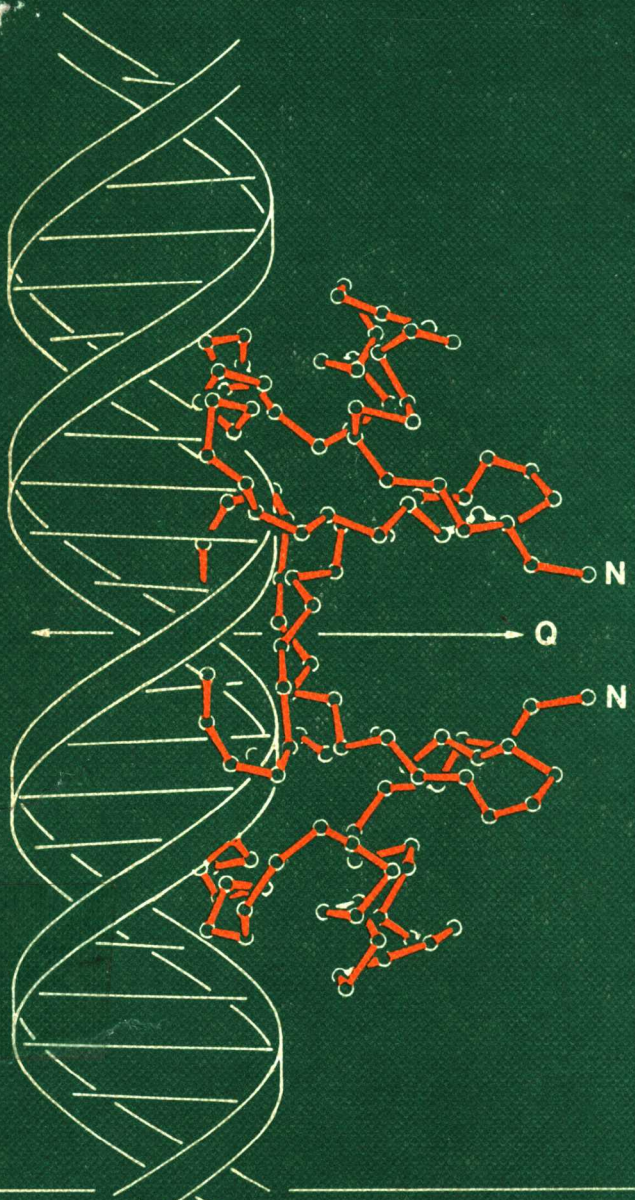
Nucleic Acids and Molecular Biology

5

Edited by
Fritz Eckstein
and
David M.J. Lilley



Springer-Verlag

Nucleic Acids and Molecular Biology

Volume 5

Edited by
Fritz Eckstein · David M.J. Lilley

With 70 Figures
Indices Volumes 1 – 5

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo
Hong Kong Barcelona
Budapest

Prof. Dr. FRITZ ECKSTEIN
Max-Planck-Institut
für Experimentelle Medizin
Abteilung Chemie
Hermann-Rein-Straße 3
W-3400 Göttingen, FRG

Prof. Dr. DAVID M. J. LILLEY
University of Dundee
Biochemistry Department
GB-Dundee DD1 4HN

Cover illustration by kind permission of B. W. Matthews

ISBN 3-540-53121-1 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-53121-1 Springer-Verlag New York Berlin Heidelberg

The Library of Congress has catalogued this serial publication as follows:
Library of Congress Catalog Card Number 90-64/8/7

Printed in Germany

Typesetting: International Typesetters Inc., Makati, Philippines
31/3145-543210 – Printed on acid-free paper

Preface to the Series

From its inception, molecular biology has always been a discipline of rapid development. Despite this, we are presently experiencing a period of unprecedented proliferation of information in nucleic acid studies and molecular biology. These areas are intimately interwoven, so that each influences the other to their mutual benefit. This rapid growth in information leads to ever-increasing specialization, so that it becomes increasingly difficult for a scientist to keep abreast of developments in all the various aspects of the field, although an upto-date knowledge of the field as a whole is highly desirable.

With this background in mind, we have conceived the present series *Nucleic Acids and Molecular Biology*. It comprises focused review articles by active researchers, who report on the newest developments in their areas of particular interest. The reviews are not intended to be exhaustive, but rather to place the most recent data into context. This format will allow our colleagues of familiarize themselves with new developments in areas outside their own immediate speciality, thus facilitating a more global view of their own work. Moreover, we hope sincerely that this will convey some of the excitement of the interdisciplinary nature of the study of nucleic acids and molecular biology.

