STRUCTURES OF TRIPLET REPEAT DNAs ASSOCIATED WITH HUMAN DISEASES

By

ADONG YU

Bachelor of Science
Jilin University
Changchun, Jilin, P. R. China
1983

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1992

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 1996

STRUCTURES OF TRIPLET REPEAT DNAs
ASSOCIATED WITH HUMAN
DISEASES

Thesis Approved:

Michael Mytar
Thesis Adviser

Franklin R. Leach

Richard C Essenberg

Ulrich Melcher

Steven M Johnson

Thomas C. Collins
Dean of the Graduate College

PREFACE

Triplet repeat DNA expansions are considered to be significant due to their association with several human inherited disorders. They are characterized by accelerating the expansion in germline cells of infected individuals in an unique non-Mendelian genetics manner. The expansion of trinucleotide repeat contained within a specific gene is coincident with disease manifestation. Continued expansion of the repeat is observed in offspring of affected individuals, resulting in increased severity of the disease. At present, ten triplet repeat expansion diseases (TREDs) have been identified. These diseases include severe human mental retardation and eight other neurological deficiency disorders. TREDs occur widely in the human population. The trinucleotides that undergo expansion within the genes associated with TREDs are CGG, CTG, CCG, CAG, GAA, and TTC respectively. Although the mechanisms whereby expansion of trinucleotide repeats results in disease manifestation are to be elucidated, a preliminary conclusion can be drawn that the presence of a long trinucleotide repeat may lead to formation of particular DNA structure when DNA replicaties. These particular DNA structures may be essential in causing the triplet repeat to undergo further expansion. In order to obtain insight into the mechanisms of triplet repeat expansion, the studies

through this dissertation were performed to investigate the structures of various triplet repeat DNAs.

Chemical methods and the enzymological method were developed to investigate the structures of various triplet repeat sequences. The research results lead to the conclusion that ss(CTG)15, ss(GTC)15, ss(CAG)15, ss(GAC)15 and ss(CGG)15 formed hairpin structures in *(a)* alignments that differed in thermal stabilities. The hairpin structure of ss(CCG)15 exhibited unusual features for a nucleic acid structure; the mispaired cytosines were protonated at physiological pH, and the sugar-phosphate backbone was highly distorted. The six triplet repeat DNA structures associated with TREDs are identified. These research results lead to four research articles published in the top peer review Journals. Another two papers have been submitted. Our initial research results were cited in "The NIH Research Journal", making our laboratory and our achievements internationally recognized.

As I complete this dissertation, I wish to express my deepest appreciation to my research adviser Dr. Michael Mitas for his intelligent guidance, constructive advice, invaluable assistance and consistant encouragement through the years. Many thanks go to my other committee members, Dr. Franklin Leach, Dr. Ulrich Melcher, Dr. Moses Vijayakumar, Dr. Steven Granham and Dr. Richard Essenberg who have played essential roles in my training and dissertation writing.

I would like to thank my lab colleague, Jeff Dill for his support, friendship, and consideration.

I am grateful to Dr. Ulrich Melcher for his reading and correcting my proposal for the qualification exam. I also sincerely thank the faculty of the Department of Biochemistry and Molecular Biology for their teaching, and

friendship. I also appreciate Dr. James Blair and the Department of Biochemistry and Molecular Biology for financial support.

Finally I wish to express my greatest appreciation to my wife Ying Fan, my daughters Disa Yu and Kathleen Yu for their love, support, understanding and sacrifice. My deepest appreciation is extended to my parents, Kede Yu and Ruiting Sun, and my sisters, for their constant understanding, support and love. I wish this dissertation will be a piece of consolation to them.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

NOMENCLATURE

A:	adenine
C:	cytosine
CD:	circular dichroism
CT:	cerebellar tomography
DEPC:	diethyl pyrocarbonate
DMS:	dimethyl sulfate
ds:	double-stranded
*Eco*SSB:	*Escherichia coli* single-stranded DNA binding protein
EMMP:	electrophoretic mobility melting profile
G:	guanine
HA:	hydroxylamine
M_{rel}:	relative electrophoretic mobility
MRI:	magnetic resonance imaging
MTase:	methyltransferase
nt:	nucleotide(s)
P1:	P1 nuclease
PCI:	phenol chloroform isoamyl alcohol
S1:	S1 nuclease
ss:	single-stranded
STRs:	simple tandem repetitive sequences
T:	thymine
THF-OOH:	2-hydroperoxytetrahydrofuran
T_i:	isomobility temperature
T_m:	melting temperature
TREDs:	triplet repeat expansion diseases
TRS:	tandem repeated DNA sequences

CHAPTER I

INTRODUCTION

The wide distribution of tandem repeat DNA sequences in the human genome

Tandemly repeated DNA sequences (TRS) are interspersed in both transcribed and nontranscribed regions of human chromosomes (Moyzis et al., 1989, Grady et al., 1992, Oberie et al., 1991). Of those TRS localized at nontranscribed regions, some have been characterized. For example, those in the centromere compose more than 10% of human DNA (Moyzis et al., 1989). These DNA arrays are known to consist of various copy numbers of α satellite (Willard & Waye, 1987), β satellite (Waye & Willard, 1989), and the three classic satellites I, II, and III (Prosser et al., 1986). The human telomere sequence $(TTAGGG)_n$ is the one that has been most extensively investigated. It was identified and cloned by screening for evolutionarily conserved repetitive DNA sequences (Moyzis et al., 1988). Studies on the human telomere have indicated its extreme conservation through vertebrates (Meyne et al., 1989). It undergoes occasional amplification, often at chromosome fusion points (Meyne et al., 1990) and it possesses an ability to form unusual DNA structures known as G quartets (Sen and Gilbert, 1988, Sundquist and Klug, 1989, Williamson et al., 1989). A G quartet is composed of four guanine residudes that lie in the same plane and are joined together by a total of eight hydrogen bonds. Each guanine in a G quartet

forms hydrogen bonds through O6′ and N7 with one of the neighbor guanines and forms the other two hydrogen bonds at its N1 and N2′ positions with another neighbor guanine residue. Both X-ray crystallographic and NMR solution studies have defined the conformation G quartets and demonstrated that they are responsible for the stability of the G-quadruplex structures formed by the alignment of four separate strands. These strands can be arranged as dimerization of a pair of hairpins, or through intramolecular folding of a single strand (Wang et al., 1993, Gupta et al., 1993, Kettani et al., 1995).

The tandemly repeated DNA sequences localized in transcribed regions vary widely in complexity, such as ribosomal RNA genes which are complicated repetitive sequences with a repetitive unit of a whole gene. In comparison, simple tandem repetitive sequences (STRs), usually referred to a subset of short repetitive sequence, are composed of less than six base pairs in repeating units, such as di- or tri- nucleotides. They are localized on both translated and 5′ or 3′ nontranslated regions of the genes. Due to their location within or near genes, these STRs were predicted to be involved in a number of functions, including gene regulation (Wang et al., 1979, Weintraub and Groudine, 1976, Hentschel, 1982), gene recombination (Yamazaki et al., 1992), signals for DNA binding proteins (Collick & Jeffrey, 1990, Richards et al, 1993) and signals for nucleosome assembly (Wang et al., 1994).

Many STRs are polymorphic in copy number in human genes (Sutherland and Richards, 1995, Weber, 1990). Therefore, this rich source of DNA polymorphism has been widely exploited for studies of the human genome. The frequency of events leading to repeat length alteration can be as high as 10^{-4}-10^{-2}, values much higher than that of classic mutation (Dallas, 1992). In general, the degree of variability of such repeats is directly related to the length of the perfect repeat (Weber, 1990).

Members of one group of the simple tandem repeats, the trinucleotide repeats can undergo an increase in copy number by a process that has recently been referred to as "dynamic mutation "(Richards & Sutherland, 1992). Dynamic mutation is a continuous process of change in genetic material that occurs over several generations (Grant et al., 1995). It is distinguished from conventional mutational events by several properties: the products of dynamic mutations have the potential to undergo further mutations, and the probability of dynamic mutation of an STR is a function of the number of perfect repeating units.

Triplet repeat DNA expansion: a new mechanism that underlies many human inherited diseases

Triplet repeat expansion diseases (TREDs) are now regarded as a unique family among inherited genetic disorders. Genetic instability in the number of tandemly repeated triplets is a major characteristic of these disorders. Typically, individuals with one of these diseases contain greater than 50 trinucleotide repeats within a specific gene, while unaffected individuals contain fewer than 30 repeats. The number of triplet repeats in disease-associated genes tends to increase in germline transmission from parent to offspring. Usually, the larger the repeat number, the higher the probability of expansion and the earlier and more severe the disease (reviewed in Ashley et al., 1995, Sutherland et al., 1995, Kunkel, 1993, Mandel, 1994, Nelson & Warren, 1993). Due to the high frequency with which the trinucleotide repeats associated with these diseases undergo further expansion in offspring of affected individuals, the mutation(s) responsible for this phenomenon has been described as "dynamic" (Sutherland and Richards, 1995) This is a non-Mendelian pattern of inheritance, in contrast

to the conventional (stable) mutations that cause most other forms of characterized genetic disorders.

Individuals with a fragile site at a specific location on the X chromosome develop mental retardation. Hence, this disease was coined "fragile X syndrome". In 1991, three groups of researchers independently reported a large amplication of (CGG)n trinucleotide repeats as the cause of fragile X syndrome (Bell et al., 1991; Vincent et al 1991; Pierreti et al., 1991). Since then, at least nine other human inherited diseases have also been found associated with triplet repeat DNA expansion. Most of these diseases are neurological disorders caused by a d(CAG)n triplet expanded in the translated regions of their respective genes.

Of all possible triplet repeat sequences, only three have been associated with human disease, which are the CTG/CAG, CCG/CGG, and most recently GAA/TCC. Most trinucleotide repeats expanded in human genetic disorders occur in exons or 5′ and 3′ untranslated regions (UTRs). Expansion of a simple tandem repeat within an intron associated with a human genetic disorder was only found with GAA (Campuzano et al., 1996).

(CGG)n and (CTG)n triplet repeats are found in close proximity to numerous gene loci (Riggins et al., 1992; Li et al., 1993). The CCG repeats are responsible for one group of fragile sites on 5′- or 3′- *FMR1* gene of the X chromosomes and can expand to very high copy numbers, in excess of 1000 copies. Expansion of the d(CCG) repeat is accompanied by methylation of the *FMR1* promoter and of the amplified trinucleotide tract (Bell et al., 1991; Vincent et al 1991; Pierreti et al., 1991). Subsequent to d(CGG) expansion and hypermethylation, the *FMR1* gene becomes transcriptionally silent (Pierreti et al 1991; Hansen et al., 1992; Sutcliffe et al., 1992). The CAG repeats are involved in a number of neurological disorders. They can also expand to high copy

numbers when in the untranslated region of a gene (as in myotonic dystrophy) but are mostly in coding regions where their copy number is usually less than 100. Expansion of (CAG)n in the corresponding gene usually does not cause inactivity of the gene, but results in abberant translation products (Sutherland and Richards, 1995, Warren, 1996).

Characteristics of the Known Triplet Repeat Expansion Diseases

A brief summary of the characteristics of the known TREDs is given below. These include the triplet repeat sequence compositions, gene location, repeat copy numbers, proposed causes of the diseases and main clinical phenomena.

Fragile X Syndromes (type A and type E)

Fragile X Syndrome is an inherited, X chromosome-linked dominant mental retardation disorder affecting about 1 male in 1500 and 1 female in 2500 (Webb, 1989; Richards and Sutherland, 1992). In addition, 1 in 700 females is thought to be a carrier (Laxova, 1994). This syndrome is frequently associated with a folate-sensitive fragile site, Xq27.3, on chromosome X of cells of affected males (Sutherland, 1977). Fragile sites appear as nonstaining regions, chromatid gaps or as breaks in metaphase chromosome spreads. The fragile X syndrome is characterized by a substantial expansion of a d(CGG) trinucleotide repeat located in the 5′ untranslated region of a house-keeping gene, *FMR1* , termed type A fragile X syndrome and another 5' untranslated region of the *FMR1* gene, termed type E (Verkek *et al.*, 1991). Whereas normal individuals have 2 to 50 copies of the d(CGG) sequence, the trinucleotide is amplified in affected subjects to >200-2000 copies (Fu *et al.*, 1991; Kremer *et al.*, 1991; Oberle *et al.*, 1991;

Nakhahori *et al.*, 1991; Verkek *et al.*, 1991, Yu *et al.*, 1991). Expansion of the d(CGG) repeat is accompanied by hypermethylation of the *FMR1* promoter and of the amplified trinucleotide tract, and results in inactivation of the *FMR1* gene which normally encodes an RNA binding protein (Bell *et al.*, 1991, Vincent *et al.*, 1991; Pierreti *et al.*, 1991; Luo *et al.*, 1991). Fragile X syndromes are thought to be among the most common causes of mental retardation in the population.

Huntington's Disease (HD)

Huntington's disease is a dominant neurodegenerative disorder characterized by motor disturbance, loss of cognitive functions and psychiatric manifestations, and inexorable progression to death, typically 10-15 years after onset (Nasir, 1995, Destee, 1995). Huntington's disease is caused by expansion of a CAG repeat within a widely expressed gene localized to chromosome 2p. Expansion of the repeat leads to selective neuronal death. The gene encodes huntingtin, a protein of unknown function (Duyao, 1995). It has been found that the expanded repeat region of the protein specifically binds other proteins critical to glucose and oxygen metabolism in the brain (Burke et al., 1996).

A survey of Huntington's disease has shown a 1991 prevalence rate of 6.4/100,000. The prevalence rate is similar to the high prevalence rates of Huntington's disease found in most European populations (Morrison, 1995)

Myotonic Dystrophy

Myotonic dystrophy (Dystrophia myotonica or Steinert's disease) is the most common hereditary disease of the neuromuscular system in adults, with a global incidence of 1 per 7,500 - 8,000 (Harley et al, 1992, Aslanidis et al, 1992).

Its mode of inheritance is autosomal dominant. The genetic basis of myotonic dystrophy is an unstable expansion of CTG repeats found in the 3'-untranslated region of a gene on chromosome 19 that encodes a putative serine/threonine protein kinase (Bush et al, 1996).

Myotonic dystrophy is categorized by an adult and a congenital form. In the adult form, the characteristic findings are muscular atrophy in certain regions of the body (face, neck and distally in the extremities) and myotonia. Cataract, intraocular hypotension, gonadal atrophy, conduction abnormalities in the heart and hearing deficiencies appear quite often in the course of the disease (Kuhn, 1990, Shelbourne, 1993). There is a positive correlation between the CTG repeat size and clinical severity (Pizzuti, 1993).

Dentatorubral and Pallidoluysian Atrophy (DRPLA)

Dentatorubral and pallidoluysian atrophy (DRPLA) are associated with an expanded and unstable CAG repeat within a gene that encodes for atropin (Koide et al., 1994, Burke et al., 1994, Nagafuchi et al., 1994). The gene is localized on the short arm of chromosome 12 (Onodera et al., 1995). All affected individuals showed one expanded allele with the repeat number of CAG at the DRPLA locus, ranging from 58 to 82, and a normal allele, ranging from 10 to 21 (Schmitt et al., 1995).

The clinical characteristics of the disease vary. Patients with juvenile-onset were characterized by myoclonus, epilepsy and mental retardation, whereas cerebellar ataxia, choreoathetosis, and dementia were typical of adult- and senile-onset patients. The most severely affected patient, a case of maternal transmission and with the largest allele, became bedridden in a vegetative state by age 12. On the Cerebellar Tomography (CT) and Magnetic Resonance

Imaging (MRI), varying degrees of brain atrophy were present in all patients. T2-weighted MRI in patients with senile-onset showed symmetric high-signal lesions in the cerebral white matter, globus pallidus, thalamus, midbrain, and pons (Ikeuchi et al., 1995). Repeat size showed a close correlation with age of onset of symptoms and disease severity (Nagafuchi et al., 1994). Hardly anything is known about the pathological mechanisms causing neuronal cell death which leads to these symptoms (Uyama et al., 1995).

Spinocerebellar ataxia type I

Spinocerebellar ataxia type 1 (SCA1) is an autosomal dominant progressive neurodegenerative disorder characterized by ataxia, dysarthria, ophthalmoparesis, and variable degrees of amyotrophy and neuropathy (Chung, 1993). Symptoms usually develop in the third or fourth decade but anticipation has been noted in juvenile onset cases. Neuropathologic findings include severe neuronal loss in the cerebellum and brainstem as well as degeneration of spinocerebellar tracts (Zoghbi, 1995).

The SCA1 gene which maps to the short arm of human chromosome 6 was identified using a positional cloning approach (Ranum et al., 1994). The disease causing mutation is an expansion of a CAG trinucleotide repeat which lies within the coding region of a novel protein, ataxin-1, and encodes a polyglutamine tract. The number of CAG repeats varies from 6-39 repeats on normal alleles and 40-81 repeats on SCA1 alleles. The repeat has a perfect CAG configuration on expanded alleles whereas it is interrupted by 1-3 CAT units on normal alleles (Matilla et al., 1993; Chung et al., 1993).

Kennedy's disease (X-linked spinal and bulbar muscular atrophy)

Kennedy's disease (X-linked spinal and bulbar muscular atrophy) is an inherited form of motor neuron disease (Choi et al., 1993). This form of motor neuron disease principally affects the proximal limb girdle muscles as well as those involved with deglutition and phonation. Onset is usually late, in the fourth to fifth decades of life, and progression is slow. Moderate gynaecomastia and testicular atrophy are usually present, suggesting a defect in androgen receptor function. Being inherited in an X-linked recessive manner, only males are affected, with females as the unaffected carriers (Amato et al., 1993). The genetic abnormality that causes Kennedy's disease is an enlargement CAG repeats within the first exon of the androgen receptor gene, which is located on the proximal long arm of the X chromosome (Amato et al., 1993, Roling, 1993).

Machado-Joseph disease (MJD)

Machado-Joseph disease (MJD) is an autosomal dominantly inherited neuro-degenerative disorder primarily affecting the motor system. It can be divided into three phenotypes based on the variable combination of a range of clinical symptoms including pyramidal and extra-pyramidal features, cerebellar deficits, and distal muscle atrophy. MJD is thought to be caused by mutation of a single gene which has recently been mapped to a 29 cM region on chromosome 14q24.3-q32 (Twist, 1995). The mutation was identified to be an expanded CAG triplet repeat in the gene (Kawakami, 1995).

Large pedigrees have been found on the east and west coasts of the United States in which four main syndromes are described. Type I disease presents

with pyramidal and extrapyramidal findings usually in individuals in the second or third decades of life. Type II disease, which is the most common form of presentation, includes true cerebellar deficits associated with other motor features. Type III is late-onset in the fifth through the seventh decades of life presenting with pancerebellar deficits with motor and sensory polyneuropathy. A rare presentation is Type IV with parkinsonian features with mild cerebellar deficits and a distal motor sensory neuropathy or amyotrophy (Suite, 1986).

Synpolydactyly (SPD) (syndactyly type II)

Synpolydactyly (SPD) (syndactyly type II) is an autosomal dominant condition with typical abnormalities of the distal parts of both upper and lower limbs. Synpolydactyly is caused by insertion of a polyalanine stretch in the amino-terminal region of HOXD13 gene, corresponding to the insertion of $(CGG)_7(CAG)(CGG)_7$ trinucleotide repeat sequence on the gene (Muragaki et al, 1996). Due to the non-perfect multi-triplet feature of the expanded sequence, many others have not classified this disease as a TRED. The disorder is included here as TRED because it has many characteristics of TREDs except the non-perfect triplet repeat unit.

The typical characteristic clinical features in this disease is as follows (Akarsu et al., 1995, Muragaki, 1996): (1) short hands with wrinkled fatty skin and short feet; (2) complete soft tissue syndactyly involving all four limbs; (3) polydactyly of the preaxial, mesoaxial, and postaxial digits of the hands; (4) loss of the normal tubular shape of the carpal, metacarpal, and phalangeal bones, so as to give polygonal structures; (5) loss of the typical structure of the cuboid and all three cuneiform bones while the talus calcaneus and navicular bones remain intact; (6) large bony islands instead of metatarsals, most probably because of

cuboid-metatarsal and cuneiform-metatarsal fusions; and (7) severe middle phalangeal hypoplasia/aplasia as well as fusion of some phalangeal structures that are associated with the loss of normal phalangeal pattern.

Friedreich's Ataxia

Friedreich's Ataxia (Friedreich, 1863) is an autosomal recessive, degenerative disease that involves the central and peripheral nervous systems and the heart. Friedreich's Ataxia is caused by an intronic GAA triplet repeat expansion in a gene, X25, on chromosome 9q 13 (Campuzano et al., 1996).

Friedreich's ataxia is the most common hereditary ataxia, with an estimated prevalence of 1 in 50,000 and a deduced carrier frequency of 1 in 120 in European populations (Skre, 1975, Romeo et al., 1983). Friedrich's ataxia is an autosomal recessive degenerative disease characterized by a progressive gait and limb ataxia, a lack of tendon reflexes in the legs, loss of position sense, dysarthria, and pyramidal weakness of the legs (Geoffroy et al., 1976, Harding and Hewer, 1983). Hypertrophic cardiomyopathy is found in almost all patients (Harding and Hewer, 1983). Diabetes mellitus is seen in about 10% of patients, carbohydrate intolerance in an additional 20%, and a reduced insulin response to arginine stimulation in all patients (Finocchiaro et al., 1988). The age of onset is usually around puberty, and almost always before age 25. Most patients are confined to a wheelchair by their late 20s, and there is no treatment to slow the progression of the disease.

The molecular mechanisms that govern these TREDs manifestations, such as triplet repeat expansion, hypermethylation, gene transcription inactivity on some of them and those causing the disease phenomenon, are not known.

Though it is belived that trinucleotide repeats spur different diseases through different mechanisms, many predict that the repeats expand through a common mechanism.

Proposed expansion mechanisms

Two general mechanisms were proposed years ago to explain the instability of simple repetitive DNA: unequal recombination (Smith, 1973) and DNA polymerase slippage (Streisinger et al., 1966). Unequal crossing over involves pairing between misaligned tandem repeats. An exchange would create one array with an increased number of repeats and a second array with the reciprocal decrease. Due to the observation that a reciprocal decrease of repeat sequence is rarely observed in families afflicted with TREDs, unequal crossing over has been ruled out as a mechanism for triplet repeat expansion.

Polymerase slippage synthesis, a favored molecular explanation for genetic expansions of tandemly repeated triplets, is primer/template slippage in DNA replication. This idea was first proposed to explain frameshift mutations in simple sequence repeats in genomes (Streisinger et al., 1966). In DNA polymerase slippage events, the primer and template strands temporarily dissociate during DNA replication. When the strands reassociate, individual repeats are left unpaired. If the unpaired repeat is in the primer strand, continued DNA synthesis results in a larger tandem array. If the unpaired repeat is in the template strand, the same mechanism results in a deletion. Recently, Leach et al. (1995) proposed that during replication of triplet repeat regions, the primer or template strands fold into hairpin structures stabilized by intrastrand base pairing. Formation of a hairpin structure on the elongating primer strand might result in addition of repeats. Alternatively, hairpin

formation on the template strand might result in a deletion if the polymerase can bypass the hairpin. It has been suggested that the potential for slippage is greatest when the lagging strand is the template (Richards and Sutherland, 1994; Kunst and Warren, 1994). Furthermore, it was proposed that when an Okazaki fragment is initiated within a repeated region, slippage of the elongating strand can occur (Eichier et al., 1994, Kang et al., 1995, Mitas et al., 1995a).

At the time the work for this dissertation was initiated, there were no publications on structures of triplet repeat sequences. We were one of the several research groups that initiated on extensive investigation of the relevant structures associated with TREDs. The results of a hairpin structure of ss $(CTG)_{15}$ reported in our earlier publication was recognised as the first study to experimentally demonstrate hairpin structures (J. NIH Res., No. 7, 1995). Most of the trinucleotides associated with TREDs share a palindromic GC dinucleotide, while the CGG trinucleotide additionally contains a palindromic CG dinucleotide. A preliminary conclusion can be drawn that the presence of a palindromic GC or CG dinucleotide within a long trinucleotide repeat may lead to physical instabilty of the DNA structure. These expanded triplet repeats were predicted by energy minimization to fold into stable hairpin structures (Mitas et al., 1995a). These stable secondary structures are essential to the mechanisms of the triplet repeat expansion.

The exact cause of a DNA slippage event, the detailed steps of repeat expansion or contraction (which may happen according to one hypothesis) and the likelihood of such events as DNA polymerase traversing a trinucleotide repeat region need to be demonstrated experimentally. To provide insight into the mechanisms of expansion, it is necessary to understand structures of the DNA containing these sequences. The results of the studies of class I triplet repeat structures presented in this dissertation, and by other laboratories,

provide evidence that supports a slippage replication hypothesis driven by hairpin formation.

CHAPTER II

MATERIAL AND METHODS

Oligonucleotides

All oligonucleotides were synthesized on an Applied Biosystems 381 A oligonucleotide synthesizer (Foster City, CA) with the trityl group on and purified with oligonucleotide purification cartridges (Cruachem, Glasgow, UK). Sequences of oligonucleotides were:

$(CTG)_{15}$: GATCC$(CTG)_{15}$GGTACCA,

$(CAG)_{15}$: AGCTTGGTACC$(CAG)_{15}$GGATC,

$(CCG)_{15}$: GATCC$(CCG)_{15}$GGTACCA,

$(CGG)_{15}$: AGCTTGGTACC$(CGG)_{15}$GGATC,

$(GTC)_{15}$: GATCC$(GTC)_{15}$GGTACCA,

$(GAC)_{15}$: AGCTTGGTACC$(GAC)_{15}$GGATC,

$\Delta T_{II}(GTC)_{15}$: GATCC(GTC)GC$(GTC)_{13}$GGTACCA,

$\Delta A_{XIV}(GAC)_{15}$: AGCTTGGTACC$(GAC)_{13}$GC(GAC)GGATC

Plasmid DNA Preparation

Plasmids were prepared in our laboratory as previously described by Mitas et al. (1995a).

Probe preparation

3' End-labeling

For labeling of an oligonucleotide strand containing a pyrimidine-rich triplet repeat, 15 µg of purified plasmid DNAs were digested with 50 U *Hin*d III (Boehringer Mannheim) for 1 h at 37°C in a volume of 70 µl. Recessed ends were labeled at the 3'-terminus by adding 5 µl [α-^{32}P]dCTP, 5 µl [α-^{32}P]dATP (each 3000 Ci/mmol; ICN, Irvine, CA), 2.5 µl 5 mM dTTP and dGTP and 25 U Klenow enzyme (New England Biolabs). Reactions were incubated for 1 h at room temperature. Plasmid DNAs were extracted with 25:24:1 phenol:chloroform:isoamyl alcohol (PCI) and precipitated with ethanol. Resuspended DNAs were digested with 50 U *Bam*HI in a volume of 70 µl and applied directly to a Nuctrap column (Stratagene) for further removal of unincorporated [^{32}P]dNTPs. Labeling of oligonucleotides containing purine-rich triplet strands was performed in an identical manner except that the order of restriction enzymes was reversed. Probes were diluted to 2 x 10^4 d.p.m./µl with H_2O. For electrophoretic analysis, DNAs (4 x 10^4 d.p.m.) were diluted to 10 µl in buffer containing 8% glycerol, 10 mM HEPES, pH 8.5, 50 mM KCl, 1 mM EDTA. One microliter of loading dye (50% glycerol, 0.4% bromophenol blue) was added to the DNA sample prior to gel electrophoresis. During electrophoresis, 15°C tap water was circulated through a Hoeffer SE 600 series unit (San Francisco, CA).

5' End-labeling

For labeling of an oligonucleotide strand containing a pyrimidine-rich triplet repeat, 15 μg of purified plasmid DNAs were digested with 50 U *Bam*HI (Boehringer Mannheim) for 1 h at 37°C in a volume of 70 μl, and dephosphorylated with calf intestinal phosphatase. DNAs were 5' end-labeled with ^{32}P by incubation in buffer containing 1 mM DTT, 1 μl 7000 Ci/mmol [γ-^{32}P]dATP (ICN, Irvine, CA), 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 0.1 mM spermidine, 0.1 mM EDTA and 20 U T4 polynucleotide kinase (New England Biolabs). The labeled pyrimidine-rich strand (annealed to its unlabeled purine-rich complement) was liberated from the plasmid by digestion with *Hin*dIII. Unincorporated DNA was removed by size exclusion chromatography (Nuctrap column, Stratagene). Labeled DNAs were subjected to electrophoresis in a 0.8% agarose gel. Oligonucleotides containing triplet repeats were excised from gels and purified from vector DNA with glass beads (Mermaid Kit; Bio 101, La Jolla, CA). Labeling of oligonucleotides containing purine-rich triplet strands was performed in an identical manner except that the order of restriction enzymes was reversed.

Labeling of synthetic oligonucleotides used for markers of nucleotide length.

Five micrograms of column-purified synthetic oligonucleotide containing the sequence CGATA(CTG)nACGTA was 5′ end-labeled with [γ-^{32}P]ATP by incubation in buffer containing 1 mM DTT, 1 μl 7000 Ci/mmol [γ-^{32}P]ATP (ICN), 50 mM Tris-HCI (pH 7.6), 10 mM MgCl, 0.1 mM spermidine, 0.1 mM EDTA and 20 U T4 polynucleotide kinase (New England Biolabs).

Unincorporated [γ-^{32}P] ATP was removed by size exclusion chromatography (Nuctrap column, Stratagene).

DNA sequencing

Cesium chloride-purified plasmid DNAs containing various triplet repeat sequences were linearized with *Xho*I. For sequence analysis, 8 pmol of a synthetic oligonucleotide complementary to a portion of the thymidine kinase promoter of pBLCAT2 (sequence GTTCGAATTCGCCAATGACAA) was added to 1 μg linearized plasmid DNA. DNA sequence analysis was performed by the OSU recombinant DNA/Protein Resource Facility on an ABI 373 automated sequencer.

Electrophoresis analysis

The oligonucleotide containing $(CXG)_{15}$ or $(GXC)_{15}$ (X indicating A, C, G or T, respectively) was labeled at the 3' end with the use of Klenow enzyme and subjected to purification as described above. The electrophoretic analysis was described previously (Yu et al., 1995). Briefly, to obtain a homogeneous population of labeled ss$(CXG)_{15}$, 1 μM of unlabeled synthetic oligonucleotide of the same sequence as the labeled strand was added, placed in a 90°C H_2O bath for 5 minutes, and then placed at 25°C for 5 minutes. To obtain a homogeneous population of labeled ss$(CGG)_{15}$, unlabeled synthetic oligonucleotide of the same sequence as the labeled strand was added to yield a final concentration of 1 μM. The DNAs were placed in a 100°C H_2O bath for 5 minutes, and placed at 25°C for 5 minutes. For electrophoretic analysis, DNAs (4 x10^4 d.p.m.) were diluted to 10 μl in buffer containing 8% glycerol, 10 mM HEPES, pH 8.5, and 1

mM EDTA. One µl of loading dye (50% glycerol, 0.4% bromophenol blue) was added to the DNA samples prior to gel electrophoresis. Electrophoresis was performed in a Hoeffer (San Francisco, CA) SE600 series unit at various temperatures at 25 mA/gel in 45 mM Tris-borate (pH 8.5) and 1 mM EDTA (TBE). Gel plates were 14 cm (length) x 16 cm (width) x 1.5 mm (thickness). Electrophoresis was stopped when the bromophenol blue marker migrated 10 cm. Dried gels were placed between two intensifying screens (Dupont) and exposed to Fuji RX film three hours to overnight at -80°C.

P1 nuclease digestion

Unlabeled synthetic oligonucleotide (1.4 pmol) of the same sequence as the labeled strand was added to 4 X 10^3 d.p.m. (0.7 fmol) of 5′ end-labeled DNA, placed in a boiling water bath for 5 min and then cooled at room temperature for 5 min. P1 nuclease digestions were performed essentially according to the method of Wohlrab (Wohlrab, 1992). Briefly, DNA was incubated for 4 min at 37°C in 12 µl buffer containing 0.05 M NaCl, 0.05 M Tris-HCl (pH 7.4), 2 mM $MgCl_2$, 5% glycerol and various amounts of P1 nuclease (Sigma, St. Louis, MO). Reactions were stopped by the addition of 1 µl 0.5 M EDTA. Seventeen microliters of formamide loading buffer were added. DNAs were purified and subjected to polyacrylamide gel electrophoresis as described in the $KMnO_4$ oxidation studies.

Potassium permanganate oxidation

Unlabeled synthetic oligonucleotide (1.4 pmol) of the same sequence as the labeled strand was added to 4 X 10^3 d.p.m. (0.7 fmol) of 5′ end-labeled DNA,

placed in a boiling water bath for 3 min and then cooled at room temperature for 5 min. $KMnO_4$ oxidation was performed essentially according to the method of McCarthy and Rich (1991). Briefly, DNAs were incubated at room temperature in 50 mM sodium cacodylate, pH 7.0, 2 mM EDTA and various concentrations of $KMnO_4$ for 4 min. Final volume of the reaction was 50 µl. Reactions were stopped by addition of 150 µl ice-cold solution containing 98% v/v ethanol, 1% v/v β-mercaptoethanol and 0.02 µg/µl tRNA. DNA was precipitated at -70°C for 1 h after addition of 5 µl 3 M sodium acetate. Pelleted DNA was washed with 750 µl 70% ethanol and dried. DNA was resuspended in 40 µl 1 M piperidine, heated at 92°C for 30 min and dried (X 3). Dried samples were dissolved in 10 µl H_2O and 17 µl formamide loading buffer (80% formamide, 10 mM NaOH, 1 mM EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue) was added. Samples were placed in a boiling water bath for 3 min, chilled on ice for 5 min and loaded onto a 12% polyacrylamide gel containing 8 M urea. During electrophoresis, 55°C tap water was circulated through a Hoeffer SE 600 series unit. Electrophoresis was performed at 25 mA/gel in 45 mM tris-borate, 1 mM EDTA, the same as above discribed.

Dimethyl sulfate reactions

pCNG15 (15 µg) were first digested with *Hin*dIII for pyrimidine rich repeats (or *Bam*HI, for purine rich strands) and dephosphorylated with calf intestinal phosphatase. DNA was labeled at the 5′ end by incubation in buffer containing 1 mM DTT, 1 µL of 7000 Ci/mmol ATP [γ-^{32}P] (ICN), Irvine, CA), 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 0.1 mM spermidine, 0.1 mM EDTA, and 20 units of T4 polynucleotide kinase (New England Biolabs, Beverly, MA) for 1 h at 37°C. Labeled DNA was extracted with PCI and precipitated with ethanol.

Resuspended DNA was digested with 50 units of *Bam*HI (or *Hind*III) in a volume of 70 µL and applied directly to a Nuctrap column (Stratagene, La Jolla, CA) for removal of unincorporated ^{32}P-dNTPs. To separate labeled vector DNA from labeled oligonucleotide, DNAs were subjected to electrophoresis in a 2% agarose gel. The oligonucleotide containing CGG_{15} was excised from the agarose gel and purified with glass beads (Mermaid Kit, Bio 101, La Jolla, CA). Unlabeled synthetic oligonucleotide (1.4 pmol) of the same sequence as the labeled strand was added to 4 x 10^3 dpm (~0.7 fmol) of labeled DNA and the mixture placed in a boiling H_2O bath for 3 min and placed on ice for 5 min. This procedure resulted in complete conversion of all labeled DNA to the ss form (see Figure 2). Dimethyl sulfate (DMS) reactions were performed at 25°C for 15 min essentially according to the method of Maxam and Gilbert (1980). Reaction products were analyzed on a sequencing gel containing 20% polyacrylamide and 8 M urea.

Chemical modification with hydroxylamine

Single-stranded $(CCG)_{15}$ or ss$(CCG)_{20}$ were labeled as described above for the HA analysis. To separate labeled oligonucleotide from labeled vector DNA, DNAs were subjected to electrophoresis in a 2% agarose gel. Oligonucleotides containing $(CCG)_{15}$ or $(CCG)_{20}$ were excised from the agarose gel and purified with glass beads (Mermaid Kit, Bio 101, La Jolla, CA). Hydroxylamine (HA) reactions were performed essentially according to the method of Rubin and Schmid, (1980). HA was freshly prepared by titrating 4 M hydroxylamine hydrochloride (Aldrich) to the required pH (6.5-8.5) with diethylamine. Seven µl of labeled ss$(CCG)_{15}$ (2.0 x 10^4 dpm, ~0.7 fmoles) and 1.4 pmoles of synthetic oligonucleotide of the the same sequence as the labeled strand was heated to

100°C for 3 minutes and immediately placed on ice. Various amounts of HA (6 - 46µl) were added to the DNA and incubated for 25 min at various temperatures (1 - 55°C). NaCl was added to yield a final concentration of 50 mM. Stop solution (0.25 ml) containing 0.3 M sodium acetate (pH 5.2), 0.1 mM EDTA, 25 µg/ml tRNA was added, followed by 0.75 ml cold ethanol. The DNA was precipitated by centrifugation and was resuspended in 100 µl 1 M piperidine and heated for 30 min at 90°C to generate strand breaks. After removal of piperidine *in vacuo* (x 2), the DNA was resuspended in formamide loading buffer, placed in a boiling water bath for 2 minutes, immediately chilled on ice, and loaded on a DNA sequencing gel containing 20% polyacrylamide and 8 M urea.

Chemical modification with 2-hydroperoxytetrahydrofuran

THF-OOH was prepared according to the method of Liang et al. (1994). Labeled DNA (2 x 10^4 dpm, ~ 0.7 fmol) and unlabeled synthetic oligonucleotide (1.4 pmoles) of the same sequence as the labeled strand were combined in buffer containing 50 mM Tris-HCl, 50 mM NaCl and 10 mM $MgCl_2$, in a final reaction volume of 30 µl, placed in a boiling water bath for 3 min, and placed on ice. THF-OOH was added to yield final concentrations of 0.5, 1.0, and 2.0 M. Reactions were incubated for 1 h at various temperatures. The pH of the reactions was either 7.5 or 8.5. In alternative reactions, the amount of THF-OOH were decreased to 0.2 M and the reaction time was increased to 10 hrs. Incubations were stopped by cooling the reaction in ice and precipitating the DNA at -70°C by adding 0.23 ml stop solution and 0.75 ml ice cold EtOH. The precipitated DNA was washed with EtOH and dried *in vacuo*. To generate strand breaks, the THF-OOH treated DNA was heated at 90°C for 15 min in 30

µl of 10 mM Tris-HCl (pH 7.0), and 1 mM EDTA. The DNA was precipitated, washed and dried, and treated with piperidine as described above for the HA reactions.

Chemical modification with diethyl pyrocarbonate

Labeled DNA (2 x 10^4 dpm, ~0.7 fmol) and unlabeled synthetic oligonucleotide (1.4 pmoles) of the same sequence as the labeled strand were combined in buffer containing 50 mM sodium cacodylate, pH 7.0, and 2 mM EDTA, in a final reaction volume of 50 µl, placed in a boiling water bath for 3 min, and chilled on ice, then preincubated at room temperature (23°C) for 5 min prior to chemical modification. Diethyl pyrocarbonate (DEPC) (Sigma) was added to yield final concentrations of 0.135 M to 1.674 M. Reactions were vortexed vigorously at the start of the DEPC reaction and after 5 and 10 min. Reactions are allowed to proceed for 15 min at various temperatures. The pH of the reactions was varied between 6.5 and 8.5. Reactions were stopped by cooling the reaction in ice and precipitating the DNA at -70°C by adding 50 µl stop solution containing 1.5 M of sodium acetate, 200 µg/ml tRNA, pH 5.2, and 0.75 ml ice cold EtOH. The precipitated DNA was washed with EtOH and dried *in vacuo*. The dried, modified DNA was precipitated by centrifugation and was resuspended in 100 µl 1 M piperidine and heated for 30 min at 90°C to generate strand breaks. After removal of piperidine *in vacuo* (x 2), the DNA was resuspended in formamide loading buffer, placed in a boiling water bath for 2 minutes, immediately chilled on ice, and loaded on a DNA sequencing gel containing 20% polyacrylamide and 8 M urea.

Chemical modification with hydrazine

Labeled DNA (2 x 10^4 dpm., ~0.7 fmol) and unlabeled synthetic oligonucleotide (1.4 pmoles) of the same sequence as the labeled strand were combined in buffer containing 50 mM sodium cacodylate, pH 7.0, and 2 mM EDTA, in a final reaction volume of 25 µl, placed in a boiling water bath for 3 min, and chilled on ice, then preincubated at room temperature (23°C) for 5 min prior to chemical modification. 95% hydrazine (Sigma) was added to yield final concentrations of 6.25, 15.5, and 25 M (10 - 40 µl). Reactions were mixed by gently tapping the tube with a finger at the start of the hydrazine reaction, and were allowed to proceed for 10 min at 23°C. Reactions were stopped by cooling the reaction in ice and precipitating the DNA at -70°C by adding 200 µl "hydrazine stop solution" containing 0.3 M of sodium acetate, 0.1 mM EDTA, 100 µg/ml yeast tRNA, pH 7.0, and 0.75 ml ice cold EtOH. The precipitated DNA was carried through EtOH wash, piperidine breakage, and removal of piperidine as described above in the dimethyl sulfate modification. The DNA was resuspended in formamide loading buffer, placed in a boiling water bath for 2 minutes, immediately chilled on ice, and loaded on a DNA sequencing gel containing 20% polyacrylamide and 8 M urea.

CHAPTER III

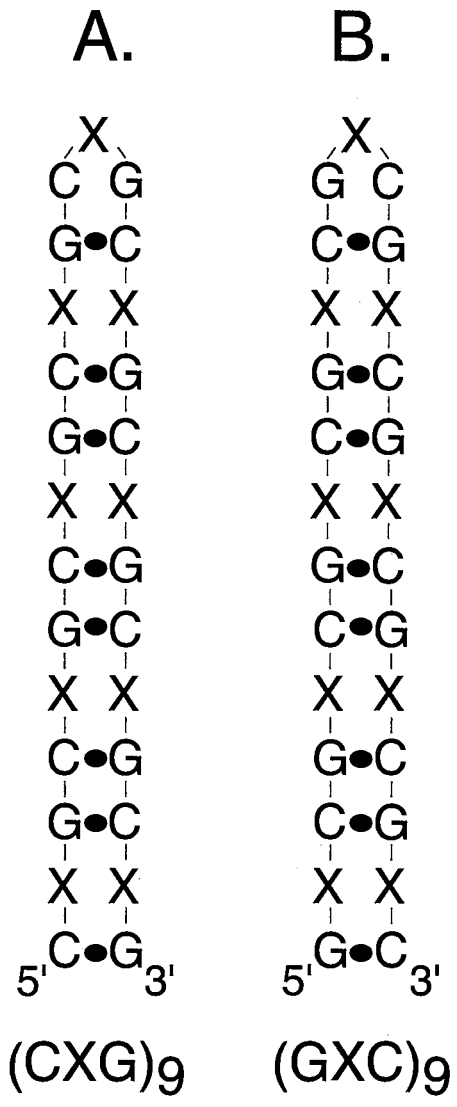
CLASSIFICATION OF TRINUCLEOTIDE REPEATS

Trinucleotide repeat DNA sequences are simple tandem repetitive sequences. They fall into the "microsatellite" category of repetitive sequences in the human genome (Jeffreys et al., 1985). Although trinucleotide repeat sequences have been found to be widely scattered in the human genome, no definitive function has been ascribed to them (Sutherland and Richards, 1995). To aid in correlating potential structures of triplet repeat nucleic acids with their function and their propensity to undergo expansion events, Mitas *et al.* (1995a) described a sequence-based classification system for them. All possible triplet repeats were classified into six groups according to presence or absence of homopolymeric sequence, presence or absence of a palindromic dinucleotide and the G+C content. Class I triplet repeats, containing (CGG)n/(CCG)n, (CAG)n/(CTG)n, and (GAC)n/(GTC)n, were defined by the presence of a GC or CG palindrome, and exhibited the lowest base stacking energies, the lowest slippage synthesis rates, and were associated with triplet repeat expansion diseases. Since the GC or CG dinucleotide is palindromic, ss Class I triplet repeats could form imperfect but stable hairpin structures with mismatched third bases (Figure 1). All six complementary single strands of class I triplet repeats potentially formed stable hairpin structures, as determined by energy minimization.

Investigation of the potential structures of each single strand repeat sequence in the class I repeat members is necessary to provide experimental evidence to verify the theoretical prediction. It is also possible that common characteristics in the structures of these triplet repeats associated with diseases might reveal some phenomena relevant to a possible common expansion mechanism. I will describe the investigation of the structures of all the six triplet repeat sequences in the following chapters. Our work (Mitas et al., 1995a, Mitas et al., 1995b, Yu et al., 1995a; Yu et al., 1995b), as well as that reported by others (Gacy et al., 1995, Chen et al., 1995, Smith et al., 1995, Mariappan et al., 1996, Petruska et al., 1996), supports the hairpin structures. These hairpin structures vary in stability with melting temperatures that range from 30°C up to 75°C in low ionic strength.

