, Le Bret M, van der Martel GA, van Boom JH, Fazakerley GV. Solution conformation of an oligonucleotide containing a GG mismatch determined by nuclear magnetic resonance and molecular mechanics. *Nucleic Acids Res* 1991, **19**, 6771–6779.
136. Borden K, Jenkins TC, Skelly JV, Brown T, Lane AN. Conformational properties of the GG mismatch in $d(CGCGAATTGGCG)_2$ determined by NMR. *Biochemistry* 1992, **31**, 5411–5422.
137. Skelly JV, Edwards KJ, Jenkins TC, Neidle S. Crystal structure of an oligonucleotide duplex containing GG base pairs: influence of mispairing on DNA backbone conformation. *Proc Natl Acad Sci USA* 1993, **90**, 804–808.
138. Venczel EA, Sen D. Synapsable DNA. *J Mol Biol* 1996, **257**, 219–224.
139. Rippe K, Fritsch V, Westhof E, Jovin TM. Alternating d(GA) sequences form a parallel-stranded DNA homoduplex. *EMBO J* 1992, **11**, 3777–3786.
140. Huertas D, Bellsolell L, Casanovas JM, Coll M, Azorin F. Alternating $d(GA)_n$ DNA sequences form antiparallel stranded homoduplexes stabilized by the formation of GA base pairs. *EMBO J* 1993, **12**, 4029–4038.
141. Suda T, Mishima Y, Asakura H, Kominami R. Formation of parallel-stranded DNA homoduplex by d(GGA) repeat oligonucleotides. *Nucleic Acids Res* 1995, **23**, 3771–3777.
142. Veselkov AG, Malkov VA, Frank-Kamenitskii MD, Dobrynin VN. Triplex model of chromosome ends. *Nature* 1993, **364**, 496.
143. Radhakrishnan I, Patel DJ. DNA triplexes: solution structures, hydration sites, energetics, interactions, and function. *Biochemistry* 1994, **33**, 11405–11416.
144. Frank-Kamenetskii MD, Mirkin SM. Triplex DNA structures. *Annu Rev Biochem* 1995, **64**, 65–95.
145. Olivas WM, Maher LJ. Competitive triplex/quadruplex equilibria involving guanine-rich oligonucleotides. *Biochemistry* 1995, **34**, 278–284.
146. 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.
147. Panyutin IG, Kovalsky OI, Budowsky EI, Dickerson RE, Rikhirev ME, Lipanov AA. G-DNA: a twice-folded DNA structure adopted by single-stranded oligo(dG) and its implications for telomeres. *Proc Natl Acad Sci USA* 1990, **87**, 867–870.
148. Panyutin IG, Kovalsky OI, Budowsky EI. Magnesium-dependent supercoiling-induced transition in $(dG)_n.(dC)_n$ stretches and formation of a new G-structure by $(dG)_n$ strand. *Nucleic Acids Res* 1989, **17**, 8257–8271.
149. Williamson JR. G-Quartet structures in telomeric DNA. *Annu Rev Biophys Biomol Struct* 1994, **23**, 703–730.
150. Rhodes D, Giraldo R. Telomere structure and function. *Curr Opin Struct Biol* 1995, **5**, 311–322.
151. Gellert M, Lipsett MN, Davies DR. Helix formation of guanylic acid. *Proc Natl Acad Sci USA* 1962, **48**, 2013–2018.
152. Pinnavaia TJ, Marshall CL, Mettler CM, Fisk CL, Miles HT, Becker ED. Alkali metal ion specificity in the solution of ordering of a nucleotide 5′-guanosine monophosphate. *J Am Chem Soc* 1978, **100**, 3625–2627.
153. Sen D, Gilberrt W. Guanine quartet structures. *Methods Enzymol* 1992, **211**, 191–199.
154. Sen D, Gilbert W. Cationic switches in the formation of DNA structures containing guanine-quartets. In Sarma RH, Sarma MA, eds. *Structure and Function,* 1st edn, Vol. 1. New York, Adenine Press, 1992, 43–52.
155. Hardin CC, Henderson E, Watson T, Prosser JK. Monovalent cation-induced structural transitions in telomeric DNAs: G-DNA folding intermediates. *Biochemistry* 1991, **30**, 4460–4472.
156. Hardin CC, Watson T, Corregan M, Bailey C. Cation-dependent transition between the quadruplex and Watson–Crick hairpin form of $d(CGCG_3GCG)$. *Biochemistry* 1992, **31**, 833–841.