This series is planned to appear annually. This period will allow us to return to important topics with sufficient frequency to cover new developments as they emerge.

FRITZ ECKSTEIN
DAVID M. J. LILLEY

Introduction to Volume 5

This volume begins with a review of the structures adopted by oligoguanine sequences commonly found in the telomeres of eukaryotic chromosomes. These sequences have been shown to adopt a variety of novel structures, depending upon the nature of cations present, resulting from the ability of guanine bases to form tetrad structures. Some of these structures are based on a parallel association of DNA strands, and the following chapter discusses parallel-stranded duplex DNA. Scanning tunnelling microscopy potentially offers a number of advantages for the study of DNA structure, including high resolution, and some examples of what has been achieved so far are presented in the third chapter.

We then turn to the related topics of repair and replication of DNA, beginning with reviews of repair of DNA damage caused by UV light, and a review of the relatively poorly understood area of base mismatch repair in eukaryotes. This is followed by a discussion of recent advances in the understanding of the initiation of replication at bacterial origins, and in-depth studies of the mechanism of DNA polymerase I.

This naturally leads us into the exciting and rapidly moving area of DNA protein interactions, which has been a major subject of interest throughout this series. The helix-turn-helix was the first protein motif for sequence-specific DNA recognition to be described (and is displayed on our cover as a result); now very many examples have been described from the original bacterial repressors to eukaryotic DNA-binding proteins important in development. From this basis of information, some general rules for sequence recognition are described. The SPKK motif is another protein element involved in DNA binding, found in histones and RNA polymerase. It becomes more and more difficult to find general rules for proteins that interact with DNA. Restriction enzymes continue to provide examples of very high sequence specificity, while other nucleases may exhibit much lower sequence specificity. The mechanism of *ECo*RI is described, an enzyme that binds predominantly in the major groove, followed by the structure of DNaseI which binds in the minor groove.

Enzymes acting on DNA may have consequences that were quite unforeseen a short time ago. Jim Wang and Leroy Liu turned our ideas about transcription and DNA topology around when

they suggested that the complex of RNA polymerase and associated transcript may be unable to rotate about the DNA. The consequence of this is the generation of domains of positive supercoiling ahead of the translocating polymerase, and negative supercoiling behind it. The ramifications of the twin supercoiled-domain model are potentially enormous.

This brings us finally to examples of the control of gene expression. HIV provides some very interesting examples of gene control in a retrovirus of particular concern for our species, and incidentally offers new and significant examples of RNA-protein interaction. Finally, for those interested in the study of gene expression in eukaryotes, transgenic animals are becoming an important tool.

We gratefully thank the authors of this Volume for the continued high standard of reviews. After five years we believe that this series has found its identity in view of molecular biology from the perspective of nucleic acid structure and function. We hope and expect to maintain the standard established in the next half decade.

Spring 1991

FRITZ ECKSTEIN
DAVID M. J. LILLEY

Contents

The Structures of Telomeric DNA

W.I. SUNDQUIST

1 Introduction

Linear eukaryotic chromosomes terminate with specialized sequences which are essential for complete DNA replication. These telomeric sequences (and their associated proteins) also appear to function in chromosome stability, organization, and association (for general reviews see Blackburn and Szostak 1984; Zakian 1989; Zakian et al. 1990; Blackburn 1990a; Greider 1990). All known eukaryotic telomeres consist of simple repeating sequences in which one DNA stand of the double helix contains short tracts of guanines alternating with short tracts of A/T-rich sequences (Table 1 for representative telomeric sequences). In every case, the guanine-rich strand is oriented 5′ to 3′ toward the end of the chromosome. This strand extends to produce a 3′ single-stranded overhang of about two repeating units in species where the exact telomeric terminus is known (Klobutcher et al. 1981; Henderson and Blackburn 1989). Although telomeric sequences from different organisms are not identical, the remarkable conservation of the G-tract sequence motif suggests that the structures and functions of telomeres are strongly conserved across the eukaryotes.