Figure 1. Potential hairpin structures of ss oligonucleotides containing nine Class I triplet repeats.

X represents A, T, C or G base. Watson-Crick base pairs are indicated by filled ovals.


A.
B.
X
C G
G C
X X
C G
G C
X X
C G
G C
X X
C G
G C
X X
C G
5' 3'
(CXG)9
X
G C
C G
X X
G C
C G
X X
G C
C G
X X
G C
C G
X X
G C
5' 3'
(GXC)9

CHAPTER IV

SINGLE STRAND $(CTG)_{15}$ ADOPTS A HAIRPIN STRUCTURE

(CTG)n/(CAG)n repeated sequences are particularly interesting because expansion of them is associated with several neurological disorders. To date, at least seven TREDs caused by expansion of (CTG)n/(CAG)n repeat sequences of individual genes have been reported. A characteristic of (CAG)n expansion is that the mutated sequences expand to less than one hundred repeat copy numbers, where they are located within the coding region of a gene. This is compared to hundreds or over a thousand copies where the repeat sequences are located in non-coding regions of other genes (Sutherland and Richards, 1995). To understand the mechanism whereby (CTG)n/(CAG)n preferably undergoes expansion and uniquely causes several neurological diseases, it is essential to characterize the secondary structures of each single strand.

Chemical modification of nucleic acids is one of the methods widely used to investigate structure in solution. Many reagents which react to target bases with high specificity and are sensitive to the accessibility of the target groups have been developed. Many of these reagents preferentially attack the bases in the exposed single stranded regions in the nucleic acid structure and do not modify the bases with the target group when they participate in hydrogen bond formation. These properties of the base-specific reagents are of great help in identifying in solution bases participating in secondary or tertiary structure by efficently distinguishing the specific base that participates not only in hydrogen

bond formation, but also in base stacking interactions. Chemical probes are particular valuable as they obviate concerns arising from the possible perturbation caused by large molecules, such as enzymes, frequently used for structural analysis. In addition, chemical modification methods are able to probe relatively large sequences and avoid the complicated steps of crystallization needed for X-ray diffration. In the studies presented in this and the following chapters, several base specific chemical probes were used to investigate the structures of class I single stranded triplet repeat sequences. The investigation was based on the prediction that all class I triplet repeats have a potential to form stable hairpin structures (Mitas et al., 1995a). For the following studies, oligonucleotides containing fifteen repeat units were utilized. Non-repetitive sequences were also included in the termini of the oligonucleotides to help prevent "slippage" of the hairpin structure. To explore the possibility of hairpin formation by ss class I triplet repeats, $d(CTG)_{15}$ was first investigated.

The electrophoretic mobility of oligonucleotides containing single strand $(CTG)_{15}$ is concentration-independent

To analyze DNA containing CTG repeats, a double stranded (ds) oligonucleotide containing 15 CTG triplet repeats was cloned into a plasmid as described in Materials and Methods. Oligonucleotides liberated from the plasmid were utilized for all studies.

To detect whether the oligonucleotides exhibited properties of intramolecular hairpins or some type of intermolecular structure under the experimental conditions, various studies were performed. First, the molecular composition of the structure(s) formed with ss triplet repeat $(CTG)_{15}$ was investigated by performing electrophoretic studies with labeled ss $(CTG)_{15}$

mixed with various amounts of unlabeled ss synthetic oligonucleotide containing ss $(CTG)_{15}$ (Figure 2). If ss $(CTG)_{15}$ formed a stable intramolecular hairpin structure, increasing the concentration of unlabeled synthetic ss $(CTG)_{15}$ should have no effect on the amount of hairpin formed.

In the absence of added unlabled ss synthetic oligonucleotide, a predominant species of DNA with a relatively fast electrophoretic mobility was observed, corresponding to ss DNA (Figure 2). Addition of a 10^5-fold molar excess (final DNA concentration 7 μM) of unlabeled ss synthetic oligonucleotide of the same sequence as ss $(CTG)_{15}$ did not result in formation of slow migrating complexes (as anticipated), indicating that ss $(CTG)_{15}$ formed a stable unimolecular structure.

A minor, slower migrating species of DNA was also detected (Figure 2). Addition of a 10^3-fold molar excess of complementary unlabeled ss synthetic oligonucleotide resulted in conversion of the fast migrating species to the slow migrating species (Figure 3), indicating that the slow migrating species was the ds form of $(CTG)_{15}$.

To demonstrate that the labeled ss synthetic oligonucleotide containing $(CTG)_{15}$ was not degraded and contained CTG repetitive sequences, a third assay was performed. A control experiment was performed with labeled ss $(CAG)_{15}$ (Figure 4). Addition of increasing amounts of unlabeled ss synthetic oligonucleotide containing $(CTG)_{15}$ to the labeled ss $(CAG)_{15}$ probe resulted in complete conversion of the fast migrating ss form to the slow migrating ds form.

$KMnO_4$ oxidization shows a sensitive region in the middle of ss $(CTG)_{15}$

Potassium permanganate reacts with DNA primarily via oxidation of the 5-6 double bond of pyrimidines (T>>C). Base pairing by forming hydrogen

bonds and base stacking affords protection from this reaction by prohibiting $KMnO_4$ from attacking above or below the plane of the base. The unpaired or unstacked thymines are oxidized preferentially, resulting in strand cleavage upon subsequent treatment with piperidine (Hayatsu and Ukita, 1967, Rubin and Schmid, 1980).

To investigate the structures of the intramolecular ss $(CTG)_{15}$, $KMnO_4$ oxidiation experiments were performed. The structures of two non-class I triplet repeats, ss $(ATC)_{15}$ and ss $(GAT)_{15}$, were also investigated. Electrophoretic analysis revealed that neither of these sequences contained secondary structure. Both ss $(ATC)_{15}$ and ss $(GAT)_{15}$ showed uniform oxidation of all thymines (Figure 5, and Figure 6), indicating that these residues did not participate in base pairing or base stacking interactions. These results indicated that ss $(ATC)_{15}$ and ss $(GAT)_{15}$ do not form hairpin structures.

Treatment of ss $(CTG)_{15}$ with $KMnO_4$/piperidine resulted in a single cleavage product within the triplet repeat region at T_{28} (predicted loop region C27-T28-G29) (Figure 7). The results indicated that ss $(CTG)_{15}$ forms a structure that contains a single loop located in the middle of the triplet repeat region and a stem region where thymines were involved in base paring and/or base stacking interactions.

P1 digestion

ss $(CTG)_{15}$ versus ss $(ATC)_{15}$

Nuclease sensitivity studies were performed with P1, a single-stranded-specific endonuclease that exhibits no apparent sequence specificity (Wohlrab, 1992). Experiments were performed at 37°C with increasing amounts of P1

nuclease. For comparison, ss $(ATC)_{15}$ was also tested for P1 nuclease sensitivity. The results with ss $(ATC)_{15}$ show that P1 nuclease uniformly cleaved the triplet repeat region to a minor extent (Figure 8), demonstrating that ss $(ATC)_{15}$ does not form a hairpin.

Treatment of ss $(CTG)_{15}$ with P1 produced a fragment of size similar to that generated by $KMnO_4$/piperidine treatments, indicating a single strand region in the middle of the sequence in which the bases were unpaired and unstacked. (Figure 9). However, triplets I-VII and IX-XV were not digested. The results of P1 digestion are in agreement with $KMnO_4$ studies and are consistent with a hairpin structure of $(CTG)_{15}$ in which the thymines in the stem are base paired and/or stacked. To further analyze the results of P1 digestion and $KMnO_4$ oxidation, it was necessary to determine the precise sites of cleavage.

Determination of cleavage sites

To further investigate the putative hairpin loop structure of ss $(CTG)_{15}$, P1 nuclease digestions were performed at 50 mM Na^+ and pH 7.4. In this study, probe ss $(CTG)_{15}$ was labeled near the 3'-terminus with Klenow enzyme. After reacting with P1 nuclease, the products were applied to a 20% polyacrylamide-8 M urea sequencing gel. To generate size markers, the DNAs were separately reacted with DMS and cleaved with piperidine. DMS reacts specifically with guanine by methylating N^7, thus producing size markers for each triplet.

Incubation of ss $(CTG)_{15}$ with P1 nuclease resulted in a single digestion product in the loop of ss $(CTG)_{15}$. The product co-migrated with the DMS product generated from cleavage of G_{VII} (the VII band is faint, but clearly detectable on the autoradiograph) (Figure 10). Therefore, the single site of cleavage within the triplet repeat region was the G_{VII}-C_{VIII} phosphodiester

bond. The results of nuclease cleavage of ss $(CTG)_{15}$ are consistent with a hairpin structure and suggest extensive base pairing and or/base stacking interactions between the T•T mismatches.

$KMnO_4$ oxidation at various temperatures reveals thermal stability of the $(CTG)_{15}$ hairpin

To further investigate the properties of the hairpin conformation of ss$(CTG)_{15}$ and the stability of the hairpin, $KMnO_4$ oxidation experiments were performed in 50 mM Na^+ with and without 150 mM KCl at 40, 50, 60 and 70°C. In the presence or absence of KCl, the thymines in the stems of ss$(CTG)_{15}$ were not oxidized by $KMnO_4$ at 40°C and 50°C (Figure 11 and 12). In the absence of added KCl, significant oxidation of the thymines in the presumed stem region of ss$(CTG)_{15}$ occurred at 60°C (Figure 13). In the presence of KCl, significant oxidation required an incubation temperature of 70°C (Figure 14). These results indicate that the ss$(CTG)_{15}$ hairpin was relatively heat stable. This stability is probably contributed by formation of hydrogen bonds between the T-T pairs (Figure 15). K^+ stabilizes the T-T interactions and the ss$(CTG)_{15}$ hairpin.

Discussion

The results of P1 nuclease digestion and $KMnO_4$ oxidation provided experimental evidence to support the prediction of a hairpin structure for ss $(CTG)_{15}$. The results of $KMnO_4$ oxidation at various temperatures, as well as the EMMP data from our lab, demonstrated that ss $(CTG)_{15}$ hairpin is relatively heat stable (Tm 48°C in low ionic strength) (Yu et al., 1995a). Increasing the concentration of potassium ions stabilized the hairpin structure. For instance, at

150 mM K^+, the Tm value of ss $(CTG)_{15}$ was enhanced to 60~70°C. The hairpin structure of ss $(CTG)_{15}$ is with an alignment that triplet I paired with triplet XV, triplet II paired with triplet XIV,... . In each pair of triplets, cytosine in one triplet binds with guanine in the complementary triplet to form a Watson-Crick base pair. The T•T mismatch in each triplet pair interacts in the helix by base-stacking and/or base-pairing interactions. Computer modeling simulations performed in collaboration with Dr. Haworth indicated that the thermal stability of ss $(CTG)_{15}$ hairpin was enhanced by formation of T•T base pairs (Figure 15). [^{1}H]NMR studies performed in other research groups have shown that the T-T pairs in hairpins formed from (CTG)n are hydrogen bonded (Gacy et al., 1995, Smith et al., 1995, Marriapin et al., 1996). The results are in agreement with our prediction.

Figure 2. Structure of ss$(CTG)_{15}$ is concentration-independent.

A ds oligonucleotide containing $(CTG)_{15}$ was excised from plasmid pCTG15 as described in Materials and Methods. Strands end-labeled with polynucleotide kinase were ss$(CTG)_{15}$. Prior to gel electrophoresis, unlabeled synthetic oligonucleotide containing $(CTG)_{15}$ at the indicated concentration was added to the labeled ss $(CTG)_{15}$, placed in a boiling water bath for 10 min and then cooled at room temperature for 15 min. DNA samples were applied to a native 8% polyacrylamide gel.

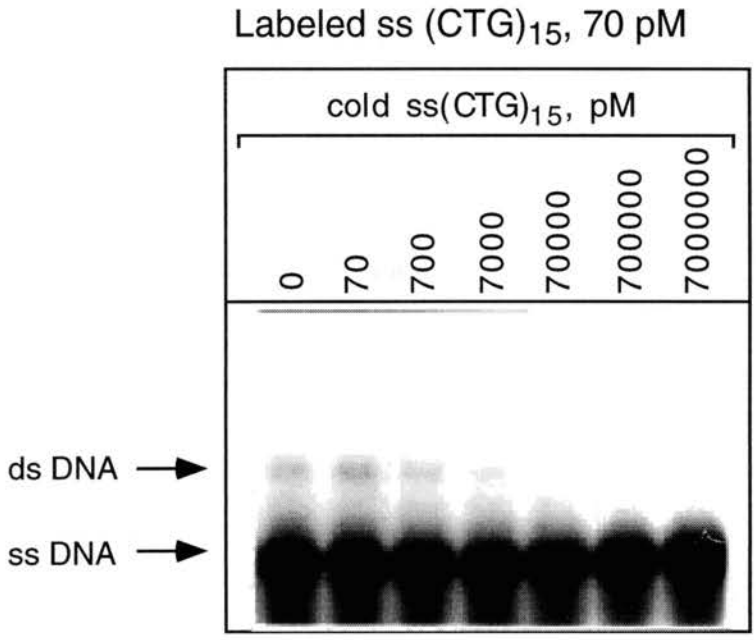

Labeled ss (CTG)15, 70 pM
cold ss(CTG)15, pM
0
70
700
7000
70000
700000
7000000
ds DNA
ss DNA

Figure 3. Labeled ss$(CTG)_{15}$ with added cold ss$(CAG)_{15}$.

Strands end-labeled with polynucleotide kinase were ss$(CTG)_{15}$. Prior to gel electrophoresis, unlabeled synthetic oligonucleotide containing $(CAG)_{15}$ at the indicated concentration was added to the labeled ss $(CTG)_{15}$, placed in a boiling water bath for 10 min and then cooled at room temperature for 15 min. DNA samples were applied to a native 8% polyacrylamide gel.

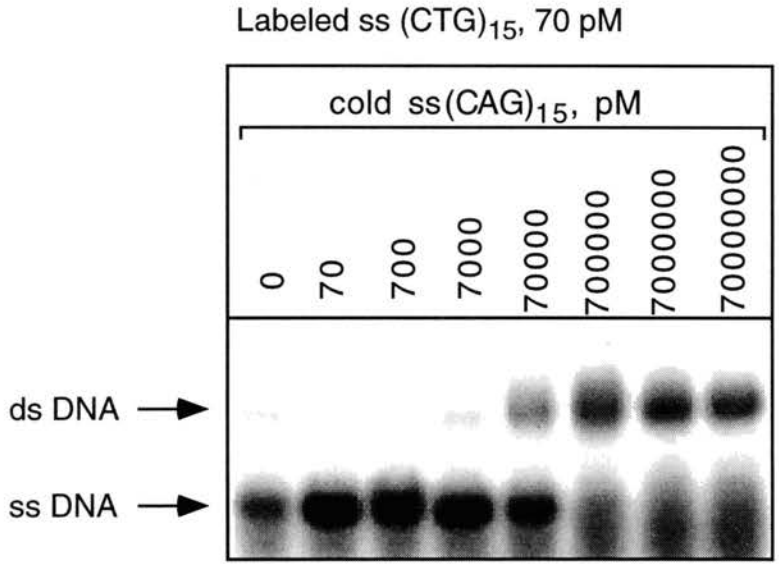

Labeled ss (CTG)15, 70 pM
cold ss(CAG)15, pM
0
70
700
7000
70000
700000
7000000
70000000
ds DNA
ss DNA

Figure 4. Labeled ss$(CAG)_{15}$ with added cold ss$(CTG)_{15}$

Strands end-labeled with polynucleotide kinase were ss $(CAG)_{15}$. Prior to gel electrophoresis, unlabeled synthetic oligonucleotide containing $(CTG)_{15}$ at the indicated concentration was added to the labeled ss (CAG)15, placed in a boiling water bath for 10 min and then cooled at room temperature for 15 min. DNA samples were applied to a native 8% polyacrylamide gel.

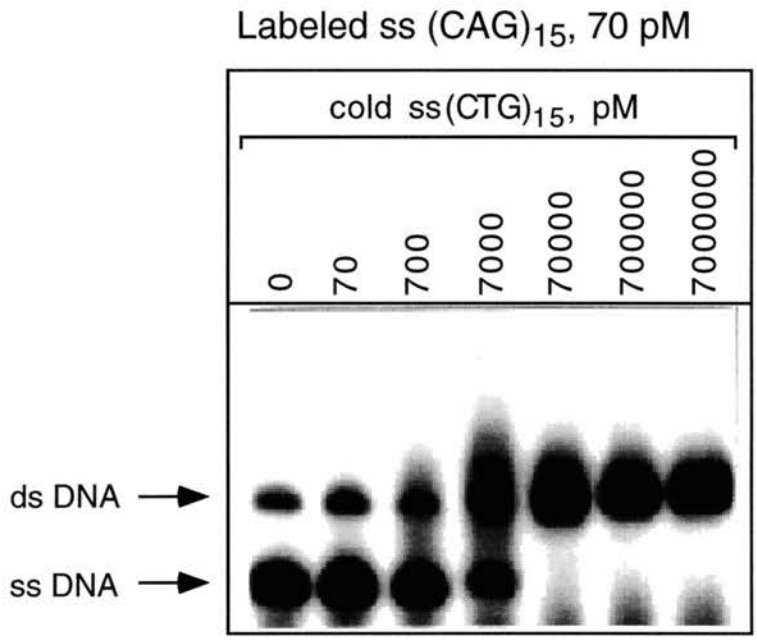

Labeled ss (CAG)15, 70 pM
cold ss(CTG)15, pM
0
70
700
7000
70000
700000
7000000
ds DNA
ss DNA

Figure 5. $KMnO_4$ uniformly oxidizes all thymines of $ss(ATC)_{15}$.

A ss oligonucleotide containing $(ATC)_{15}$ was prepared as described in Materials and Methods and labeled at the 5' end with polynucleotide kinase. Oxidation reactions were performed at room temperature as described in Materials and Methods at the above indicated concentration of $KMnO_4$. Reactions were performed with and without piperidine as indicated. The marker lane contained synthetic 5' end-labeled ss oligonucleotides with the sequence 5'-CGATA$(CTG)_n$ACGTA-3', where n = 1, 3, 5, or 7, or 9, corresponding to lengths of 13, 19, 25, 31 and 37 nt respectively. The gel contained 12% polyacrylamide and 8 M urea. In the lane containing 250 μM $KMnO_4$ the second signal from the bottom corresponds to the thymine in the first ATC repeat.

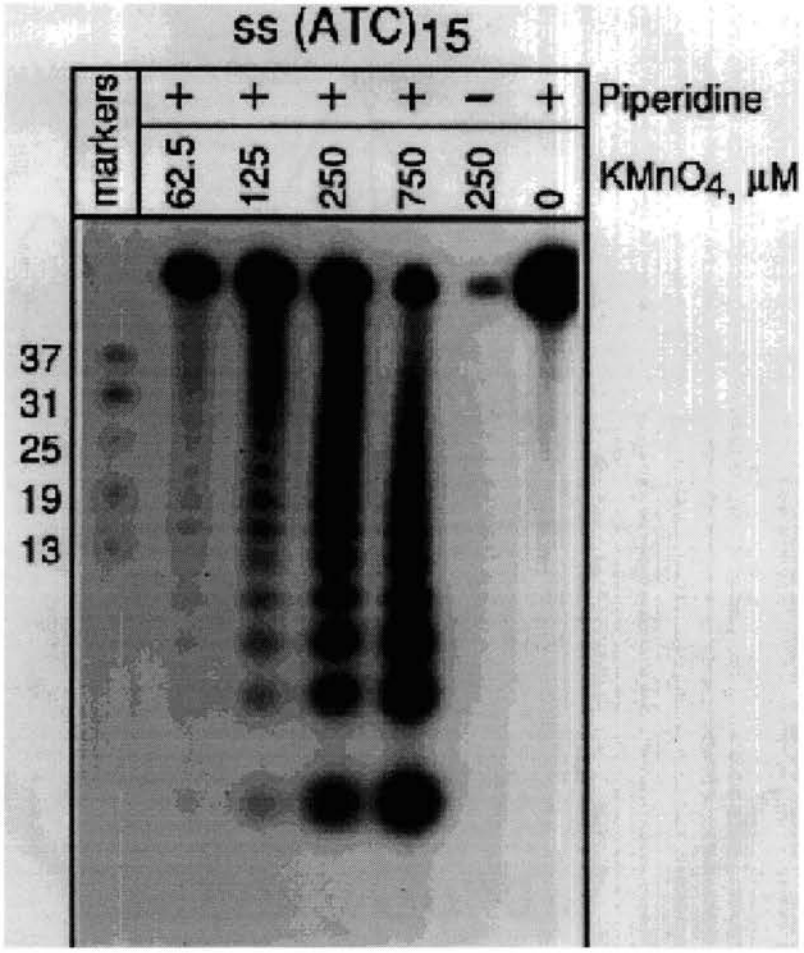

ss (ATC)15
markers
+
+
+
+
−
+
Piperidine
62.5
125
250
750
250
0
KMnO4, µM
37
31
25
19
13

Figure 6. $KMnO_4$ uniformly oxidizes all thymines of $ss(GAT)_{15}$.

A ss oligonucleotide containing $(GAT)_{15}$ was prepared as described in Materials and Methods and labeled at the 5' end with polynucleotide kinase. Oxidation reactions were performed at room temperature as described in Materials and Methods at the above indicated concentration of $KMnO_4$. Reactions were performed with and without piperidine as indicated. The marker lane contained synthetic 5' end-labeled ss oligonucleotides with the sequence 5'-CGATA$(CTG)_n$ACGTA-3', where n = 1, 3, 5 or 7, corresponding to lengths of 13, 19, 25, 31 and 37 nt respectively. The gel contained 12% polyacrylamide and 8 M urea. In the lane containing 250 μM $KMnO_4$ the first signal from the bottom corresponds to the thymine in the first GAT repeat.

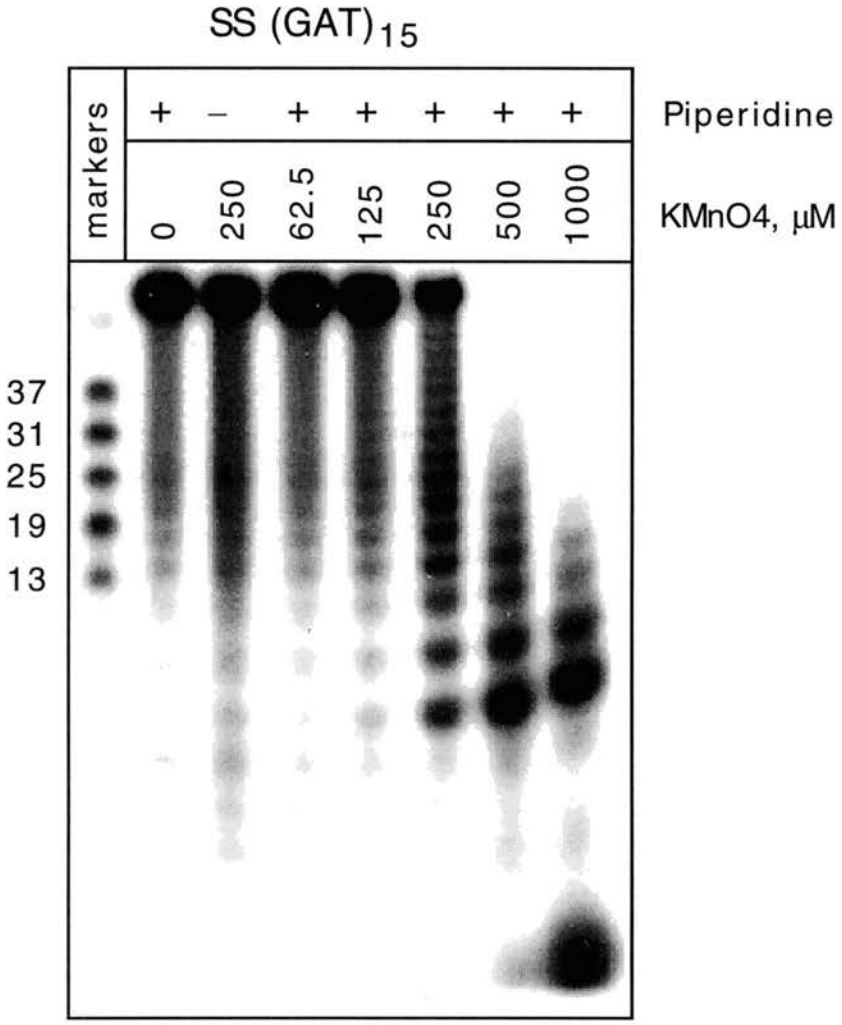

SS (GAT)15
markers
+
–
+
+
+
+
+
Piperidine
0
250
62.5
125
250
500
1000
KMnO4, μM
37
31
25
19
13

Figure 7. $KMnO_4$ oxidizes a single thymine in the middle of the triplet repeat region of ss$(CTG)_{15}$.

A ss oligonucleotide containing $(CTG)_{15}$ was prepared as described in Materials and Methods and labeled at the 5' end with polynucleotide kinase. Oxidation reactions were performed at room temperature as described in Materials and Methods at the indicated concentration of $KMnO_4$. Reactions were performed with and without piperidine as indicated. The gel contained 12% polyacrylamide and 8 M urea. The marker lane contained synthetic 5' end-labeled ss oligonucleotides with the sequence 5'-CGATA$(CTG)_n$ACGTA-3', where n = 1, 3, 5 or 7, corresponding to lengths of 13, 19, 25, 31 and 37 nt respectively.

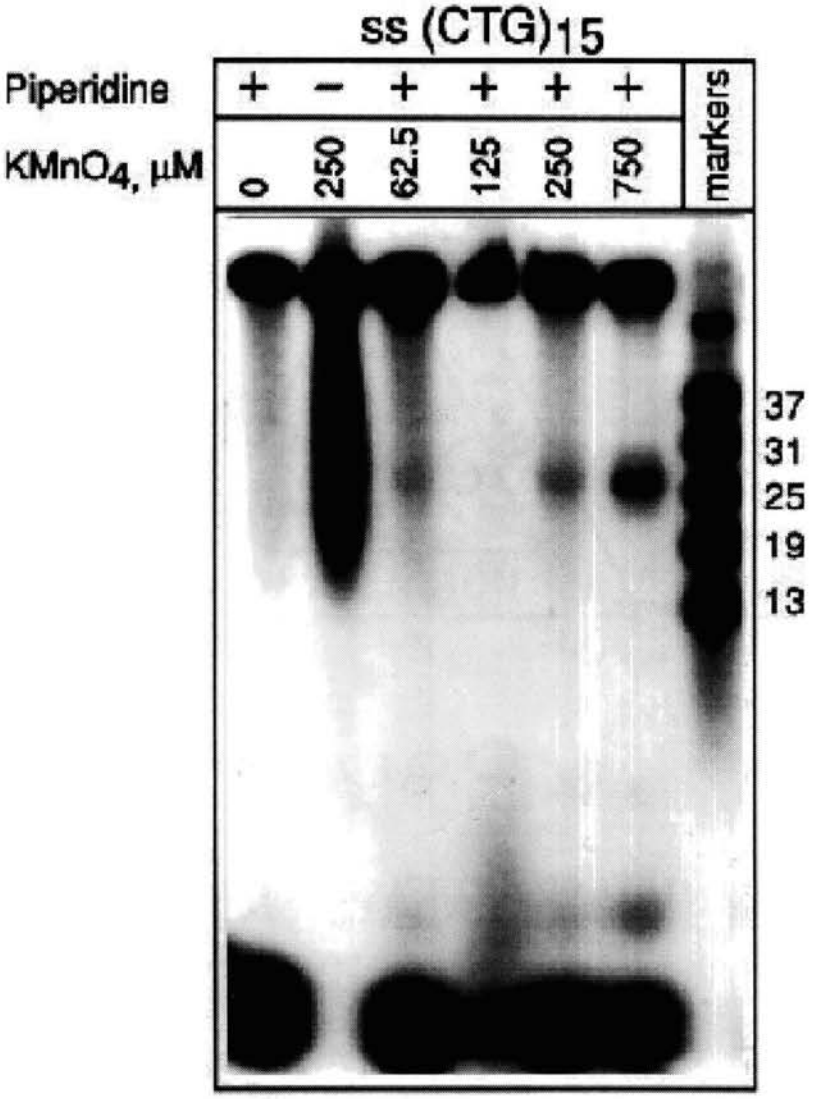

ss (CTG)15
Piperidine
+ − + + + +
KMnO4, μM
0 250 62.5 125 250 750
markers
37
31
25
19
13

Figure 8. P1 nuclease digestion of ss $(ATC)_{15}$.

P1 nuclease digestions were performed at 37°C with ss $(ATC)_{15}$ as described in Materials and Methods. The amounts of P1 nuclease used to digest ss$(ATC)_{15}$ were (from left to right) 0, 2.6, 7.5, 25 and 75×10^{-4} U. Marker lane contained synthetic 5' end-labeled ss oligonucleotides with the sequence 5'-CGATA$(CTG)_n$ACGTA-3', where n = 1, 3, 5 or 7, corresponding to lengths of 13, 19, 25, 31 and 37 nt respectively. The gel contained 12% polyacrylamide and 8 M urea.

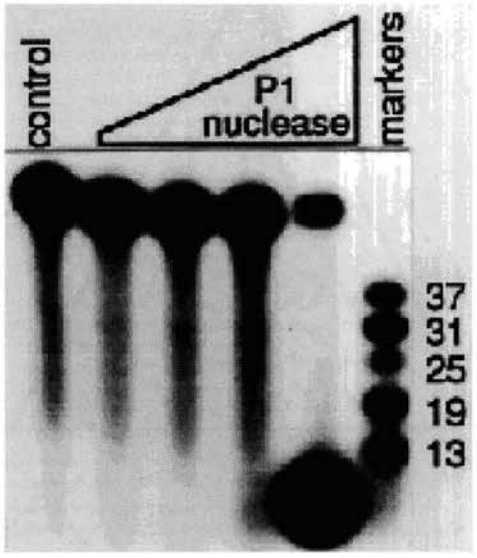

control
P1
nuclease
markers
37
31
25
19
13

Figure 9. P1 nuclease digestion of ss$(CTG)_{15}$.

P1 nuclease digestions were performed at 37 °C with ss$(CTG)_{15}$ as described in Materials and Methods. The amounts of P1 nuclease used to digest ss$(CTG)_{15}$ were (from left to right) 3.8, 12 and 34 × 10^{-3} U respectively. Marker lane contained synthetic 5' end-labeled ss oligonucleotides with the sequence 5'-CGATA$(CTG)_n$ACGTA-3', where n = 1, 3, 5 or 7, corresponding to lengths of 13, 19, 25, 31 and 37 nt respectively. The gel contained 12% polyacrylamide and 8 M urea.

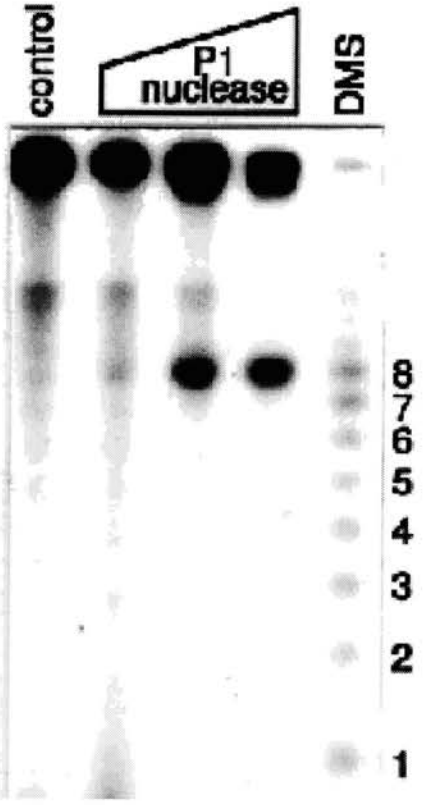

control
P1
nuclease
DMS
8
7
6
5
4
3
2
1

Figure 10. Determination of the P1 nuclease digestion sites on ss$(CTG)_{15}$.

The oligonucleotide containing $(CTG)_{15}$ purified from pCTG15 was labeled on the CTG-containing strand only with Klenow enzyme and ^{32}P. One micromolar unlabeled synthetic oligonucleotide of the same sequence as the labeled strand was added to the oligonucleotide liberated from the plasmid, placed in a 100°C water bath for 3 min and immediately chilled on ice. The amounts of P1 nuclease used to digest ss $(CTG)_{15}$ (from left to right) at 37°C in 50 mM Na^+ were 1.15×10^{-2}, 3.46×10^{-2} and 0.104 U respectively. DMS (21 mM) reactions were performed as described in Materials and Methods. Roman numerals represent triplet repeat numbers. The nucleotides labeled with ^{32}P are marked with an asterisk. Arrows indicate sites of P1 nuclease cleavage.

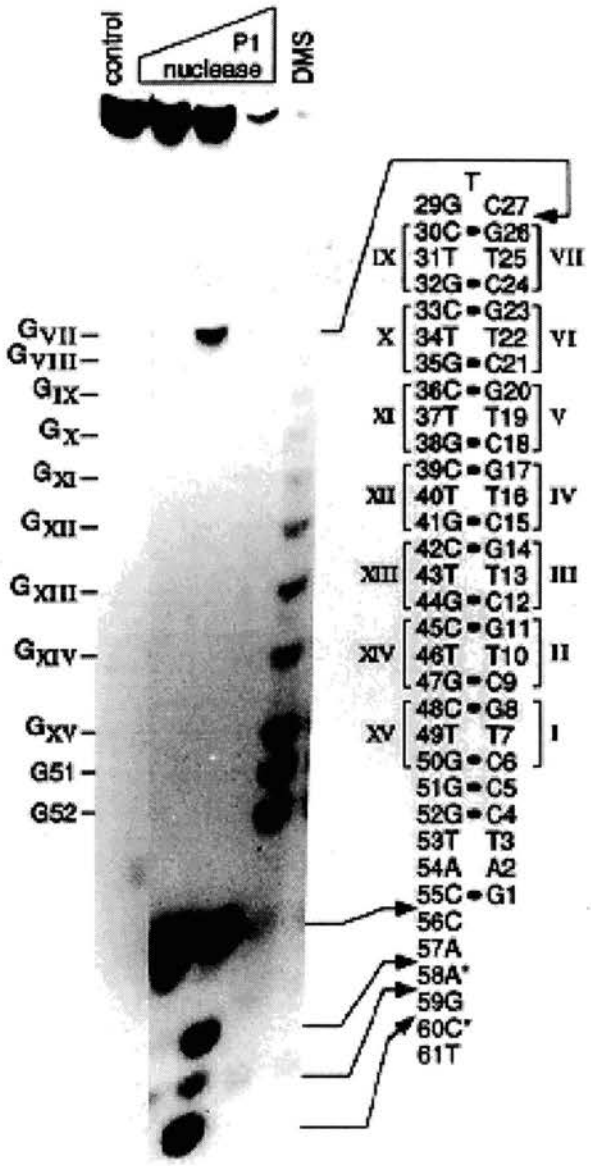

control
P1 nuclease
DMS
T
29G C27
30C • G26
31T T25
32G • C24
33C • G23
34T T22
35G • C21
36C • G20
37T T19
38G • C18
39C • G17
40T T16
41G • C15
42C • G14
43T T13
44G • C12
45C • G11
46T T10
47G • C9
48C • G8
49T T7
50G • C6
51G • C5
52G • C4
53T T3
54A A2
55C • G1
56C
57A
58A*
59G
60C*
61T
IX
X
XI
XII
XIII
XIV
XV
VII
VI
V
IV
III
II
I
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G51
G52

Figure 11. $KMnO_4$ oxidation of ss $(CTG)_{15}$, at 40 ^oC.

$KMnO_4$ oxidations were performed with ss$(CTG)_{15}$ as described in Materials and Methods at 40 ^oC in 50 mM Na^+, no K^+ (left), and in 50 mM Na^+, 150 mM K^+ (right) . From left to right the amounts of $KMnO_4$ in the reaction mixtures were 0, 0.13, 0.25 and 0.75 mM. The marker lanes contained DNAs reacted with DMS (10.5 and 21 mM of DMS from left to right). The positions of the G residues in the triplet repeats are indicated to the right in the figures. The deduced hairpin structure of the DNA sequences is shown to the left in the figures. Roman numerals depict a particular triplet repeat. Conventional arabic numerals depict the position of a nucleotide with respect to the 5'-end. $KMnO_4$ oxidation sites of particular thymine residues are indicated by arrows.

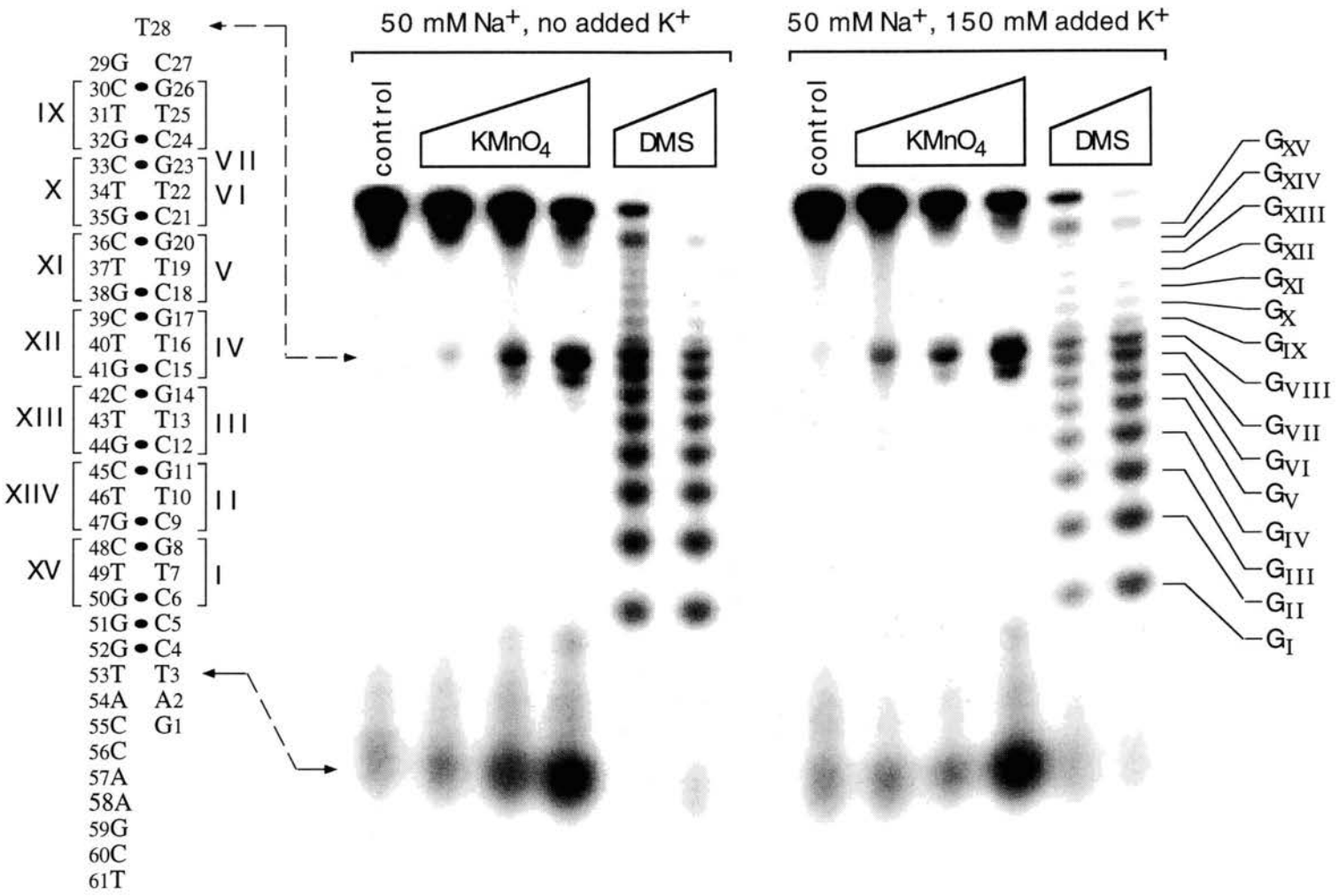

T28
29G C27
IX
30C • G26
31T T25
32G • C24
VII
X
33C • G23
34T T22
VI
35G • C21
XI
36C • G20
37T T19
V
38G • C18
XII
39C • G17
40T T16
IV
41G • C15
XIII
42C • G14
43T T13
III
44G • C12
XIIV
45C • G11
46T T10
II
47G • C9
XV
48C • G8
49T T7
I
50G • C6
51G • C5
52G • C4
53T T3
54A A2
55C G1
56C
57A
58A
59G
60C
61T
50 mM Na+, no added K+
control
KMnO4
DMS
50 mM Na+, 150 mM added K+
control
KMnO4
DMS
GXV
GXIV
GXIII
GXII
GXI
GX
GIX
GVIII
GVII
GVI
GV
GIV
GIII
GII
GI

Figure 12. $KMnO_4$ oxidation of ss $(CTG)_{15}$, at 50 °C.

$KMnO_4$ oxidations were performed with ss $(CTG)_{15}$ as described in Materials and Methods at 50 °C in 50 mM Na^+, no K^+ (left), and in 50 mM Na^+, 150 mM K^+ (right) . From left to right the amounts of $KMnO_4$ in the reaction mixtures were 0, 0.13, 0.25 and 0.75 mM. See the legend to Figure 11 for other details.

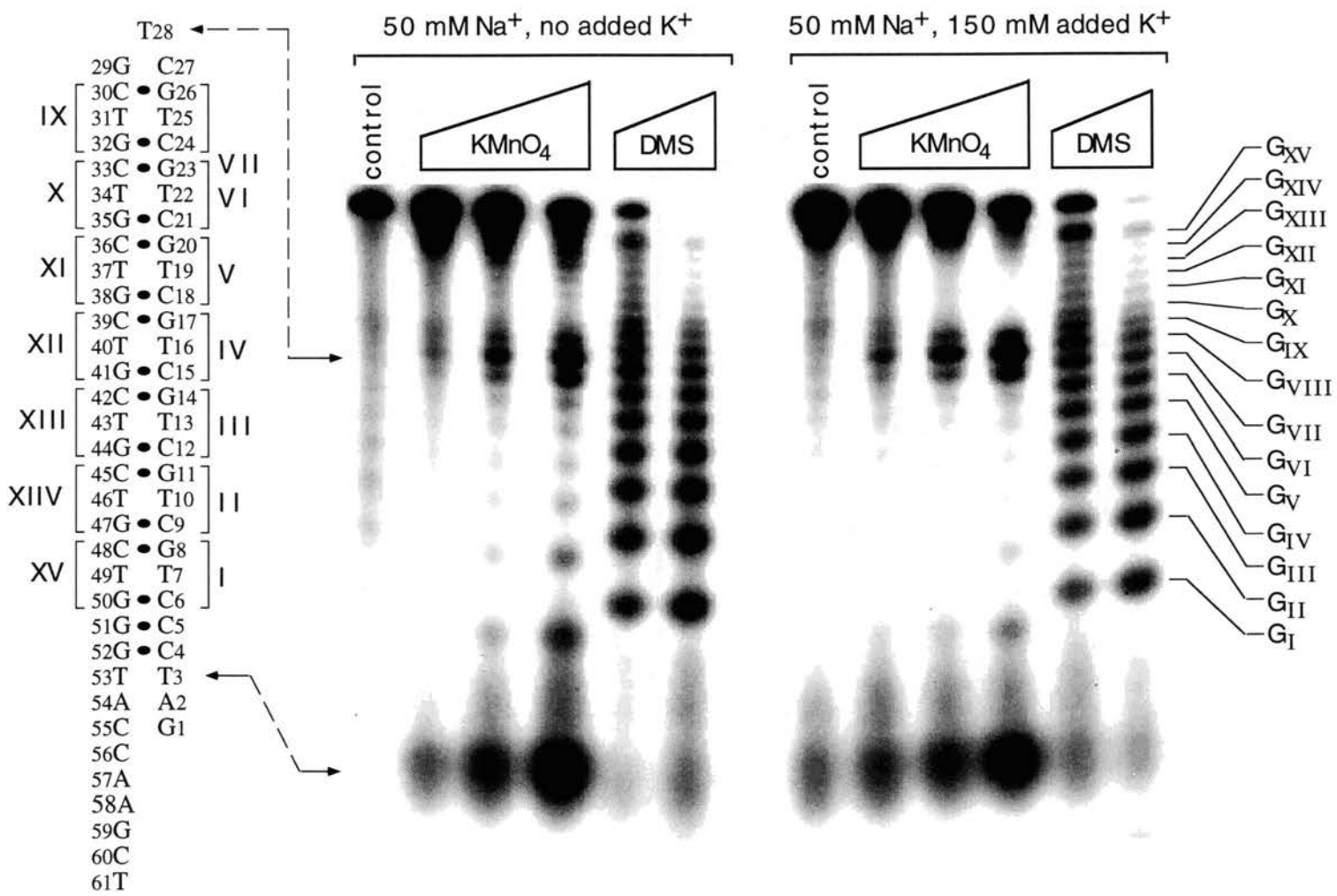

T28
29G C27
IX
30C • G26
31T T25
32G • C24
VII
X
33C • G23
34T T22
35G • C21
VI
XI
36C • G20
37T T19
38G • C18
V
XII
39C • G17
40T T16
41G • C15
IV
XIII
42C • G14
43T T13
44G • C12
III
XIIV
45C • G11
46T T10
47G • C9
II
XV
48C • G8
49T T7
50G • C6
I
51G • C5
52G • C4
53T T3
54A A2
55C G1
56C
57A
58A
59G
60C
61T
50 mM Na+, no added K+
control
KMnO4
DMS
50 mM Na+, 150 mM added K+
control
KMnO4
DMS
GXV
GXIV
GXIII
GXII
GXI
GX
GIX
GVIII
GVII
GVI
GV
GIV
GIII
GII
GI

Figure 13. $KMnO_4$ oxidation of ss $(CTG)_{15}$, at 60 °C.