157. Venczel EA, Sen D. Parallel and antiparallel G-DNA structures from a complex telomeric sequence. *Biochemistry* 1993, **32**, 6220–6228.
158. Ross WS, Hardin CC. Ion-induced stabilization of the G-DNA quadruplex: free energy perturbation studies. *J Am Chem Soc* 1994, **116**, 6070–6080.
159. Kim J, Cheong C, Moore PB. Tetramerization of an RNA oligonucleotide containing a GGGG sequence. *Nature* 1991, **351**, 331–332.
160. Acevedo OL, Dickinson LA, Macke TJ, Thomas CA Jr. The coherence of synthetic telomeres. *Nucleic Acids Res* 1991, **19**, 3409–3419.
161. Chen FM. Sr^{2+} facilitates intermolecular G-quadruplex formation of telomeric sequences. *Biochemistry* 1992, **31**, 3769–3776.
162. Lee JS. The stability of polypurine tetraplexes in the presence of mono- and divalent cations. *Nucleic Acids Res* 1990, **18**, 6057–6060.
163. Jin R, Gaffney BL, Wang C, Jones RA, Breslauer KJ. Thermodynamics and structure of a DNA tetraplex: a spectroscopic and calorimetric study of the tetramolecular complexes of $d(TG_3T)$ and $d(TG_3T_2G_3T)$. *Proc Natl Acad Sci USA* 1992, **89**, 8832–8836.
164. Wang Y, Patel D. Guanine residues in $d(T_2AG_3)$ and $d(T_2G_4)$ form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry* 1992, **31**, 8112–8119.
165. Wang Y, Patel D. Solution structure of a parallel-stranded G-quadruplex DNA. *J Mol Biol* 1993, **234**, 1171–1183.
166. Wang Y, Patel D. Solution structure of the human telomeric repeat $d[AG_3(T_2AG_3)_3]$ G-tetraplex. *Structure* 1993, **1**, 263–282.
167. Wang Y, Patel D. Solution structure of the *Tetrahymena* telomeric repeat $d(T_2G_4)_4$ G-tetraplex. *Structure* 1994, **2**, 1141–1156.
168. Gupta G, Garcia AE, Guo Q, Lu M, Kallenbach NR. Structure of a parallel-stranded tetramer of the *Oxytricha* telomeric DNA sequence dT_4G_4. *Biochemistry* 1993, **32**, 7098–7103.
169. Aboul-ela F, Murchie AIH, Norman DG, Lilley DMJ. NMR study of parallel-stranded tetraplex formation by the hexadeoxynucleotide $d(TG_4T)$. *J Mol Biol* 1994, **243**, 458–471.

170. Laughlan G, Murchie AIH, Normand DG, *et al.* The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science* 1994, **265**, 520–524.
171. Cheong C, Moore PB. Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry* 1992, **31**, 8406–8414.
172. Kang CH, Zhang X, Ratliff R, Moyzis R, Rich A. Crystal structure of four-stranded *Oxytricha* telomeric DNA. *Nature* 1992, **356**, 126–131.
173. Smith FW, Feigon J. Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides. *Nature* 1992, **356**, 164–168.
174. Mohanty D, Bansal M. Conformational polymorphism in telomeric structures: loop orientation and interloop pairing in ($G_4T_nG_4$). *Biopolymers* 1994, **34**, 1187–1211.
175. Schultze P, Smith FW, Feigon J. Refined solution structure of the dimeric quadruplex formed from the *Oxytricha* telomeric oligonucleotide d(GGGGTTTTGGGG). *Structure* 1994, **2**, 221–223.
176. Wang KY, Swaminathan S, Bolton PH. Tertiary structure motif of *Oxytricha* telomere DNA. *Biochemistry* 1994, **33**, 7517–7527.
177. Wang Y, Jin R, Gaffney B, Jones RA, Breslauer KJ. Characterization by ^{1}H NMR of glycosidic conformations in the tetramolecular complex formed by d(GGTTTTTGG). *Nucleic Acids Res* 1991, **19**, 4619–4622.
178. Schultze P, Macaya RF, Feigon J. Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). *J Mol Biol* 1994, **235**, 1532–1547.
179. Padmanabhan K, Padmanabhan KP, Ferrara JD, Sadler JE, Tulinsky A. The structure of alpha-thrombin inhibited by a 15-mer single- stranded DNA aptamer. *J Biol Chem* 1993, **268**, 17651–17654.