Much of what is known about telomeres and telomere-related processes has emerged from studies on lower eukaryotes, especially the ciliated protozoa. Many ciliates are highly enriched in telomeres and their associated proteins, owing to the developmentally regulated fragmentation and amplification of their chromosomes into multiple short linear DNA fragments which have telomeres at both ends (Gall 1986). In extreme examples, these DNA fragments may be gene-sized and may be present in

Table 1. Representative telomeric DNA sequences[a]

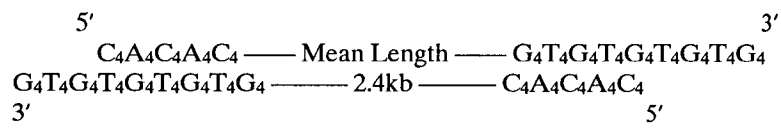
Organism	Telomeric repeat	Reference
Tetrahymena thermophila	TTGGGG	Blackburn and Gall (1978)
Oxytricha nova	TTTTGGGG	Klobutcher et al. (1981)
Arabidopsis thaliana	TTTAGGG	Richards and Ausubel (1988)
Homo sapiens	TTAGGG	Moyzis et al. (1988)
Dictyostelium discoideum	AG_{1-8}	Emery and Weiner (1981)
Saccharomyces cerevisiae	TG_{1-3}	Shampay et al. (1984)
		Walmsley et al. (1984)

[a] Sequences of a single telomeric repeat are given 5′ to 3′ toward the end of the chromosome. For more complete lists, see Zakian (1989) and Blackburn (1990a).

Structural Studies Division, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Nucleic Acids and Molecular Biology, Vol. 5
ed. by F. Eckstein and D.M.J. Lilley

thousands of copies per cell, as is the case for macronuclear DNA from the hypotrichous ciliate *Oxytricha* (shown schematically).



It has long been recognized that novel mechanisms must exist for complete replication of the ends of linear DNA since Okazaki fragment-primed DNA replication should result in incomplete replication of 3′ terminal template sequences upon removal of the RNA primer (Watson 1972). The eukaryotic solution to this fundamental problem was largely revealed with the discovery, in *Tetrahymena*, of telomerase, an enzyme which adds G-rich telomeric repeats to the 3′ ends of chromosomes in a template-independent fashion (Greider and Blackburn 1985, 1987). Telomerase activity has subsequently been detected in other lower eukaryotes (Zahler and Prescott 1988; Shippen-Lentz and Blackburn 1989) and also in human HeLa cell extracts (Morin 1989). The telomerases studied to date have an essential RNA component which, where known, contains sequences which are complementary to the telomeric repeat (Greider and Blackburn 1989; Shippen-Lentz and Blackburn 1990). This RNA sequence has been shown to serve, in vivo, as the template for the addition of *Tetrahymena* telomeric DNA repeats (Yu et al. 1990).

Telomeres in the macronucleus of *Oxytricha* also interact with a heterodimeric protein which binds specifically at the 3′ terminal telomeric overhang (Lipps et al. 1982; Gottschling and Zakian 1986; Price and Cech 1989; Raghuraman and Cech 1989; Hicke et al. 1990). Although the generality and precise functions of this protein have not yet been determined, possible roles include regulation of the extension of telomeric DNA by telomerase (Yu et al. 1990), protection of the 3′ telomeric overhang (Gottschling and Zakian 1986), and/or promotion of telomere-telomere associations (Lipps et al. 1982; Raghuraman and Cech 1989).

A number of other telomere-related processes have been described but are not yet understood in detail. Yeast telomeres can undergo unusual recombinations which may be involved in telomere generation and/or the maintenance of telomere length (Pluta and Zakian 1989; Wang and Zakian 1990). The synthesis of an RNA strand which is complementary to an exogenous, single-stranded, guanine-rich telomeric DNA strand has also been observed in crude extracts of *Oxytricha* macronuclei (Zahler and Prescott 1989). Such telomere-specific RNA molecules could serve to prime the synthesis of the cytosine-rich DNA strand of macronuclear DNA fragments. Telomeres frequently appear to be specifically localized in the nucleus and to mediate interchromosomal associations during meiosis and mitosis (reviewed in Blackburn and Szostak 1984). These observations may reflect mechanisms for chromosomal organization, recombination, and/or aberrant telomere-telomere fusion events (although some of these activities may alternatively involve subtelomeric chromosomal regions). Finally, several investigations indicate that telomeres may play crucial roles in chromosomal stability (Müller 1938; McClintock 1941), cell division (Yu et al. 1990), and senescence (Lundblad and Szostak 1989; Harley et al. 1990), although the absolute requirement for telomeres in these processes is uncertain (Levis 1989; Biessmann et al. 1990).

This review focuses on the recently discovered ability of telomeric sequences to form unusual DNA structures in vitro. Although telomeric and subtelomeric sequences can extend into chromosomes for thousands of base pairs, the DNA structures considered here are formed by sequences at the tips of telomeres, especially the single-stranded, guanine-rich telomeric overhangs. A variety of different guanine-guanine interactions allow these single-stranded telomeric overhangs to form monomeric, dimeric and tetrameric structures, which are reviewed here in detail. The structures which may be adopted by double stranded telomeric repeats are also discussed.