$KMnO_4$ oxidations were performed with ss $(CTG)_{15}$ as described in Materials and Methods at 60 °C in 50 mM Na^+, no K^+ (left), and in 50 mM Na^+, 150 mM K^+ (right) . From left to right the amounts of $KMnO_4$ in the reaction mixtures were 0, 0.13, 0.25 and 0.75 mM. See the legend to Figure 11 for other details.

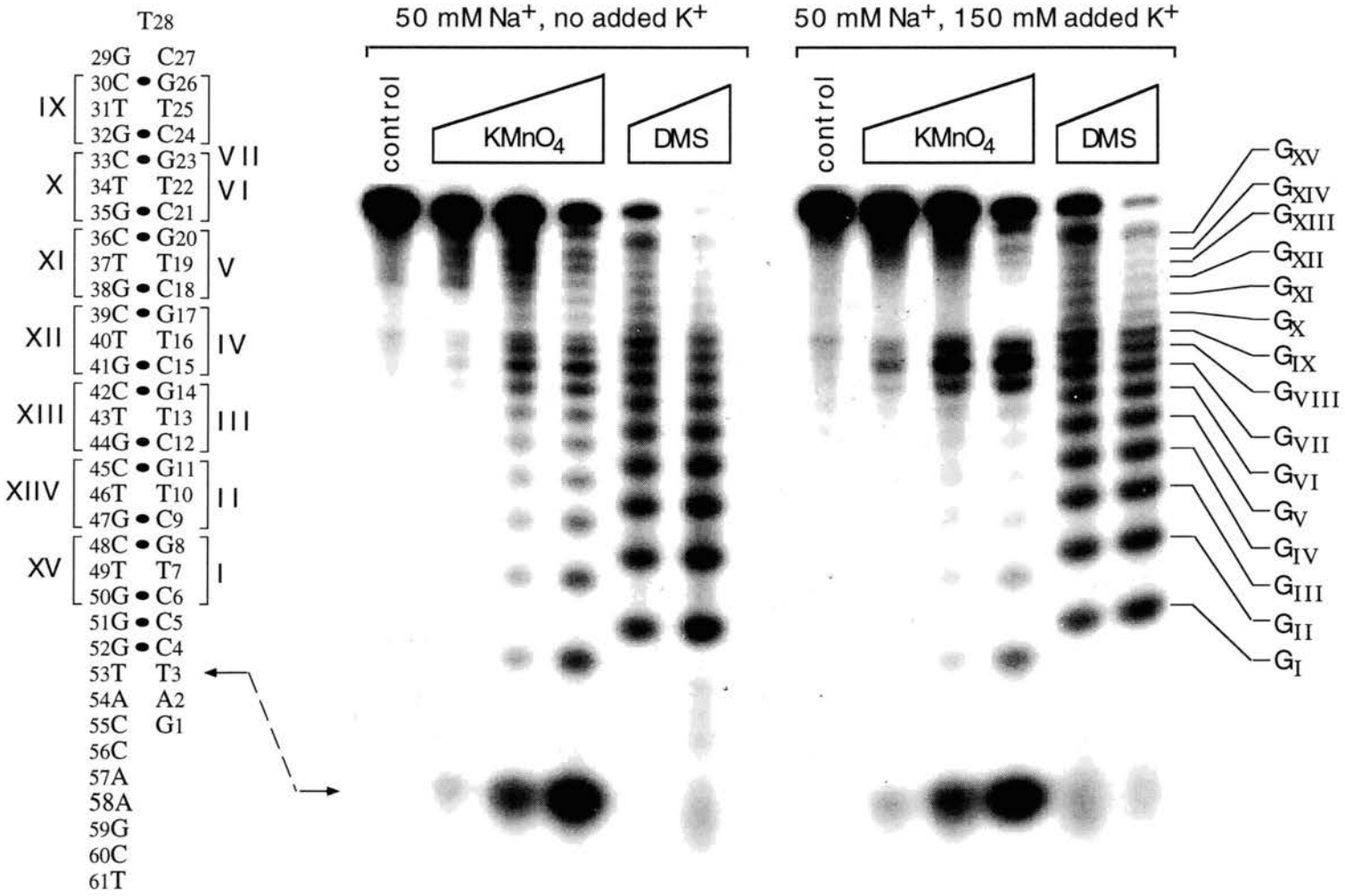

T28
29G C27
IX
30C G26
31T T25
32G C24
VII
X
33C G23
34T T22
35G C21
VI
XI
36C G20
37T T19
38G C18
V
XII
39C G17
40T T16
41G C15
IV
XIII
42C G14
43T T13
44G C12
III
XIIV
45C G11
46T T10
47G C9
II
XV
48C G8
49T T7
50G C6
I
51G C5
52G C4
53T T3
54A A2
55C G1
56C
57A
58A
59G
60C
61T
50 mM Na+, no added K+
control
KMnO4
DMS
50 mM Na+, 150 mM added K+
control
KMnO4
DMS
GXV
GXIV
GXIII
GXII
GXI
GX
GIX
GVIII
GVII
GVI
GV
GIV
GIII
GII
GI

Figure 14. $KMnO_4$ oxidation of ss $(CTG)_{15}$, at 70 ^oC.

$KMnO_4$ oxidations were performed with ss$(CTG)_{15}$ as described in Materials and Methods at 70 ^oC in 50 mM Na^+, no K^+ (left), and in 50 mM Na^+, 150 mM K^+ (right) . From left to right the amounts of $KMnO_4$ in the reaction mixtures were 0, 0.13, 0.25 and 0.75 mM. See the legend to Figure 11 for other details.

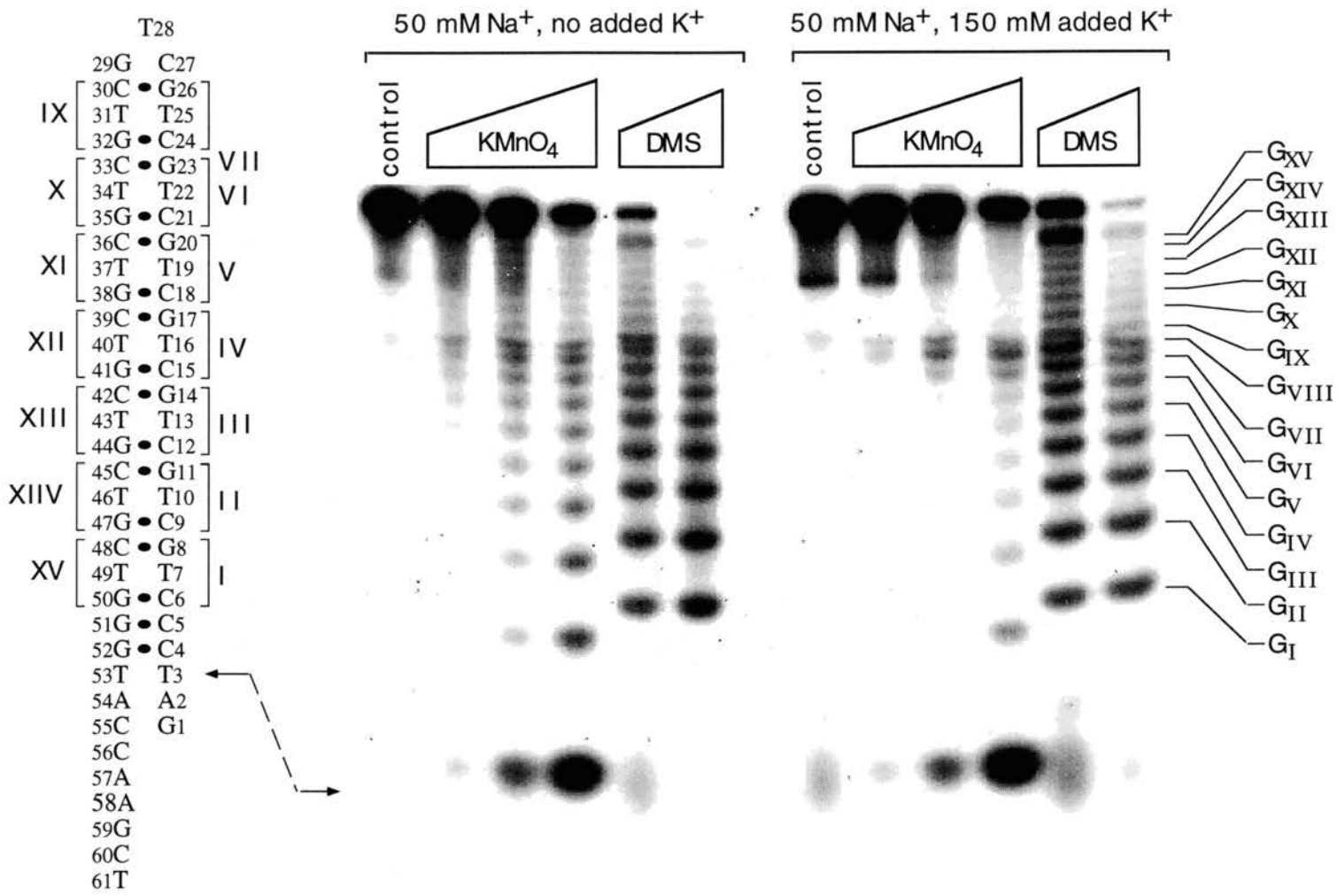
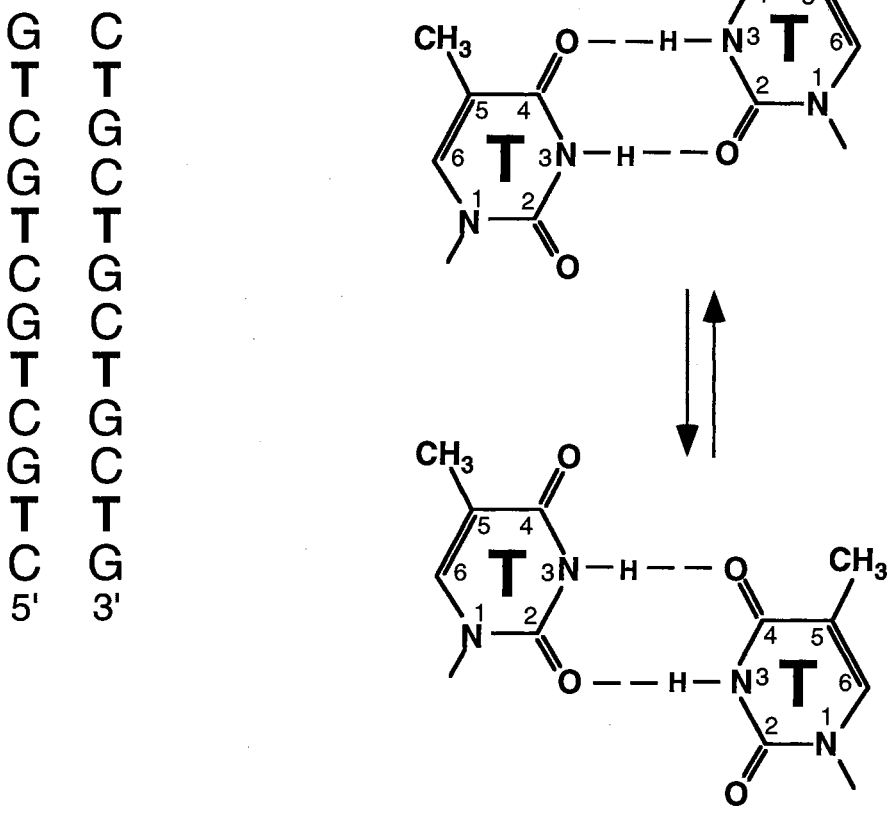

T28
29G C27
IX 30C • G26
31T T25
32G • C24
VII
X 33C • G23
VI
34T T22
35G • C21
XI 36C • G20
V
37T T19
38G • C18
XII 39C • G17
IV
40T T16
41G • C15
XIII 42C • G14
III
43T T13
44G • C12
XIIV 45C • G11
II
46T T10
47G • C9
XV 48C • G8
I
49T T7
50G • C6
51G • C5
52G • C4
53T T3
54A A2
55C G1
56C
57A
58A
59G
60C
61T
50 mM Na+, no added K+
control
KMnO4
DMS
50 mM Na+, 150 mM added K+
control
KMnO4
DMS
GXV
GXIV
GXIII
GXII
GXI
GX
GIX
GVIII
GVII
GVI
GV
GIV
GIII
GII
GI

Figure 15. T•T bases pair by hydrogen bonds.

Dash line indicates hydrogen bond. A structure containing T-T mismatch as determined from Marriapin et al., (1996) and Smith et al., (1996) is shown to the left in the figure.


G T C G T C G T C G T C 5'
C T G C T G C T G C T G 3'
CH3
O
H
N3
T
N
1
2
4
5
6

CHAPTER V

SINGLE-STRANDED d(GTC)$_{15}$ FORMS A HAIRPIN THAT IS LESS STABLE THAN THE d(CTG)$_{15}$ HAIRPIN

(GTC)n/(GAC)n triplet repeats belong to class I triplet repeat sequences which are GC rich and contain a GC- or CG- palindrome (Mitas et al., 95a). The other members of this category, (CCG)n/(CGG)n and (CTG)n/(CAG)n, are associated with 9 of the 10 known TREDs. All the single strand repeat sequences of this category have the potential to form hairpin structures *in vivo* (Mitas et al., 1995a).

That (GTC)n/(GAC)n sequences do not associate with TREDs and have rare frequencies in gene bank databases (Mitas et al., 1995a) may result from unique structural features. Preliminary energy minimization calculations suggested that hairpins formed from GTC triplet repeats were not as stable as hairpins formed from CTG repeats (Mitas et al., 1995a).

Here we investigate the structure of ss (GTC)$_{15}$ and compare it to that formed by ss (CTG)$_{15}$. Chemical modification of normal and mutated ss (GTC)$_{15}$ repeat sequences indicated that the T-T mismatches in the hairpin stem of (GTC)$_{15}$, like that in the hairpin stem of (CTG)$_{15}$, are base stacked and/or base paired

ss $(GTC)_{15}$ exhibits a concentration-independent, intramolecular structure.

To investigate whether ss $(GTC)_{15}$ exhibited properties of an intramolecular hairpin or some type of intermolecular structure, the molecular composition of the structure(s) formed with ss$(GTC)_{15}$ was examined by performing electrophoretic studies. The assays were similar to the previous studies on ss $(CTG)_{15}$. Electrophoretic analysis of labeled ss $(GTC)_{15}$ or ss $(GAC)_{15}$ mixed with various concentrations of unlabeled $(GTC)_{15}$, were performed. Both experimental results demonstrated that ss $(GTC)_{15}$ exhibited a concentration-independent, intramolecular structure (Yu, et al., 1995a).

$KMnO_4$ oxidizes T3 and T_{VIII} of ss$(GTC)_{15}$

As mentioned in Chapter IV, $KMnO_4$ preferentially oxidizes unpaired or unstacked thymines, resulting in strand cleavage upon subsequent treatment with piperidine. Treatment of ss$(GTC)_{15}$ with $KMnO_4$/piperidine in 50 mM Na^+, pH 7.0, at 30°C resulted in major cleavage products at T_3 and T_{VIII} (Figure 16), a result consistent with a hairpin structure of ss $(GTC)_{15}$. Minor cleavage products were also observed at T_{VII} and T_{IX}. The results suggest that ss$(GTC)_{15}$, like ss$(CTG)_{15}$, forms a hairpin containing base paired and/or base stacked thymines in the stem.

To further investigate the hairpin structure of $(GTC)_{15}$, $KMnO_4$ oxidation experiments were performed with a ss oligonucleotide of the same sequence, but lacking a single thymine within triplet II (named $\Delta T_{II}(GTC)_{15}$). Due to the deletion in $\Delta T_{II}(GTC)_{15}$, the thymine in triplet XIV does not have a nucleotide with which it can pair. Therefore, if $\Delta T_{II}(GTC)_{15}$ formed the hairpin as predicted in Figure 16, the thymine in triplet XIV, as well as the thymine in

triplet VIII (the loop apex), should be sensitive to $KMnO_4$ oxidation. The $KMnO_4$ oxidation results (Figure 17) are in agreement with those expected; the thymine in triplets VIII and XIV were preferentially oxidized, providing further evidence for a hairpin conformation.

$KMnO_4$ oxidation at various temperatures reveals reduced thermal stability of the $(GTC)_{15}$ hairpin

To investigate thermal stability of the ss $(GTC)_{15}$ hairpin, $KMnO_4$ oxidation experiments were performed in 50 mM Na^+ with and without 150 mM KCl at 40, 50, 60 and 70°C. In the presence or absence of KCl, the thymines in the stems of ss$(GTC)_{15}$ were not oxidized by $KMnO_4$ at 40°C (Figure 18). In the absence of added KCl, significant oxidation of the thymines in the presumed stem region of ss$(GTC)_{15}$ occurred at 50°C (Figure 19). In the presence of KCl, significant oxidation required an incubation temperature of 60-70°C (Figure 20 and 21). Stabilization of the T-T interactions by added KCl is consistent with a hairpin structure for ss$(GTC)_{15}$.

P1 nuclease digestion of ss$(GTC)_{15}$

P1 nuclease reactions was performed with ^{32}P labeled $(GTC)_{15}$ to further examine its structure. To generate size markers, the DNAs were separately reacted with dimethyl sulfate and $KMnO_4$ and cleaved with piperidine. In comparing the size of the P1 products to the markers, it should be mentioned that piperidine completely removes the 5' base oxidized by $KMnO_4$ or methylated by DMS. Hence, the signals in the marker lanes do not contain a thymine or guanine respectively at the 5'-terminus.

Incubation of ss$(GTC)_{15}$ with P1 nuclease did not result in cleavage of the phosphodiester bonds between C_{VIII} and C55, indicating that the nucleotides participated in base-pairing and/or base stacking interaction (Figure 22). At the 3'-terminus the C55- C56 phosphodiester bond was extensively degraded. In the vicinity of the loop the major product of P1 nuclease digestion co-migrated with the $KMnO_4$ product generated from cleavage of T_{VII}. Therefore, the major site of P1 cleavage within the triplet repeat region was the T_{VII}-C_{VII} phosphodiester bond. Minor sites of P1 cleavage in the loop region included the C_{VII}-G_{VIII}, G_{VIII}-T_{VIII}, and T_{VIII}-C_{VIII} phosphodiester bonds. Surprisingly, the T_{VIII}-C_{VIII} phosphodiester bond was not the major site of P1 cleavage. This result is surprising since T_{VIII} was readily oxidized by $KMnO_4$. The structure of ss$(GTC)_{15}$ most consistent with these results is shown in Figure 22, a hairpin structure, of which triplet I is base paired to triplet XV, triplet II to triplet XIV, etc.

Unexpected results were that the T53-A54, and A54-C55 phosphodiester bonds of ss$(GTC)_{15}$ were not cleaved by P1 nuclease, but as shown in the experiment above, T3 was readily oxidized by $KMnO_4$. In the hairpin structure of ss$(GTC)_{15}$ T53 is opposite T3. Hence, since T3 is not base paired to T53, the phosphodiester bonds should be cleaved by P1. The inability of P1 to cleave the phosphodiester bonds of T53 suggests that base stacking interactions and not base pairing interactions inhibit cleavage. The base stacking interaction may arise from neighboring C·G base pairs that form between C4-G52 and G1-C55 (as shown in Figure 22).

The results of nuclease cleavage of ss$(GTC)_{15}$ are consistent with hairpin structures and suggest extensive base pairing and/or base stacking interactions between the T•T mismatches.

ss$(GTC)_{15}$ has the same base compositions as ss$(CTG)_{15}$, but with the opposite orientation of the phosphodiester bonds. The results presented in this chapter demonstrate that ss$(GTC)_{15}$ forms a hairpin structure that is stable at physiologic salt and temperature. To futher compare the stabilities of the two hairpin conformations, electrophoretic mobility melting profiles of ss$(GTC)_{15}$ and ss$(CTG)_{15}$ were performed in our laboratory (Yu et al., 1995a). The results were in agreement with the hairpin conformations of both ss$(GTC)_{15}$ and ss$(CTG)_{15}$ and were consistent with the results from the $KMnO_4$ oxidation reactions.

Discussion

In 50 mM Na^+, 150 mM K^+, significant oxidation of the thymines in the stem of ss$(GTC)_{15}$ required an incubation temperature of 60-70°C (Figs 20 and 21). These results indicate that under physiological conditions, the T•T mismatches in a ss $(GTC)_{15}$ hairpin are extensively stacked. Aside from differences in temperature dependencies, the reactivities of $KMnO_4$ oxidation of the thymines in ss$(GTC)_{15}$ were similar to those of ss$(CTG)_{15}$ (Fig 11-14). Since ^1H NMR studies have shown that the thymines in the homoduplex d$(CTG/CTG)_3$ are base paired (Gao et al., 1995, Gupta et al., 1996), we suspect that the thymines in ss$(GTC)_{15}$ are also base paired, presumably through hydrogen bonds formed between N3 and O4', O2' and N3 as show in Figure 15.

(GTC)n/(GAC)n is unique from (CTG)n/(CAG)n and (CGG)n/(CCG)n. In both of the latter, one of the complementary strands, ss (CAG)n and ss (CCG)n, forms a hairpin much less stable than the hairpin of its complementary strand (Mitas et al, 1995b; A. Yu et al, 1995a). Expansions of (CTG)n/(CAG)n and (CGG)n/(CCG)n sequences are associated with ten TREDs, however, no

TREDs are associated with expansion of (GTC)n/(GAC)n sequences (Mitas et al., 1995a, Sutherland et al., 1995). Sequence search in Gen Bank found that (GTC)n/(GAC)n was present at a much lower frequency in relevant genes (Mitas et al, 95a; Stallings, 1994), indicating the existence of selections against (GTC)n/(GAC)n in evolution which may relate to their tendency to form stable hairpin or other secondary structures during replication. These structures may interfere with polymerase processing in DNA replication, resulting in severe deficiency in the individuals owning the loci.

Particular triplet repeat regions in DNA sequences have the potential to undergo expansion during replication. This phenomenon has been revealed recently by both *in vivo* and *in vitro* assays (Ohshima et al., 1996, Kang et al., 1996). Ohshima et al., reported that large copy numbers of repeat $(CTG)_{140}$ were expanded to $(CTG)_{170}$, and a low level of GTC expansion occurred when the integrated plasmid replicated in *E. coli.*. Other experimental results in our laboratory demonstrated that $(GTC)_{15}/(GAC)_{15}$ sequences preferentially undergo mutation (expansion and deletion) following plasmid replication in *E. coli.* A model that is able to accurately coordinate with the expansion mechanism needs to be established and proved by experimental evidence. Investigating the properties of replication of the triplet repeat sequences with known structures are significant for exploring the expansion mechanism.

Figure 16. $KMnO_4$ oxidation of ss $(GTC)_{15}$, at 30 °C.

$KMnO_4$ oxidations were performed with ss $(GTC)_{15}$ as described in Materials and Methods, at 30 °C in 50 mM Na^+, no K^+ (left), and with 50 mM Na^+, 150 mM K^+ (right) . From left to right the amounts of $KMnO_4$ in the reaction mixtures were 0, 0.13, 0.25 and 0.75 mM. The marker lane contained DNAs reacted with DMS (10.5 and 21 mM of DMS from left to right). The positions of the G residues in the triplet repeats are indicated to the right in the figure. The deduced hairpin structure of the DNA sequences is shown to the left in the figures. Roman numerals depict a particular triplet repeat. Conventional arabic numerals depict the position of a nucleotide with respect to the 5'-end. $KMnO_4$ oxidation sites of particular thymine residues are indicated by arrows.

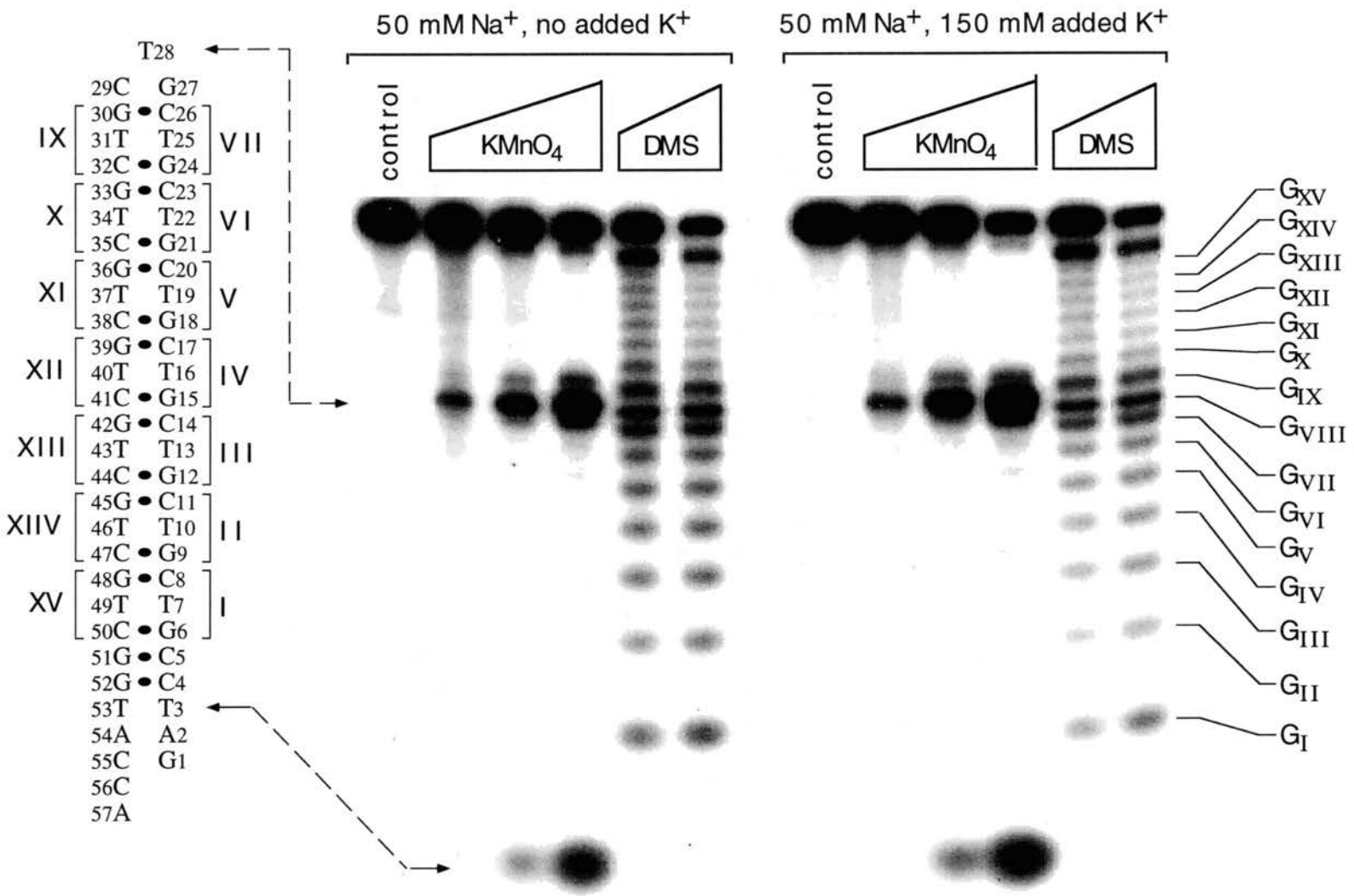

50 mM Na+, no added K+
50 mM Na+, 150 mM added K+
control
KMnO4
DMS
control
KMnO4
DMS
T28
29C G27
IX 30G • C26 VII
31T T25
32C • G24
X 33G • C23 VI
34T T22
35C • G21
XI 36G • C20 V
37T T19
38C • G18
XII 39G • C17 IV
40T T16
41C • G15
XIII 42G • C14 III
43T T13
44C • G12
XIIV 45G • C11 II
46T T10
47C • G9
XV 48G • C8 I
49T T7
50C • G6
51G • C5
52G • C4
53T T3
54A A2
55C G1
56C
57A
GXV
GXIV
GXIII
GXII
GXI
GX
GIX
GVIII
GVII
GVI
GV
GIV
GIII
GII
GI

Figure 17. $KMnO_4$ oxidation of ss ΔT_{II} $(GTC)_{15}$ at 30 °C.

$KMnO_4$ oxidations were performed with ss ΔT_{II} $(GTC)_{15}$ as described in Materials and Methods, at 30 °C in 50 mM Na^+ with ss ΔT_{II} $(GTC)_{15}$. From left to right the amounts of $KMnO_4$ in the reaction mixtures were 0, 0.13, 0.25 and 0.75 mM. The marker lane contained DNAs reacted with DMS (10.5 and 21 mM of DMS from left to right). The positions of the G residues in the triplet repeats are indicated to the left in the figure. The deduced hairpin structure of the DNA sequence is shown to the right in the figure. Roman numerals depict a particular triplet repeat. Conventional arabic numerals depict the position of a nucleotide with respect to the 5'-end. $KMnO_4$ oxidation sites of particular thymine residues are indicated by arrows.

control
DMS
KMnO4
GXV
GXIV
GXIII
GXII
GXI
GX
GIX
GVIII
GVII
GVI
GV
GIV
GIII
GII
GI
T27
28C G26
IX 29G • C25 VII
30T T24
31C • G23
X 32G • C22 VI
33T T21
34C • G20
XI 35G • C19 V
36T T18
37C • G17
XII 38G • C16 IV
39T T15
40C • G14
XIII 41G • C13 III
42T T12
43C • G11
XIIV T45 44G • C10 II
46C • G9
XV 47G • C8 I
48T T7
49C • G6
50G • C5
51G • C4
52T T3
53A A2
54C G1
55C
56A

Figure 18. $KMnO_4$ oxidation of $ss(GTC)_{15}$, at 40 ^oC.

$KMnO_4$ oxidations were performed with $ss(GTC)_{15}$ as described in Materials and Methods at 40 ^oC in 50 mM Na^+, no K^+ (left), and in 50 mM Na^+, 150 mM K^+ (right) . From left to right the amounts of $KMnO_4$ in the reaction mixtures were 0, 0.13, 0.25 and 0.75 mM. The marker lane contained DNAs reacted with DMS (10.5 and 21 mM of DMS from left to right). The positions of the G residues in the triplet repeats are indicated to the right in the figures. The deduced hairpin structure of the DNA sequences is shown to the left in the figures. Roman numerals depict a particular triplet repeat. Conventional arabic numerals depict the position of a nucleotide with respect to the 5'-end. $KMnO_4$ oxidation sites of particular thymine residues are indicated by arrows.

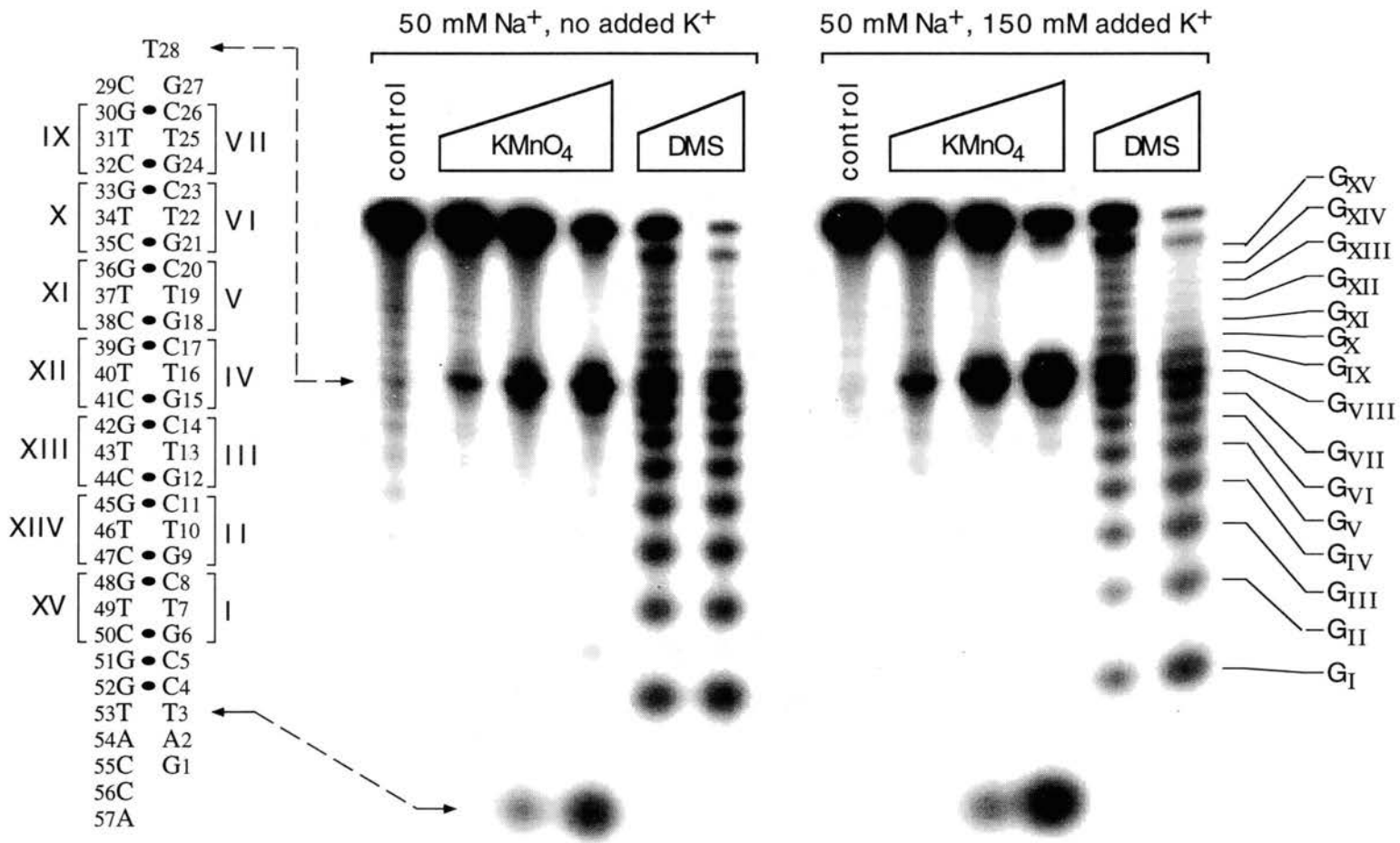

50 mM Na+, no added K+
50 mM Na+, 150 mM added K+
control
KMnO4
DMS
control
KMnO4
DMS
T28
29C G27
IX 30G • C26 VII
31T T25
32C • G24
X 33G • C23
34T T22 VI
35C • G21
XI 36G • C20
37T T19 V
38C • G18
XII 39G • C17
40T T16 IV
41C • G15
XIII 42G • C14
43T T13 III
44C • G12
XIIV 45G • C11
46T T10 II
47C • G9
XV 48G • C8
49T T7 I
50C • G6
51G • C5
52G • C4
53T T3
54A A2
55C G1
56C
57A
GXV
GXIV
GXIII
GXII
GXI
GX
GIX
GVIII
GVII
GVI
GV
GIV
GIII
GII
GI

Figure 19. $KMnO_4$ oxidation of ss$(GTC)_{15}$ at 50 ^oC.

$KMnO_4$ oxidations were performed with ss$(GTC)_{15}$ as described in Materials and Methods at 50 ^oC in 50 mM Na^+, no K^+ (left), and in 50 mM Na^+, 150 mM K^+ (right). From left to right the amounts of $KMnO_4$ in the reaction mixtures were 0, 0.13, 0.25 and 0.75 mM. See the legend to Figure 18 for other details.

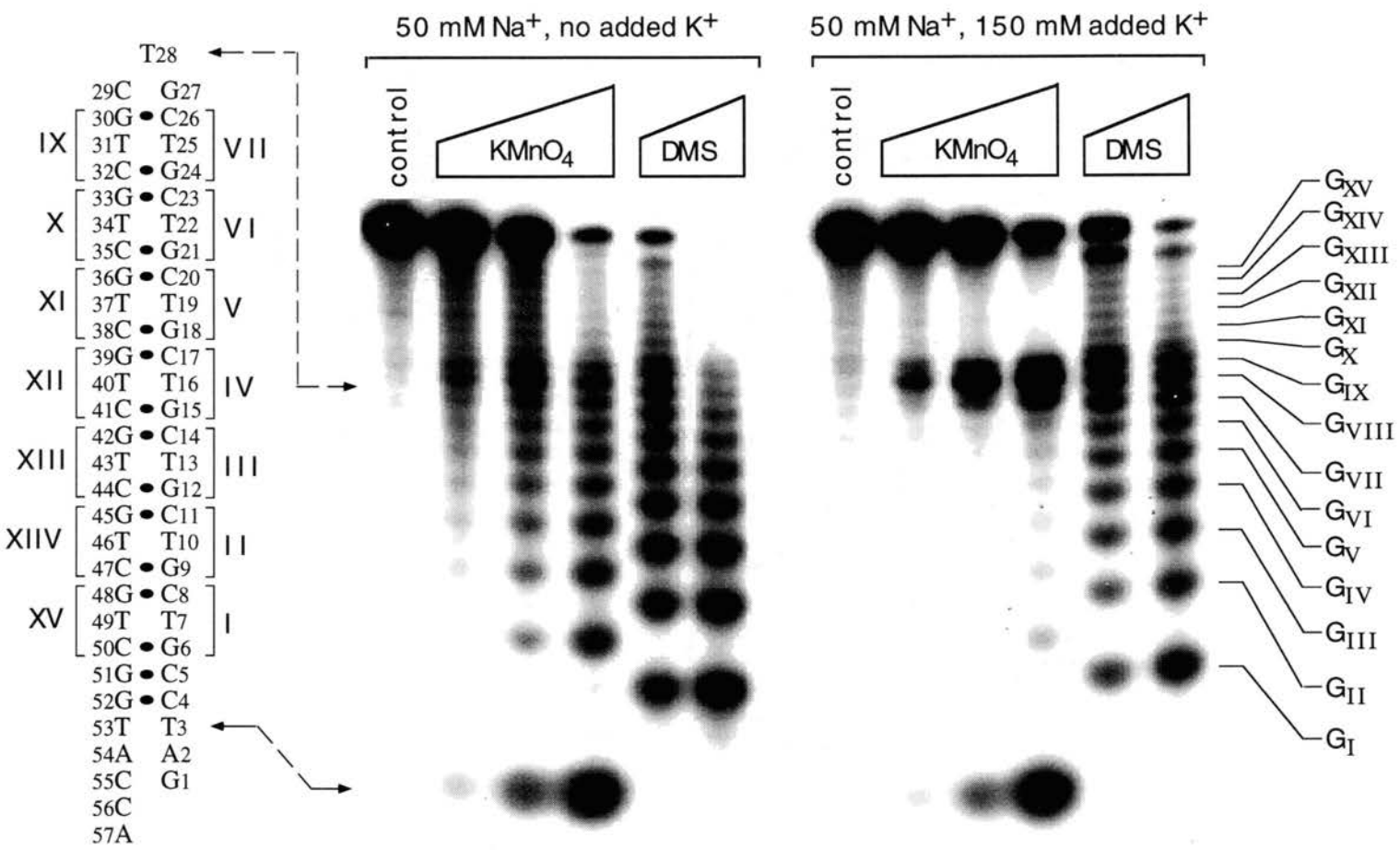

50 mM Na+, no added K+
50 mM Na+, 150 mM added K+
control
KMnO4
DMS
control
KMnO4
DMS
T28
29C G27
30G • C26
31T T25
32C • G24
33G • C23
34T T22
35C • G21
36G • C20
37T T19
38C • G18
39G • C17
40T T16
41C • G15
42G • C14
43T T13
44C • G12
45G • C11
46T T10
47C • G9
48G • C8
49T T7
50C • G6
51G • C5
52G • C4
53T T3
54A A2
55C G1
56C
57A
IX
X
XI
XII
XIII
XIIV
XV
VII
VI
V
IV
III
II
I
GXV
GXIV
GXIII
GXII
GXI
GX
GIX
GVIII
GVII
GVI
GV
GIV
GIII
GII
GI

Figure 20. $KMnO_4$ oxidation of $ss(GTC)_{15}$ at 60 ^oC.

$KMnO_4$ oxidations were performed with $ss(GTC)_{15}$ as described in Materials and Methods at 60 ^oC in 50 mM Na^+, no K^+ (left), and in 50 mM Na^+, 150 mM K^+ (right). From left to right the amounts of $KMnO_4$ in the reaction mixtures were 0, 0.13, 0.25 and 0.75 mM. See the legend to Figure 18 for other details.

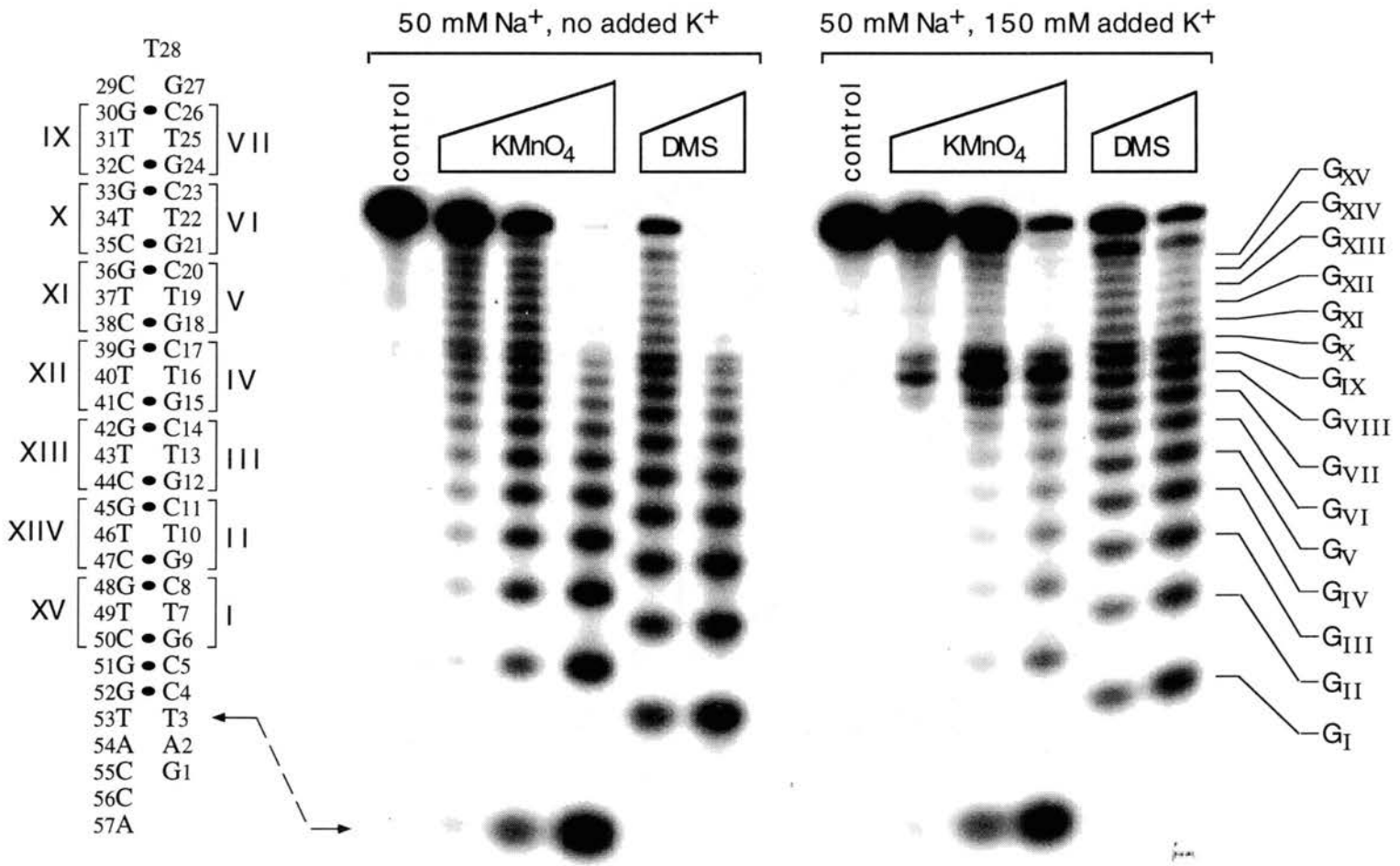

T28
29C G27
IX 30G • C26 VII
31T T25
32C • G24
X 33G • C23 VI
34T T22
35C • G21
XI 36G • C20 V
37T T19
38C • G18
XII 39G • C17 IV
40T T16
41C • G15
XIII 42G • C14 III
43T T13
44C • G12
XIIV 45G • C11 II
46T T10
47C • G9
XV 48G • C8 I
49T T7
50C • G6
51G • C5
52G • C4
53T T3
54A A2
55C G1
56C
57A
50 mM Na+, no added K+
control
KMnO4
DMS
50 mM Na+, 150 mM added K+
control
KMnO4
DMS
GXV
GXIV
GXIII
GXII
GXI
GX
GIX
GVIII
GVII
GVI
GV
GIV
GIII
GII
GI

Figure 21. $KMnO_4$ oxidation of ss$(GTC)_{15}$ at 70 ^oC.