180. Pilch DS, Plum GE, Breslauer KJ. The thermodynamics of DNA structures that contain lesions or guanine tetrads. *Curr Opin Struct Biol* 1995, **5**, 334–342.
181. Lu M, Guo Q, Kallenbach NR. Thermodynamics of G-tetraplex formation by telomeric DNAs. *Biochemistry* 1993, **32**, 598–601.
182. Guo Q, Lu M, Markey LA, Kallenbach NR. Interaction of the dye ethidium bromide with DNA containing guanine repeats. *Biochemistry* 1992, **31**, 2451–2455.
183. Guo Q, Lu M, Kallenbach NR. Adenine affects the structure and stability of telomeric sequences. *J Biol Chem* 1992, **267**, 15293–15300.
184. Fang G, Cech TR. Characterization of a G-quartet formation reaction promoted by the beta-subunit of the *Oxytricha* telomere-binding protein. *Biochemistry* 1993, **32**, 11646–11657.
185. Wyatt JR, Davis PR, Freier SM. Kinetics of G-quartet-mediated tetramer formation. *Biochemistry* 1996, **35**, 8002–8008.
186. Raghuraman MK, Cech TR. Effect of monovalent cation-induced telomeric DNA structure on the binding of *Oxytricha* telomeric protein. *Nucleic Acids Res* 1990, **18**, 4543–4552.
187. Shida T, Yokoyama K, Tamai S, Sekiguchi J. Self-association of telomeric short oligodeoxyribonucleotides containing a dG cluster. *Chem Pharm Bull* 1991, **39**, 2207–2211.
188. Miura T, Benevides JM, Thomas GJJ. A phase diagram for sodium and potassium ion control of polymorphism in telomeric DNA. *J Mol Biol* 1995, **248**, 233–238.
189. Miura T, Thomas GJJ. Structural polymorphism of telomere DNA: interquadruplex and duplex–quadruplex conversions probed by Raman spectroscopy. *Biochemistry* 1994, **33**, 7848–7856.
190. Gehring K, Leroy J-L, Guéron M. A tetrameric DNA structure with protonated cytosine:cytosine base pairs. *Nature* 1993, **363**, 561–565.
191. Kang C, Berger I, Lockshin C, Ratliff R, Moyzis R, Rich A. Stable loop in the crystal structure of the intercalated four-stranded cytosine-rich metazoan telomere. *Proc Natl Acad Sci USA* 1995, **92**, 3874–3878.
192. Ahmed S, Kintanar A, Henderson E. Human telomeric C-strand tetraplexes. *Nature Struct Biol* 1994, **1**, 83–88.
193. Leroy JL, Guéron M, Mergny JL, Hélène C. Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. *Nucleic Acids Res* 1994, **22**, 1600–1606.
194. Leroy JL, Gehring K, Kettanni A, Guéron M. Acid multimers of oligodeoxycytidine strands: stoichiometry, base-pair characterization, and proton exchange properties. *Biochemistry* 1993, **32**, 6019–6031.
195. Lingner J, Cooper JP, Cech TR. Telomerase and DNA end replication: no longer a lagging strand problem? *Science* 1995, **269**, 1533–1534.
196. Leroy JL, Guéron M. Solution structures of the i-motif tetramers of d(TCC), d(^{5me}CCT) and d(T^{5me}CC): novel NOE connections between amino protons and sugar protons. *Structure* 1995, **3**, 101–120.
197. Voloshin ON, Veseljov AG, Belotserkovskii BO, *et al.* An eclectic DNA structure adopted by human telomere sequence under superhelical stress and low pH. *J Biomol Struct Dyn* 1992, **9**, 643–652.
198. Lacroix L, Mergny J-L, Leroy J-L, Hélène C. Inability of RNA to form the i-motif: implication for tetraplex formation. *Biochemistry* 1996, **35**, 8715–8722.
199. Zhong Z, Shiue L, Kaplan S, de Lange T. A mammalian factor that binds telomeric TTAGGG repeats *in vitro*. *Mol Cell Biol* 1992, **12**, 4834–4843.
200. Chong L, van Steensel B, Broccoli D, *et al.* A human telomeric protein. *Science* 1995, **270**, 1663–1667.
201. Liu ZP, Tye BK. A yeast protein that binds to vertebrate telomeres and conserved yeast telomeric junctions. *Genes Dev* 1991, **5**, 49–59.
202. Berman J, Tachibana CY, Tye B-K. Identification of a telomere-binding activity from yeast. *Proc Natl Acad Sci USA* 1986, **83**, 3713–3717.