Although various unusual telomeric DNA structures have been well characterized in vitro, their biological relevance remains uncertain. For example, some of these structures are not recognized by purified cellular components which interact directly with telomeres, indicating that unusual DNA structures are not utilized in these specific interactions (*vide infra*). Nevertheless, the possibility remains that non-Watson-Cricktelomeric structures may yet find functions in telomere-related processes, especially in mediating the various interchromosomal associations of telomeres. The structures observed in vitro also appear to have the potential to assist in the telomeric functions of chromosomal replication, protection, organization and association in the absence of protein cofactors. This raises the possibility that such nucleic acid structures may have performed telomeric functions prior to the development of specialized proteins.

2 Guanine Self-Association

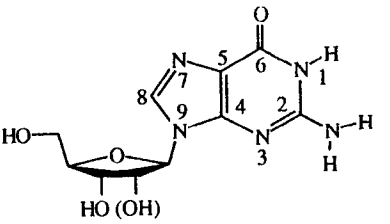
As noted above, many of the unusual structures formed by guanine-rich telomeric DNA sequences involve specific guanine-guanine interactions. Such interactions may also be emerging as a recurring theme in RNA structure and recognition (Kim et al. 1974; Robertus et al. 1974; Michel et al. 1989; Blackburn 1990b; Sundquist, unpubl.). The different possible guanine-guanine interactions, many of which were first elucidated in studies of guanine derivatives and their homopolymers, are reviewed in this section.

Guanosine is the only nucleotide that can form self-ordered structures in solution, a phenomenon that was first described by Bang in 1910:

> One boils a solution of α-guanylic acid supplemented with 2–3% KOH for 30 minutes on a waterbath, adds acetic acid until the solution is acid, and filters while the solution is still warm; then one sees the filtrate form a jelly on cooling down, provided the solution is sufficiently concentrated. (Bang 1910, translated from the German by E. Kaufmann)

This propensity of guanosine (and its derivatives) to self-associate reflects several of the unique properties of the guanine base. In particular, guanine can participate in strong stacking interactions and has multiple hydrogen bonding and metal binding sites which provide the potential for a variety of different interactions.

As shown in Fig. 1, four different G-G base pairs may be formed by the major tautomeric form of guanine at neutral pH (Donohue 1956). Base pair G:G 1 has been observed crystallographically in the structure of RNA^{phe} (Kim etal. 1974; Robertus et al. 1974), where such a base pair is formed between m^7G and G^{22}. The same pair has been observed crystallographically in the structures of guanine, guanosine and cyclic



2'-Deoxyguanosine (guanosine)

diguanylic acid (Bugg et al. 1968, Egli et al. 1990). Chemical probing experiments indicate that G:G 1 base pairs also form in triple helical DNA structures of the purine-purine-pyrimidine type (Kohwi and Kohwi-Shigematsu 1988). Guanine base pairs of the second type (G:G 2) have also been observed crystallographically in the intermolecular packing of B-DNA (Wing et al. 1980), cyclic diguanylic acid (Egli et al. 1990) and derivatives of m^7G (Ishida et al. 1988).

Guanine base pairs of the third and fourth types (G:G 3 and G:G 4) have not yet been observed crystallographically, although it seems likely that these pairs will form under the appropriate conditions. For example, mutagenesis and binding studies indicate that the recognition of guanosine in self-splicing group I introns probably occurs through the formation of a G:G 3 base pair (Michel et al. 1989). Guanine pairs of the fourth type have not yet been shown experimentally, but theoretical calculations indicate that these pairs may actually have a more stable hydrogen bonding configuration than any of the other guanine pairs (Hobza and Sandorfy 1987 and references therein).