$KMnO_4$ oxidations were performed with ss$(GTC)_{15}$ as described in Materials and Methods at 70 ^oC in 50 mM Na^+, no K^+ (left), and in 50 mM Na^+, 150 mM K^+ (right). From left to right the amounts of $KMnO_4$ in the reaction mixtures were 0, 0.13, 0.25 and 0.75 mM. See the legend to Figure 18 for other details.

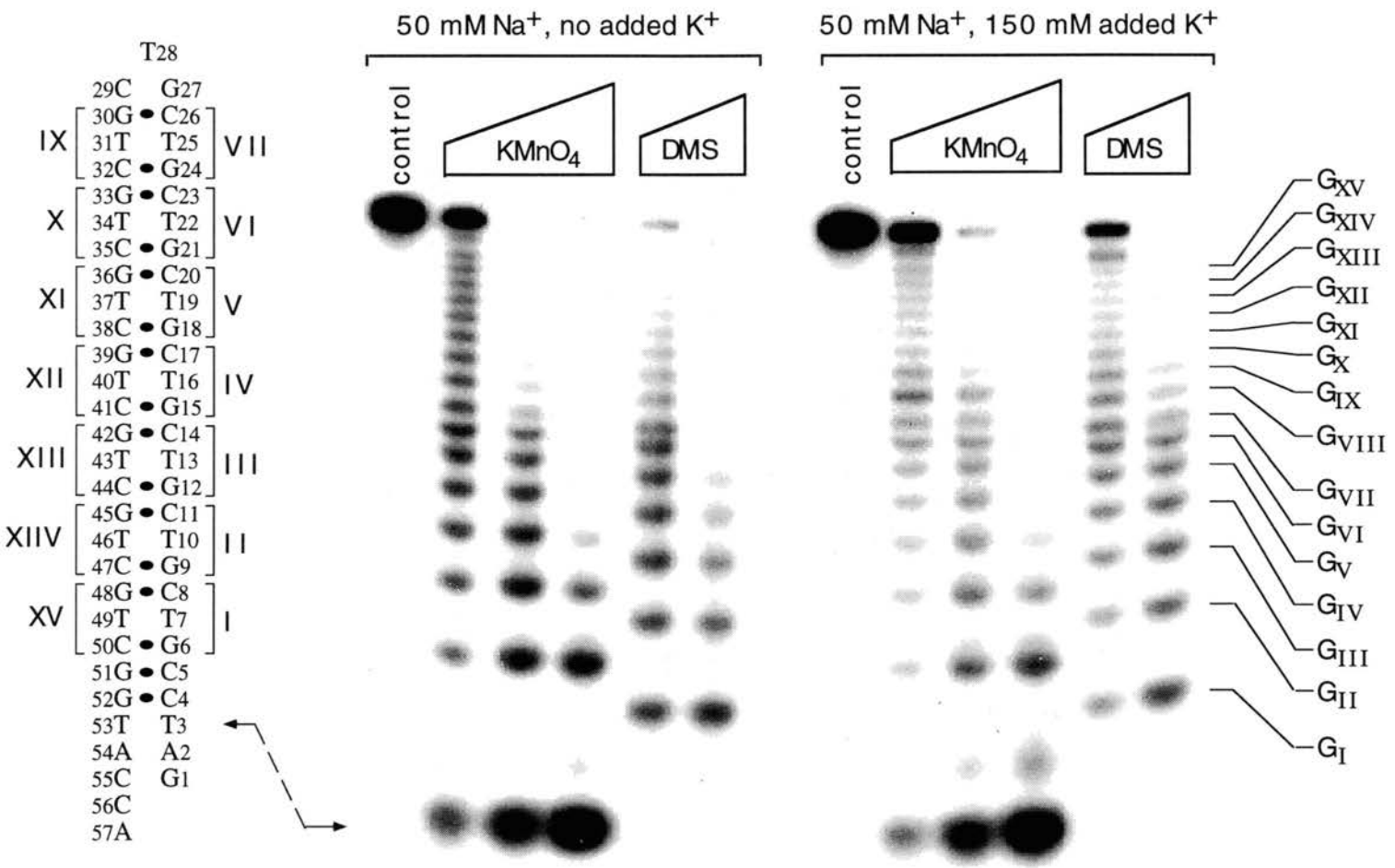
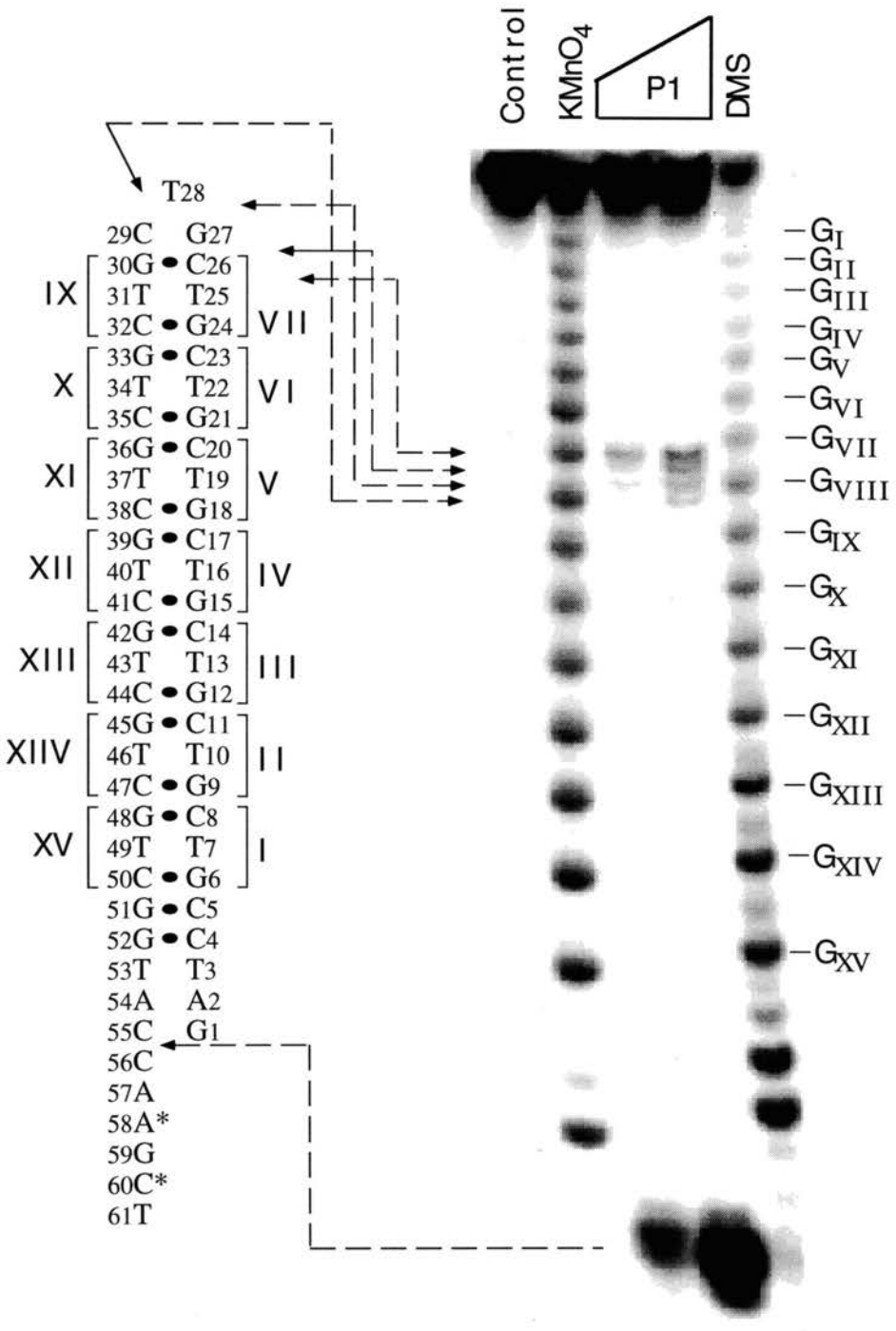

T28
29C G27
IX 30G • C26 VII
31T T25
32C • G24
X 33G • C23 VI
34T T22
35C • G21
XI 36G • C20 V
37T T19
38C • G18
XII 39G • C17 IV
40T T16
41C • G15
XIII 42G • C14 III
43T T13
44C • G12
XIIV 45G • C11 II
46T T10
47C • G9
XV 48G • C8 I
49T T7
50C • G6
51G • C5
52G • C4
53T T3
54A A2
55C G1
56C
57A
50 mM Na+, no added K+
control
KMnO4
DMS
50 mM Na+, 150 mM added K+
control
KMnO4
DMS
GXV
GXIV
GXIII
GXII
GXI
GX
GIX
GVIII
GVII
GVI
GV
GIV
GIII
GII
GI

Figure 22. P1 nuclease digestion of ss$(GTC)_{15}$.

The oligonucleotide containing $(GTC)_{15}$ purified from pGTC15 was labeled on the GTC-containing strand only with Klenow enzyme and ^{32}P. One micromolar unlabeled synthetic oligonucleotide of the same sequence as the labeled strand was added to the oligonucleotide, placed in a 100 ^{o}C water bath for 3 min and immediately chilled on ice. The amounts of P1 nuclease used to digest ss$(GTC)_{15}$ (from left to right) at 37 ^{o}C in 50 mM Na^{+} were 1.15×10^{-2} and 3.46×10^{-2} U respectively. $KMnO_4$ (250 μM) oxidation of ss$(GTC)_{15}$ was performed as described in Materials and Methods at 50^{o}C. DMS (21mM) reactions were performed as described in Materials and Methods. Roman numerals represent triplet repeat numbers. The nucleotides labeled with ^{32}P are marked with an asterisk. Arrows indicate sites of P1 nuclease cleavage.


Control
KMnO4
P1
DMS
T28
29C G27
30G C26
31T T25
32C G24
33G C23
34T T22
35C G21
36G C20
37T T19
38C G18
39G C17
40T T16
41C G15
42G C14
43T T13
44C G12
45G C11
46T T10
47C G9
48G C8
49T T7
50C G6
51G C5
52G C4
53T T3
54A A2
55C G1
56C
57A
58A*
59G
60C*
61T
IX
X
XI
XII
XIII
XIIV
XV
VII
VI
V
IV
III
II
I
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV

CHAPTER VI

SINGLE-STRANDED $(CGG)_{15}$ FORMS A HEAT-STABLE HAIRPIN THAT CONTAINS G^{syn}•G^{anti} BASE PAIRS

Repeats of CGG are present in many genes (Mitas 1995c; Riggins et al., 1992, Gacy et al., 1995) and are often located in the 5' untranslated regions. These repeats may form hairpins and play a role in gene regulation (Mitas et al., 1995b). Expansions of CGG/CCG sequences were found to be associated with the appearence of fragile sites on the expanded loci of chromosomes (*FRAXA, FRAXE, FRAXF,* and *FRA16A*). Fragile sites are non staining gaps in metaphase chromosomes caused by non condensed chromatin or, more rarely, chromosomal breaks (A. Weiner, personal commun.). Fragile sites frequently colocalize with recurrent cancer breakpoints (Miro et al., 1987), with preferential targets for viral integration (Popescu et al., 1990), and with sites of recent chromosome translocations in primate lineages (Miro et al., 1987). The best characterized fragile sites are induced by chemicals, drugs, or chromosomal alterations which interfere directly with DNA replication, and only indirectly with subsequent chromatin packing (Li and Weiner, 1996). The fragile sites caused by (CGG)n/(CCG)n triplet repeat expansion are characteristic of folic acid deprivation and accompanying hypermethylation of the *FMR1* promoter and of the amplified trinucleotide tract (Bell *et al.*, 1991, Vincent *et al.*, 1991; Pierreti *et al.*, 1991; Luo *et al.*, 1991). Subsequent to d(CGG) expansion and hypermethylation, the *FMR1* gene becomes transcriptionally silent, (Pierreti *et*

al., 1991, Hansen *et al.*, 1992, Sutcliffe *et al.*, 1992) and the replication of a chromosomal segment spanning ≥ 150 kb 5′ and < 34 kb 3′ from the d(CGG)n stretch is delayed (Hansen *et al.*, 1993).

The molecular mechanisms that govern (CGG)n amplification and hypermethylation and that link d(CGG) expansion to the transcriptional suppression of *FMR1* and to its delayed replication, are not known. To investigate the possibility of a hairpin structure, we conducted studies of a DNA sequence containing ss $(CGG)_{15}$. The results provide evidence that ss$(CGG)_{15}$ forms an extremely heat-stable "slipped" hairpin containing G^{syn}•G^{anti} base pairs.

ss $(CGG)_{15}$ exhibits a concentration-independent, intramolecular structure

To investigate the structure of ss $(CGG)_{15}$, electrophoretic studies were performed in our laboratory (Mitas et al., 1995b). The assays were similar to the previous studies on that of ss $(CTG)_{15}$. Electrophoretic analysis of labeled ss $(CGG)_{15}$ was performed in the presence of various amounts of an unlabeled DNA of the same sequence. The results demonstrated that ss $(CGG)_{15}$ exhibited a concentration-independent, intramolecular structure (Mitas et al., 1995b).

Dimethyl sulfate reactions with ss $(CGG)_{15}$

Computer modeling performed by I. Haworth at USC indicated that ss $(CGG)_{15}$ had the potential to form hairpin structures with two alignments, *(a)* or *(b)*, and also to form tetraplex structures based on refolding the *(a)* or *(b)* hairpin structures (Mitas et al., 1995b). Figure 23 illustrates the four potential hairpin

structures and the two potential tetraplex structures. In theory, the structures should be distinguishable by differential reactivity with DMS.

Dimethyl sulfate (DMS) reacts specially with N^7 of guanine bases (Nancarroow et al., 1994). Base pairing and/or stacking will interfere with the modification. In normal Watson-Crick base pairs, N^1, $N^{2'}$ and $O^{6'}$ of guanine residues form hydrogen bonds with respectively N^3, $O^{2'}$ and $N^{4'}$ of cytosine residues of the complementary strand. N^7 atoms of guanine residues are not involved in Watson-Crick base pairing. If N^7 of guanine forms a hydrogen bond (ie. non-Watson-Crick base pairing), the guanine residue will not react with DMS. To investigate the possibility that $ss(CGG)_{15}$ formed a structure that contained non-Watson-Crick base pairs, DMS reactions were performed with $ss(CGG)_{15}$ at various concentrations of KCl. I were interested in investigating the structure of ss $(CGG)_{15}$ in various potassium concentrations because K^+ particularly stabilizes tetraplex structures by stabilizing G quartets (Hardin et al., 1992). Analysis of the DMS reactivities of the two Gs within a given triplet (the 5'G of a given triplet is herein designated $\underline{G}$, while the 3'G is designated G) should provide clues regarding which of the six intramolecular structures was adopted by $ss(CGG)_{15}$ (Figure 23). For example, if ss $(CGG)_{15}$ formed a tetraplex in alignment *(a)* (structure 5 in Figure 23), the DMS reactivity of the 5'$\underline{G}$ within a given triplet should approximate zero since the N7 of a G in a G quartet will not react with DMS. In contrast, the DMS reactivity of the 3'G within the same triplet should be some measurable quantity. Therefore, the 5'$\underline{G}$/3'G DMS reactivity ratio within a given triplet for structure 5 is $0/y = 0$. Structures 2 and 4-6 are characterized by unique 5'$\underline{G}$/3'G DMS reactivity ratios. The 5'$\underline{G}$/3'G DMS reactivity ratios of structure 1 and 3 are identical to one another, but different from the rest (Figure 23).

A hairpin structure observed at various concentrations of KCl

DMS reactions were performed with ss $(CGG)_{15}$ at 10 °C in 50 mM Na^+ and various concentrations of KCl (Figure 24). With no added KCl, the G residues in the middle of the triplet region (G31-37) were hypersensitive to DMS, indicating that this region formed a loop. DMS hypersensitivity was not observed at triplets IV or XII. These results are consistent with a hairpin structure of ss $(CGG)_{15}$. As the concentration of KCl was increased, the relative reactivities of the Gs in the loop region decreased, indicating that the loop region was significantly stabilized by salt.

The autoradiograph shown in figure 24 was scanned with a densitometer to determine the relative intersities of the adjacent G residues within a triplet repeat. At 0.75 M KCl, the reactivity of the 5'G, relative to that of the 3' G within a CGG triplet, was 1.99 ± 0.16. DMS reactivities similar to those in 0.75 M KCl were observed at 0.4 or 0.2 M KCl (Table 1). These results are inconsistent with a tetraplex structure of ss $(CGG)_{15}$ and suggest that , within the stem region, the 3'Gs of the CGG triplets formed $G^{syn} \bullet G^{anti}$ base pairs characteristic of a hairpin in the *(b)* alignment (structure 4 in Figure 23).

In the absence of added KCl, the 5'G/3'G ratio of DMS reactivities was 1.4 ± 0.13. These results suggest that only ~40% of the G-G mismatches were base-paired or, alternatively, that ss $(CGG)_{15}$ was in alignment *(b)* 70% of the time and alignment *(a)* 30% of the time.

The hairpin structure of ss $(CGG)_{15}$ is observed at temperatures between 1 and 37°C.

The DMS experiments described above were performed at 10°C. To determine the effect of temperature on the deduced hairpin structure of ss$(CGG)_{15}$, DMS experiments were performed with ss$(CGG)_{15}$ at 1, 5, 10, 23 and 37°C. in 50 mM Na^+ and 100 mM K^+. I were particularly interested in results at lower temperatures, since these conditions might favor tetraplex formation. Reactions performed at lower or higher temperatures yielded results similar to those reactions conducted at 10 °C (Figure 25), suggesting that the hairpin structure of ss $(CGG)_{15}$ was stable over a range of temperatures. The results did not support a tetraplex structure.

P1 nuclease cleaves the middle of the triplet repeat region in ss $(CGG)_{15}$.

To further characterize the secondary structure of ss $(CGG)_{15}$, digestions were performed with single-strand specific P1 nuclease in 50 mM Na^+ at pH 7.5 (Figure 26). The results show significant cleavage of the G32-C33 and G34-G35 phosphodiester bonds. Cleavage of phosphodiester bonds in triplets IV or XII was not observed. These results are consistent with a hairpin structure of ss $(CGG)_{15}$ and are not consistent with a tetraplex structure (Figure 23).

In the nontriplet repeat region of the DNA, only minor cleavages of the G58-A59 and A59-T60 phosphodiester bonds were observed. These results indicate extensive base-pairing and/or base-stacking interactions within the nontriplet repeat region. Due to the specific alignment of the hairpin structure, A59 and C61 can form base pairs with T8 and G6, respectively (as shown in Figure 26).

DMS cleavage of ss $(CGG)_{18}$ at pH 7.5

ss $(CGG)_{18}$ was labeled at 3′ end with α–^{32}P dATP and α–^{32}P dCTP by Klenow enzyme as described in Material and Methods. The purified probes were subjected to DMS modification; subsequent piperidine cleavage was as described in previous studies except that the probes were treated at 100°C, 1 min, then at 80°C for 2 min before adding DMS, instead of boiled for 3 min. After mixing with unlabeled ss $(CCG)_{18}$, the probes were reacted with various amounts of DMS. To investigate the structure of ss$(CGG)_{18}$, DMS modification was performed with ss $(CGG)_{15}$ at 37°C in 50 mM Na^+, pH 7.5. In each repeat unit, the 5′G shows higher sensitivity than the 3′ G to DMS and all G's show a certain degree of sensitivity to DMS (Figure 27). The results indicated that no tetraplex structure had formed. ss$(CGG)_{18}$ had a hairpin structure similar to the hairpin of ss$(CGG)_{15}$. The loop region was not particularly sensitive to DMS, which may indicate the hairpin loop of ss$(CGG)_{18}$ was more compact compared to that of ss$(CGG)_{15}$.

The Tm of the ss $(CGG)_{15}$ hairpin is 75°C

To study the stability of the ss $(CGG)_{15}$ hairpin, an electrophoretic mobility melting profile (EMMP), a technique developed in this laboratory, was performed (Mitas et al., 1995b). EMMPs, like temperature gradient gel electrophoresis (Ke & Wartell, 1993; Wartell et al., 1990), can be used to estimate the melting temperature (Tm) of a DNA by determination of the midpoint of its electrophoretic phase transition (Yu et al., 1995). The EMMP data of ss $(CGG)_{15}$ and ss $(CTG)_{15}$ were obtained and plotted. In agreement with previous data, the Tm of ss $(CTG)_{15}$ was 48°C. The EMMP assay and the annealing

experiments indicated a Tm of ss $(CGG)_{15}$ that was 75°C, 27°C higher than the Tm of ss $(CTG)_{15}$. The results are also consistent with the Tm reported by Gacy et al.. The results indicate that G•G base pairs contribute a significant amount of stability to the hairpin structure of ss $(CGG)_{15}$.

Discussion

The results presented in the above studies suggest that ss $(CGG)_{15}$ formed a stable intramolecular hairpin that contained G•G base pairs in the stem. CGG triplets on the 5′ side of the stem preferentially base-paired with GCG triplets on the 3′ side of the stem, giving rise to an alignment referred to as *(b)* in Figure 23. The results of chemical modification with DMS (Figures 24 and 25) suggested that a given G residue within a G•G base pair alternated between *anti* and *syn* conformations. The conclusion that $G^{syn} \bullet G^{anti}$ base pairs are contained within the stem region of the ss $(CGG)_{15}$ hairpin is supported by ^1H NMR studies on stable duplex structures of ss $(GGC)_{4\text{-}6}$ (Chen et al., 1995). The conclusion that a given G residue with a G•G pair alternated between *syn* and *anti* is consistent with the rapid (~14000 s^{-1}) transition of $G^{syn} \bullet G^{anti}$ base pairs in duplex DNA (Lane & Peck, 1995).

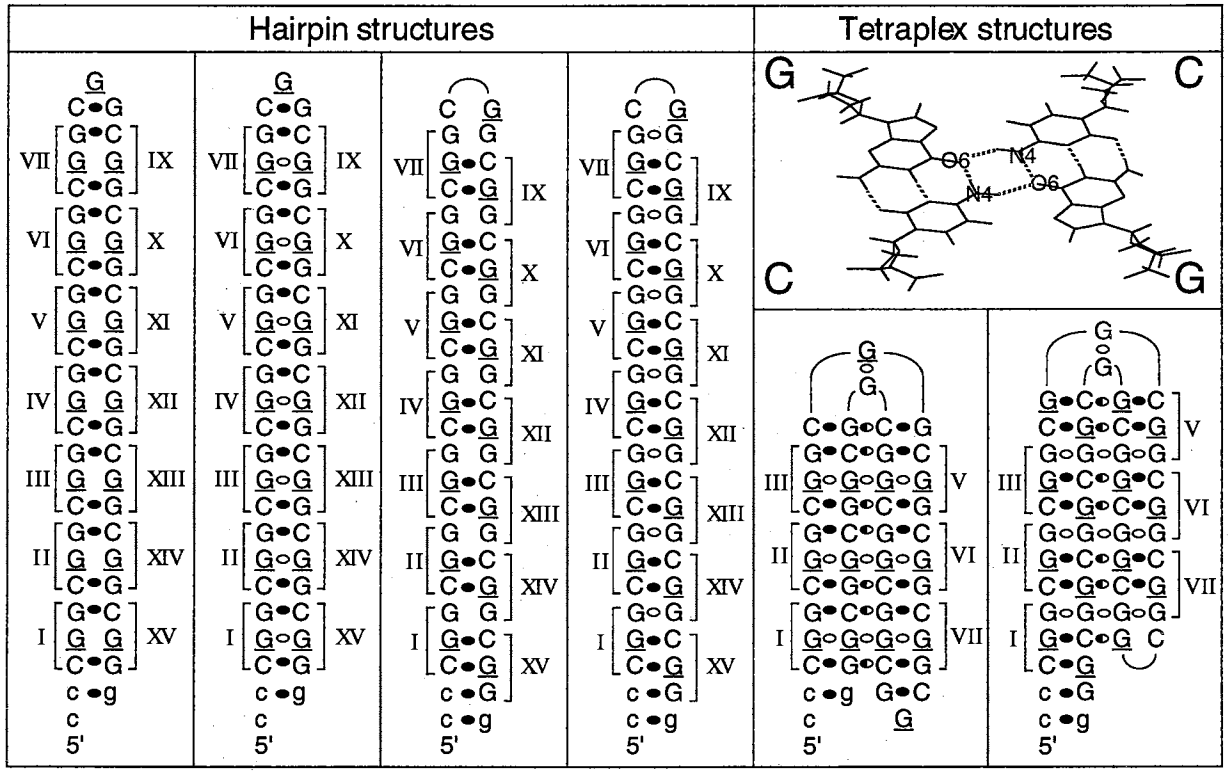
At ≥ 200 mM KCl, the DMS reactivities of the 5′Gs/3Gs were ~2.0 (Table 1), indicating that ss $(CGG)_{15}$ exclusively formed a hairpin in the *(b)* alignment. At KCl concentrations ≤ 100 mM, the DMS reactivities of the 5′Gs/3′Gs were 1.82, indicating that the hairpin in alignment *(b)* (structure 4 in Figure 23) was in equilibrium with another structure. Owing to the stability of the $G^{syn} \bullet G^{anti}$ base pair, we think it is more likely that , at KCl ≤ 100 mM, structure 4 was in equilibrium with structure 2. Molecular dynamics simulations suggested that structure 4 contained a *syn-anti* base pair in the loop region ($35G_{VIII}$ and $32G_{VII}$).

In contrast, simulations of the loop in the structure 2 predicted that $35G_{VIII}$ was stacked on top of $32G_{VII}$. We proposed that, at higher salt concentrations the $32G_{VII}$-$35G_{VIII}$ syn-trans pair is favored while, at lower salt concentrations, the $35G_{VIII}$-$32G_{VII}$ stack is favored. Hence, at lower salt concentrations the $35G_{VIII}$-$32G_{VII}$ stack may tend to drive the duplex toward the (a) aliginment.

Our results indicate that hairpins formed from repeats of $ss(CGG)_{15}$ are more stable than alternative tetraplex structures. Furthermore, our results showed that a longer $ss(CGG)_{18}$ repeat also formed the hairpin with a similar alignment and is more compact than that of ss $(CGG)_{15}$. However, I do not exclude the possibility that in ss (CGG)n, here, if n > 20, the intramolecular tetraplex structure, a tetraplet containing G4 or CGCG quartets, may be formed because of a dominant stability of G4 and CGCG quartet contributions.

Figure 23. Potential hairpin and tetraplex strucures of ss(C<u>G</u>G)15.

Four potential hairpin and two potential tetraplex structures of ss(C<u>G</u>G)$_{15}$ are shown. Nontriplet repeat sequences of ss(C<u>G</u>G)$_{15}$ anticipated to form base pairs are in lowercase letters. Triplet repeat numbers are indicated by Roman numerals. Watson-Crick C-G base pairs are indicated by filled ovals. Non-Watson-Crick C-G base pairs are indicated by open ovals. To distinguish the 5′G in a triplet repeat from the 3′G, the 5′G is underlined. In hairpin alignment *(a)*, the 5′<u>G</u>s are opposite one another. In hairpin alignment *(b)*, the 3′Gs are opposite one another. Non-Watson-Crick C-G base pairs (described below) in the tetraplex structures are indicated by half-filled ovals. In the tetraplex structures, the non-Watson-Crick C-G base pairs that form between the left-most strand (triplets I-III) and the right-most strand (triplets V-VII) are not shown. Tetraplex structure 5 was constructed by the folding in half the hairpin in alignment *(a)*. Tetraplex structure 6 was constructed in a similar manner from the hairpin in the *(b)* alignment. Molecular-moldeling studies (Mitas et al., 1995) revealed favorable interactions between the G•C base pairs, as shown in the upper right corner. Vertically oriented H bonds are from Watson-Crick base pairs derived from the hairpin. Horizontally oriented H bonds arise from hairpin-hairpin interactions. Energy minimization also indicated favorable G•G base pairs in the tetraplex loops. Hairpin alignments other than those shown contained fewer Watson-Crick base pairs and were not considered thermodynamically favorable. Expected features of structures 1-6 are listed below each structure. The variable y refers to an arbitrary measurement of dimethyl sulfate (DMS) reactivity of a given G residue within a triplet. For simplicity, it was assumed that a G residue within a G•G base pairs in a hairpin would alternate between the *syn* and *anti* conformations. Therefore, the DMS reactivity of a G residue that alternated between the two conformations should be 0.5, relative to the DMS reactivity of a G residue within a C•G base pair.



	1	2	3	4	5	6
structure number						
hairpin alignment	*(a)*	*(a)*	*(b)*	*(b)*	*(a)*	*(b)*
G-G base pairs?	no	yes	no	yes	yes	yes
# of potential H bonds	48	62	48	62	88	88
DMS reactivity: 5'G/3'G	y/y = 1.0	0.5y/y = 0.5	y/y = 1.0	y/0.5y = 2.0	0/y = 0	y/0 ~ infinity
triplets susceptible to P1	VIII	VIII	VIII	VIII	IV, VIII, XII	IV, VIII, XII

Table 1. Dimethyl sulfate reactivity of adjacent G residues in the stem of ss(CGG)15*

KCI, M	DMS reactivities 5'G/3'G
0	1.40 ± 0.13
0.05	1.56 ± 0.15
0.10	1.82 ± 0.22
0.20	2.03 ± 0.21
0.40	1.96 ± 0.14
0.75	1.99 ± 0.16

* The autoradiograph shown in Figure 24 was scanned with a PDI densitometer (x3). For reaction products generated at the specified KCl concentration, the optical density of a band derived from cleavage of the 5'G within a given triplet repeat was determined and divided by the optical density of the 3'G within the same triplet repeat. Optical density values obtained from triplet repeat I-II were used for data analysis. Values in the table represent the mean ± standard deviation.

Figure 24. Dimethyl sulfate reactions with ss$(CGG)_{15}$ at various K^+ concentrations.

DMS methylation was performed at 10 °C with ss $(CGG)_{15}$ as described in Materials and Methods. Reaction mixtures contained 50 mM Na^+ and various concentrations of K^+ (as indicated). For a given concentration of K^+, the concentration of DMS reacted with ss $(CGG)_{15}$ was (from left to right) 0 mM, 5.3 mM, 21 mM, 84 mM, and 0.42 M (with the exception of 0.2 m K^+, where the concentrations of DMS were 21 mM, 84 mM, and 0.42 M). Reaction mixtures were applied to a sequencing gel containing 20% polyacrylamide and 8 M urea. The deduced structure of ss$(CGG)_{15}$ in 0.75 m K^+ is shown to the right of the gel. Open ovals denote G•G base pairs. Filled ovals denote G•G base pairs.

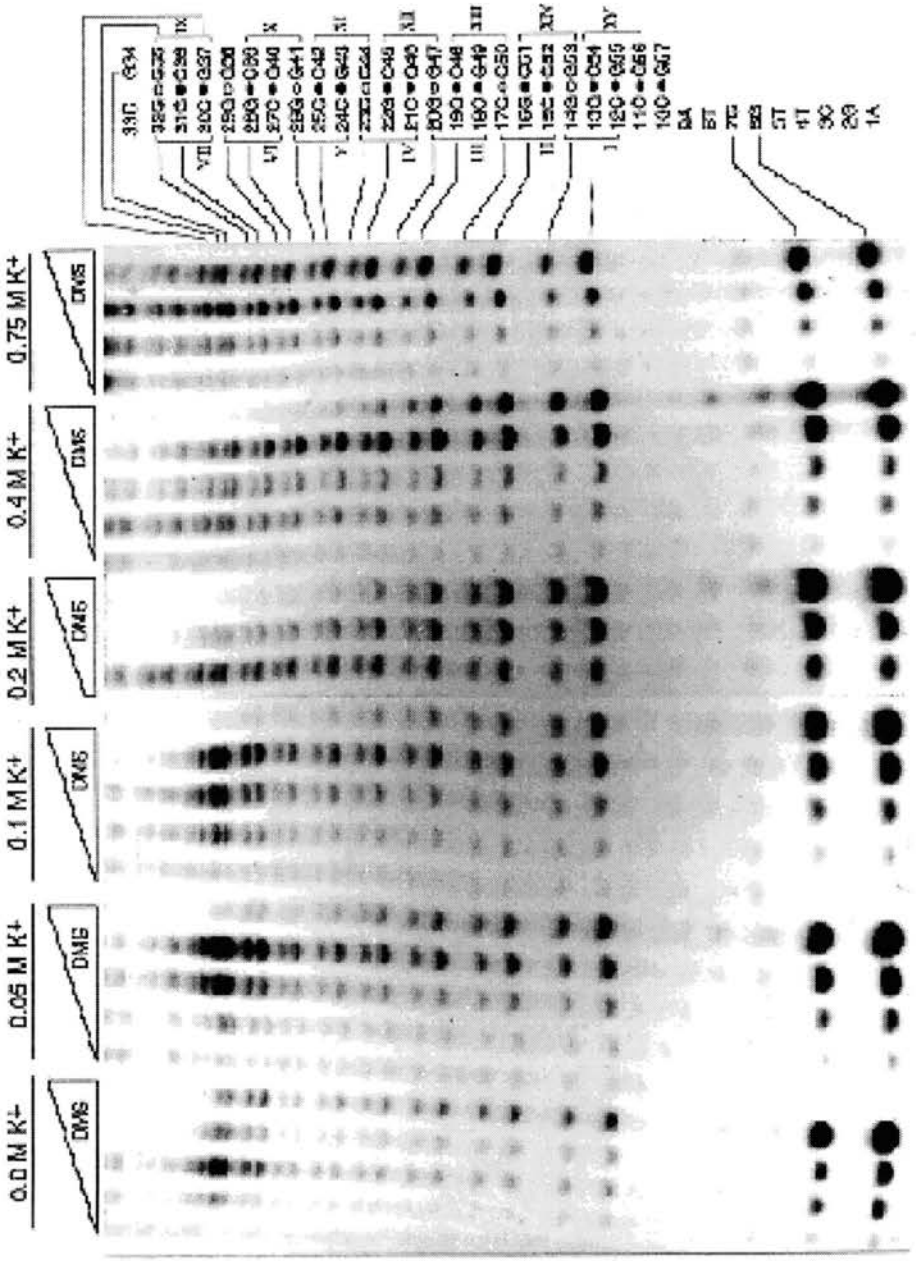

0.0 M K+
0.05 M K+
0.1 M K+
0.2 M K+
0.4 M K+
0.75 M K+
DM8
DM6
DM4
DM2
DM1
DM0
1A
2D
3T
4T
5T
6T
7G
8A
9T
10A
11A
12A
13A
14A
15A
16A

Figure 25. Dimethyl sulfate reactions with ss$(CGG)_{15}$ at various temperatures.

DMS methylation of ss$(CGG)_{15}$ was performed in 50 mM Na^+ and 100 mM KCl at the indicated temperature as described in Materials and Methods. For a given temperature, the concentration of DMS reacted with ss $(CGG)_{15}$ was (from left to right) 0 mM, 5.3 mM, 21 mM, 84 mM, and 0.42 M. Reaction mixtures were applied to a sequencing gel containing 20% polyacrylamide and 8 M urea.

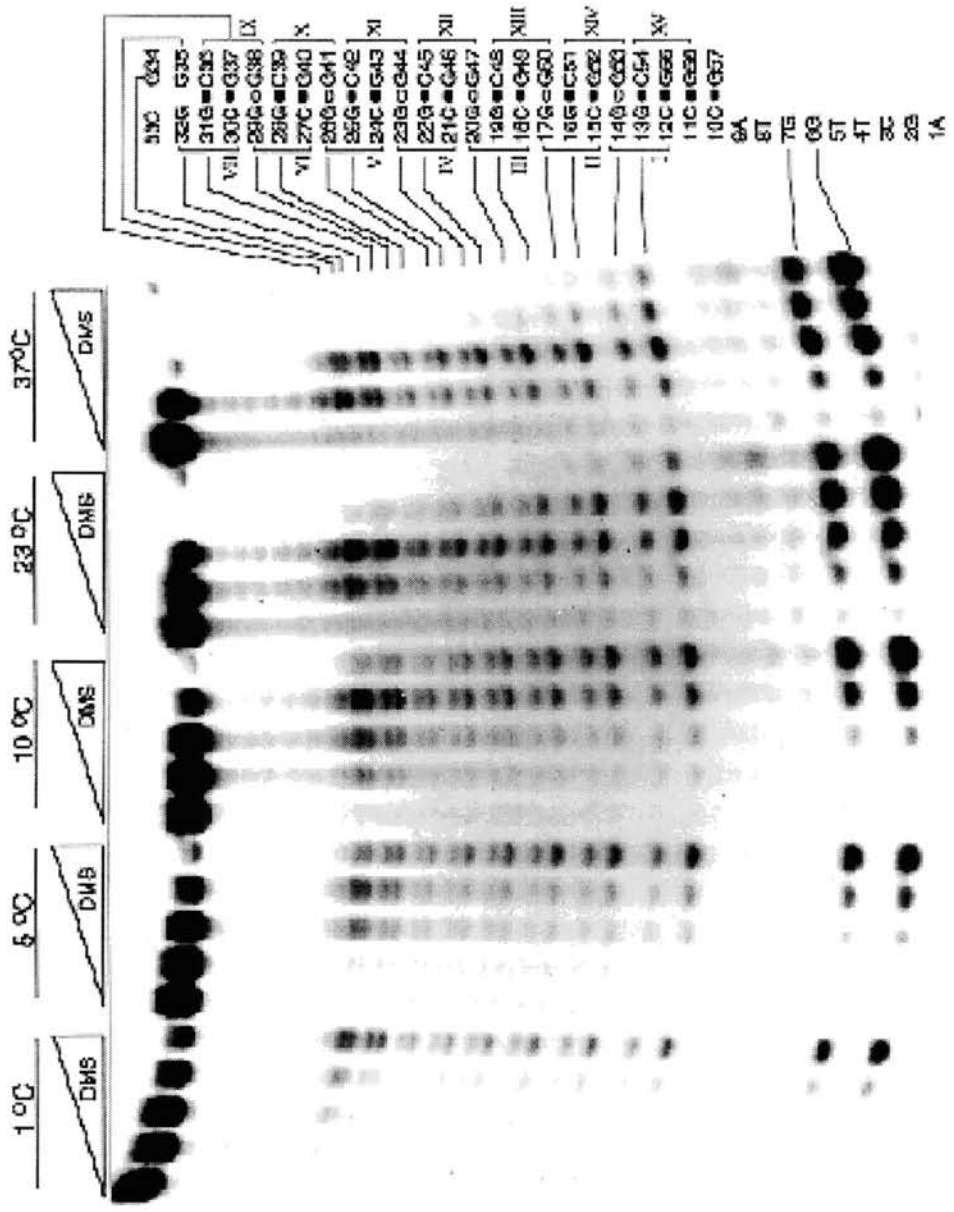

1 °C
5 °C
10 °C
23 °C
37°C
DMS
DMS
DMS
DMS
DMS
IX
X
XI
XII
XIII
XIV
XV
9A
8T
7G
6G
5T
4T
3C
2G
1A

Figure 26. P1 nuclease digestion of ss$(CGG)_{15}$.

The oligonucleotide containing $(CCG)_{15}$ purified from pCCG15 was labeled on the CGG-containing strand only with ^{32}P. The unlabeled synthetic oligonucleotide (1 μmol) of the same sequence as the labeled strand was added to the oligonucleotide, and the mixture was placed in a 100 ^{o}C H_2O bath for 3 min and immediately chilled on ice. The amounts of P1 nuclease used to digest ss$(CGG)_{15}$ (from left to right) at 37 ^{o}C in 50 mM Na^+ were 1.15×10^{-2} and 3.46×10^{-2} unit, respectively. Dimethyl sulfate (21 mM) reactions were performed as described in Materials and Methods. Roman numerals represent triplet repeat numbers. Arrows indicate sites of P1 nuclease cleavage.

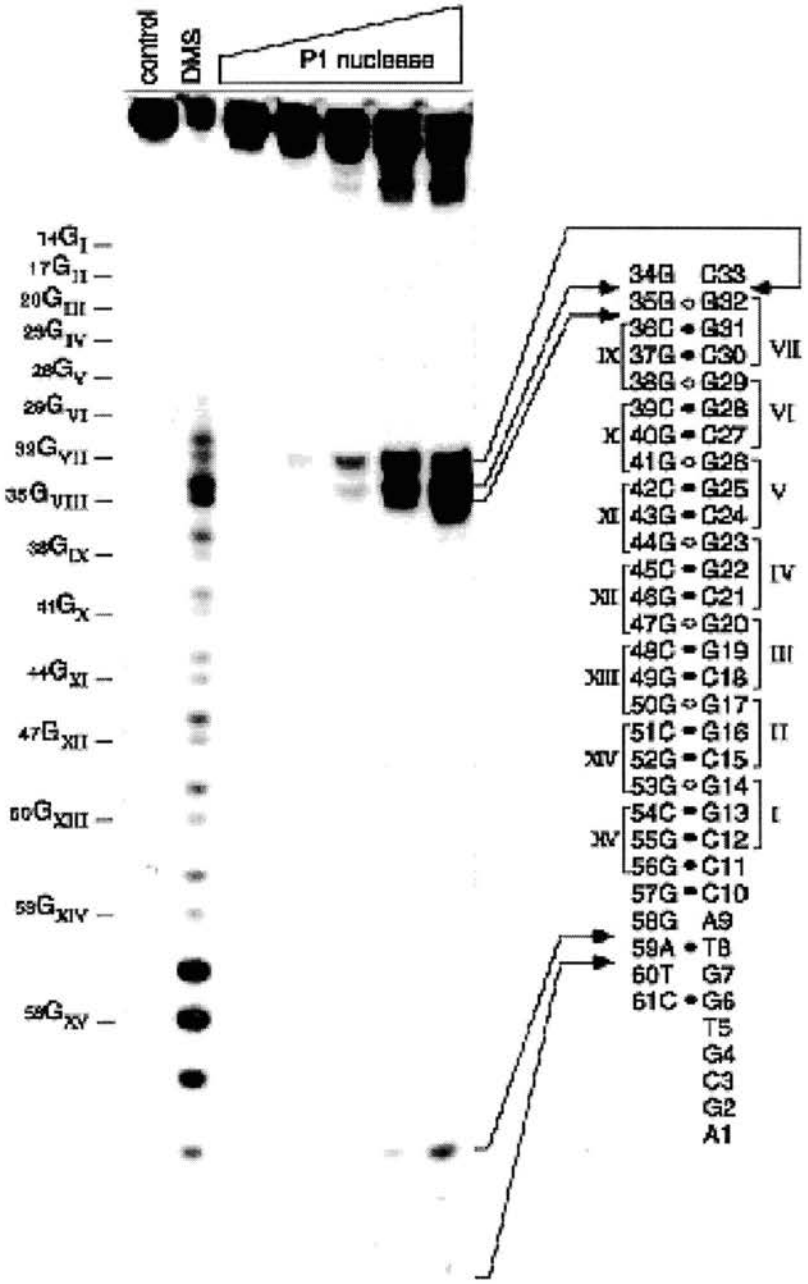
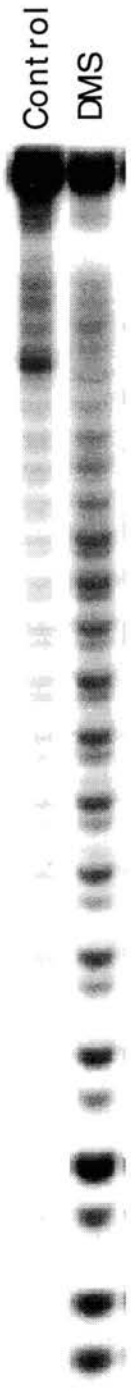

control
DMS
P1 nuclease
14G I
17G II
20G III
25G IV
28G V
29G VI
33G VII
35G VIII
38G IX
41G X
44G XI
47G XII
50G XIII
53G XIV
56G XV
34G C33
35G G32
36C G31
37G C30
38G G29
39C G28
40G C27
41G G26
42C G25
43G C24
44G G23
45C G22
46G C21
47G G20
48C G19
49G C18
50G G17
51C G16
52G C15
53G G14
54C G13
55G C12
56G C11
57G C10
58G A9
59A T8
60T G7
61C G6
T5
G4
C3
G2
A1
VII
VI
V
IV
III
II
I

Figure 27. DMS modification of ss$(CGG)_{18}$

DMS methylation was performed at 37°C with ss $(CGG)_{18}$ as described in Materials and Methods. Reaction mixtures contained 50 mM Na^+, pH 7.5. The final concentration of DMS was 21 mM, and no DMS in control. Reaction mixtures were applied to a sequencing gel containing 20% polyacrylamide and 8 M urea.