203. Shore D, Nasmyth K. Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. *Cell* 1987, **51**, 721–732.
204. Zakian VA. *Saccharomyces* telomeres: function, structure, and replication. In Blackburn EH, Greider CW, eds. *Telomeres*, 1st edn. Cold Spring Harbor Laboratory Press, 1995, 107–137.
205. Price CM, Cech TR. Telomeric DNA-protein interactions of *Oxytricha* macronuclear DNA. *Genes Dev* 1987, **1**, 783–793.
206. Raghuraman MK, Cech TR. Assembly and self-association of *Oxytricha* telomeric nucleoprotein complexes. *Cell* 1989, **59**, 719–728.
207. Price CM, Cech TR. Properties of the tolemeric DNA-binding protein from *Oxytricha nova*. *Biochemistry* 1989, **28**, 769–774.
208. Coren JS, Epstein EM, Vogt VM. Characterization of a telomere-binding protein from *Physarum polycephalum*. *Mol Cell Biol* 1991, **11**, 2282–2290.
209. Cardenas ME, Bianchi A, de Lange T. A *Xenopus* egg factor with DNA-binding properties characteristic of terminus-specific telomeric proteins. *Genes Dev* 1993, **7**, 870–882.
210. Sheng H, Hou Z, Schierer T, Dobbs DL, Henderson E. Identification and characterization of a putative telomere end-binding protein from *Tetrahymena thermophila*. *Mol Cell Biol* 1995, **15**, 1144–1153.
211. Brigati C, Kurtz S, Balderes D, Vidali G, Shore D. An essential yeast gene encoding a TTAGGG repeat-binding protein. *Mol Cell Biol* 1993, **13**, 1306–1314.
212. Santori F, Domini P. *In vitro* identification of a protein of *Saccharomyces cerevisiae* that interacts specifically with the G-rich DNA strand of the telomere. *Res Microbiol* 1994, **145**, 519–530.
213. Petracek ME, Konkel LMC, Kable ML, Berman J. A Chlamydomonas protein that binds single-stranded G-strand telomere DNA. *EMBO J* 1994, **13**, 3648–3658.
214. Ishikawa F, Matunis MJ, Dreyfuss G, Cech TR. Nuclear proteins that bind the pre-mRNA 3′ splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)$_n$. *Mol Cell Biol* 1993, **13**, 4301–4310.
215. McKay SJ, Cooke H. A protein which specifically binds to single stranded TTAGGGn repeats. *Nucleic Acids Res* 1992, **20**, 1387–1391.
216. Lin J-J, Zakian VA. Isolation and characterization of two *Saccharomyces cerevisiae* genes that encode proteins that bind to (TG$_{1-3}$)*n* single strand telomeric DNA *in vitro*. *Nucleic Acids Res* 1994, **22**, 4906–4913.
217. Walsh K, Gualberto A. MyoD binds to the guanine tetrad nucleic acid structure. *J Biol Chem* 1992, **267**, 13714–13718.

218. Weisman-Shomer P, Fry M. QUAD, a protein from hepatocyte chromatin that binds selectively to guanine-rich quadruplex DNA. *J Biol Chem* 1993, **268**, 3306–3312.
219. Pearson AM, Rich A, Krieger M. Structure and stability of sodium and potassium complexes of dT_4G_4 and polynucleotide binding to macrophage scavenger receptors depends on the formation of base-quartet-stabilized four-stranded helices. *J Biol Chem* 1993, **268**, 3546–3554.
220. Schierer T, Hederson E. A protein from *Tetrahymena thermophila* that specifically binds parallel-stranded G4-DNA. *Biochemistry* 1994, **33**, 2240–2246.
221. Frantz JD, Gilbert W. A novel yeast gene product, G4p1, with a specific affinity for quadruplex nucleic acids. *J Biol Chem* 1995, **270**, 20692–20697.
222. Gualberto A, Patrick RM, Walsh K. Nucleic acid specificity of a vertebrate telomere-binding protein: evidence for G:G base pair recognition at the core binding site. *Genes Dev* 1992, **6**, 815–824.
223. Chung IK, Mehta VB, Spitzner JR, Muller MT. Eukaryotic topoisomerase II cleavage of parallel stranded DNA tetraplexes. *Nucleic Acids Res* 1992, **20**, 1973–1977.
224. Fang G, Cech TR. The β subunit of *Oxytricha* telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell* 1993, **74**, 875–885.