In addition to forming G-G base pairs, guanine also has the potential to form cyclic base tetrads (Fig. 2). Guanine base tetrad formation was first proposed by Davies

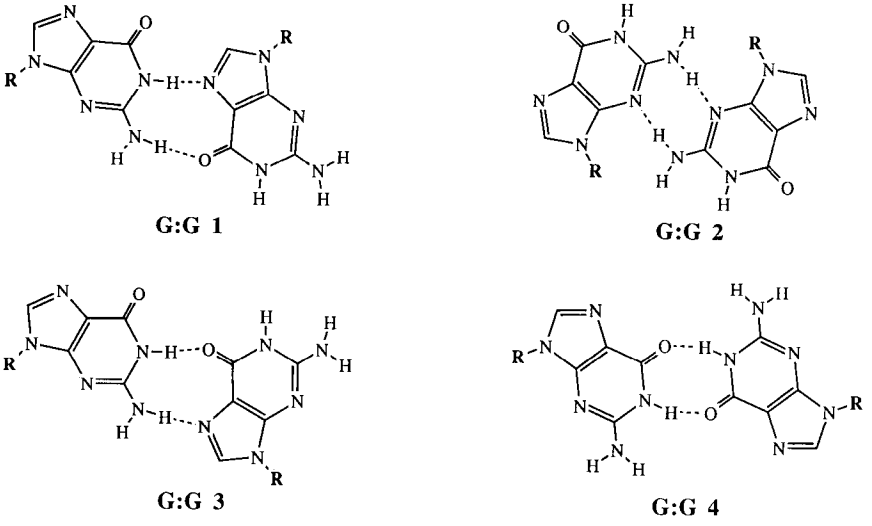


Fig. 1. Structures of the four possible G:G base pairs

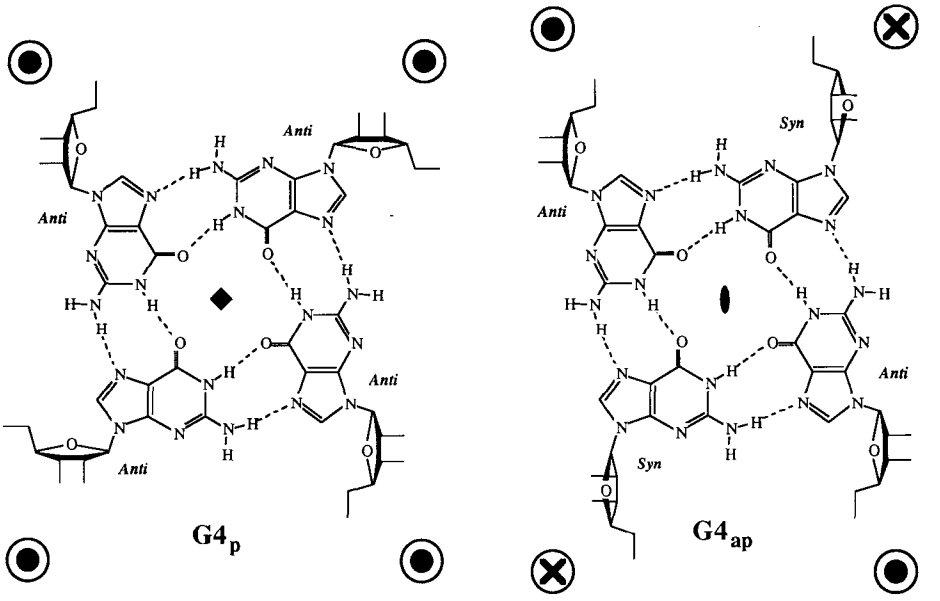


Fig. 2. Structures of guanine base tetrads showing the cyclic hydrogen bonding scheme and two possible orientations of the backbone chains. Guanines in tetrad $G4_p$ are shown in the *anti* conformation with parallel sugar-phosphate chains projecting 5′ to 3′ out of (⊙) the plane of the paper (an all *syn* parallel tetrad is also possible). Guanines in tetrad $G4_{ap}$ are shown with alternating *syn* and *anti* conformations; the antiparallel sugar-phosphate chains therefore have alternating orientations 5′ to 3′ out of (⊙) or into (⊗) the plane of the paper. Other antiparallel tetrads are also possible (see Appendix)

and coworkers to explain the self-ordered helical structures formed by 3′ GMP (Gellert et al. 1962). Infinite stacking of such 3′ GMP tetrads was shown to generate a quadruple helical structure which was consistent with fibre diffraction data. Subsequent investigations have revealed that most other guanosine derivatives and polymers can associate into similar quadruple helices (Tougard et al. 1973; Arnott et al. 1974; Zimmerman et al. 1975; Zimmerman 1976; Mariani et al. 1989), although a related one-start helical structure is also possible for 5′ GMP (Sasisekharan et al. 1975).