Control
DMS

CHAPTER VII

THE TRINUCLEOTIDE REPEAT SEQUENCE d(CCG)$_{15}$ FORMS A HAIRPIN CONTAINING PROTONATED CYTOSINES AND A DISTORTED HELIX AT PHYSIOLOGIC pH

In the description of the sequence-based classification system, ss (CCG)n was estimated to form the hairpin structure with the least stability among the six members of class I triplet repeats (Mitas et al., 1995a). It would be meaningful if structural features of ss(CCG)n could be found which could reasonably explain important phenomena associated with TREDs, such as fragile sites or hypermethylation. d(CCG)n sequences can adopt hairpin alignments that contain either CpG or GpC base pair steps. This is in contrast to d(CTG)n or d(CAG)n, which form hairpins containing only GpC base pair steps. It has been reported that when the 5′ nucleotide in a d(GCC)$_{5\text{-}11}$ sequence is a guanine, a hairpin containing multiple GpC base pair steps is formed (Chen et al., 1995, Mariappan et al., 1996). This hairpin alignment is particularly important with respect to potential recognition by the human DNA (cytosine-5) methyltransferase (MTase). MTase is the only enzyme identified to function in human DNA methylation. It requires the substrate cytosine in a CpG base step and the G paired with C of the other strand (Figure 28) (Smith et al., 1987 Baker, et al., 1991, Smith et al., 1991). In a d(CCG)n hairpin containing GpC base pair steps, the methylatable cytosine of a CpG dinucleotide is mispaired with

another cytosine. This alignment is predicted to be the preferred substrate for MTase activity (Chen et al., 1995).

^{1}H NMR studies do not reveal any hydrogen bonds between the C-C mismatches of $(GCC)_{5\text{-}11}$ hairpins (Chen et al., 1995, Mariappan et al., 1996). In another study, Gao and colleages recently discovered that the $d(CCG)_2$ formed a duplex $d[CCG)_2]_2$ that contained two CpG base pair steps rather than a single GpC base pair step. In this completely new DNA structure, the cytosines within the C-C mismatch, which were located in the center of the helix, were not stacked within the helix, but were extrahelical, pointed away from each other, and symmetrically located in the minor groove (Gao et al., 1995).

To understand the mechanisms of gene hypermethylation and sequence amplification, it is essential to determine the structures of oligonucleotides containing relatively large numbers of CCG repeats. We provide experimental evidence that oligonucleotides containing 15 or 20 CCG repeats exclusively form hairpin structures in a *(b)* alignment. However, the hairpins exhibit highly unusual features: the mispaired cytosines were protonated at a relatively high pH, and, the sugar-phosphate backbone was highly distorted. The results described in this chapter are completely consistent with a DNA structure containing multiple pairs of extrahelical cytosines.

ss $(CCG)_{15}$ exhibits a structural transition between pH 7.5 and pH 8.5

Electrophoretic mobility studies of the six triplet repeat sequences in class I revealed that $ss(CCG)_{15}$ showed an alteration in its electrophoretic mobility between pH 8.5 and pH 7.5. Further examinations at various pH values by electrophoretic and CD studies demonstrated that the pKa of $ss(CCG)_{15}$ was 7.7 + 0.20 (Yu et al., 1996). These data demonstrate that at least some fraction of the

cytosines in ss$(CCG)_{15}$ were protonated at or near pH 7.7. Since CG base pairs are not stabilized by cytosine protonation, it was concluded that the protonation in ss$(CCG)_{15}$ was limited to the C-C mispairs (Mitas and Gray, unpublished results).

Recent molecular modeling studies (I. S. Haworth , unpublished results) have indicated that protonation of N3 of cytosine in CCG sequences can stabilize two different structures. In the first theoretical structure, $(CCG)_n$ sequences fold into a tetraplex by virtue of formation of C•$^+$C pairs that are arranged in parallel. The second theoretical structure is an extension of the e-motif recently described by Gao and colleagues (Gao et al., 1995), in which pairs of extrahelical cytosines, which are formally separated by two C•G base-pairs, stack together in the minor groove of a hairpin. Since circular dichroism (CD) data provided no indication of C•$^+$C base pairs that were arranged in parallel (Maria Barron, D. M. Gray, unpublished results), at pH 7.5, the structure of ss $(CCG)_{15}^+$ might be a hairpin rather than a tetraplex. Potential *(a)* and *(b)* alignment hairpin structures of ss$(CCG)_{15}$ are shown in Figure 29. I refer to the structure containing partially protonated C-C mismatches as ss$(CCG)_{15}^+$, and the one unprotonated as ss$(CCG)_{15}$.

At pH 8.5, ss $(CCG)_{15}$ forms a hairpin structure in a *(b)* alignment

In order to investigate the possibility that ss $(CCG)_{15}$ adopted hairpin conformations at neutral and mildly basic pH, chemical modifications were performed with hydroxylamine (HA) (Singer & Grunberger, 1983; Johnston & Rich, 1985; Johnston, 1992) and 2-hydroperoxytetrahydrofuran (THF-OOH) (Liang et al, 1994). These reagents preferentially react with single-stranded Cs. If ss$(CCG)_{15}$ adopted a hairpin conformation, one of the Cs within a given CCG

triplet should be base-paired to a G (ie. a C•G pair), while the other C should pair with a C (Figure 29). Since it is not known whether the mispaired cytosines in a (CCG)n-containing hairpin are H-bonded, the mispaired cytosines will be designated as a "C-C" pair. The C of the C•G pair should react poorly with HA or THF-OOH. Therefore, incubation of ss $(CCG)_{15}$ with HA or THF-OOH should reveal which of the two Cs within a triplet are base-paired to G. If ss $(CCG)_{15}$ were to adopt a hairpin in the *(a)* alignment (Figure 29), the cytosine 5' to the nearest G (designated as 5'G) should be much more reactive with HA or THF-OOH compared to the cytosine 3' to the nearest G (Fig. 29). In contrast, if ss $(CCG)_{15}$ were to adopt a hairpin in the *(b)* alignment, the 3'Cs should be much more reactive with HA compared to the 5'Gs (Fig. 29).

Incubation of ss $(CCG)_{15}$ with HA (Figure 30) or THF-OOH (Figure 31, 32) at pH 8.5 or pH 8.0 revealed high reactivity of the 3'Cs, indicating that the 3'Cs were not base-paired to Gs. Reactivities of the 5'Cs were not observed, with the exception of C28, a nucleotide located in the presumed loop region of ss $(CCG)_{15}$ (Figure 30). Also, both Cs in triplet I were non-reactive with HA or THF-OOH, a result consistent with a hairpin in the *(b)* alignment.

Hydroxylamine highly reacted with C55 and C56, two Cs not part of the triplet repeat region. C55 was opposite to A2 in the hairpin structure of ss $(CCG)_{15}$. Since C55 can not form H-bonds to A2 at pH 8.5, the HA reactivity of C55 served as a control for a single-stranded C. The HA reactivity of C55 was comparable to C48, a cytosine with a C-C pair. The fact that the HA reactivities of C55 and C48 were similar provided evidence that there were no H-bonds in the C-C pairs of ss $(CCG)_{15}$ at pH 8.5.

The sugar-phosphate backbone of ss$(CCG)_{15}$ is distorted at pH 8.5

To further investigate the intramolecular structure of ss $(CCG)_{15}$, P1 nuclease reactions were performed at various temperatures in 50 mM NaCl. At pH 8.5, 25 °C, major sites of P1 nuclease cleavage in the triplet repeat region were the G26-C27, C28-G29, G29-C30, and C31-G32 phosphodiesters (Figure 33, 34). This result indicated that a fold was present near triplet VIII and provided further evidence for a hairpin conformation. At 55 °C, the pattern of P1 nuclease cleavage was more uniform, indicating that the hairpin structure was significantly denatured at this temperature. This result is consistent with UV studies, which indicated that the T_m of ss $(CCG)_{15}$ in a higher salt concentration (150 mM) was ~54 °C (Mitas unpublished). The primary site of P1 nuclease cleavage at the 3' terminus was the C56-A57 phosphodiester. This result was in contrast to P1 nuclease cleavages of ss$(CTG)_{15}$ and ss$(GTC)_{15}$, where the major site of cleavage in both of these oligonucleotides was the C55-C56 phosphodiester (Yu et al., 1995b). In ss $(CTG)_{15}$ and ss $(GTC)_{15}$, the base-pairing arrangements at the termini are exactly the same as would be observed in an *(a)* hairpin alignment of ss$(CCG)_{15}$. Therefore, if ss $(CCG)_{15}$ were to adopt a hairpin in an *(a)* alignment, the major site of cleavage at the 3' terminus should be the C55-C56 phosphodiester. The finding that the major site of P1 cleavage at the 3' terminus of ss $(CCG)_{15}$ was the C56-A57 phosphodiester further supports a hairpin in a *(b)* alignment, and not an *(a)* alignment.

The 5' C-G 3' phosphodiester linkages in triplets X-XIV of ss$(CCG)_{15}$ were also cleaved to a minor extent (Fig. 33, 34). This result was significant since minor cleavages of the phosphodiesters in the stem regions of oligonucleotides containing all other Class I triplet repeat sequences have not been observed at pH 7.5 (Mitas et al., 1995a,b, Yu et al., 1995a,b). The same cleavage patterns

were also observed at low temperature (1 °C - 25 °C) (Figure 35). However, what is more important is that the cleaved phosphodiesters were part of the C•G base-pairs, and not the C-C mismatches. The ability of P1 nuclease to cleave the phosphodiester linkage within two C•G base-pairs provides direct evidence that the sugar-phosphate backbone was distorted.

P1 nuclease reactions with a control ss $\Delta T_{II}(GTC)_{15}$ sequence indicate a lack of cleavage in the hairpin stem

Detection of P1 nuclease susceptibility at the CpG base-pair step in the ss $(CCG)_{15}$ hairpin might have been due to experimental (eg. high pH) rather than structural conditions. To provide a control for P1 nuclease studies performed at pH 8.5, a sequence variant of ss $(GTC)_{15}$ was analyzed (Figure 36). The hairpin formed by ss $(GTC)_{15}$ is similar to that formed by ss $(CCG)_{15}$ at pH 8.5 in a number of respects. First, both hairpins contain CpG base-pair steps. Second, their thermal stabilities are low relative to ss $(CTG)_{15}$ (Table 1), and third, the mismatched bases in the ss $(CCG)_{15}$ and ss $(GTC)_{15}$ hairpins are pyrimidines. The sequence variant $\Delta T_{II}(GTC)_{15}$ is comparable to ss$(GTC)_{15}$ except that the thymine in triplet II is missing (Figure 36). In the hairpin structure of $\Delta T_{II}(GTC)_{15}$, the extrahelical thymine in triplet XIV does not have a nucleotide with which it can pair, and hence, is readily oxidized by $KMnO_4$ (Figure 17) (Yu et al., 1995b).

At pH 8.5, 50 mM Na^+, P1 nuclease did not cleave the stem region of $\Delta T_{II}(GTC)_{15}$ (including the region corresponding to the extrahelical thymine) at 25 and 35 °C (Figure 36), indicating that the sugar-phosphate backbone was not distorted. These results indicated that cleavages of the CpG dinucleotides in the

ss$(CCG)_{15}$ hairpin at pH 8.5 were due to major distortion of the sugar phosphate backbone.

The structure of ss $(CCG)_{15}^+$ at pH 6.5 is a hairpin in a *(b)* alignment

The results of electrophoretic mobility, UV absorbance, and CD studies (Mitas, M. and Gray, D. M., unpublished data) indicated that the bases of ss$(CCG)_{15}$ were more highly stacked at pH 6.5-7.5. In order to investigate the base-pairing arrangement below neutral pH, HA reactions at various temperatures (1- 55 °C) were performed in 50 mM Na^+ at pH 6.5. We were particularly interested to determine whether the hairpin at pH 6.5 might have switched to an *(a)* alignment. In this alignment, it is possible to form antiparallel C•$^+$C pairs similar to those described by Brown et al. (1990).

Similar to results obtained at pH 8.5 (Figure 30), reactions performed at or below 45 °C revealed high reactivity of the 3'Cs, indicating that ss$(CCG)_{15}$ formed a hairpin in the *(b)* alignment (Figure 37). At ≤ 35 °C, all cytosines in triplets VIII and IX reacted with HA, indicating that the bases in the loop region of ss$(CCG)_{15}$ were not paired. Identical results were obtained at 37 °C at pH 7.0 with HA (Figure 38) or THF-OOH (Figure 31, 32) or at pH 8.3, 25 °C, 400 mM or 600 mM NaCl with HA (Figure 39). At 55 °C, near uniform reactivities of adjacent cytosines in ss$(CCG)_{15}$ were observed, indicating that all C•G base-pairs were melted at this temperature (Figure 37). The result is consistent with UV absorbance studies (Mitas unpublished), which indicated that the melting temperature of ss $(CCG)_{15}^+$ (indicating those at pH <7.7) in 150 mM NaCl was 57.5 °C.

The results demonstrated that the base-pairing arrangements at pH 8.5 and 6.5 were identical. Since the results of CD studies provided no indication of

parallel C•+C pairs at pH 6.5 (Gray, D. M., unpublished data) and since it is not possible to form antiparallel C•+C pairs within a *(b)* alignment at pH 6.5 (Gao et al., 1995), the combined data suggest that ss $(CCG)_{15}$ does not contain C•+C pairs in any arrangement at pH 7.5-6.5.

Base stacking and base flipping at pH 6.5

To further investigate the structure of ss$(CCG)_{15}$ at or slightly below neutral pH, P1 nuclease digestions were performed at pH 6.5 and 7.5. Incubation of ss$(CCG)_{15}$ with P1 nuclease at pH 6.5 resulted in extensive cleavage of the loop region (Figure 40). This result was in complete agreement with the HA results (Figure 37), which indicated that C31 was not extensively paired to G26. Substantial cleavages of the G41-C42 and C42-C43 phosphodiester linkages in the center of the helix stem were also observed. These phosphodiesters flank a cytosine (C42) of a C-C mismatch. Little or no cleavages of the C40-G41 or C43-G44 phosphodiester linkages were observed, indicating that C42, and only C42 was extruded, or flipped away from the 3' portion of the helix. Cleavages of the G41-C42 and C42-C43 phosphodiesters were also observed over a wide range of salt concentrations (0-400 mM NaCl) (Figure 41) and over a range of temperatures (37-57 °C) at pH 7.5 (Figure 42), indicating that the flipped cytosine configuration was heat and salt stable. Little or no cleavages of the G14-C15 and C15-C16 phosphodiesters were observed, indicating that C15, the base opposite that of C42 in the formal hairpin, was not flipped out of the helix.

In contrast to the P1 nuclease results at pH 8.5 (Figure 33, 35, 43), little or no cleavages of the CpG phosphodiester linkage were observed in the stem of the hairpin at pH 6.5 (Figure 40) or pH 7.5 (Figure 41, 42, 43), indicating that the

CpG base pair steps were better stacked compared to pH 8.5. However, between triplets IX-XII, minor cleavages of all phosphodiesters were observed at pH 6.5, most prominantly at 25°C. (Figure 40). Cleavages of these phosphodiesters were also observed at pH 6.0, 7.0, and 7.5 (Figure 41, 42 and 43). These results provide evidence that the sugar phosphate backbone of ss $(CCG)_{15}^+$ was distorted.

A $(CCG)_n$-containing sequence designed to form a hairpin in the *(a)* alignment forms a hairpin in the alternative *(b)* alignment

The results presented above demonstrate that ss$(CCG)_{15}$ and ss$(CCG)_{15}^+$ (indicating those at pH < 7.7) formed hairpins in the *(b)* alignment. One possibility was that ss$(CCG)_{15}$ folded into a *(b)* alignment because of the influence of the loop structure and/or the sequences in the non-triplet repeat region. To address this possibility, we designed an oligonucleotide (Figure 44) with the following features: (i) a greater number of CCG repeats (20), which diminished the influence of loop structures and the non-triplet repeat regions, (ii) terminal sequences that were designed to force the triplet repeat region to fold into an *(a)* hairpin alignment (ie. four consecutive Watson-Crick base-pairs are formed), and (iii) an *(a)* alignment that contained a four (rather than a three) nucleotide loop. Energy minimizations predicted that the former loop size was more stable than the latter (I.S.H., R.R., unpublished results).

To investigate the structures of ss$(CCG)_{20}$, its electrophoretic mobility was analyzed as a function of pH (Mitas, M. & Christy, M., unpublished results). Similar to ss$(CCG)_{15}$, the M_{rel} of ss$(CCG)_{20}$ and ss$(C^mCG)_{20}$ increased as the pH was lowered from 8.5 to 7.5. Sharp transitions occurred between pH 8.5 and 8.3, indicating that the pK_a of the sequences was 8.4. A gradual transition occurred

between pH 8.3 and 7.5. These results indicate that a fraction of the cytosines in ss$(CCG)_{20}$ were protonated above pH 8.0.

To investigate the base-pairing arrangement(s) of ss $(CCG)_{20}$ and ss $(CCG)_{20}^{+}$, chemical modifications were performed with HA at pH 7.5 or 8.3 at 37 °C, 50 mM Na^{+}. The results (Figure 45) show that the 3'Cs in triplets XII-XIX were more reactive compared to the 5'Cs, providing evidence for a hairpin in the *(b)* alignment. However, the reactivities of the cytosines in triplet I and in the non-triplet repeat region were not consistent with the *(b)* alignment hairpin depicted in Fig. 44. For example, C66, which was predicted to form a base-pair with G10 in the *(b)* alignment (and hence poorly react with HA), was nearly as reactive as C65. This result suggested that C66 only formed "weak" H-bonds with some base, presumably G10. Further, C68, which was predicted to pair with C8 (and hence react well with HA) exhibited HA reactivity that was lower compared to C65, a cytosine within a C-C mismatch. This result suggested that C8 formed H-bonds with a guanine, presumably G67. Therefore, the results suggested that the bases in the non-triplet repeat region were partially paired according to an "*(a)*" alignment, while those in the triplet repeat region were paired according to a *(b)* alignment. These apparently conflicting data can be reconciled if one of the cytosines in triplet I becomes extrahelical (Figure 45).

The identity of the potential extrahelical base was determined by discerning which of the two cytosines in triplet I reacted with HA. The position of the band at triplet I (marked with an asterisk in Figure 45) in the reactions performed at pH 7.5 and pH 8.3 was consistent with reactivity of C8 (and not C9), indicating that C8 might be extrahelical. The extrahelical configuration of this base would allow for formation of three base-pairs in the non-triplet repeat region and a *(b)* hairpin alignment of triplets II-XIX.

To further investigate the hairpin structures of ss$(CCG)_{20}$ and ss$(CCG)_{20}^+$, P1 nuclease studies were performed at pH 7.5 or 8.5 at 37 °C, 50 mM Na^+. At pH 7.5, the major sites of P1 nuclease cleavage at the 3'-end were the A74-C75 and C75-A76 phosphodiesters (Figure 46), a result consistent with the hairpin structure shown in Figure 46. Similar to results obtained with ss$(CCG)_{15}$ at pH 8.5, minor cleavages of the phosphodiesters within the CpG dinucleotide were observed (Figure 46), indicating that the sugar phosphate backbone was distorted. At pH 7.5, minor cleavages of all phosphodiesters in the helix region were observed. However, unlike for ss$(CCG)_{15}^+$, there was no evidence for a flipped cytosine in the central portion of the helix.

Discussion

I have provided evidence that the ss$(CCG)_{15}$ forms a hairpin in the *(b)* alignment, whereby the cytosines 3' to the nearest guanine were mismatched with one another. Chemical modification and P1 nuclease cleavage experiments revealed that the *(b)* alignment was stable in a variety of salt concentrations (0-400 mM NaCl) and pH values (6.0-8.5) (Figures 30, 31, 32, 33, 34, 37, 40, 41). Although we did not investigate the structure of ss$(CCG)_{20}$ as thoroughly as ss$(CCG)_{15}$, the results described in this report indicate that ss$(CCG)_{20}$ (Figure 44, 45, and 46) also adopted a hairpin in the *(b)* alignment.

In theory, ss$(CCG)_{15}$ or ss$(CCG)_{20}$ could have adopted the alternative *(a)* alignment, whereby the cytosines 5' to the nearest guanine were mismatched with one another. The *(a)* alignment contains GpC base-pair steps, whereas the *(b)* alignment contains CpG base-pair steps. The stacking energies of GpC and CpG base-pair steps are -14.59 and -9.69 kcal/mole, respectively (Saenger, 1984). Therefore, if one assumes that the interactions between the mismatched Cs in *(a)*

or *(b)* alignments are identical, one would also assume that an *(a)* alignment should be more stable than a *(b)* alignment. Our failure to detect an *(a)* alignment of ss$(CCG)_{15}$ or $(CCG)_{20}$ suggests that when the number of repeats is ≥ 15, the interactions of the mismatched Cs in a *(b)* alignment are more stable compared to the interactions of the mismatched Cs in a *(a)* alignment. When the number of repeats is below 15, we suspect that the loop structure and/or end effects favor an *(a)* alignment, as previously observed in ^{1}H NMR studies for the sequences d$(GCC)_{5\text{-}7}$ (Chen et al., 1995, Mariappan et al., 1996a).

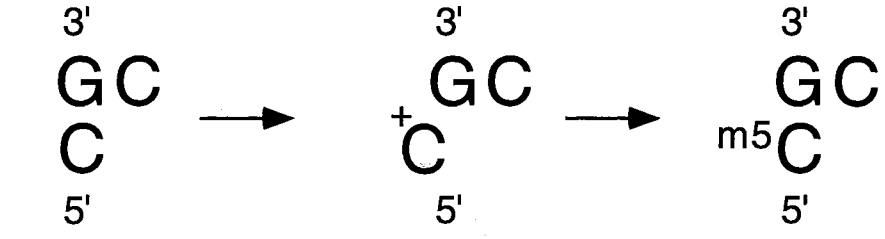
In vitro studies with the human MTase (Smith et al., 1994, Chen et al., 1995, Baker, et al., 1991, Smith et al., 1991) and two bacterial cytosine MTases (Klimasauskas and Robert, 1995, Yang et al., 1995) have shown that when the methylatable cytosine was paired with a base other than guanine, the rate of methylation was generally high. It has been suggested that the high rate of methylation was a reflection of the ease with which the methylatable cytosine flipped away from the helix, a requirement for methylation at the 5 position (Mitas et al., 1996). ^{1}H NMR studies do not reveal any H-bond within the C-C mismatches of $(GCC)_{5\text{-}11}$ hairpins (Chen et al., 1995, Mariappan et al., 1996a), suggesting that the methylatable cytosine might also be free to flip out of the helix and undergo rapid methylation.

We suspect that the high rate of *in vitro* methylation of d$(CCG)_{15}$ observed in the previous study might be due to one of two factors. First, d$(CCG)_{15}$ may exist in solution in a number of conformations, of which the *(b)* alignment hairpin is the predominant form. The *(b)* alignment hairpin might be ignored by the human MTase, while a minor population of *(a)* alignment hairpins might be rapidly methylated as a consequence of the increased length of the hairpin stem. In support of this possibility, it has been shown that optimal rates of methylation by the human MTase are achieved when DNA fragment sizes are at

least 20 bp (Laayoun & Smith, 1995), which is the approximate helix length of $d(CCG)_{15}$. The second and more intriguing possibility is that due to the distorted nature of sugar-phosphate backbone, the *(b)* alignment hairpin may be an excellent substrate for the MTase. In support of this possibility, cytosines within extrahelical CpG dinucleotides (ie., an extremely distorted helix) are rapidly methylated by the human MTase (Laayoun & Smith, 1995).

Figure 28. Substrate specificity of human (cytocine-5) methyltransferase

Specific conformation of the substrate for 5′- MTase is shown (left). 5′-CpG-3′ step is required. The 3′ G pairs with C in Watson-Crick base pairing. The 5′ C should be unpaired or non-Watson-Crick base paired.



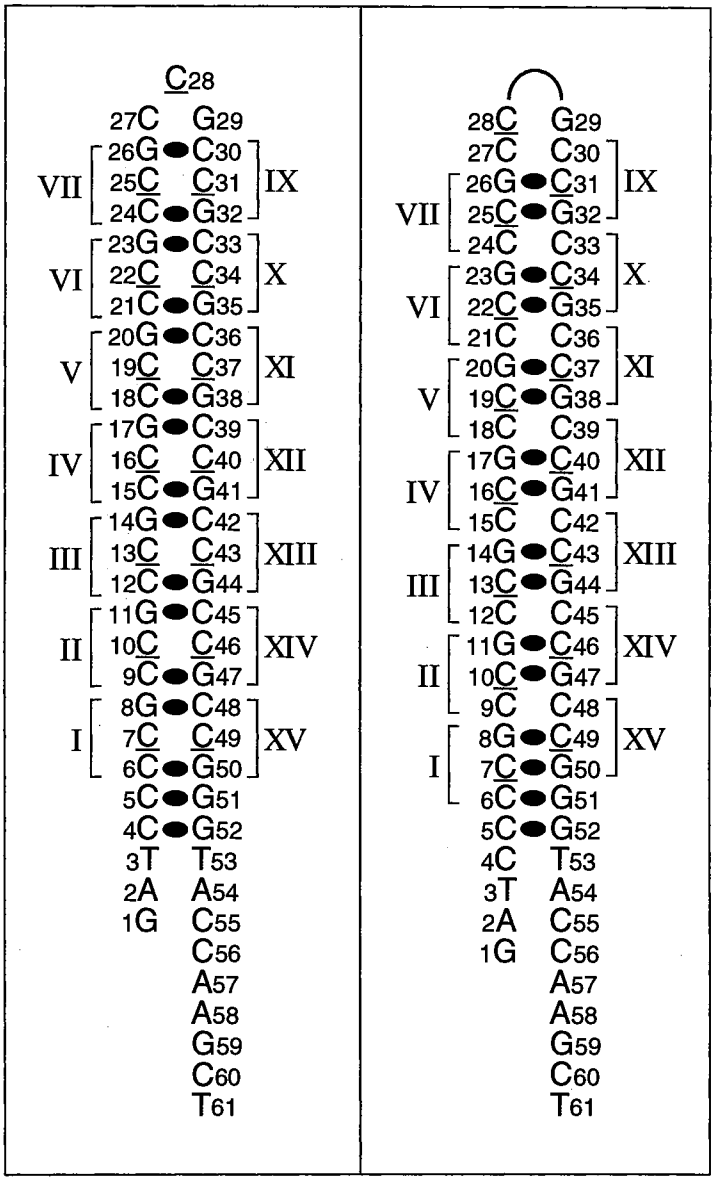
recognition motif

intermediate

product

Figure 29. Potential hairpin alignments of ss $(CCG)_{15}$

Two potential hairpin alignments of ss $(CCG)_{15}$ are shown. Filled ovals indicate Watson-Crick C•G base-pairs. Roman numerals depict a particular triplet repeat. Conventional Arabic numerals depict the position of a nucleotide with respect to the 5'-end. Various features of the hairpins are listed.



Hairpin alignment	*(a)*	*(b)*
HA reactivity	5'C > 3'C	3'C > 5'C
Loop size	3 nt	4 nt
# of Watson-Crick base pairs	16	16

Figure 30. Chemical modification with hydroxylamine at pH 8.5 reveals a hairpin in a *(b)* alignment

Single-stranded $(CCG)_{15}$ was incubated with hydroxylamine (HA) at 37 °C, pH 8.5 in 50 mM Na^+ as described in Materials and Methods and applied to a 20% polyacrylamide gel containing 8 M urea. The amount of HA in the reaction mixture was 6.3 M. The deduced hairpin structure of $ss(CCG)_{15}$ is shown on the left. The ^{32}P labels in the oligonucleotide are 5' to A58 and C60. Roman numerals indicate triplet repeat numbers. The positions of the reactive Cs within the deduced hairpin structure are shown. Dimethyl sulfate (21 mM) reactions were performed as described in Materials and Methods.

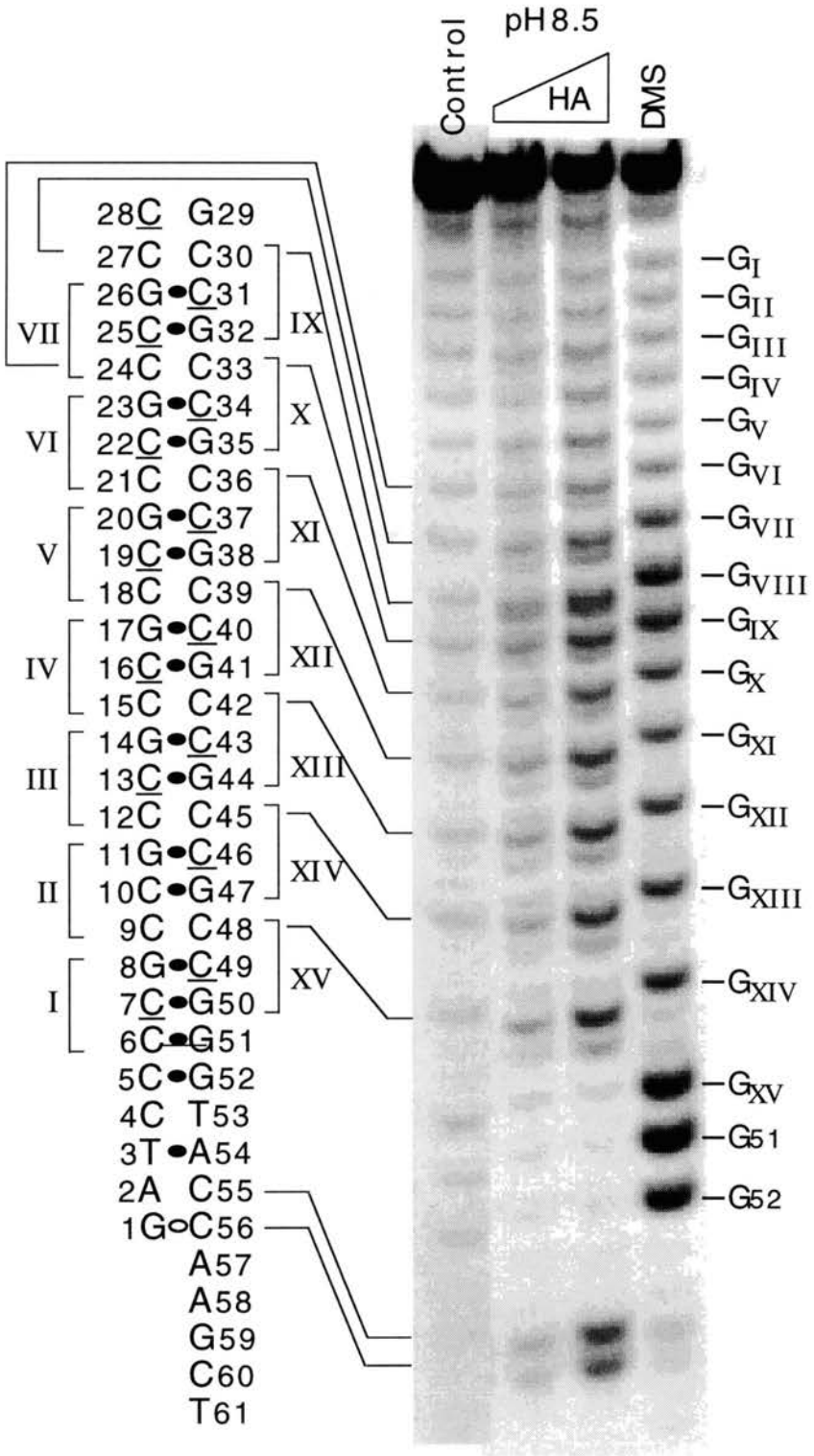

Control
pH 8.5
HA
DMS
28C G29
27C C30
26G•C31
25C•G32
24C C33
23G•C34
22C•G35
21C C36
20G•C37
19C•G38
18C C39
17G•C40
16C•G41
15C C42
14G•C43
13C•G44
12C C45
11G•C46
10C•G47
9C C48
8G•C49
7C•G50
6C•G51
5C•G52
4C T53
3T•A54
2A C55
1G•C56
A57
A58
G59
C60
T61
VII
VI
V
IV
III
II
I
IX
X
XI
XII
XIII
XIV
XV
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G51
G52

Figure 31. Chemical modification with THF-OOH at various pH reveals a hairpin in a *(b)* alignment.

Single-stranded $(CCG)_{15}$ was incubated with 2-hydroperoxy tetrahydrofuran (THF-OOH) at 37 ^{o}C, pH 6.0 and 7.0 in 50 mM Na^{+} as described in Materials and Methods and applied to a 20% polyacrylamide gel containing 8 M urea. The amounts of THF-OOH in the reaction mixtures were (from left to right) in a final concentrations of 0.2, 0.5, 1.0 and 2.0 M. The ^{32}P labels in the oligonucleotide are 5' to A58 and C60. Dimethyl sulfate (21 mM) reactions were performed as described in Materials and Methods. Each guanine is shown on the left. Roman numerals indicate triplet repeat numbers.

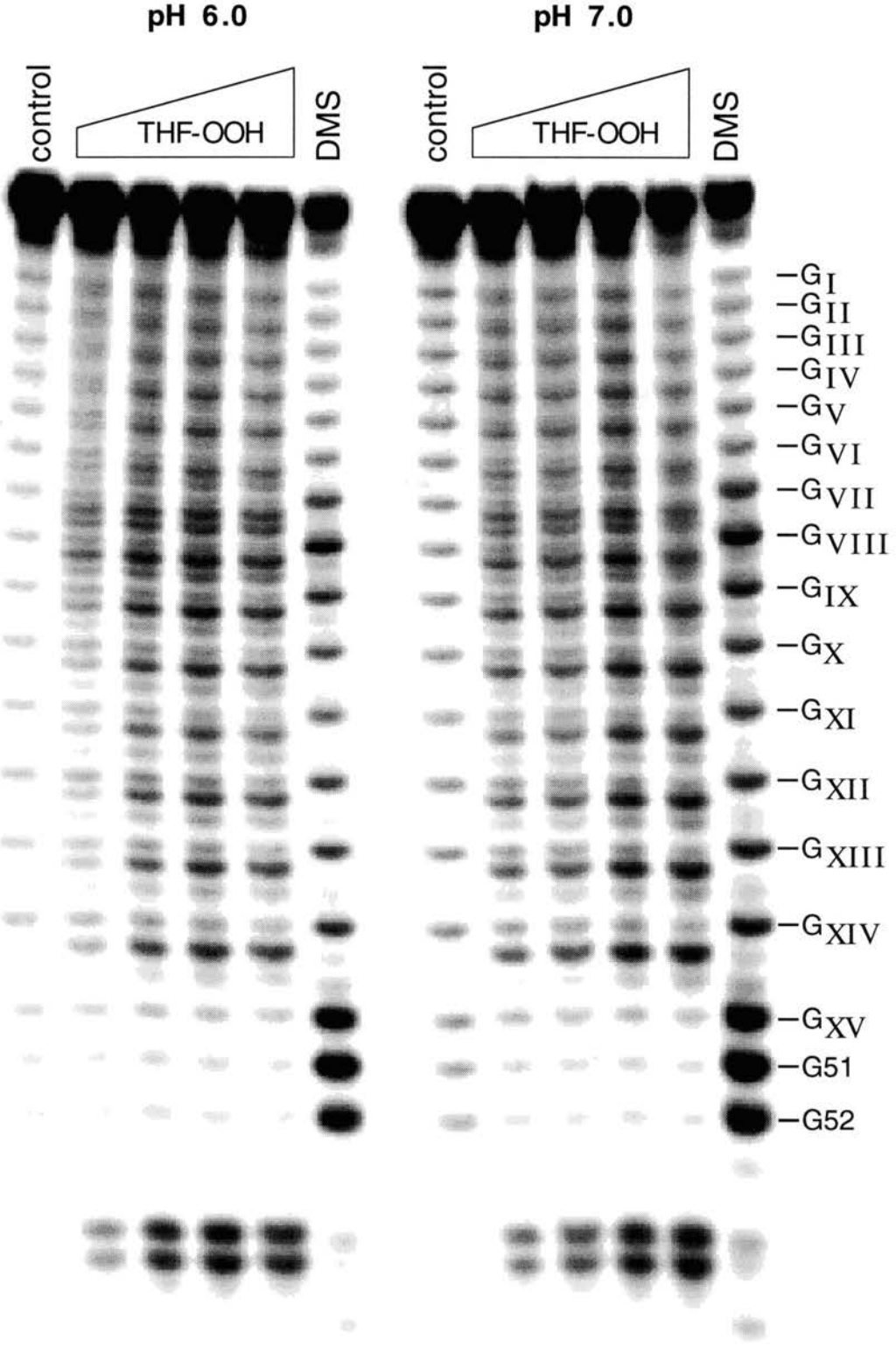

pH 6.0
pH 7.0
control
THF-OOH
DMS
control
THF-OOH
DMS
–GI
–GII
–GIII
–GIV
–GV
–GVI
–GVII
–GVIII
–GIX
–GX
–GXI
–GXII
–GXIII
–GXIV
–GXV
–G51
–G52

Figure 32. THF-OOH modification of ss $(CCG)_{15}$ at pH 6.5 and 8.0.

Single-stranded $(CCG)_{15}$ was incubated with THF-OOH at 37 ^{o}C, pH 6.5 and 8.0 in 50 mM Na^{+} as described in Materials and Methods and applied to a 20% polyacrylamide gel containing 8 M urea. The amounts of THF-OOH in the reaction mixtures were (from left to right) in a final concentrations of 0.2, 0.5, 1.0 and 2.0 M. The ^{32}P labels in the oligonucleotide are 5' to A58 and C60. Dimethyl sulfate (21 mM) reactions were performed as described in Materials and Methods. Each guanine is shown on the left. Roman numerals indicate triplet repeat numbers.

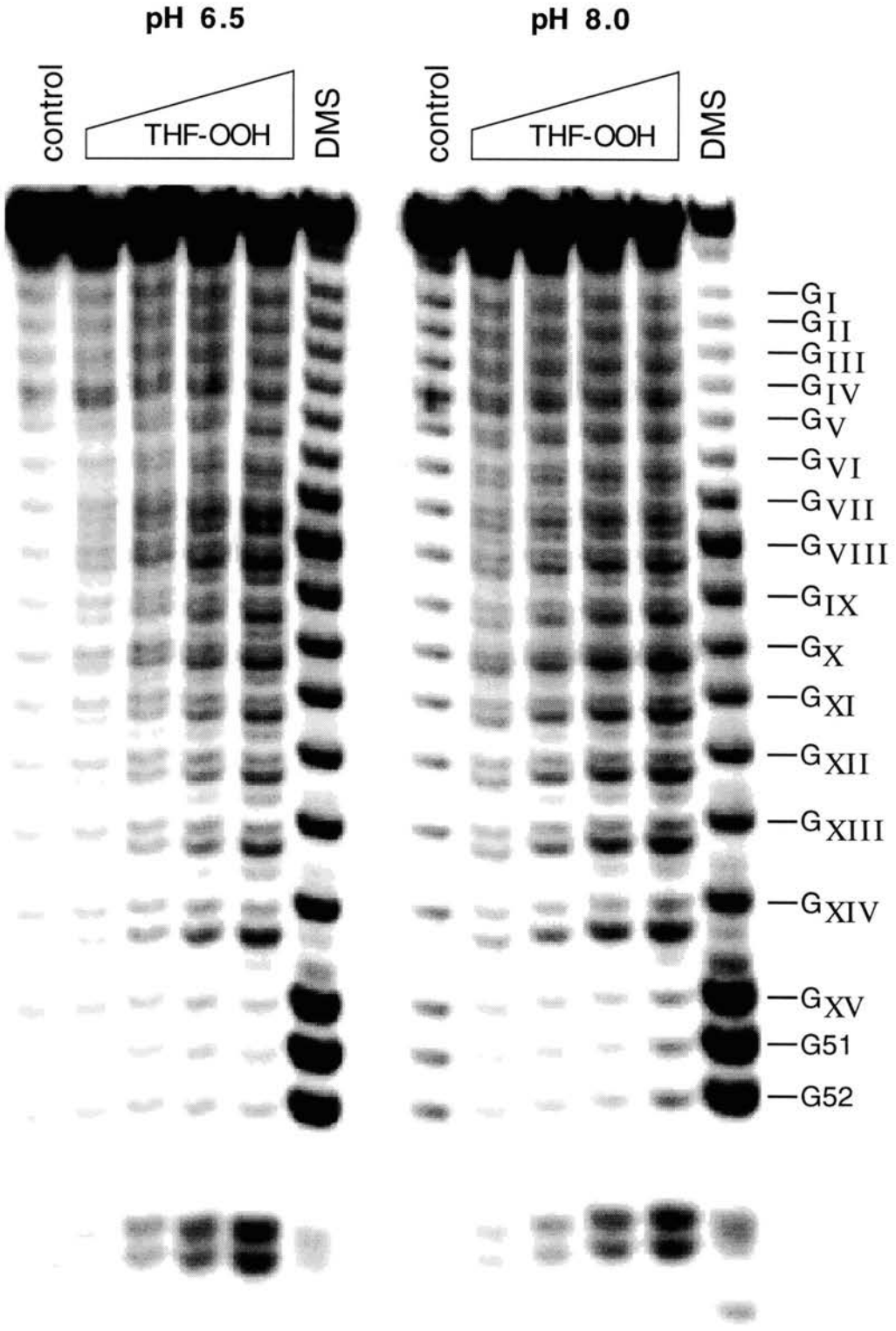

pH 6.5
pH 8.0
control
THF-OOH
DMS
control
THF-OOH
DMS
—GI
—GII
—GIII
—GIV
—GV
—GVI
—GVII
—GVIII
—GIX
—GX
—GXI
—GXII
—GXIII
—GXIV
—GXV
—G51
—G52

Figure 33. P1 nuclease digestion of ss $(CCG)_{15}$ at pH 8.5

Single-stranded $(CCG)_{15}$ was incubated with P1 nuclease for 5 minutes as described in Materials and Methods at the indicated temperatures. Buffer did not contain Na^+. Reaction products were applied to a 20% polyacrylamide gel containing 8 M urea. The amounts of P1 nuclease added at the respective temperatures were: 25-45 °C, 0.116 (left lane) and 0.31 (right lane) U; 55 °C, 0.035 (left lane) and 0.116 U (right lane). Control lane was incubated at 25 °C in the absence of P1 nuclease. Dimethyl sulfate reactions were performed as described in Materials and Methods. Roman numerals represent triplet repeat numbers. Arrows indicate sites of P1 nuclease cleavage in the ss$(CCG)_{15}$ hairpin.