225. Giraldo R, Suzuki M, Chapman L, Rhodes D. Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. *Proc Natl Acad Sci USA* 1994, **91**, 7658–7660.
226. Giraldo R, Rhodes D. The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J* 1994, **13**, 2411–2420.
227. Liu Z, Frantz JD, Gilbert W, Bik-Kwoon T. Identification and characterization of a nuclease activity specific for G4 tetrastranded DNA. *Proc Natl Acad Sci USA* 1993, **90**, 3157–3161.
228. Liu Z, Gilbert W. The yeast KEM1 gene encodes a nuclease specific for G4 tetraplex DNA: implication of *in vivo* functions for this novel DNA structure. *Cell* 1994, **77**, 1–10.
229. Interthal H, Bellocq C, Baehler J, Bashkirov V, Edelstein S, Heyer W-D. A role of Sepl (=Keml, Xrnl) as a microtubule-associated protein in *Saccharomyces cerevisiae*. *EMBO J* 1995, **14**, 1057–1066.
230. Murchie AIH, Lilley DMJ. Tetraplex folding of telomere sequences and the inclusion of adenine bases. *EMBO J* 1994, **13**, 993–1001.
231. Wellinger RJ, Zakian VA. Lack of positional requirements for autonomously replicating sequence elements on artificial yeast chromosomes. *Proc Natl Acad Sci USA* 1989, **86**, 973–977.
232. Weinert TA, Hartwell LH. The *RAD9* gene controls the cell cycle response to DNA damage in *Saccharomyces cerevisiae*. *Science* 1988, **241**, 317–322.
233. Wellinger RJ, Wolf AJ, Zakian VA. Structural and temporal analysis of telomere replication in yeast. *Cold Spring Harbor Symp Quant Biol* 1993, **58**, 725–731.
234. Ashley T, Ward DC. A "hot spot" of recombination coincides with an interstitial telomeric sequence in the Armenian hamster. *Cytogenet Cell Genet* 1993, **62**, 169–171.
235. Katinka MD, Bourgain FM. Interstitial telomeres are hotspots for illegitimate recombination with DNA molecules injected into the macronucleus of *Paramecium primaurelia*. *EMBO J* 1992, **11**, 725–732.
236. Hastie ND, Allshire RC. Human telomeres: fusion and interstitial sites. *Trends Genet* 1989, **5**, 326–331.
237. Gilson E, Mueller T, Sogo J, Laroche T, Gasser S. RAP1 stimulates single- to double-stranded association of yeast telomeric DNA: implications for telomere-telomere interactions. *Nucleic Acids Res* 1994, **22**, 5310–5320.
238. Bouffler S, Silver A, Cox R. The role of DNA repeats and associated secondary structures in genomic instability and neoplasia. *BioEssays* 1993, **15**, 409–412.
239. Budarf ML, Blackburn EH. S1 nuclease sensitivity of a double-stranded telomeric DNA sequence. *Nucleic Acids Res* 1987, **15**, 6273–6292.
240. Huertas D, Lipps H, Azorin F. Characterization of the structural conformation adopted by (TTAGGG)n telomeric DNA repeats of different length in closed circular DNA. *J Biomol Struct Dyn* 1994, **12**, 79–90.
241. Gravholt CH, Friedrich U, Caprani M, Jorgensen AL. Breakpoints in Robertsonian translocations are localized to satellite III DNA by fluorescence *in situ* hybridization. *Genomics* 1992, **14**, 924–930.
242. Knapp S. Guanine-rich (GGNNGG) elements at chromosomal breakpoints interact with a loop-forming, single-stranded DNA-binding protein. *Oncogene* 1994, **9**, 1501–1505.
243. Zahler AM, Williamson JR, Cech TR, Prescott DM. Inhibition of telomerase by G-quartet DNA structures. *Nature* 1991, **350**, 718–720.
244. Shippen DE, Blackburn EH, Price CM. DNA bound by the *Oxytricha* telomere protein is accessible to telomerase and other DNA polymerases. *Proc Natl Acad Sci USA* 1994, **91**, 405–409.

Acknowledgements—We thank various colleagues for their input in the ideas and concepts described here and R. Lemire for typing. Work in the authors' laboratories was supported by grants from the Canadian MRC (MT-12616) and the NCI (007103) to R.J.W., and a grant from NSERC (OGP0105785) to D.S. R.J.W. is a chercheur-boursier Junior II of the FRSQ and D.S. is a research scholar of the BC Health Research Foundation.