Two very similar quadruple helical models were proposed independently for the structure of poly(G) on the basis of fibre diffraction and model building studies (Arnott et al. 1974; Zimmerman et al. 1975). One of these structural models is shown in Fig. 3. In both models, poly(G) forms a right-handed quadruple helix in which each helical step consists of a guanine base tetrad. The tetrad steps are stacked at a distance of 3.4 Å with a rotation per tetrad of 31° to yield a helix with a repeat of 11.5 bases and a pitch of 39.2°. The four poly(G) strands in this structure have the same (parallel) orientation and all guanines adopt an *anti* conformation, resulting in tetrads of the first type ($G4_p$) as shown in Fig. 2.

The self-association of guanine derivatives and polymers has also been studied in solution using a number of different techniques (reviewed in Chantot and Guschlbauer 1972, Saenger 1984). These studies have generally supported the tetrad model and have further revealed that size-specific, monovalent cation binding is integrally in-

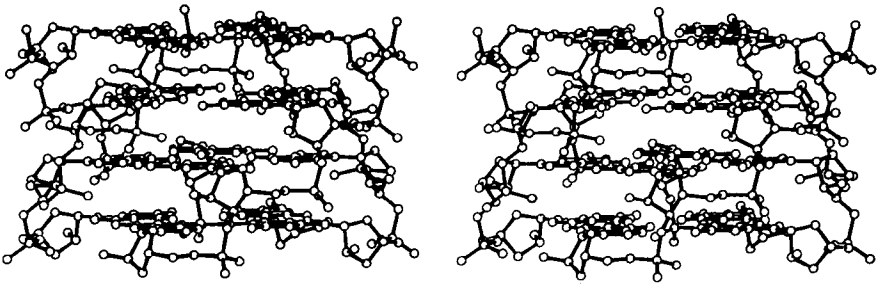


Fig. 3. Stereo drawing of the structure of quadruple helical poly(rG) derived from fibre diffraction and model building studies (Zimmerman et al. 1975). The parallel quadruplex incorporates the guanine tetrad $G4_p$ (shown in Fig. 2). The helix is oriented 5′ to 3′ from *top* to *bottom* and the axis is tipped by 5° towards the viewer so that the tetrads are seen nearly on edge. *Antiparallel* quadruple helices may be also be constructed from guanine tetrads like $G4_{ap}$, with the restriction that guanines which pair antiparallel strands must have different glycosidic conformations (e.g., *syn/anti* or *anti/syn*)

volved in the association of guanine base tetrads into quadruple helical structures. The specific cation binding site(s) of guanine quadruple helices appear to be of optimal size for potassium binding (the cation originally employed by Bang) since the stability of quadruplexes decreases sharply as the size of the monovalent cation is decreased (K > Na > Li) or increased (K ≥ Rb > Cs; Pinnavaia et al. 1978; Miles and Frazier 1978; Howard and Miles 1982).

Miles and coworkers have postulated that the cation-specific stability of quadruple helical structures arises because potassium can bind in chelation "cages" between adjacent tetrads in the axial channel of the quadruplex (Fig. 4). In this model, the eight O6 ligands which define the chelation cage are donated by guanines in the tetrads above and below the metal binding site. These eight-coordinate axial cages have an ideal size and ligand stoichiometry for potassium chelation (a calculated from the fibre diffraction model of poly(G), Zimmerman et al. 1975). Other monovalent cations, such as sodium, would be expected to fit less well into a chelation cage of this size and hence stabilize these structures to a lower degree (Pinnavaia et al. 1978; Sundquist and Klug 1989). Analogous size-selective cation binding has been well documented for various crown ether (and related) ionophoric ligand systems where the binding affinities of different monovalent cations can differ by many orders of magnitude, depending upon size complementarity between the cation and the ligand cage (Pedersen 1988 and references therein).

There are two different proposals for the binding site of sodium in quadruple helical structures. Since the axial cages between guanine tetrads appear to be too large for the optimal binding of sodium, it has been proposed that sodium instead binds within the tetrad planes to give a heme-like structure (Howard and Miles 1982; Williamson et al. 1989). Although the NaO bond lengths calculated from such a model are reasonable (~2.3 Å), sodium binding in the tetrad plane would require four coordinate, square planar coordination which is not typically observed in sodium model complexes. In fact, eight coordinate ligand cages which preferentially bind potassium usually also bind sodium in the *same site*, albeit with reduced affinities and, in some cases, with a reduction in coordination number (see Gandour et al. 1986). There are strong structural