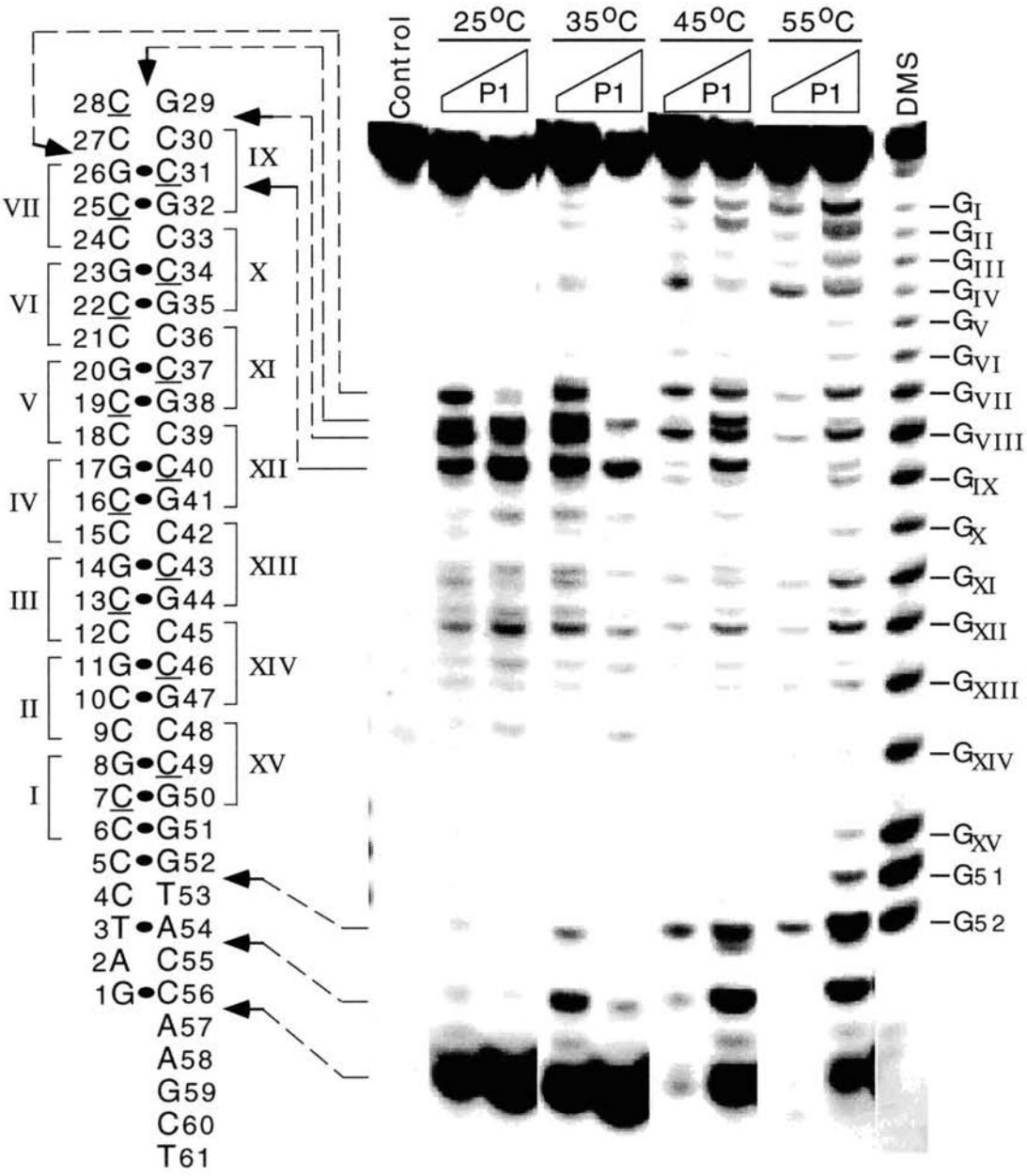

Control
25°C
35°C
45°C
55°C
P1
P1
P1
P1
DMS
28C G29
27C C30
26G•C31
25C•G32
24C C33
23G•C34
22C•G35
21C C36
20G•C37
19C•G38
18C C39
17G•C40
16C•G41
15C C42
14G•C43
13C•G44
12C C45
11G•C46
10C•G47
9C C48
8G•C49
7C•G50
6C•G51
5C•G52
4C T53
3T•A54
2A C55
1G•C56
A57
A58
G59
C60
T61
IX
X
XI
XII
XIII
XIV
XV
VII
VI
V
IV
III
II
I
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G51
G52

Figure 34. P1 nuclease digestion of ss $(CCG)_{15}$ at pH 8.5, 50 mM Na^+

Single-stranded $(CCG)_{15}$ was incubated with P1 nuclease for 5 minutes as described in Materials and Methods at the indicated temperatures. Buffer contained 50 mM Na^+. Reaction products were applied to a 20% polyacrylamide gel containing 8 M urea. The amounts of P1 nuclease added at the respective temperatures were: 37-47°C, 0.116 (left lane) and 0.31 (right lane) U; 52 °C, 0.035 (left lane) and 0.116 U (right lane). Control lane was incubated at 25 °C in the absence of P1 nuclease. Dimethyl sulfate reactions were performed as described in Materials and Methods. Roman numerals represent triplet repeat numbers. Arrows indicate sites of P1 nuclease cleavage in the ss $(CCG)_{15}$ hairpin.

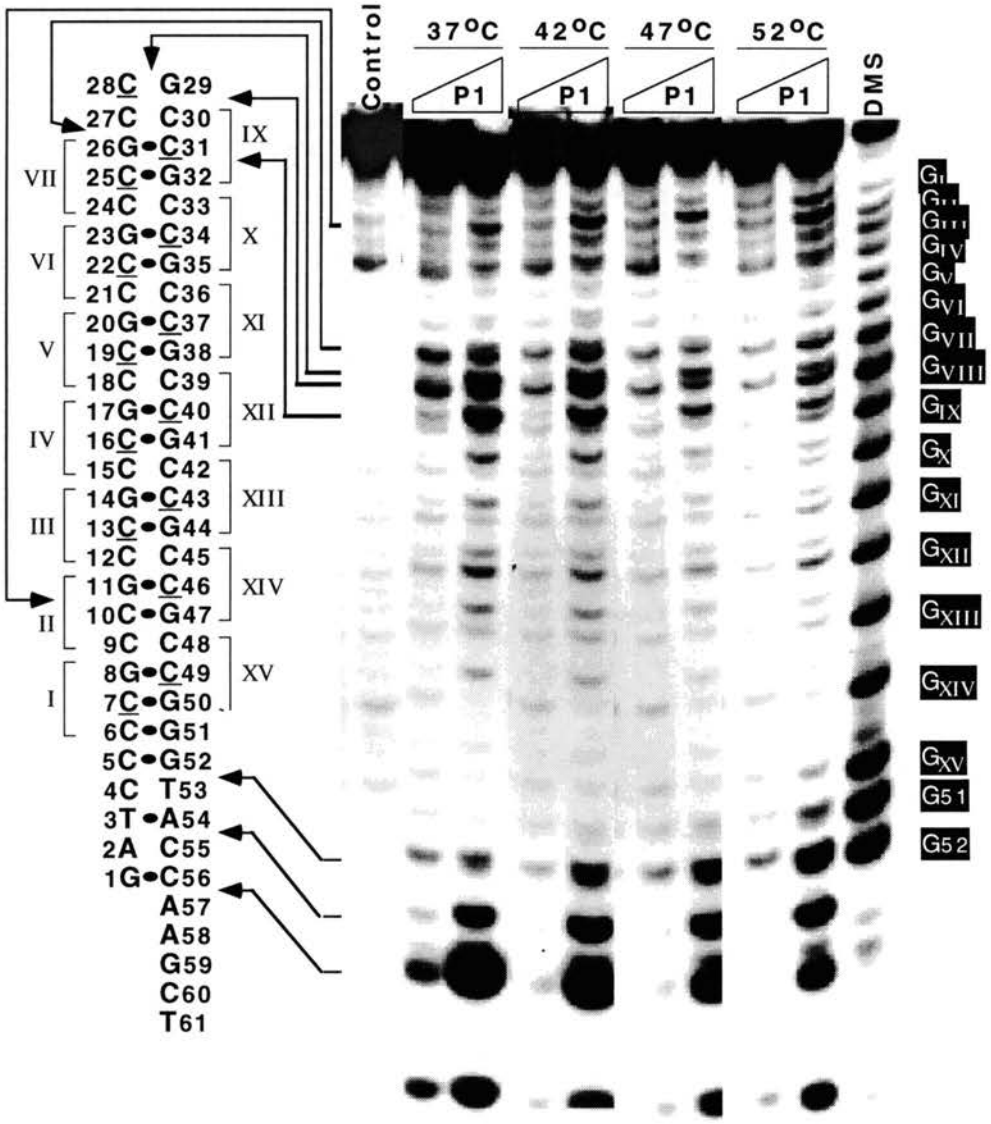

Control
37°C
42°C
47°C
52°C
P1
P1
P1
P1
DMS
28C G29
27C C30
26G•C31
25C•G32
24C C33
23G•C34
22C•G35
21C C36
20G•C37
19C•G38
18C C39
17G•C40
16C•G41
15C C42
14G•C43
13C•G44
12C C45
11G•C46
10C•G47
9C C48
8G•C49
7C•G50
6C•G51
5C•G52
4C T53
3T•A54
2A C55
1G•C56
A57
A58
G59
C60
T61
IX
X
XI
XII
XIII
XIV
XV
VII
VI
V
IV
III
II
I
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G51
G52

Figure 35. P1 nuclease digestion of ss $(CCG)_{15}$ at low temperatures, pH 8.5

Single-stranded $(CCG)_{15}$ was incubated with P1 nuclease at the indicated temperatures as described according to the legend to Fig. 33. Buffer contained 50 mM Na^+, pH 8.5. The amount of P1 nuclease added per reaction was 1.3×10^{-3} U (left lane), 3.9×10^{-3} U, 1.17×10^{-2} U, 3.5×10^{-2} U and 0.105 U (right lane). Incubation times are as follows: 1 °C, 12 h; 10 °C, 1 h; 20 °C, 30 min; 25 °C, 15 min. Control lane was incubated at 25 °C temperature in the absence of P1 nuclease. Arrows indicate sites of P1 nuclease cleavage. Dimethyl sulfate reactions were performed as described in Materials and Methods.

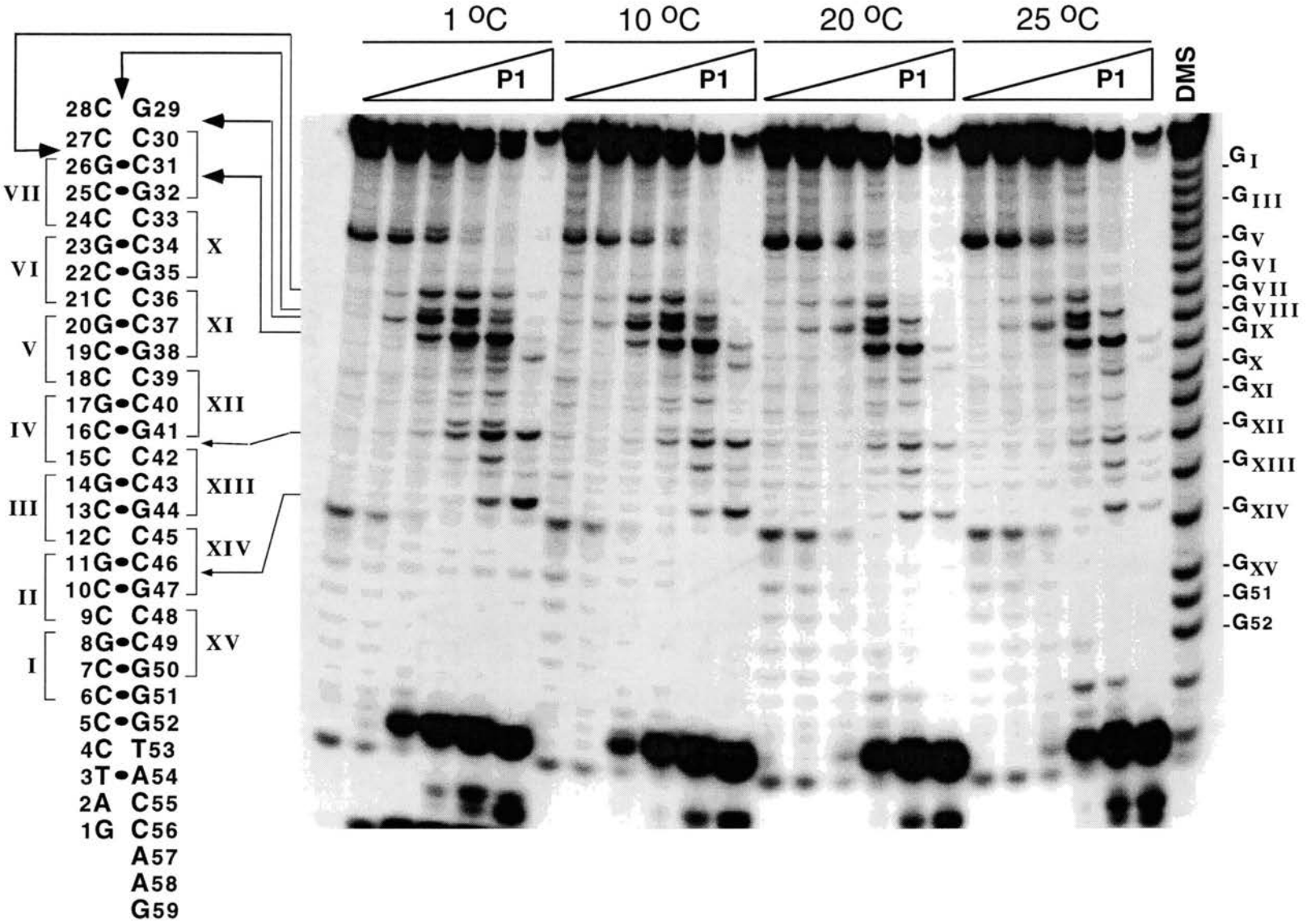

1 °C
10 °C
20 °C
25 °C
P1
P1
P1
P1
DMS
28C G29
27C C30
26G•C31
25C•G32
24C C33
23G•C34
22C•G35
21C C36
20G•C37
19C•G38
18C C39
17G•C40
16C•G41
15C C42
14G•C43
13C•G44
12C C45
11G•C46
10C•G47
9C C48
8G•C49
7C•G50
6C•G51
5C•G52
4C T53
3T•A54
2A C55
1G C56
A57
A58
G59
VII
VI
V
IV
III
II
I
X
XI
XII
XIII
XIV
XV
GI
GIII
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G51
G52

Figure 36. P1 nuclease digestion of $\Delta T_{II}(GTC)_{15}$ at pH 8.5

Single-stranded $\Delta T_{II}(GTC)_{15}$ was incubated with P1 nuclease at the indicated temperatures as described in the legend to Fig. 33. The amounts of P1 nuclease added to each of the reactions was 0.10 U. Control lane was incubated at 25 ^{o}C in the absence of P1 nuclease. The signal in the control lane that corresponds to G_V in the marker lane is a contaminant of ds DNA. Quantitative analysis of the control lane with a Storm Molecular Dynamics PhosphorImager (Sunnyvale, CA) revealed that the contaminant represented <5% of the total DNA.

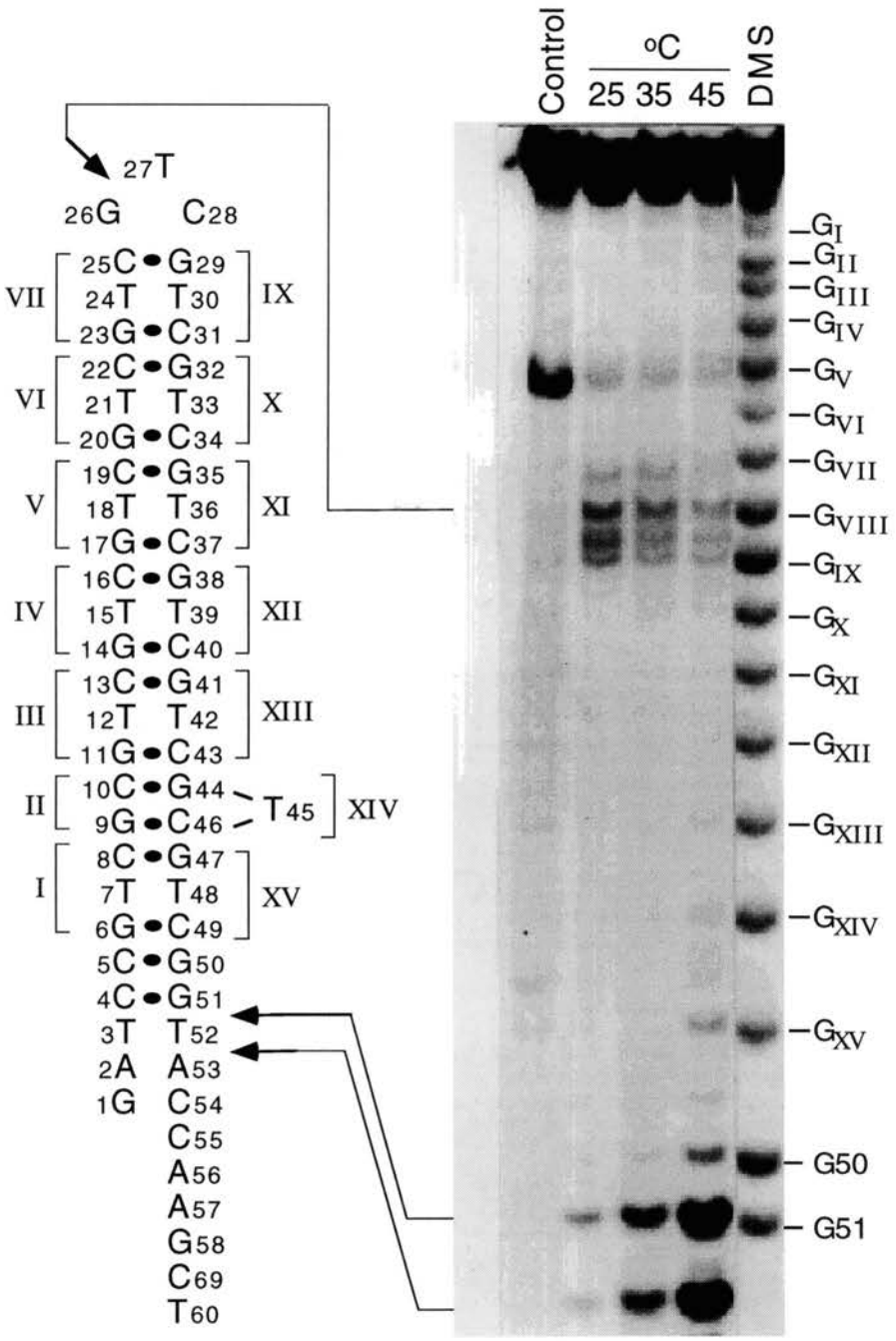

Control
°C
25 35 45
DMS
27T
26G C28
25C G29
24T T30
23G C31
22C G32
21T T33
20G C34
19C G35
18T T36
17G C37
16C G38
15T T39
14G C40
13C G41
12T T42
11G C43
10C G44
9G C46
T45
8C G47
7T T48
6G C49
5C G50
4C G51
3T T52
2A A53
1G C54
C55
A56
A57
G58
C69
T60
VII
IX
VI
X
V
XI
IV
XII
III
XIII
II
XIV
I
XV
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G50
G51

Figure 37. Hydroxylamine modification of ss $(CCG)_{15}$ at pH 6.5

Single-stranded $(CCG)_{15}$ was incubated with HA at the indicated temperatures in buffer containing 50 mM Na^+, pH 6.5 as described according to the legend to Fig. 30. Concentrations of HA in the reactions were as follows: 1, 10, 20, 35, 45, and 55 ^oC, 4.27 M; 15 and 25 ^oC, 5.68 M; 30 ^oC, 3.03 M. Control lane was incubated at 25 ^oC in the absence of HA. The deduced hairpin structure of ss $(CCG)_{15}$ is shown on the left. Dimethyl sulfate (21 mM) reactions were performed as described in Materials and Methods.

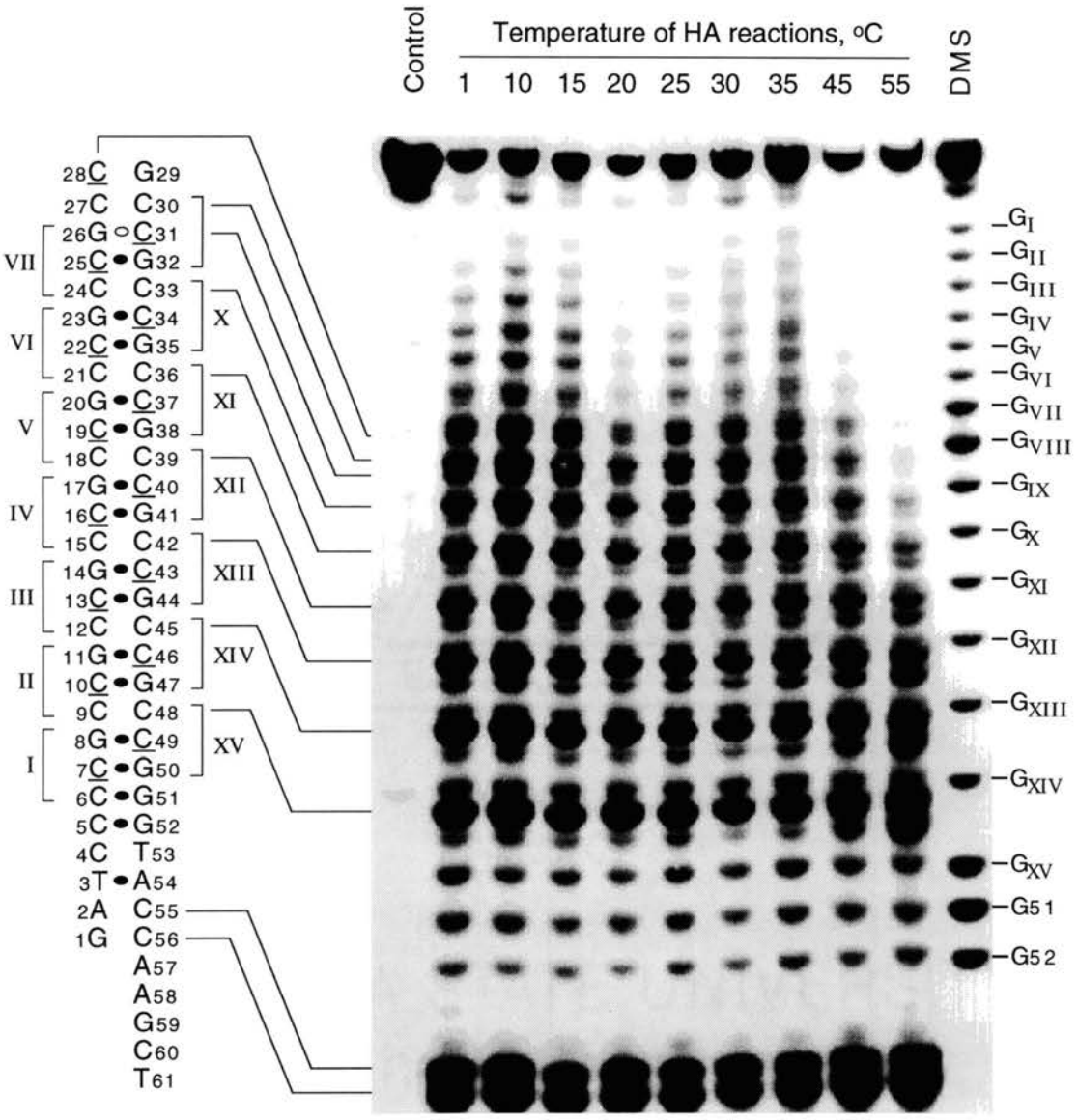

Control
Temperature of HA reactions, °C
1 10 15 20 25 30 35 45 55
DMS
28C G29
27C C30
26G C31
25C G32
24C C33
23G C34
22C G35
21C C36
20G C37
19C G38
18C C39
17G C40
16C G41
15C C42
14G C43
13C G44
12C C45
11G C46
10C G47
9C C48
8G C49
7C G50
6C G51
5C G52
4C T53
3T A54
2A C55
1G C56
A57
A58
G59
C60
T61
VII
VI
V
IV
III
II
I
X
XI
XII
XIII
XIV
XV
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G51
G52

Figure 38. Hydroxylamine modification of ss $(CCG)_{15}$ at pH 7.0

Single-stranded $(CCG)_{15}$ was incubated with HA at 37°C in buffer containing 50 mM Na^+, pH 7.0 as described according to the legend to Fig. 30. Concentrations of HA in the reactions were as follows: 3.03 M 4.27 M, and 5.68 M. Control lane was incubated in the absence of HA. Dimethyl sulfate (21 mM) reactions were performed as described in Materials and Methods.

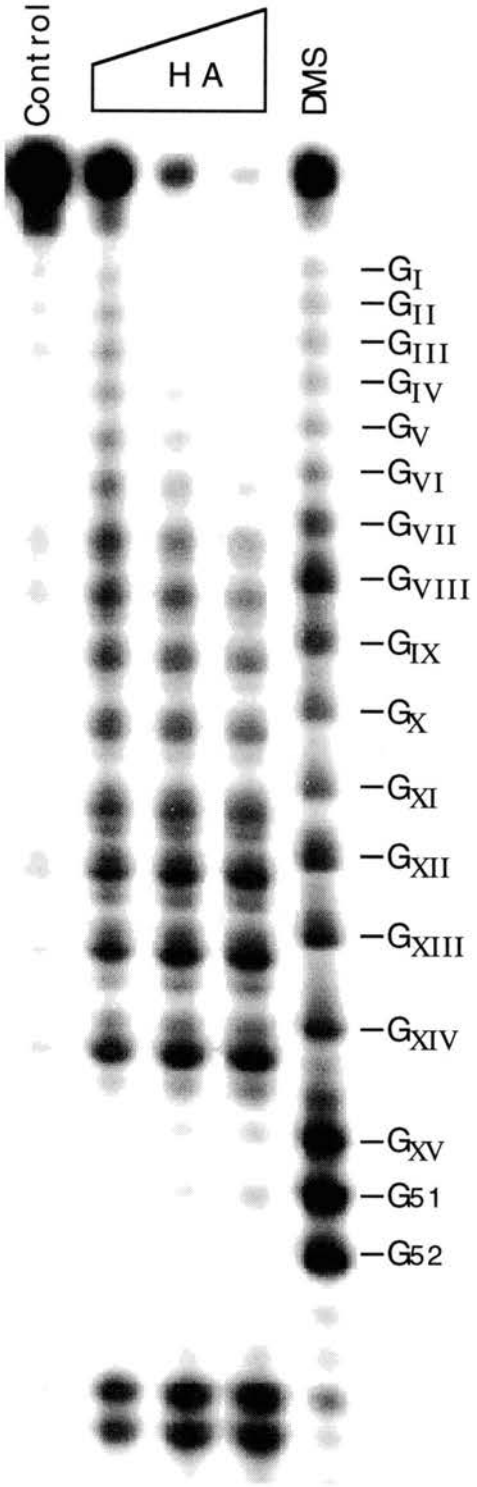

Control
HA
DMS
–GI
–GII
–GIII
–GIV
–GV
–GVI
–GVII
–GVIII
–GIX
–GX
–GXI
–GXII
–GXIII
–GXIV
–GXV
–G51
–G52

Figure 39. Hydroxylamine modification of ss $(CCG)_{15}$ in 400 mM and 600 mM Na^+, pH8.3

Single-stranded $(CCG)_{15}$ was incubated with HA at 25°C in buffer containing 400 mM and 600 mM Na^+, pH 8.3 as described according to the legend to Fig. 30. Concentrations of HA in the reactions were (from left to right) as follows: 3.03 M, 4.27 M, and 5.68 M. Control lane was incubated at 25 oC, 600 mM Na^+ in the absence of HA. Dimethyl sulfate (21 mM) reactions were performed as described in Materials and Methods.

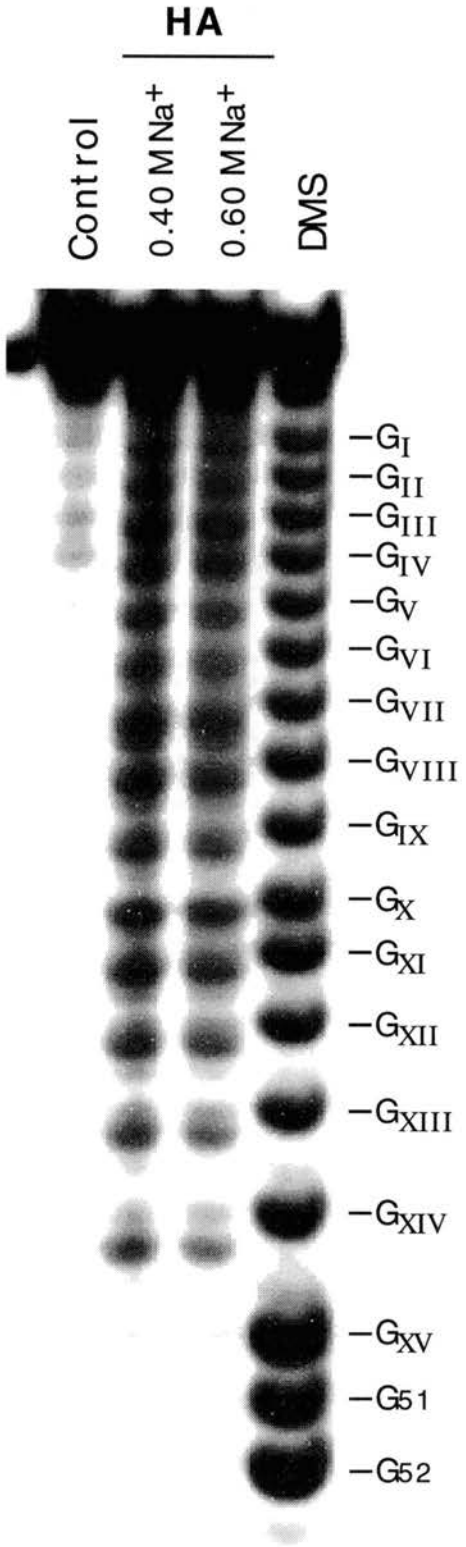

HA
Control
0.40 M Na+
0.60 M Na+
DMS
–GI
–GII
–GIII
–GIV
–GV
–GVI
–GVII
–GVIII
–GIX
–GX
–GXI
–GXII
–GXIII
–GXIV
–GXV
–G51
–G52

Figure 40. P1 nuclease digestion of ss $(CCG)_{15}$ at pH 6.5

Single-stranded $(CCG)_{15}$ was incubated with P1 nuclease at the indicated temperatures as described according to the legend to Fig. 33. Buffer contained 50 mM Na^+, pH 6.5. The amount of P1 nuclease added per reaction was 3.5×10^{-2} U. Incubation times are as follows: control, 5 h; 1 °C, 5 h; 10 °C, 25 min; 20 °C, 15 min; 25 °C, 9 min. Control lane was incubated at 25 °C in the absence of P1 nuclease. Dimethyl sulfate reactions were performed as described in Materials and Methods.

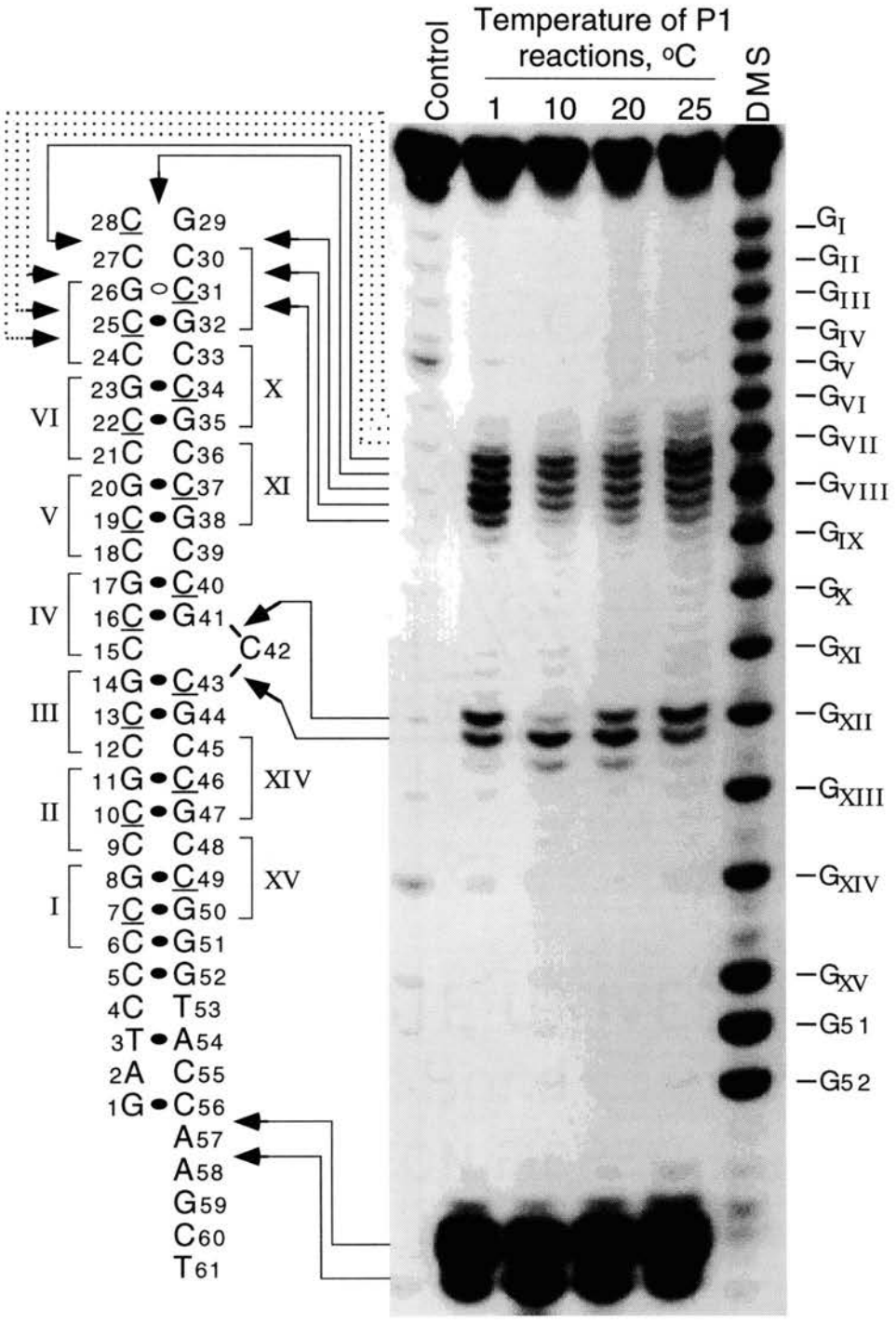

Control
Temperature of P1 reactions, °C
1
10
20
25
DMS
28C G29
27C C30
26G C31
25C G32
24C C33
23G C34
22C G35
21C C36
20G C37
19C G38
18C C39
17G C40
16C G41
15C C42
14G C43
13C G44
12C C45
11G C46
10C G47
9C C48
8G C49
7C G50
6C G51
5C G52
4C T53
3T A54
2A C55
1G C56
A57
A58
G59
C60
T61
VI
V
IV
III
II
I
X
XI
XIV
XV
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G51
G52

Figure 41. P1 nuclease digestion of ss $(CCG)_{15}$ at pH 7.5, various salt concentrations

Single-stranded $(CCG)_{15}$ was incubated with P1 nuclease for 5 minutes as described in Materials and Methods at 37°C in the indicated concentration of NaCl. Reaction products were applied to a 20% polyacrylamide gel containing 8 M urea. The amounts of P1 nuclease added at the respective salt concentration were: 3.9×10^{-3} U (left lane), 1.17×10^{-2} U, 3.5×10^{-2} U and 0.105 U (right lane). Control lane was incubated at 37 °C in the absence of P1 nuclease and Na^+. Dimethyl sulfate reactions were performed as described in Materials and Methods. Roman numerals represent triplet repeat numbers.

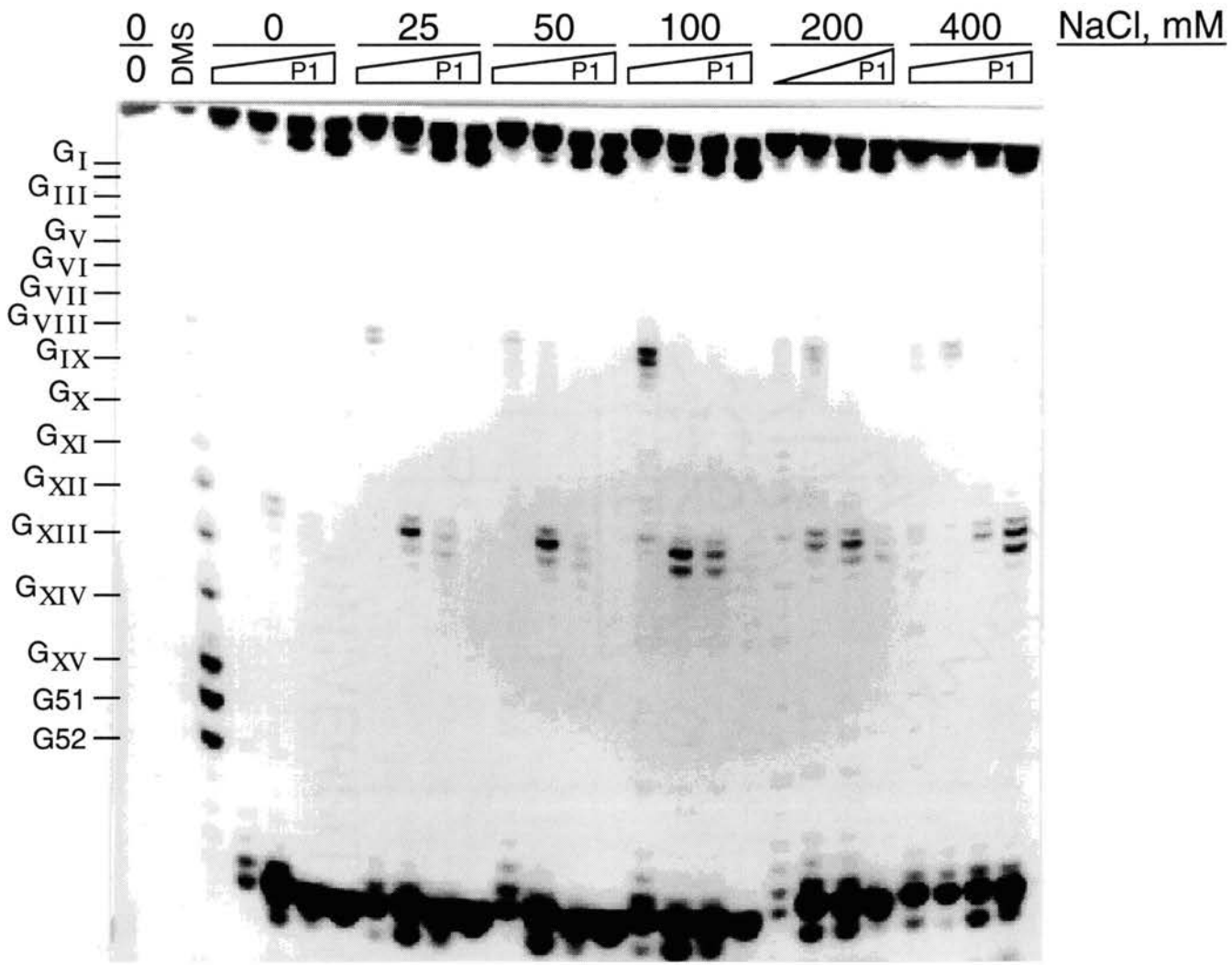

0
0
DMS
0
P1
25
P1
50
P1
100
P1
200
P1
400
P1
NaCl, mM
G I
G III
G V
G VI
G VII
G VIII
G IX
G X
G XI
G XII
G XIII
G XIV
G XV
G51
G52

Figure 42. P1 nuclease digestion of ss$(CCG)_{15}$ at pH 7.5, 50 mM Na^+

Single-stranded $(CCG)_{15}$ was incubated with P1 nuclease for 5 minutes as described in Materials and Methods at the indicated temperatures. Buffer contained 50 mM Na^+. Reaction products were applied to a 20% polyacrylamide gel containing 8 M urea. The amounts of P1 nuclease added at the respective temperatures were: 37-47°C, 0.035 U; 52-57 °C, 0.116 U. Control lane was incubated at 37 °C in the absence of P1 nuclease. Dimethyl sulfate reactions were performed as described in Materials and Methods. Roman numerals represent triplet repeat numbers.

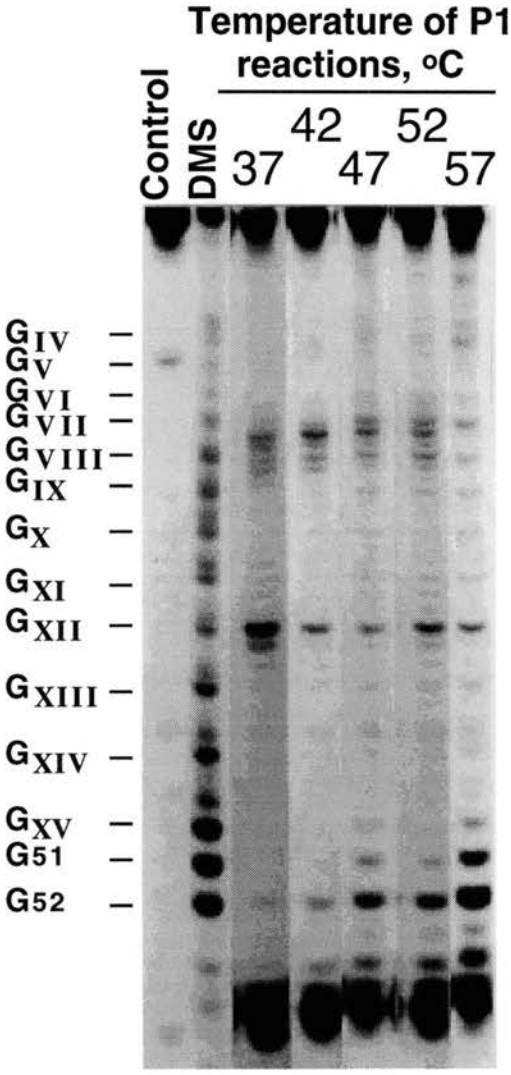

Temperature of P1 reactions, °C
Control
DMS
37
42
47
52
57
G IV
G V
G VI
G VII
G VIII
G IX
G X
G XI
G XII
G XIII
G XIV
G XV
G51
G52

Figure 43. P1 nuclease digestion of ss $(CCG)_{15}$ at various pHs

Single-stranded $(CCG)_{15}$ was incubated with P1 nuclease at 37 ^{o}C, as described in Materials and Methods at the indicated pH. Buffer contained 50 mM Na^{+}. Time of incubations were 5 min (pH 7.5), 7 min (pH 8.0) and 10 min (pH 8.5). Reaction products were applied to a 20% polyacrylamide gel containing 8 M urea. The amounts of P1 nuclease added at the respective pHs were: 4.3×10^{-4} U (left lane), 1.3×10^{-3} U, 3.9×10^{-3} U, 1.17×10^{-2} U, 3.5×10^{-2} U and 0.105 U (right lane). Control lane was incubated at 37 ^{o}C in the absence of P1 nuclease. Dimethyl sulfate reactions were performed as described in Materials and Methods. Roman numerals represent triplet repeat numbers. Arrows indicate sites of P1 nuclease cleavage in the ss$(CCG)_{15}$ hairpin.