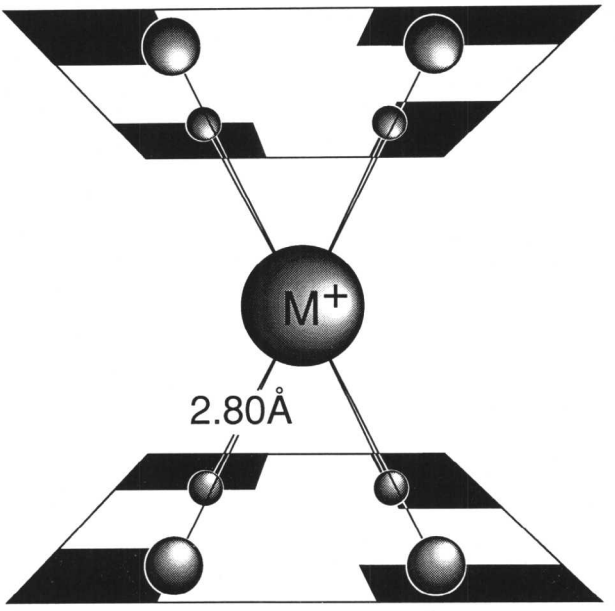


Fig. 4. Schematic illustration (in perspective view) showing how an axial metal binding site is created by stacking guanine tetrads into a quadruple helix. Each guanine surrounding the axial site (*shaded*) contributes one O6 ligand to form an eight-coordinate chelation cage. For clarity, the stacked quartets are depicted without helical twist. The average M-O bond distance, which is independent of helical twist, is derived from the fibre diffraction model for the quadruple helical structure of poly(rG) (Zimmerman et al. 1975)

precedents in model complexes for the idea that sodium can bind and "rattle around" in a potassium-specific cage of fixed size (see Cram et al. 1986). Sundquist and Klug have therefore proposed that both sodium and potassium bind in the axial cages between tetrads in quadruple helical structures, but that potassium binds with greater affinity and hence stabilizes these structures; to a greater extent (special sodium binding sites do seem possible at the ends of quadruple helices, however, where water molecules could serve as additional ligands; (see Bouhoutsos-Brown et al. 1982).

In general, guanine quadruple helical structures are remarkably stable in spite of the steric and electrostatic repulsions which might be expected in a quadruple helical nucleic acid structure. For example, the poly(G) quadruple helix remains intact at 100 °C, even at quite low ionic strength (Fresco and Massouli 1963). This intrinsic stability must arise from the highly organized structure of the quadruplex, which allows a number of energetically favourable interactions. Specifically, guanine quadruple helical structures are stabilized by four interstrand hydrogen bonds to each guanine, multiple metal ion coordination sites, and (presumably) strong stacking interactions between each of the tetrads. This high degree of organization is apparently reflected in the slow kinetics of poly(G) association (Howard et al. 1977) and also in the highly cooperative self assembly of 5′ GMP/Na^+ which is strongly enthalpically driven (($\Delta H = -17 \pm 2$ kcal mol^{-1} and $\Delta S = -0.051 \pm 0.006$ kca mol^{-1} $logK^{-1}$; Borzo et al. 1980). In contrast, the enthalpy of self-association for other purines is typically more positive than –7 kcal mol^{-1} (Saenger 1984).

3 Evidence for Unusual Telomeric DNA Structures

Given the structural polymorphism possible for guanine, it is perhaps not surprising that guanine-rich telomeric DNA sequences can adopt a variety of different structures in vitro. The first indication that telomeric DNA could adopt such unusual structures arose from the observation that macronuclear DNA (not devoid of protein) isolated from hypotrichous ciliates could associate into higher order structures in vitro (Lipps 1980; Lipps et al. 1982). Electron microscopic and biochemical studies suggested that this association involved the pairing of the telomeric ends of macronuclear DNA. The resulting complexes appeared as "cusps" in which the two double helical DNA fragments projected from the same side of the cohered telomeres. Telomeric complexes isolated from the ciliate *Oxytricha nova* could be dissociated by high temperature, high pH, low ionic strength, or protease treatment, but were remarkably resistant to nuclease treatment (Lipps et al. 1982).

A subsequent study revealed that, although macronuclear telomeric DNA is stoichiometrically bound to a protein in vivo, purified macronuclear DNA (shown schematically in the introduction) can slowly associate in vitro under high salt conditions *in the absence of any protein* (Oka and Thomas 1987). Nuclease digestion and oligonucleotide competition studies were used to demonstrate that this association is telomere-specific and that the 3′ telomeric overhang participates in complex formation. Telomeric complexes were highly resistant to exonuclease digestion, but were cleaved specifically by T7 gene 3 endonuclease and endonuclease S1, indicating some single-stranded character near (or within) the telomeric structure. Preformed telomeric complexes showed a striking increase in thermal stability of more than 25 °C in the presence of potassium (versus sodium). As previously noted, such a dramatic alkali metal-specific stability is not typically observed for most helical DNA structures, but had been observed for the quadruple helical structures formed by guanine derivatives and their polymers.