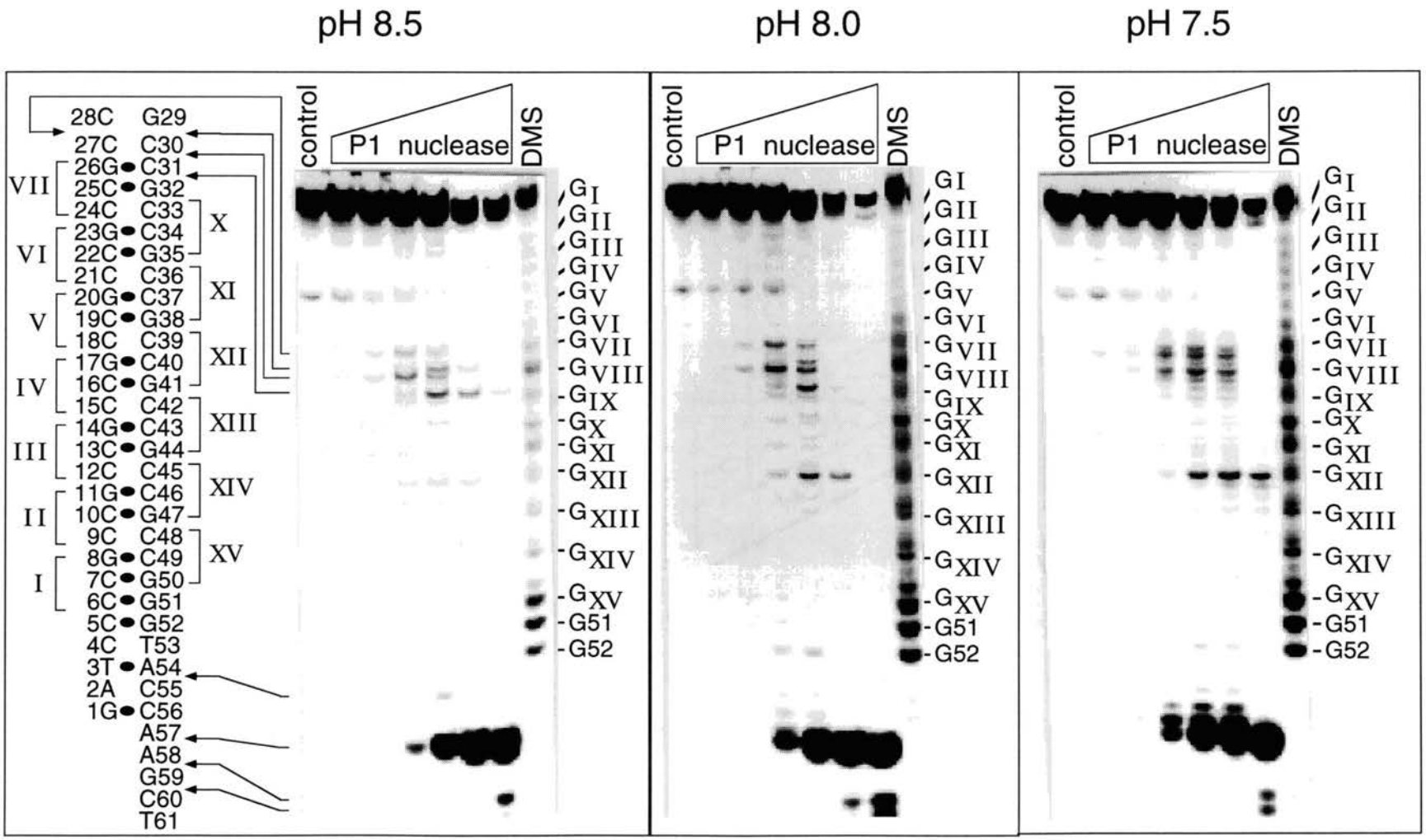
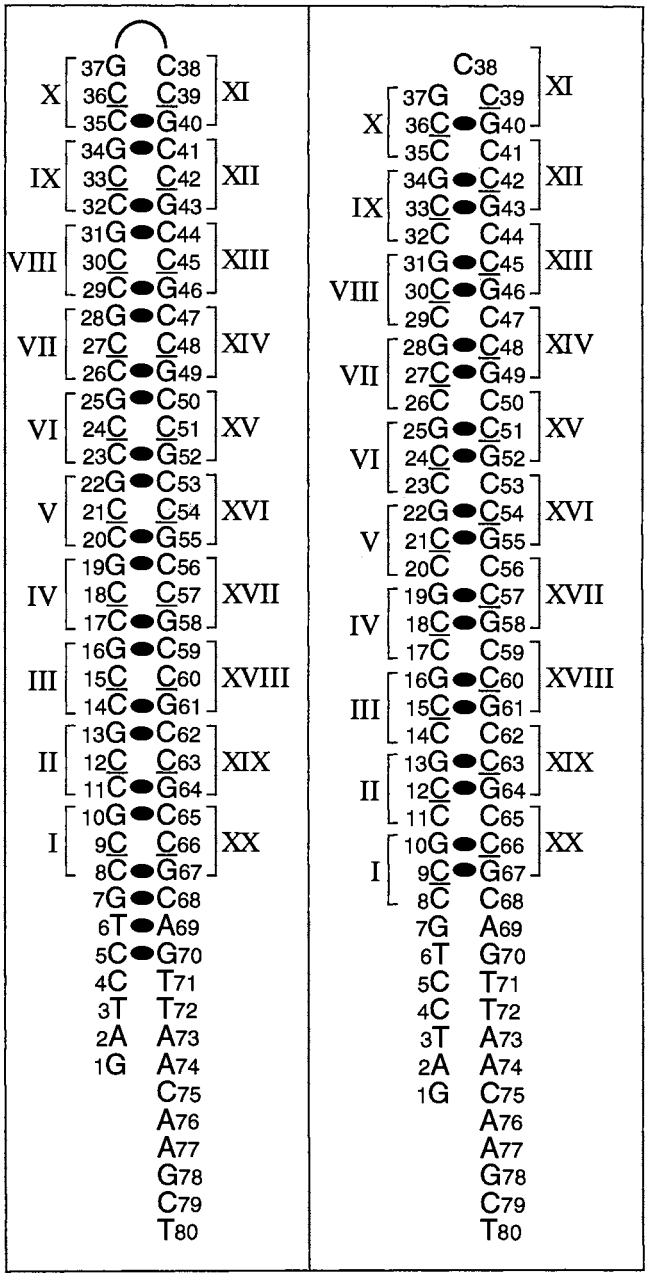

pH 8.5
pH 8.0
pH 7.5
control
P1 nuclease
DMS
28C G29
27C C30
26G C31
25C G32
24C C33
23G C34
22C G35
21C C36
20G C37
19C G38
18C C39
17G C40
16C G41
15C C42
14G C43
13C G44
12C C45
11G C46
10C G47
9C C48
8G C49
7C G50
6C G51
5C G52
4C T53
3T A54
2A C55
1G C56
A57
A58
G59
C60
T61
VII
VI
V
IV
III
II
I
X
XI
XII
XIII
XIV
XV
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G51
G52

Figure 44. Potential hairpin alignments of ss$(CCG)_{20}$

Two potential hairpin alignments of ss$(CCG)_{20}$ are shown. Filled ovals indicate Watson-Crick C•G base-pairs. Roman numerals depict a particular triplet repeat. Conventional Arabic numerals depict the position of a nucleotide with respect to the 5'-end. Various features of the hairpins are listed.



Hairpin alignment	*(a)*	*(e)*
HA reactivity	5'C > 3'C	3'C > 5'C
Loop size	4 nt	3 nt
# of Watson-Crick base pairs	22	19

Figure 45. Chemical modification of ss $(CCG)_{20}$ with Hydroxylamine

Single-stranded $(CCG)_{20}$ was incubated with hydroxylamine (HA) at 37 ^{o}C in 50 mM Na^{+} as described according to the legend to Fig. 30. The amount of HA in each reaction was 6.3 M. Time of incubations were 25 min (pH 7.5) and 40 min (pH 8.3). The deduced hairpin structure of $ss(CCG)_{15}$ is shown on the left. The ^{32}P labels in the oligonucleotide are 5' to A77 and C79. The band corresponding to C8 is marked with an "*" to the left of the figure. Dimethyl sulfate (21 mM) reactions were performed as described in Materials and Methods.

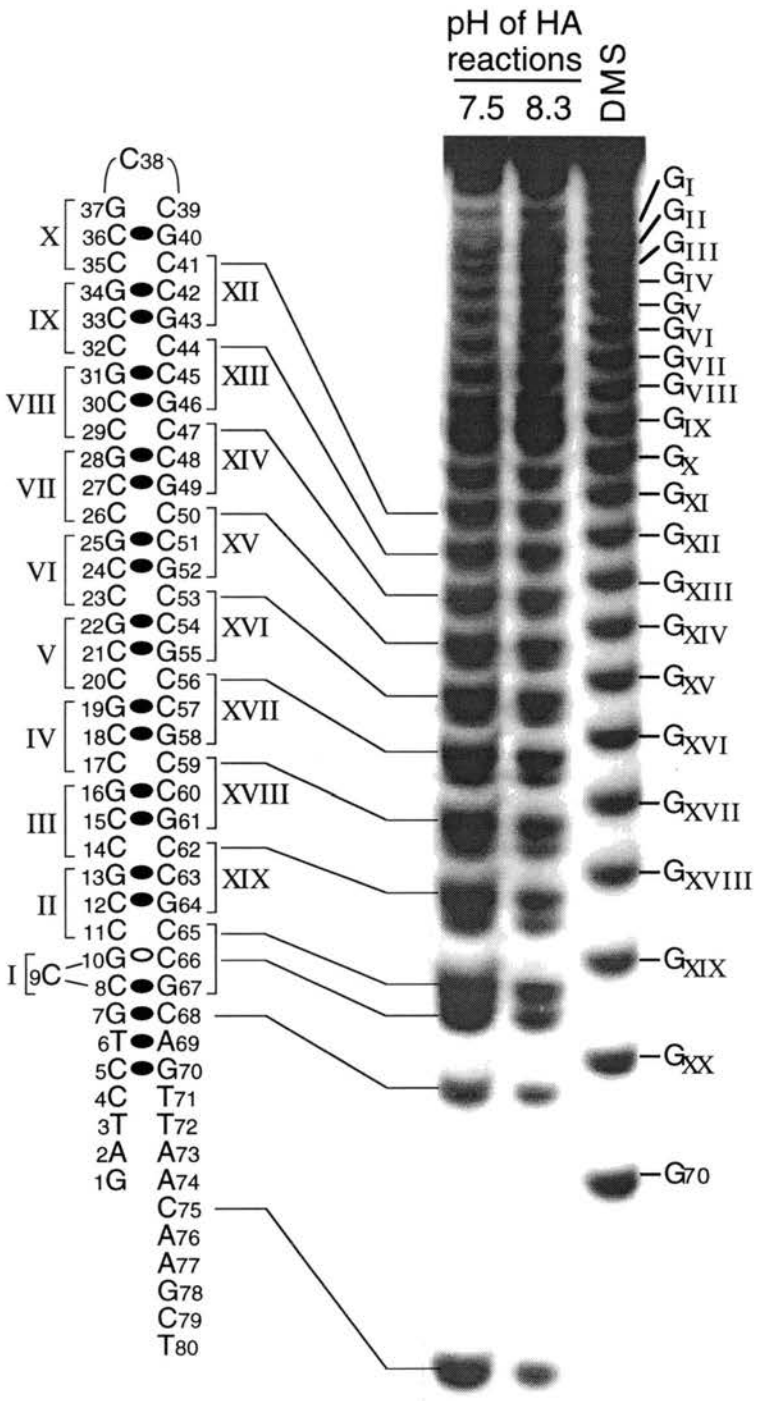
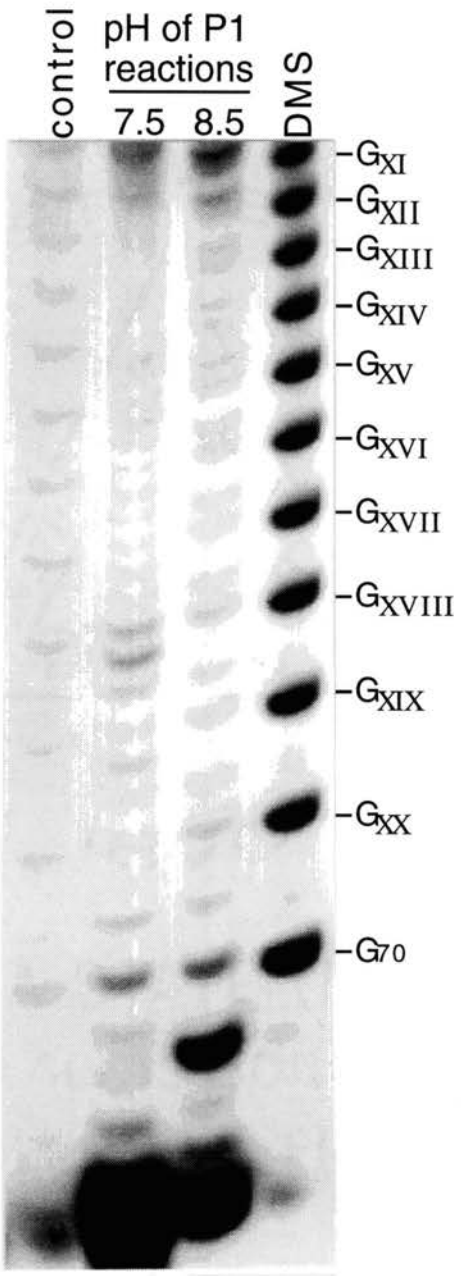

pH of HA reactions
7.5
8.3
DMS
C38
37G C39
36C G40
35C C41
34G C42
33C G43
32C C44
31G C45
30C G46
29C C47
28G C48
27C G49
26C C50
25G C51
24C G52
23C C53
22G C54
21C G55
20C C56
19G C57
18C G58
17C C59
16G C60
15C G61
14C C62
13G C63
12C G64
11C C65
10G C66
9C
8C G67
7G C68
6T A69
5C G70
4C T71
3T T72
2A A73
1G A74
C75
A76
A77
G78
C79
T80
X
IX
VIII
VII
VI
V
IV
III
II
I
XII
XIII
XIV
XV
XVI
XVII
XVIII
XIX
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
GXVI
GXVII
GXVIII
GXIX
GXX
G70

Figure 46. P1 nuclease analysis of ss $(CCG)_{20}$

Single-stranded $(CCG)_{20}$ was incubated with P1 nuclease for 5 min at the 37 ^{o}C (pH as indicated in the figure) as described according to the legend to Fig. 33. The amounts of P1 nuclease added were 0.10 (pH 7.5) and 0.31 (pH 8.5) U. Time of incubations were 10 min (pH 7.5) and 27 min (pH 8.5). The deduced hairpin structure of ss$(CCG)_{15}$ is shown on the left. Dimethyl sulfate reactions were performed as described in Materials and Methods.


control
pH of P1 reactions
7.5
8.5
DMS
GXI
GXII
GXIII
GXIV
GXV
GXVI
GXVII
GXVIII
GXIX
GXX
G70

CHAPTER VIII

THE PURINE-RICH TRINUCLEOTIDE REPEAT SEQUENCES ss $(CAG)_{15}$ AND ss $(GAC)_{15}$ FORM HAIRPIN

We have described the hairpin structures of ss $(CTG)_{15}$ and ss $(GTC)_{15}$ in previous chapters. Their complementary strands, ss $(CAG)_{15}$ and ss $(GAC)_{15}$ were predicted by enegy minimization to form hairpin structures (Mitas et al., 1995a.) Studies of the properties of the putative hairpin structures of $(CAG)_{15}$ and $(GAC)_{15}$ and comparison of their stabilities might provide information regarding TREDs. The structures of single-stranded oligonucleotides containing $(CAG)_{15}$ or $(GAC)_{15}$ were examined (Yu et al., 1995b).

P1 nuclease digestions indicate that the structures of ss$(CAG)_{15}$ and ss$(GAC)_{15}$ are hairpins

To investigate the structures of ss$(CAG)_{15}$ and ss$(GTA)_{15}$, digestions were performed with single-strand-specific P1 nuclease. Incubation of ss$(CAG)_{15}$ with P1 nuclease did not result in cleavage of the phosphodiester bonds between triplets IX and XV, indicating that the nucleotides within these triplets participated in base-pairing and/or base-stacking interactions (Figure 47). The sites of P1 cleavage were G32-C33, C33-A34 phosphodiester bonds. A minor site of P1 cleavage was the G35-C36 phosphodiester bond. These results indicate that ss $(CAG)_{15}$ formed a hairpin structure.

Similar to ss $(CAG)_{15}$, incubation of ss $(GAC)_{15}$ with P1 did not result in cleavage of phosphodiester bonds between triplet IX and XV (Figure 48). The main sites of P1 nuclease cleavage were G33-G36 phosphodiester bonds. These results provide direct evidence for a hairpin structure for ss $(GAC)_{15}$.

Further examination of the structure was performed with a mutated ss $(GAC)_{15}$, termed ΔA_{XIV} $(GAC)_{15}$, in which an adenine residue in triplet XIV was deleted, which caused the A of triplet repeat unit II to be extruded out of the helix of the proposed hairpin stem region. P1 nuclease digestion showed the same cleavage sites with its wild type sequence (Figure 49). The main cleavage sites were G33-G36 phosphodiester bonds. P1 nuclease failed to cleave the phosphodiesters of the extrahelical adenine in repeat unit II, indicating that the extrahelical adenine might have been involved in base stacking interactions.

DEPC modification shows a stable hairpin structure of ss $(GAC)_{15}$.

To investigate the properties of potential A-A mismatches in the proposed hairpin structures of ss $(CAG)_{15}$ and ss $(GAC)_{15}$, diethyl pyrocarbonate (DEPC) was used. Diethyl pyrocarbonate reacts with adenine residues to produce ring-opened products (Leonard et al., 1971, Vincze et al., 1973). The major site of attack is the nitrogen at position 7 accessed via the major groove of B-DNA. DEPC is however a sterically bulky molecule, and its degree of reactivity is severely restricted by the accessibility of the target molecules. However, at sufficiently high concentrations, DEPC produces an adenine ladder in reactions with B-DNA (Herr et al, 1982); when DNA is in a non-B DNA conformation, the reactivity to DEPC is largely enhanced (Nordheim et al., 1984, Johnston and Rich, 1985, Runkel and Nordheim, 1986). It has been observed that adenine bases in the unpaired loop region of cruciform structures were hyperreactive to

DEPC (Furlong and Lilley, 1986, Scholten and Nordheim, 1986). We suspected that DEPC might be a good reagent to analyze the putative hairpin structures of ss $(CAG)_{15}$ and ss $(GAC)_{15}$, since the adenine in the loop might be much more accessible to DEPC compared to those in the stem.

I first examined the reactivity of DEPC with ΔA_{XIV} $(GAC)_{15}$. In 50 mM Na^+ at 30°C, the adenines in triplets II (the triplet opposite triplet XIV in the hairpin structure), VIII and IX (the triplets in the loop region), were highly reactive with DEPC (Figure 50). In contrast, all other adenines in the stem region of ΔA_{XIV} $(GAC)_{15}$ only partially reacted with DEPC, indicating that N^7 of adenine was involved in base-pairing and/or base stacking interactions. The results correspond to a hairpin structure with an alignment such that triplet repeat I pairs with triplet repeat XV, triplet II pairs with triplet XIV, ..., and triplet repeat VIII as the hairpin loop region. The hairpin structure is in agreement with the hairpin structure proposed from P1 nuclease studies.

Potassium (K^+) stabilizes ss $(GAC)_{15}$ hairpin.

To investigate the nature of the A-A interaction in the hairpin, the heat stability of the hairpin was investigated. DEPC modification of ΔA_{XIV} $(GAC)_{15}$ was carried out in various temperatures (Figure 50). In 50 mM Na^+, no added K^+, pH 7.0, the adenines in triplets II, VIII and IX were highly reactive with DEPC in a temperature range from 30°C to 50°C. When the temperature was increased to 60°C or higher, the sensitivity of the adenines to DEPC through all repeats tended to be the same, indicating the hairpin became disrupted at elevated temperatures. The results suggest that the melting temperature of the hairpin was between 50°C~ 60°C under this condition (Figure 50).

Similar experiments were performed in 50 mM Na^+, 150 mM K^+, pH 7.0. The adenines in triplets II, VIII and IX were highly reactive with DEPC, even up to 60°C (Figure 51). When the DEPC modifying reaction was carried out in a similar condition with 500 mM K^+, the hairpin structure was disrupted only at 70°C, suggesting that the hairpin melted between 60°C~70°C. K^+ significantly stabilized hairpin structures of ΔA_{XIV} $(GAC)_{15}$ (Figure 52).

DEPC modification demonstrates that ss $(CAG)_{15}$ forms a less stable hairpin.

DEPC modification with ss $(CAG)_{15}$ was performed to obtain more evidence to support the hairpin structure derived from P1 nuclease studies. In 50 mM Na^+, 150 mM K^+, pH 7.0, 30°C, ss $(CAG)_{15}$ showed two highly sensitive regions to DEPC (Figure 53). One region contained the adenines in triplets VII and VIII. Another highly sensitive site was 9A in the terminal unpaired region. Other adenines in the repeat region were only partially reactive to DEPC. The results demonstrate that ss $(CAG)_{15}$ forms a hairpin structure with an alignment similar to the hairpin of ss $(GAC)_{15}$.

To further study the nature of A-A pairs in the hairpin structure, DEPC modifications of ss$(CAG)_{15}$ were carried out in 50 mM Na^+, 150 mM K^+, pH 7.0, at various temperatures (Figure 53). At 30°C, the hairpin was stable, with an observation that triplet VII and VIII, as well as 9A, were highly sensitive to DEPC. At 40°C, however, the adenines of the hairpin stem region become more sensitive to DEPC. When temperature increased to 50°C or more, all adenines of ss $(CAG)_{15}$ had the same sensitivity to DEPC. The results demonstrated that ss $(CAG)_{15}$ form hairpin structure with a lower stability compared to ss $(GAC)_{15}$.

The above results demonstrated that ss $(GAC)_{15}$ and ss $(CAG)_{15}$ formed hairpin structures with similar alignments but diferent thermal stabilities. To estimate the melting temperatures of the ss $(GAC)_{15}$ and ss $(CAG)_{15}$ hairpins, the ratio of stem/loop band densities as a function of temperature was plotted. ss $(GAC)_{15}$ has a melting temperature of 54°C (Figure 54). In contrast, the melting curve of ss $(CAG)_{15}$ is flat and extends over a wide temperature range. This melting curve does not give a definite Tm, indicating the A-A pairs in the ss $(CAG)_{15}$ are not stable.

Discussion

We concluded from above results that both ss $(GAC)_{15}$ and ss $(CAG)_{15}$ formed hairpin structures. The ss $(GAC)_{15}$ hairpin structure was more stable than the hairpin structure of ss $(CAG)_{15}$, with a 10°C difference in their melting temperatures.

Our studies show that ss $(GAC)_{15}$ forms a hairpin of thermal stability equivalent to ss $(CTG)_{15}$, while the ss $(CAG)_{15}$ hairpin has a lower thermal stability. The results of these studies, in combination with EMMP data (A. Yu et al., 1995b) and energy minimization (Mitas et al., 1995a) suggest that A-A pairs in ss $(GAC)_{15}$ are more stable than A-A pairs in ss $(CAG)_{15}$. Potassium permanganate oxidation (Mitas et al., 1995a, A. Yu et al., 1995a) and [^{1}H]NMR studies (Gacy et al., 1995) (G. Gupta, Los Alamos National Laboratory; X. Gao, Univ. Houston, pers. comm.) have shown that the T-T pairs in hairpins formed from (CTG)n are hydrogen bonded (Fig. 55). ^{1}H NMR studies of the duplex structure of GCGACGC have confirmed that the the A-A pairs are H-bonded(Mariappan, S., I. Haworth, M. Mitas, G. Gupta, Unpublished results). Therefore, it is likely that the A-A pairs in ss$(GAC)_{15}$ contain hydrogen bonds.

These hydrogen bonds, plus the base stacking energy, contribute to the thermal stability of ss $(GAC)_{15}$ hairpin. In comparison, the A·A bases of ss $(CAG)_{15}$ in stem region may not paired by H-bonds.

Figure 47. P1 nuclease digestion of ss$(CAG)_{15}$

Sigle-stranded $(CAG)_{15}$ was digested with various amounts (between 1.15×10^{-2} and 3.46×10^{-2} U) of P1 nuclease as described in Materials and Methods. For size markers, dimethyl sulfate (DMS) reactions were performed at 21 mM with the DNA essentially according to the method of Maxam and Gilbert (1980). Roman numerals represent triplet repeat numbers. Arrows indicates sites of P1 nuclease cleavage. Reaction products were applied to a 20% polyacrylamide sequencing gel containing 8 M urea.

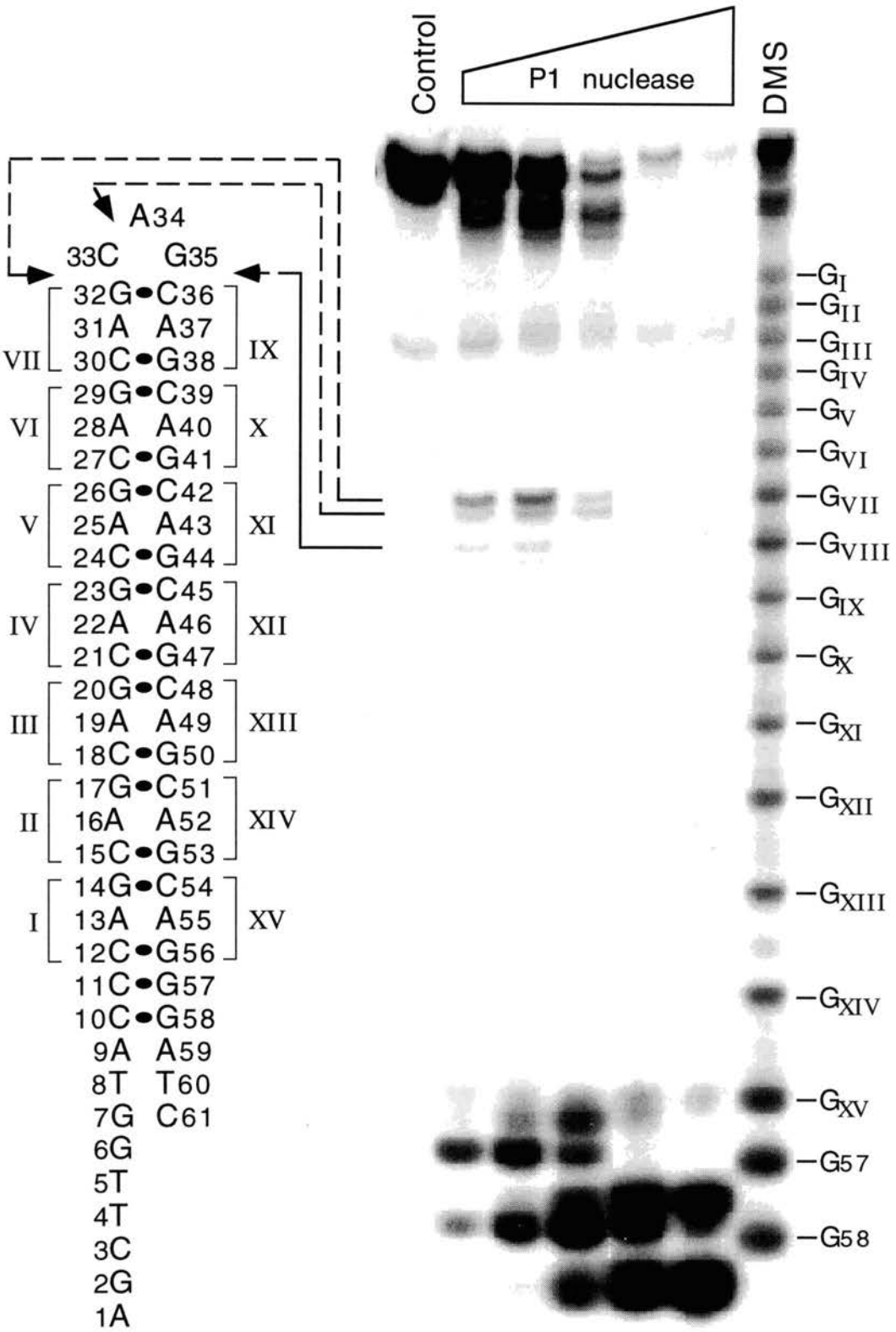

Control
P1 nuclease
DMS
A34
33C G35
32G•C36
31A A37
VII 30C•G38 IX
29G•C39
VI 28A A40 X
27C•G41
26G•C42
V 25A A43 XI
24C•G44
23G•C45
IV 22A A46 XII
21C•G47
20G•C48
III 19A A49 XIII
18C•G50
17G•C51
II 16A A52 XIV
15C•G53
14G•C54
I 13A A55 XV
12C•G56
11C•G57
10C•G58
9A A59
8T T60
7G C61
6G
5T
4T
3C
2G
1A
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G57
G58

Figure 48. P1 nuclease digestion of ss$(GAC)_{15}$

Sigle-stranded $(GAC)_{15}$ was digested with various amounts (from left to right: 1.28×10^{-3}, 3.83×10^{-3}, 1.15×10^{-2}, 3.46×10^{-2} and 0.104 U) of P1 nuclease as described in Materials and Methods. See the legend to Figure 47 for other details.

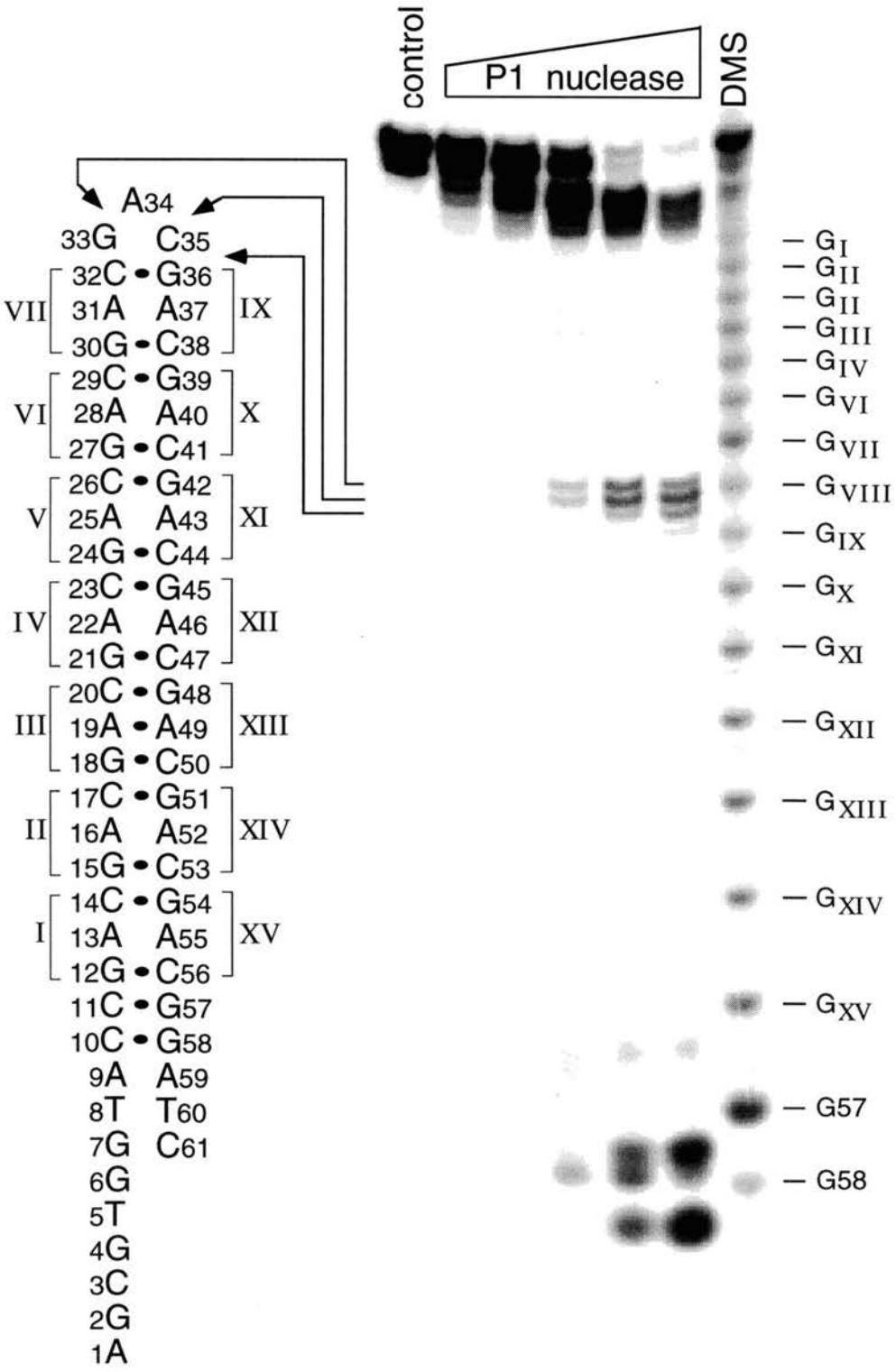

control
P1 nuclease
DMS
A34
33G C35
32C • G36
31A A37
30G • C38
29C • G39
28A A40
27G • C41
26C • G42
25A A43
24G • C44
23C • G45
22A A46
21G • C47
20C • G48
19A • A49
18G • C50
17C • G51
16A A52
15G • C53
14C • G54
13A A55
12G • C56
11C • G57
10C • G58
9A A59
8T T60
7G C61
6G
5T
4G
3C
2G
1A
VII
IX
VI
X
V
XI
IV
XII
III
XIII
II
XIV
I
XV
GI
GII
GII
GIII
GIV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G57
G58

Figure 49. P1 nuclease digestion of ss ΔA_{XIV} $(GAC)_{15}$

Sigle-stranded $(GAC)_{15}$ was digested with various amounts (from left to right: 1.28×10^{-3}, 3.83×10^{-3}, 1.15×10^{-2}, 3.46×10^{-2} and 0.104 U) of P1 nuclease as described in Materials and Methods. See the legend to Figure 47 for other details.

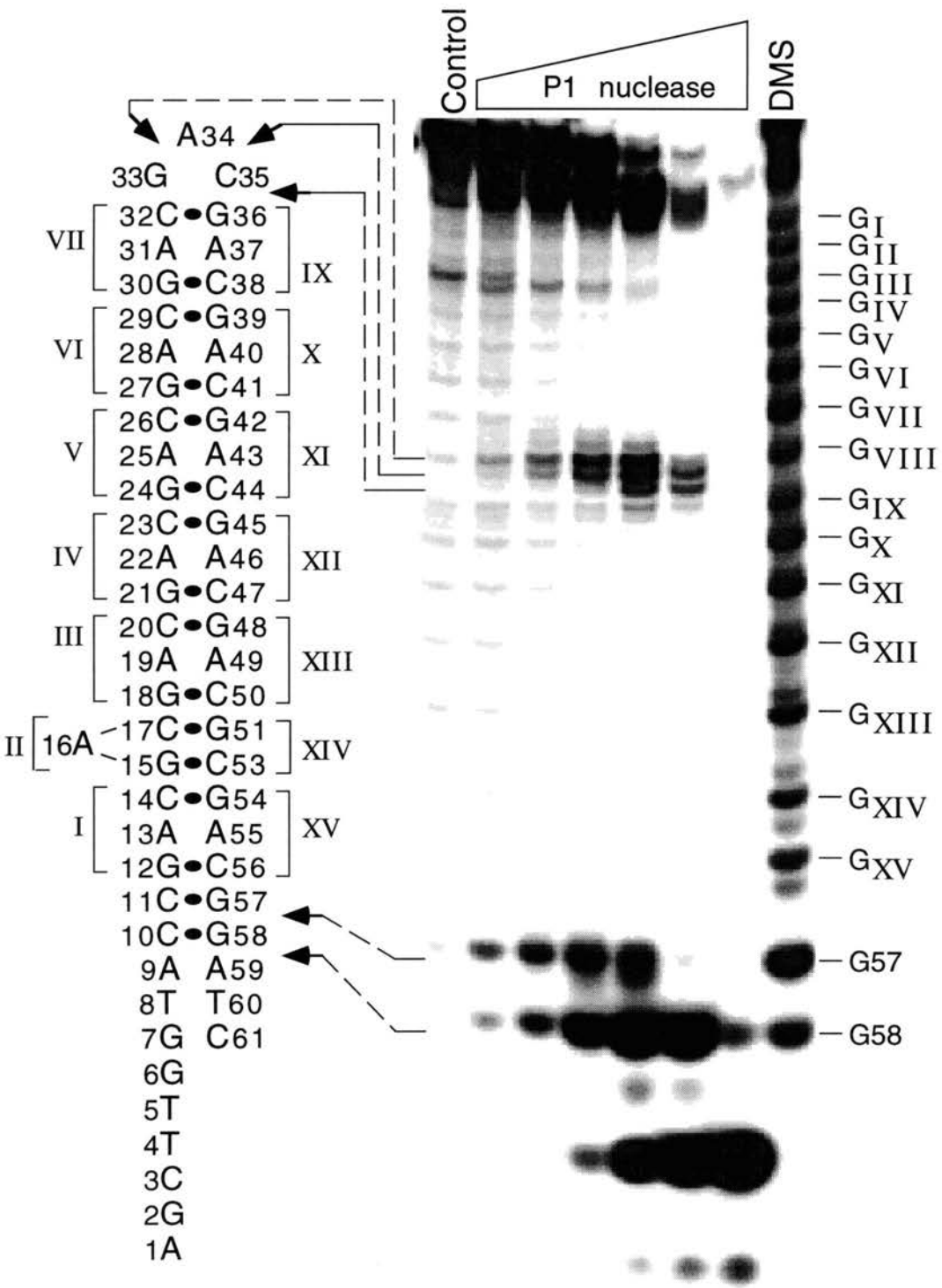

Control
P1 nuclease
DMS
A34
33G C35
VII
32C•G36
31A A37
30G•C38
IX
VI
29C•G39
28A A40
27G•C41
X
V
26C•G42
25A A43
24G•C44
XI
IV
23C•G45
22A A46
21G•C47
XII
III
20C•G48
19A A49
18G•C50
XIII
II
16A
17C•G51
15G•C53
XIV
I
14C•G54
13A A55
12G•C56
XV
11C•G57
10C•G58
9A A59
8T T60
7G C61
6G
5T
4T
3C
2G
1A
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G57
G58

Figure 50. Diethyl pyrocarbonate modification of ΔA_{XIV} $(GAC)_{15}$ in 50 mM Na^+

DEPC modification of ΔA_{XIV} $(GAC)_{15}$ was performed as described in materials and methods at the indicated temperatures. The concentration of the monovalent cations in the reactions were 50 mM Na^+, no added K^+. The concentrations of DEPC in the reactions were 0.96 (30 and 40°C), and 0.27 (50-70°C) M. The control reaction was performed at 30°C and did not contain DEPC. Positions of guanine residues determined by chemical modification with 21 mM dimethyl sulfate (DMS) are shown on the right. Arrows indicate adenine residues preferentially modified at 30°C. The deduced structure of $\Delta A_{XIV}(GAC)_{15}$ at 30°C is shown on the left.

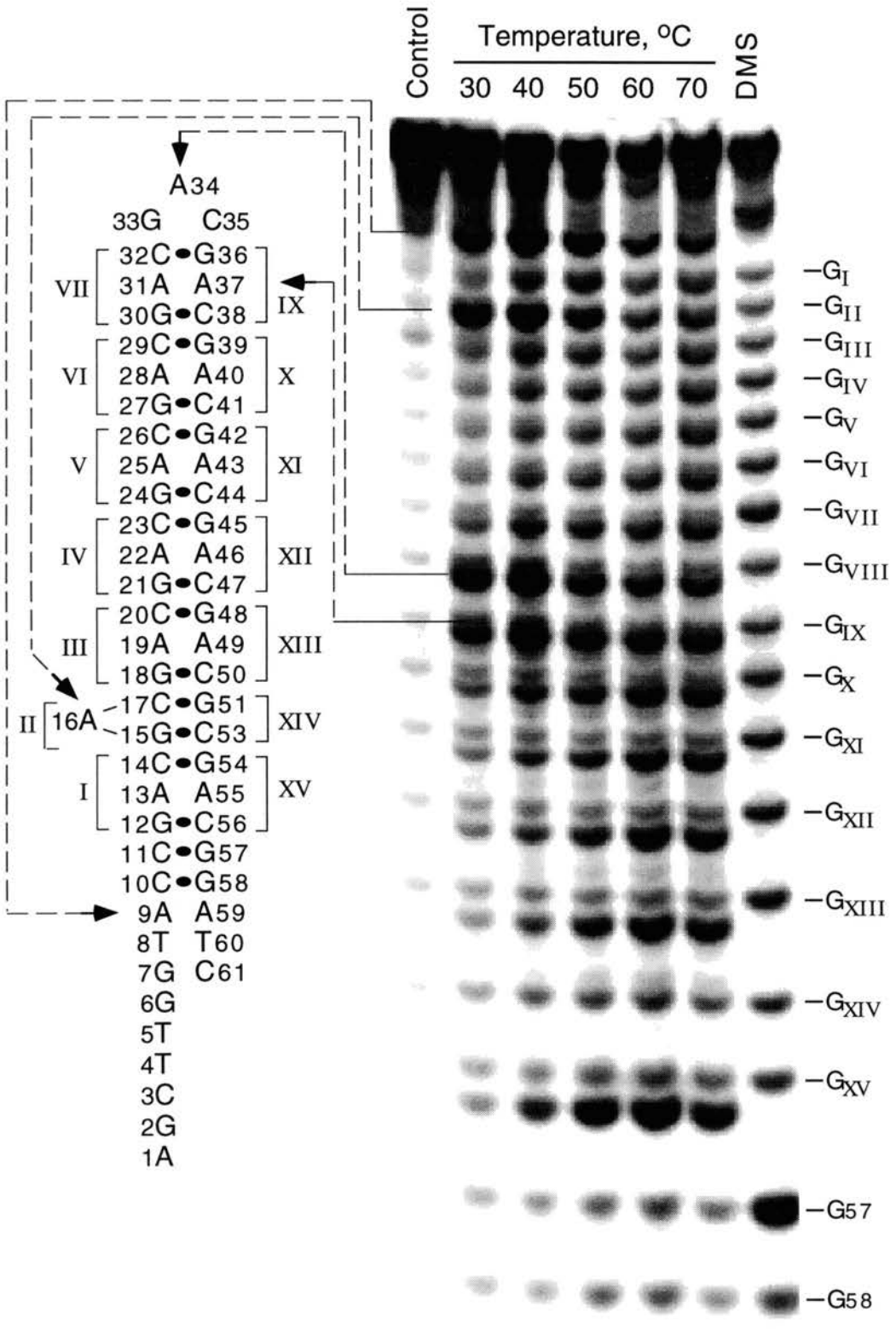

Control
Temperature, °C
30 40 50 60 70
DMS
A34
33G C35
VII 32C•G36 31A A37 30G•C38 IX
VI 29C•G39 28A A40 X 27G•C41
V 26C•G42 25A A43 XI 24G•C44
IV 23C•G45 22A A46 XII 21G•C47
III 20C•G48 19A A49 XIII 18G•C50
II 16A 17C•G51 15G•C53 XIV
I 14C•G54 13A A55 XV 12G•C56
11C•G57
10C•G58
9A A59
8T T60
7G C61
6G
5T
4T
3C
2G
1A
–GI
–GII
–GIII
–GIV
–GV
–GVI
–GVII
–GVIII
–GIX
–GX
–GXI
–GXII
–GXIII
–GXIV
–GXV
–G57
–G58

Figure 51. Diethyl pyrocarbonate modification of ΔA_{XIV} $(GAC)_{15}$ in 50 mM Na^+, 150 mM K^+.

DEPC modification of ΔA_{XIV} $(GAC)_{15}$ was performed as described in the legend to figure 50, except that the salt concentration was 50 mM Na^+, 150 mM K^+.

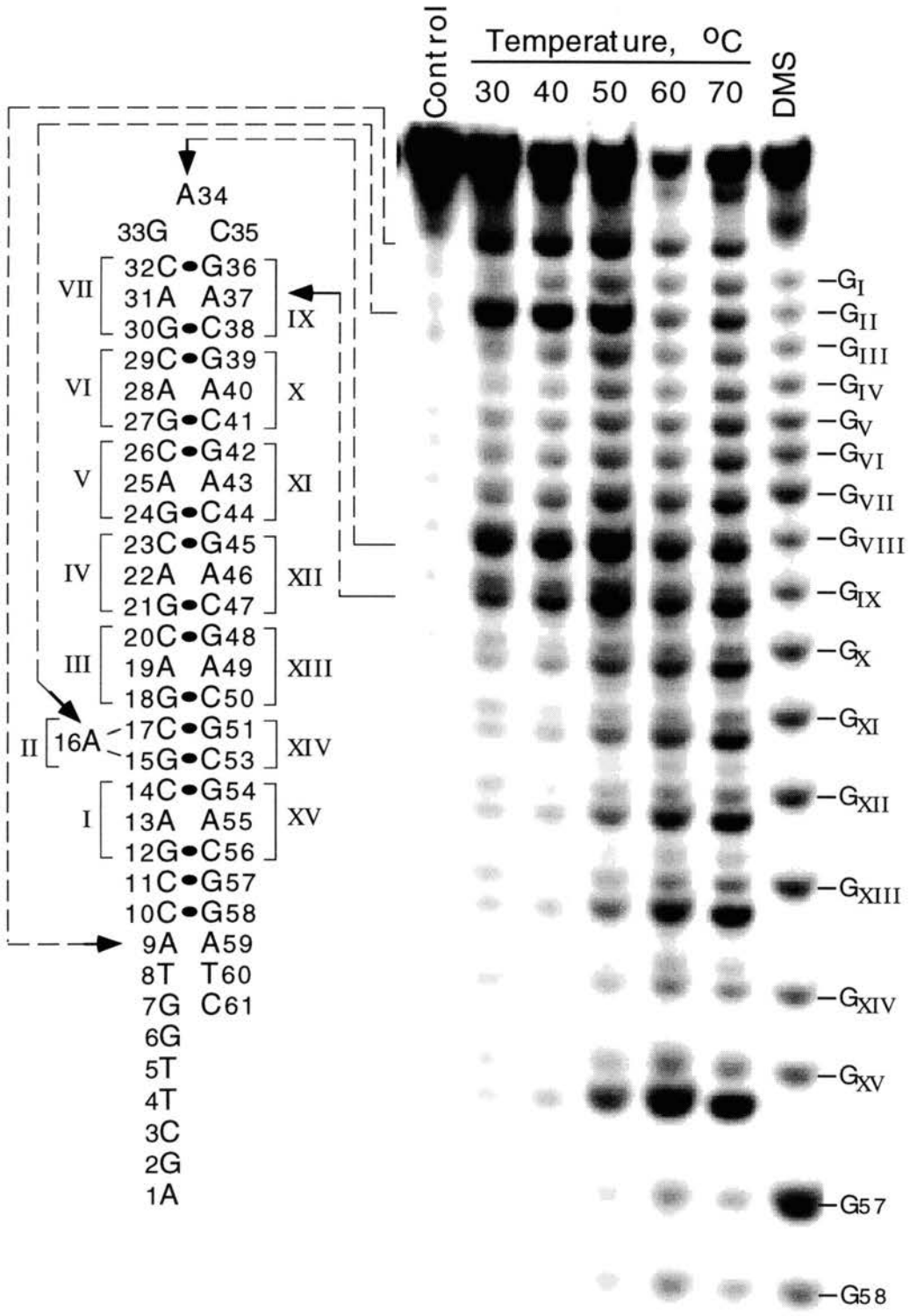

Control
Temperature, °C
30 40 50 60 70
DMS
A34
33G C35
VII
32C•G36
31A A37
30G•C38
IX
VI
29C•G39
28A A40
27G•C41
X
V
26C•G42
25A A43
24G•C44
XI
IV
23C•G45
22A A46
21G•C47
XII
III
20C•G48
19A A49
18G•C50
XIII
II
16A
17C•G51
15G•C53
XIV
I
14C•G54
13A A55
12G•C56
XV
11C•G57
10C•G58
9A A59
8T T60
7G C61
6G
5T
4T
3C
2G
1A
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G57
G58

Figure 52. Diethyl pyrocarbonate modification of $\Delta A_{XIV}(GAC)_{15}$ in 50 mM Na^+, 500 mM K^+.

DEPC modification of $\Delta A_{XIV}(GAC)_{15}$ was performed as described in the legend to figure 50, except that the salt concentration was 50 mM Na^+, 500 mM K^+.

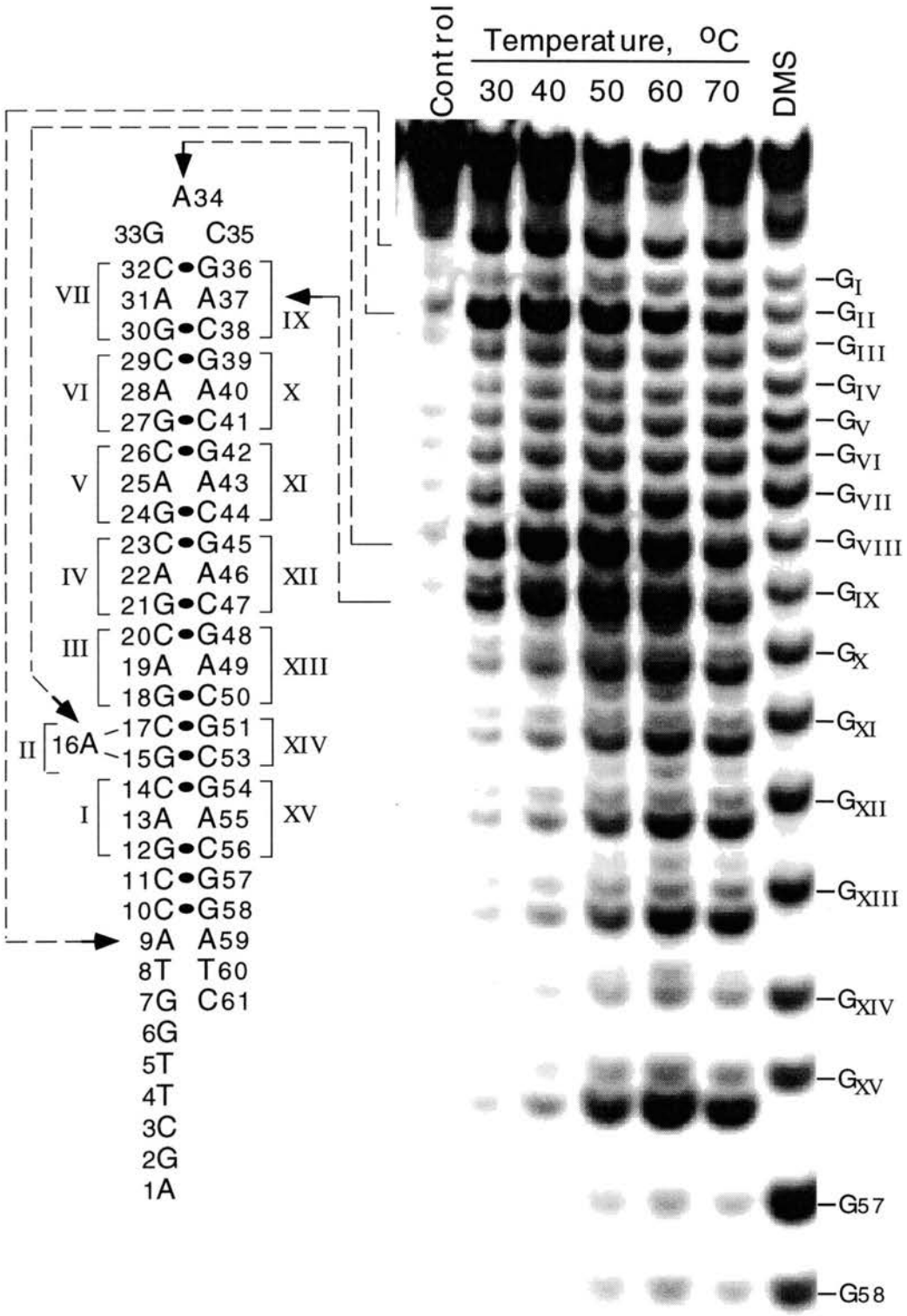

Control
Temperature, °C
30 40 50 60 70
DMS
A34
33G C35
32C•G36
VII 31A A37 IX
30G•C38
29C•G39
VI 28A A40 X
27G•C41
26C•G42
V 25A A43 XI
24G•C44
23C•G45
IV 22A A46 XII
21G•C47
20C•G48
III 19A A49 XIII
18G•C50
II 16A 17C•G51 XIV
15G•C53
14C•G54
I 13A A55 XV
12G•C56
11C•G57
10C•G58
9A A59
8T T60
7G C61
6G
5T
4T
3C
2G
1A
–GI
–GII
–GIII
–GIV
–GV
–GVI
–GVII
–GVIII
–GIX
–GX
–GXI
–GXII
–GXIII
–GXIV
–GXV
–G57
–G58

Figure 53. Diethyl pyrocarbonate modification of ss $(CAG)_{15}$ in 50 mM Na^+, 150 mM K^+

DEPC modification of ss $(CAG)_{15}$ was performed as described in materials and methods at the indicated temperatures. The concentration of the monovalent cations in the reactions were 50 mM Na^+, 150 mM K^+. The concentrations of DEPC in the reactions were 0.96 (30 and 40°C), and 0.27 (50-70°C) M. The control reaction was performed at 30°C and did not contain DEPC. Positions of guanine residues determined by chemical modification with 21 mM dimethyl sulfate (DMS) are shown on the right. Arrows indicate adenine residues preferentially modified at 30°C. The deduced structure of ss $(CAG)_{15}$ at 30°C is shown on the left.

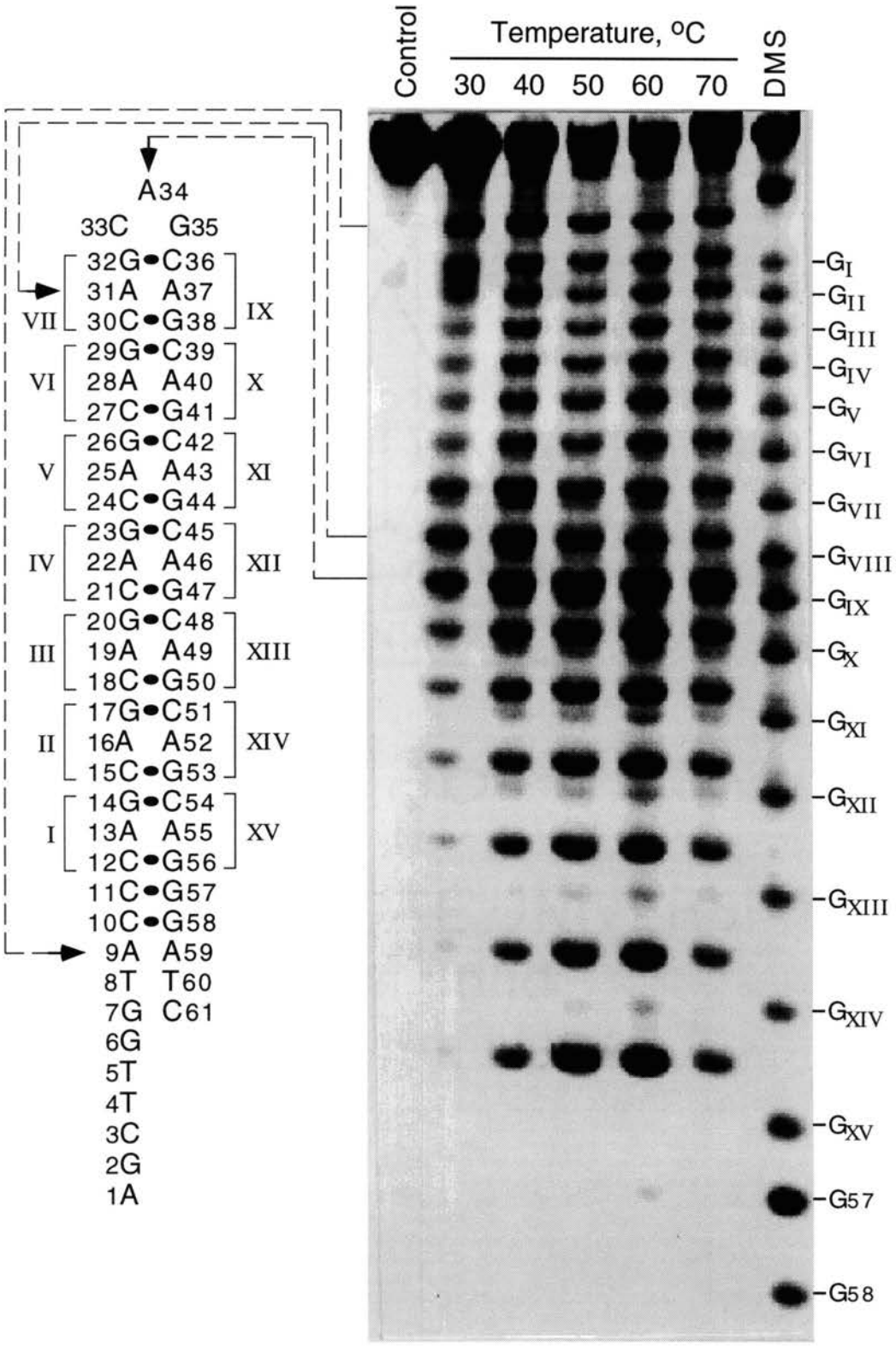

Control
Temperature, °C
30 40 50 60 70
DMS
A34
33C G35
32G•C36
31A A37
VII 30C•G38 IX
29G•C39
VI 28A A40 X
27C•G41
26G•C42
V 25A A43 XI
24C•G44
23G•C45
IV 22A A46 XII
21C•G47
20G•C48
III 19A A49 XIII
18C•G50
17G•C51
II 16A A52 XIV
15C•G53
14G•C54
I 13A A55 XV
12C•G56
11C•G57
10C•G58
9A A59
8T T60
7G C61
6G
5T
4T
3C
2G
1A
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G57
G58

Figure 54. Melting profile of $\Delta A_{XIV}(GAC)_{15}$ and $ss(CAG)_{15}$

The autoradiograph shown in Figures 50 and 53 were scanned with a PDI densitometer (x3). In addition, autoradiographs of DEPC reactions performed with $\Delta A_{XIV}(GAC)_{15}$ in 50 mM Na^+/150 mM K^+ (Figur 51), and 50 mM Na^+/500 mM K^+ (Figur 52) were also scanned. For reaction products generated at the specified temperature, the optical densities of adenines in the loops and stems of $\Delta A_{XIV}(GAC)_{15}$ and $ss(CAG)_{15}$ were determined. For loop and stem analyses, the adenines in triplets eight and ten, respectively, were used. Values in the figure represent the mean of the data. An optical density of 0.15 was subtracted from each autoradiograph.

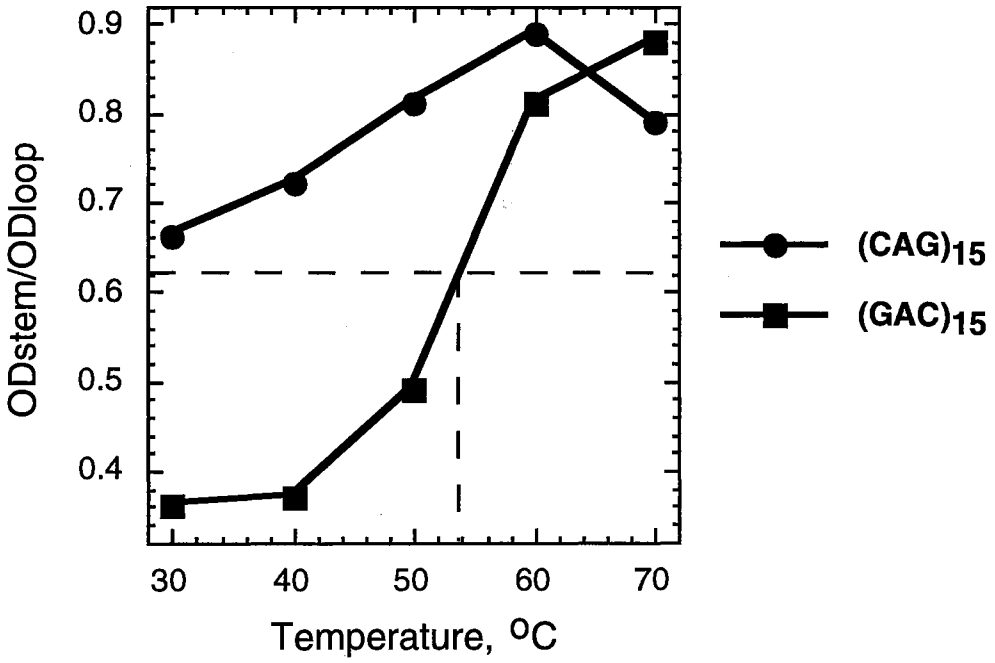
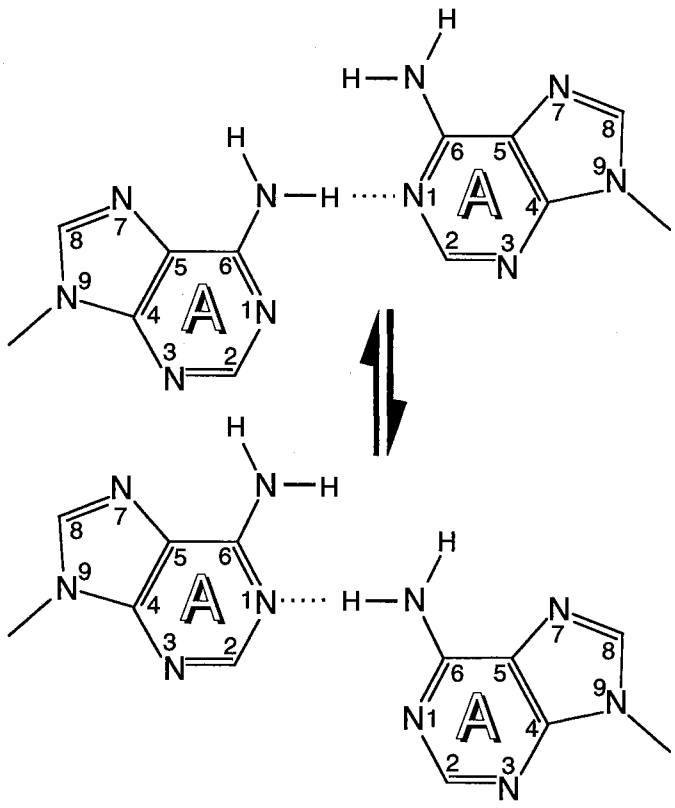

0.9
0.8
0.7
0.6
0.5
0.4
ODstem/ODloop
30
40
50
60
70
Temperature, °C
(CAG)15
(GAC)15

Figure 55. The potential A•A base pair by hydrogen bond

Dashed lines indicate hydrogen bonds. The structure of the A•A pair was determined from 1H NMR data on the duplex sequence $(GCGACGC)_2$ (S. Marriapin, G. Gupta, M. Mitas, unpublished results).



CHAPTER IX

SUMMARY

Triplet repeat expansions are considered to be significant due to their association with several human inherited disorders. They are characterized by accelerating the expansion in germline cells (Ashley and Warren, 1995) of infected individuals in an unique non-Mendelian genetics manner. The expansion of the trinucleotide repeat contained within a specific gene is coincident with disease manifestation. Continued expansion of the repeat is observed in offspring of affected individuals, resulting in increased severity of the disease. At present, ten TREDs have been identified. Due to the high frequency with which the trinucleotide repeats associated with these diseases undergo further expansion in the successive generations, the mutation(s) responsible for this phenomenon has been described as "dynamic", a novel characteristic in genetics, which differs from the classical Mendelian genetics.

The trinucleotides that undergo expansion within the genes associated with TREDs are CGG, CTG, CCG, CAG, and most recently GAA, TTC. Most of these trinucleotides share a palindromic GC dinucleotide, while the CGG and CCG trinucleotide additionally contains a palindromic CG dinucleotide. Although the mechanisms whereby expansion of trinucleotide repeats results in disease manifestation are to be eluciated, a preliminary conclusion can be drawn that the presence of a long trinucleotide repeat may lead to physical instability of the DNA structure. It is reasonable to think that these expanded triplet

repeats may form a stable structure at the ss DNA level. An understanding of these secondary structures is essential to unraveling the mechanisms of the triplet repeat expansion.

To aid in correlating potential structures of triplet repeat nucleic acids with their function and their propensity to undergo expansion events, a sequence-based classification system for them was described by our laboratory. Class I repeats, which were defined by the presence of a GC or CG palindrome, were uniquely associated with nine of the ten known triplet repeat expansion diseases. All six complementary single strands of class I triplet repeats potentially formed stable hairpin structures. We have investigated the conformations of all class I triplet repeat sequences. Our experimental data support hairpin structures. Our data support a common model of these hairpin helix, in each pair of triplets, two C, G bases form the normal Watson-Crick base pairs, the mismatch bases are base stacked and base paired by forming hydrogen bonds in stable hairpins of $(CGG)_{15}$, $(CTG)_{15}$, $(GTC)_{15}$ and $(GAC)_{15}$. Moreover, evidence for a novel hairpin structure with mismatch bases extruding out of the helix of the hairpin stem region (e-motifs) was found in ss (CCG)n triplet at physiological pH.

To explore the possibility of hairpin formation by ss Class I triplet repeats, studies were first performed with a ss oligonucleotide containing 15 prototypic CTG repeats. Electrophoretic mobility of an oligonucleotide containing ss $(CTG)_{15}$ was concentration-independent. Single-stranded oligonucleotides predicted not to form hairpins migrated slower than their ds forms. In contrast, ss $(CTG)_{15}$ migrated faster than its ds form. In support of a hairpin structure, P1 nuclease cleaved at the predicted loop region of single-stranded $(CTG)_{15}$ and $KMnO_4$ oxidized a single thymine in the triplet repeat region of ss $(CTG)_{15}$.

These data demonstrate that ss $(CTG)_{15}$ forms a hairpin containing base paired and /or stacked thymines in the stem.

The structure of a single-stranded oligonucleotide containing $(GTC)_{15}$ was examined, and compared to ss $(CTG)_{15}$ in parallel studies. Electrophoretic mobility, $KMnO_4$ oxidation and P1 nuclease studies demonstrate that, similar to ss $(CTG)_{15}$, ss $(CTG)_{15}$ forms a hairpin containing base paired and/or stacked thymines in the stem. Electrophoretic mobility melting profiles performed by a co-worker in our laboratory revealed that in ~1 mM Na^+, the melting temperatures of ss $(GTC)_{15}$ and $(CTG)_{15}$ were 38°C and 48°C respectively. The loop regions of ss $(GTC)_{15}$ and $(CTG)_{15}$ were cleaved by single-strand-specific P1 nuclease at the T25-C29 and G26-C27 phosphodiester bonds respectively (where the loop apex of the DNAs is T28). Molecular dynamics simulations by our collaborator suggested that in ss $(GTC)_{15}$ the loop was bent towards the major groove of the stem, apparently causing an increased exposure of the T25-C29 region to solvent. In ss $(CTG)_{15}$ guanine-guanine stacking caused a separation of the G26 and C27 bases, resulting in exposure of the intervening phosphodiester to solvent. The results suggest that ss$(GTC)_{15}$ and ss$(CTG)_{15}$ form similar, but distinguishable, hairpin structures.

Expansions of $d(CCG/CGG)_n$ sequences are associated with two genetic diseases which express fragile sites in the expanded loci of X chromosome and hypermethylation of the expanded d(CCG/CGG)n region within the genes. To understand DNA structure that might contribute to these phenomena, we analyzed a model single-stranded oligonucleotide containing $(CCG)_{15}$. At pH 8.5, ss $(CCG)_{15}$ formed a relatively unstable (Tm~30°C in 1 mM Na^+) hairpin containing CpG base pairs steps. The hairpin at pH 7.5, which contained protonated cytosines but no detectable C^+ C base pairs, was characterized by increased thermal stability (Tm ~37°C) , increased stacking of the CpG base pair

steps, and a single cytosine that was flipped away from the central portion of the helix. Examination of ss$(CCG)_{20}$, which was designed to adopt a hairpin containing alternative GpC base-pair steps, revealed a hairpin containing CpG base-pair steps, a pK_a of 8.4, and a distorted helix. The results suggest that DNA sequences containing $(CCG)_{n\geq 15}$ adopt hairpin conformations that contain CpG rather than GpC base-pairs steps; the mismatched cytosines are protonated at physiological pH but are not H-bonded. The results are consistent with a new DNA structure in which two extrahelical cytosines (separated by two C•G base-pairs in the formal hairpin) form a stack-pair in the minor groove of a distorted helix.

Our results suggest that single stranded $(CGG)_{15}$ forms a hairpin with the following features: (i.) a stem containing G^{syn} G^{anti} base pairs; (ii) at > 200 mM K^+, CGG repeats on the 5′ portion of the stem base paired to GCC repeats on the 3′ side (referred to as the *(b)* alignment); and (iii) heat stability (T_m = 75°C in low ionic strength). At < 100 mM K^+, dimethyl sulfate reactions indicated that the hairpin in the *(b)* alignment was in equilibrium with another structure, presumably a hairpin in the alternative *(a)* alignment. Molecular dynamics simulations suggested that the loop region of the *(a)* alignment contained two guanines stacked on top of one another. The same guanines in the *(b)* alignment were base-paired in a *syn-anti* arrangement.

The structures of ss $(CAG)_{15}$ and ss$(GAC)_{15}$ were examined. At 10°C, the electrophoretic mobilities of the two DNAs were similar to ss$(CTG)_{15}$, a DNA that forms a hairpin containing base paired and/or stacked thymines. At 37°C in 50 mM NaCl, single-strand-specific P1 nuclease cleaved the G33-G36 phosphodiesters of ss$(GAC)_{15}$, and the G32-A34, G35-A36 phosphodiesters of ss$(CAG)_{15}$. DEPC modification of the normal and mutant ss $(CAG)_{15}$ and ss $(GAC)_{15}$ demonstrated a sensitive region corresponded with the adenine in the

loop region of the hairpins. Electrophoretic mobility melting profiles (my colleague's data) indicated that the melting temperature of ss$(CAG)_{15}$ in low (~1 mM Na^+) ionic strength was 38°C. In contrast, the Tm of ss$(GAC)_{15}$ was 49°C, a value similar to the Tm of ss($CTG)_{15}$. These results provide evidence that ss$(GAC)_{15}$ and ss$(CAG)_{15}$ form similar, but distinguishable hairpin structures.

Studies on the triplet repeat structures continue in our lab. Our interest has recently turned to investigating ss (CCG)n and ss (CGG)n with larger repeat copy numbers. These sequences may form tetraplex or other structures.

REFERENCES

Akarsu, A. N., Akhan, O, Sayli, B. S., Sayli, U., Baskaya, G., Sarfarazi, M. (1995) Journal of Medical Genetics 32, 435-441.

Ashley, C. T., Jr and Warren, S. T. (1995) Annual Review Genetics. 29, 703-728.

Amato, A. A., Prior, T. W., Barohn, R. J., Snyder, P., Papp, A., Mendell, J. R., (1993) Neurology 43, 791-794.

Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Mahadevan, M., Tsilfidis, C., Chen, C., Alleman, J., Wormskamp, N. G., and Vooijs, M. (1992) Nature 355, 548-551.

Baker, D. J., Kan, J. L. C., & Smith, S. S. (1991) Genetics 74, 207-210.

Bell, M. V., Hirst, M. C., Nakahori, Y., MacKinnon, R. N., Roche, A., Flint, T. J., Jacobs, P. A., Tommerup, N., Tranebjaerg, L., Froster-Iskenius, U., Kerr, B., Turner, G., Lindenbaum, R. H., Winter, R., Pembrey, M., Thibodeau, S., and Davies, K. E. (1991) Cell 64, 861-866.

Brook, J. D., A. E. McCurrash, H. G. Harley, A. J. Buckler, D. Church, H. Aburatani, K. Hunter, V. P. Stanton, J.-P. Thirion, T. Hudson, R. Sohn, B. Zemelman, R. G. Snell, S. A. Rundle, S. Crow, J. Davies, P. Shelbourne, J. Buxton, C. Jones, V. Juvonen, K. Johnson, P. S. Harper, D. J. Shaw, and Housman, D. E. (1992). Cell 68, 799-808.

Brown, T., Leonard, G. A., Booth, E. D., and Kneale, G. (1990) J. Mol. Biol. 212, 437-440.

Burke, J. R., Wingfield, M. S., Lewis, K. E. , Roses, A. D., Lee, J. E., Hulette, C., Pericak-Vance, M. A. and Vance, J. M. (1994) Nature Genetics 7, 521-524.

Burke, J. R., Enghild, J. J., Martin, M. E., Jou, Y. S., Myers, R. M., Roses, A. D., Vance, J. M. and Strittmatter, W. J. (1996) Nature Medicine 2, 347-350.

Bush, E. W., Taft, C. S., Meixell, G. E., and Perryman, M. B. (1996) J. Biol. Chem. 271, 548-52.

Campuzano, V., Montermini, L., Koenig, M. and Pandolfo, M. (1996) Science 271, 1423-1427.

Chen, X., Mariappan, S. V. S., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M., & Gupta, G. (1995) Proc. Natl. Acad. Sci. U.S.A. 92, 5199-5203.

Choi W. T., MacLean H. E., Chu S., Warne G. L, and Zajac J. D. (1993) Aust. N. Z. J. Med. 23, 187-192.

Christiansen, J., Kofod, M. and Nielsen, F. C. (1994) Nucl. Acids Res., 22, 5709-5716.

Chung, M. Y., Ranum, L. P., Duvick, L. A., Servadio, A., Zoghbi, H. Y., and Orr, H. T. (1993) Nature Genetics 5, 254-258.

Dallas, J. F. (1992) Mam. Gen. 3: 452-456.

Destee, A., Chartier-Harlin, M. C. (1995) Presse Med 24, 312-316.

Duyao, M. P., Auerbach, A. B., Ryan, A., Persichetti, F., Barnes, G. T., McNeil, S. M., Ge, P., Vonsattel, J. P., Gusella, J. F., Joyner, A. L. (1995) Science 269, 407-410.

Edward, A., Hammond, H. A., Jin, L., Caskey, C.T., and Chakraborty, R. (1992) Genomics 12: 241-253.

Eichler, E. E., Holden, J. J. A., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., Richads, C. S., Ward, P. A. & Nelson, D. L. (1994) Nat. Genet. 8, 88-94.

Feng, Y., Zhang, F., Lokey, L. K., Chastain, J. L., Lakkis, L., Eberhart, D., and Warren, S. T. (1995) Science 268, 731-734.

Finocchiaro, G., G. Baio, P. Micossi, G. Pozza, S. Di Donato, (1988) Neurology 38, 1292.

Friedreich, N., (1863) Virchows Arch. Pathol. Anat. 26, 391.

Fu, Y. H., D. P. A. Kuhl, A. Pizzuti, M. Pierreti, S. S.Sutcliffe, S. Richards, A. J. M. H.Verkerk, J. J. A. Holden, R. G. Fenwick, Jr., W. T. Warren, B. A. Oostra, D. L. Nelson, and Caskey, C. T. (1991) Cell 67, 1047-1058.

Gacy, A. M., Geoffrey, G., Juranic, N., Macura, S., and McMurray, C. T. (1995) Cell 81, 533-540.

Grady, D. I., Ratliff, R. L., Robinson, D. L., McCanlies, E. C., Meyne, J. and Moyzis, R. K. (1992) Proc. Natl. Acad. Sci. USA 89, 1695-1699.

Geoffroy, G. (1976) Canada Journal of Neurology Science 3, 279.

Sutherland, G. R. and Richards, R. I. (1995) Proc. Natl. Acad. Sci. USA 92, 3636-3641.

Gordon A Leonard, Shude Zhang, and Mark R Peterson, (1995) Structure 3, 335-340.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985) Science, 314, 67-73.

Jones, C., L. Penny, T. Mattlna, S. Yu, E. Baker, L. Voullaire, W. Y. Langdon, G. R. Sutherland, R. L. Richards & A. Tunnacliffe. (1995) Nature 376, 145-149.

Johnson, H. B. Rich, A. (1985) Cell 42, 713-724.

A. E. Harding and R. L. Hewer, (1983) Q. J. Medicine 208: 489 The Hereditary Ataxias and Related Disorders (1984) (Churchill Livingstone, Edinburgeh)

Hansen, R. S., Gartler, S. M., Scott, C. R., Chen, S. H., and Laird, C. D. (1992) Hum. Mol. Genet., 1, 571-578.

Hansen, R. S., Canfield, T. K., Lamb, M. M., Gartler, S. M., and Laird, C. D. (1993) Cell 73, 1403-1409.

Hardin, C. C., Watson, T., Corregan, M., and Bailey, C. (1992) Biochemistry 31, 833-841.

Harley, H.G., Brook, J.D., Rundle, S.A., Crow, S., Reardon, W., Buckler, A.J., Harper, P.S., Housman, D.E., and Shaw, D.J. (1992) Nature 355, 545-546.

Hayatsu, H. and Ukita, T. (1967) Biochm. Biophys. Res. Commun., 29: 556-561.

Pieretti, M., Zhang, F., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T., and Nelson, D. L. (1991) Cell 66, 817-822.

Hirst, M. C., Grewal, P. K. & Davies, K. E. (1994) Hum. Mol. Genet., 3, 1553-1560.

The Huntington's Disease Collaborative Research Group. (1993) Cell 72, 971-983.

Kang, S., Jaworski, A., Ohshima, K. and Wells, R. D. (1995) Nature Genetics 10, 213-218.

Kawakami, H., Maruyama, H., Nakamura, S., Kawaguchi, Y., Kakizuka, A., Doyu, M., and Sobue, G. (1995) Nature Genetics 9, 344-345.

Ke, S. H., and Wartell, R. M. (1993) Nucleic Acids Research 21, 5137-5143.

Kermer, E. J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S. T., Schlessinger, D., Sutherland, G. R., and Richards, R. I. (1991) Science 252, 1711-1714.

Klimasauskas, S., and Roberts, R. J. (1995) Nucleic Acids Research 23, 1388-1395.

Koide. R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, K., Ishikawa, A., Hatashi, T., Saito, S., Tomoda, A., Miike, T., Naito, H., and Ikuta, F. (1994) Nature Genetics 6, 9-13.

Kuhn, E., Lehmann-Horn, F., and Rudel, R. (1990) Nervenarzt 61, 323-331.

Kunkle, T. A. (1993) Nature 365, 207-208.

Kunst, C. B. and Warren, S. T. (1994) Cell 77, 853-861.

Ikeuchi, T., Koide, R., Onodera, O., Tanaka, H., Oyake, M., Takano, H., Tsuji, S., (1995) Clin Neurosci 3, 23-27.

Laayoun, A., and Smith, S. S. (1995) Nucleic Acids Research 23, 1584-1589.

La Spada, A. R., Wilson, E.M., Lubahn, D.B., Harding, A.E., and Fischbeck, K.H. (1991) Nature 352, 77-79.

Lane, A. N., and Peck, B. (1995) Eur. Journal of Biochemistry 230, 1073-1087.

Laxova, R. (1994) Fragile X syndrome Advances in Pediatrics 41, 305-342.

Leach, D. R. (1994) Bioessays 16, 893-900.

Leonard, J., McDonald, J. J., Henderson, R. E. L., and Reichman, M. E. (1971) Biochemistry 10, 3335-3342.

Liang, G., Gannett, P., and Gold, B. (1995) Nucleic Acids Research 23, 713-719.

Lou, S., Robinson, C. J., Reiss, A. L., and Migeon, B. R. (1993) Somatic Cell Mol. genet., 19, 393-404.

Luckow, B. and Schutz, G. (1987) Nucleic Acids Research 15, 5490.

Mahadevan, M. C., Tsilfidis, L. Sabourin, G. Shutler, C. Amemiya, G. Jansen, C. Neville, M. Narang, J. Barcelo, K. O'Hoy, S. Leblond, J. Earle-Macdonald, P.J. DE Jong, B. Wieringa, and R.G. Korneluk, (1992) Science 255, 1253-1255.

Mariappan, S. V., Catasti, P., Chen, X., Ratcliff, R., Moyzis, R. K., Bradbury, E. M., & Gupta, G. (1996) Nucleic Acids Research 24, 784-792.

Mariappan, S. V., Garcoa, A. E. and Gupta, G. (1996) Nucleic Acids Research 24, 775-783.

Matilla T; Volpini V; Genis D; Rosell J; Corral J; Davalos A; Molins A; Estivill X. (1993) Hum Mol Genet 2, 2123-2128.

Mandel, J. L. (1994) Nature Genetics 7: 453-455.

McCarthy, J. G. and Rich, A. (1991) Nucleic Acids Research 19, 3421-3429.

Meyne, J., Ratliff, R. L., and Moyzis, R. K. (1989) Proc. Natl. acad. Sci. USA 86, 7049-7053.

Meyne, J. , Baker, R. J., Hobart, H. H., Hsu, T. C., Ryder, O. A., Ward, O. G., Wiley, J. E., Wurster-Hill, D. H., Yates, T. L. and Moyzis, R. K. (1990) Chromosoma 99, 3-10.

Miro, R., Clemente, I. C., Fuster, C. & Egozcue, J. (1987) Hum. Genet. 75, 345-349.

Mitas, M., A. Yu, J. Dill, T. J. Kamp, E. J. Chambers and I. S. Haworth (1995a) Nucleic Acids Research, 23, 1050-1059.

Mitas, M., A. Yu, J. Dill and I. S. Haworth (1995b) Biochemistry, 34, 12803-12811.

Morrison P. J., Johnston, W. P., and Nevin, N. C. (1995) Journal Medical Genetics 32, 524-530.

Moyzis, R. K., Torney, D. C., Meyne, J., Buckingham, J. M., Wu, J. R., Burks, C., Sirotkin, K. M. and Goad W. B. (1989) Genomics 4, 273-289.

Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L. and Wu, J. R. (1988) Proc. Natl. Acad. Sci. USA 85, 6622-6626.

Muragaki, Y., S. Mundlos, J. Upton and B. R. Olsen (1996) Science 272, 548-551.

Nakahori, Y., Knight, S. J. L., Holland, J., Schwartz, C., Roche, A., Tarleton, J., Wong, S., Flint, T. J., Froster-Iskenius, U., Bentley, D., Davies, K. E. and Hirst, M. C. (1991) Nucl. Acids Res., 19, 4355-4359.

Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, T., Ohsaki, E., Bundo, M., Takeda, T., Tadokoro, K., Kondo, I., Murayama, N. (1994) Nature Genetics 6, 14-18.

Nancarroow, J. K., E. Kremer, K. Holman, H. Eyre, N. A. Dogger, D. Le Paslier, D. F. Callen, G. R. Sutherland, R. I. Richards (1994) Science 264, 1938-1941.

Nasir, J., Floresco, S. B., O'Kusky, J. R., Diewert, V. M., Richman, J. M., Zeisler, J., Borowski, A., Marth, J. D., Phillips, A. G., Hayden, M. R. (1995) Cell 81, 811-823.

Nelson, D. L. and Warren, S. T. (1993) Nature Genetics 4, 107-108.

Oberie, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M. F., and Mandel, J. L. (1991) Science 252, 1097-1102.

Onodera, O., Oyake, M., Takano, H., Ikeuchi, T., Igarashi, S. and Tsuji, S. (1995) Am. J. Hum. Genet. 57, 1050-1060.

Orr, H. T., M. Y. Chung, S. Banfi, T.J. Kwiatkowski, A. Servakio, A.L. Beaudet, A.E. McCall, L.A. Duvick, L.P. Ranum, and H.Y. Zoghbi. (1993). Nature Genetics 4, 221-226.

Pentland, B. and Fox, K. A. A. (1983) Journal of Neurol. Neurosurg. Psychiatry 46, 1138.

Pieretti. M., F. Zhang, Fu, Y. H, Warren, S.T., Oostra, B,A., Caskey, C.T., and Nelson, D.L. (1991) Cell 66, 817-822.

Pizzuti, A; Friedman, D.L.; Caskey, C.T. (1993) The myotonic dystrophy gene. Arch. Neurol. 50, 1173-1179.

Popescu, N. C., Zimonjic, D. and DiPaolo, A. (1990) Hum. Genet. 84, 383-386.

Prosser, J., Frommer, J., Paul, C. and Vincent, P. C. (1986) J. Mol. Biol. 187, 145-155.

Ranum L. P., Chung M. Y., Banfi, S., Bryer, A., Schut, L. J., Ramesar, R., Duvick, L. A., McCall, A., Subramony, S. H., and Goldfarb, L. (1994) Am. J. Human Genetics 55, 244-252.

Richards. R. I., Holman. K., Yu. S and Sutherland, G. R. (1993) Plum. Mol. Genetics 2, 1429-1435.

Richards, R. I. and Sutherland, G. R. (1992) Trends Genet., 8, 249-254.

Richards, R. I. and Sutherland, G. R. (1994) Nature Genetics, 6, 114-116.

Riggins, G. J., Lokey, L. K., Chastain, J. L., Leiner, H. A., Sherman, S. L., Wilkinson, K. D., and Warren, S. T. (1992) Nature Genetics 2, 186-191.

Robinson, H., van der Marel, G. A., van Boom, J. H., Wang A. H. J. (1992) Biochemistry 31, 10510-10517.

Roling DB; La Spada AR; Fischbeck KH (1993) Kennedy's disease [letter] Neurology 43, 2424-2425.

Romeo, G. (1983) Am. J. Human Genetics 35, 523.

Rubin, C.M., and Schmid, C.W. (1980) Nucleic Acids Research 8, 4613-4619.

Schmitt, I., Epplen, J. T., and Riess, O. (1995) Human Molecular Genetics 4, 1619-1624.

Shelbourne, P., Davies, J., Buxton, J., Anvret, M., Blennow, E., Bonduelle, M., Schmedding, E., Glass, I., Lindenbaum, and R., Lane, R. (1993) New England Journal Medicine 328, 471-475.

Skre, H. (1975) Clinic Genetics 7, 287.

Smith, G. K., Jie, J., Fox, G. E. and Gao, X. (1995) Nucleic Acids Research 23, 4303-4311.

Smith, S. S., Hardy, T. A., and Baker, D. J. (1987) Nucleic Acids Research 15, 6899-6917.

Smith, S. S., Kan. J. L. C., Baker, D. J., Kaplan, B. E. and Dembek, P. (1991) J. Mol. Biol. 217, 39-51.

Smith, S. S., Laayoun, A., Lingeman, R. G.., Baker, D. J., and Riley, J. (1994) J. Mol. Biol. 243, 143-151.

Sinden, R. R. and Wells R. D. (1992) Curr. Opin. Biotechnol. 3, 612-622.

Smith, S. S., Lingeman, R. G., and Kaplan, B. E. (1992) Biochemistry 31, 850-854.

Suite, N. D., Sequeiros, J. and McKhann, G. M. (1986) Journal Neurogenetics 3, 177-182.

Sutclffe, J. S., Nelson, D. L., Zhang, F., Pieretti, M., Caskey, C.T., Saxe, D., Warren, S. T. (1992) Hum. Mol. Gen. 1, 397-400.

Sutherland, G. R. (1977) Science 197, 265-266.

Sutherland, G. R., and F. Hecht (1985) Fragile Sites on Human Chromosome, pps 80-87, (Oxford University Press, Oxford).

Sutherland, G. R. and Richards, R. I. (1995) Proc. Natl. Acad. Sci. USA 92, 3636-3641.

Twist, E. C., Casaubon, L. K., Ruttledge, M. H., Rao, V. S., Macleod, P. M., Radvany, J., Zhao, Z., Rosenberg, R. N., Farrer, L. A., and Rouleau, G. A. (1995) Journal Medical Genetics 32, 25-31.

Uyama, E., Kondo, I., Uchino, M., Fukushima, T., Murayama, N., Kuwano, A., Inokuchi, N., Ohtani, Y., and Ando, M. (1995) Journal of Neurology Science 130, 146-153.

Ververk, A. J. J. H., M. Pieretti, J. S. Sutcliffe, Y.-H. Fu, D.P.A. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M.F. Victoria, F. Zhang, B.E. Eussen, G.-J. B. van Ommen, L.A.J. Blonden, G.J. Riggins, J.L. Chastain, C.B. Kunst, H. Galjaard, C.T. Caskey, D.L. Nelson, B.A. Oostra, and S.T. Warren, (1991) Cell 65, 905-914.

Vincent, A., Heitz, D., Petit, C., Kretz, C., Oberle, I., and Mandel, J. L. (1991) Nature 369, 624-626.

Vincze, A., Henderson, R. E. L., McDonald, J. J., and Leonard, N. J. (1973) J. Am. Chem. Soc. 95, 2677-2682.

Wang, Y. H., Amirhaeri. S. Kang, S, Wells, R. D., and Griffith, J. D. (1994) Science 265, 669-671.

Warren, S. T. (1996) Science 271, 1374-1375.

Waye, J. S. and Willard, H. F. (1989) Proc. Natl. Acad. Sci. USA 86, 6250-6254.

Webb, T. (1989) in: Davies, K. E. (ed.), The Fragile X Syndrome, Oxford University Press, New York, pp 50-55.

Weber, J. L. (1990) Genomics 7, 524-530.

Winship Herr (1985) Proc. Natl. Acad. Sci. USA 82, 8009-8013.

Willard, H. F. and Waye, J. S. (1987) Trends Genetics 3, 192-198.

Williamson, J. R., Raghuraman, M. K. and Cech, T. R. (1989) Cell 59, 871-880.

Wohlrab, F. (1992) Methods Enzymol. 212B, 294-301.

Wohrle, D., Hennig, I., Vogel, W. and Steinbach, P. (1993) Nature Genetics 4, 143-146.

Woodford, K. J., Howell, R. M., and Usdin, K. (1994) J. Biol. Chem., 269, 27029-27035.

Yu, A., J. Dill, S. S. Wirth, G. Huang, V. H. Lee, I. S. Haworth, and M. Mitas (1995a) Nucleic Acids Research, 23, 2706-2714.

Yu, A., Dill, J., and Mitas, M. (1995b) Nucleic Acids Research, 23, 4055-4057.

Yu, A., M. D. Barron, M. Christy, J. Dai, B. Gold, D. M. Gray, I. S. Haworth and M. Mitas (1996) At physiologic pH, d(CCG)$_{15}$ forms a hairpin containing protonated cytosines and a distorted helix. Submitted to Biochemistry .

Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J. C., Warren, S. T., Schlessinger, D., Sutherland, G. R., and Richards, R. I. (1991) Science 252, 1179-1181.

Zoghbi, H. Y. (1995) Spinocerebellar ataxia type1. Clinic Neuroscience 3, 5-11.

VITA

Adong Yu

Candidate for the Degree of

Doctor of Philosophy

Dissertation: STRUCTURES OF TRIPLET REPEAT DNAS ASSOCIATED WITH HUMAN DISEASES

Major Field: Biochemistry & Molecular Biology

Biographical:

Personal Data: Born in Liaoning, China, January 17, 1957, the son of Kede Yu and Ruiting Sun.

Education: Graduated from Xiongyao High School, Liaoning, China, in August, 1975; received Bachelor of Science Degree in Biochemistry from Jilin University at Changchun, China in January, 1983; received Master of Science Degree in Microbiology & Molecular Genetics from Oklahoma State University, Oklahoma, U. S. A. in December, 1992; completed requirements for the Doctor of Philosophy Degree at Oklahoma State University in December, 1996.

Professional Experience: Research Assistant, Laboratory of Microbial Nitrogen Fixation, Institute of Forestry & Soil, Academia Sinica, March, 1983 to July, 1987. Research Associate, Department of Microbiology, Institute of Applied Ecology, Academia Sinica, August, 1987 to August, 1990. Teaching Assistant, Department of Microbiology & Molecular Genetics, Oklahoma State University, August, 1990 to May 1992. Graduate Research Associate, Department of Biochemistry & Molecular Biology, Oklahoma State University, February, 1993 to August, 1996.

Professional Membership: Member of American Association for The Advancement of Science; Member of American Society of Microbiology; Member of American Society of Human Genetics; Member of American Association for Cancer Research.