A second type of telomeric DNA complex was suggested by the observation (Sen and Gilbert 1988) that single- stranded, guanine-rich oligonucleotides could tetramerize to form parallel quadruple helical structures, in analogy to the structures proposed earlier for poly (G); (cf Fig. 3). The tetrameric stoichiometry of these oligonucleotide complexes was demonstrated by annealing complementary strands to the complex and observing the stepwise formation of four new complexes of reduced electrophoretic mobility (containing one, two, three, and four complementary stands, respectively). All guanines within the putative quadruple helical regions of these complexes were protected from N7 methylation by dimethylsulphate (DMS), a result which is consistent with each guanine N7 atom acting as a hydrogen bond acceptor within a guanine base tetrad (Fig. 2).

The tetramerization of these guanine-rich, but *heteropolymeric*, oligonucleotides demonstrated for the first time that discontinuities could be stably incorporated as bulges or mismatches within quadruple helical structures. Although the study of Sen and Gilbert (1988) was carried out using non-telomeric oligonucleotides (encoding sequences found at immunoglobulin switch sites), their results suggested that the guanine-rich overhangs of telomeric DNA might also form similar quadruple helices. The *tetrameric* structures observed in this study did not, however, readily explain the *pairwise* telomeric associations seen for macronuclear DNA from the hypotrichous ciliates.

This distinction suggested that at least two different types of intermolecular telomeric structure must be possible.

Yet another type of telomeric DNA structure was observed for single-stranded guanine-rich oligonucleotides containing a minimum of two telomeric repeats. Under low ionic strength conditions, these oligonucleotides folded into *intramolecular* complexes that exhibited anomalously rapid migration through native polyacrylamide gels (Henderson et al. 1987). Several lines of evidence indicated that these complexes were stabilized by G-G base interactions. In particular, folded structures were observed for different telomeric sequences whose only apparent similarity was the potential to form G-G base pairs, e.g., $(T_2G_4)_2$ (from *Tetrahymena*) and $AG_7AGAG_6AG_6$ (from *Dictyostelium*). It was therefore suggested that these oligonucleotides formed hairpin-like structures stabilized by G-G base pairing.

Another important observation made in this study was that oligonucleotides containing four telomeric repeats could fold into especially stable and compact intramolecular structures. One such four-repeat oligonucleotide, $(T_2G_4)_4$, was studied in detail. Visible spectroscopic studies revealed that this oligonucleotide folded reversibly through a multistep pathway with a T_m of approximately 40 °C in low-salt (20 mM Na^+) buffer. ^{31}P and ^1H NMR spectroscopic studies also indicated a discrete, ordered structure at low temperature which unfolded into a random coil configuration at 40 °C. Nearly all (~23) of the amino protons in the folded structure were observable in the ^1H NMR spectrum (in H_2O), indicating that these protons are either stabilized by direct hydrogen bonding interactions in the structure or are "Trapped" in structures which are not readily accessible to the solvent. Furthermore, at least four (and possibly as many as six) of the guanines in this folded conformation of $(T_2G_4)_4$ were in the *syn* conformation, as indicated by NOE's between the guanine H8 protons and their sugar H1′ protons.

In summary, initial investigations clearly indicated that telomeric DNA sequences have the potential to fold into a number of different unusual structures in vitro. Different types of telomeric DNA structures included two- and four-repeat intramolecular foldback structures (observed in studies of telomeric oligonucleotides), intermolecular complexes between two telomeres (observed for the cohered ends of macronuclear DNA), and intermolecular tetrameric quadruple helical complexes (observed for guanine-rich oligonucleotides). Although these complexes were all different, each appeared to be stabilized by guanine-guanine interactions and it therefore seemed reasonable to expect that common structural principles would connect the different types of structures. This is indeed proving to be the case, and the following sections of this review summarize what is now known about each type of telomeric DNA structure.

4 Simple Foldback Structures

As described above, oligonucleotides containing two or more guanine-rich telomeric repeats can fold intramolecularly to form what appear to be hairpin-like structures stabilized predominantly by G-G base pairing (Henderson et al. 1987). The specific hydrogen bonding scheme (Fig. 1) utilized by these G-G pairs is not yet certain, however, nor is it clear that ^1H NMR spectral studies of the four-repeat sequence $(T_2G_4)_4$ are