

**The Trouble With Triples: Elucidating the Behavior of Trinucleotide Repeats in
Chromatin**

By

Catherine Burke Volle

A.B. Mount Holyoke College, 2006

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in the Division of Biology and Medicine at Brown
University

Providence, Rhode Island

May 2014

© Copyright 2014 by Catherine Burke Volle. All rights reserved.

This dissertation by Catherine Burke Volle is accepted in the present form by the
Division of Biology and Medicine as satisfying the dissertation requirement for the
degree of Doctor of Philosophy.

Date _____

Sarah Delaney Ph.D., Advisor

Recommended to the Graduate Council

Date _____

Susan Gerbi, Ph.D., Reader (Chair)

Date _____

Richard Freiman, Ph.D, Reader

Date _____

Anatoly Zhitkovich, Ph.D, Reader

Date _____

Catherine Freudenreich, Ph.D, Reader
Tufts University

Approved by the Graduate Council

Date _____

Peter Weber, Ph.D
Dean of the Graduate School

Catherine Burke Volle

Education

Brown University, Providence, Rhode Island, September 2008- May 2013
Doctoral Candidate in Molecular and Cellular Biology and Biochemistry

Mount Holyoke College, South Hadley, Massachusetts, September 2002- May 2006
A.B. in Biology May 2006, *cum laude* and with High Honors in Biochemistry

Professional Experience

Mount Holyoke College, South Hadley, Massachusetts
Research Technician, June 2006-July 2008, Department of Chemistry, laboratory of Professor Megan Nunez

Salado Veterinary Hospital, Salado, Texas
Veterinary Technician, May 2004-August 2004, clinic of Dr. Jonathan Kendall

Research Experience

Graduate Student, Brown University, Laboratory of Sarah Delaney (2009-present). Construct nucleosome core particles (NCPs) from isolated chromatin and DNA containing trinucleotide repeats and determine if the DNA structure is altered by interaction with the histone core

Rotation Student, Brown University, Laboratory of Sarah Delaney (2009). Synthesize, purify, and probe the structure of CAG trinucleotide repeats containing the oxidative damaged DNA base 8-oxo-7,8-dihydroguanine.

Rotation Student, Brown University, Laboratory of Tricia Serio (2009). Express and purify the Sup35 yeast prion protein and determine its ability to form linear aggregates.

Rotation Student, Brown University, Laboratory of Richard Bennett (2009). Create a cloning library of *Candida albicans* genes using the Gateway cloning system.

Rotation Student, Brown University, Laboratory of Wolfgang Peti (2008-2009). Examine the interaction between the bacterial quorum-sensing protein SdiA and the chaperone DnaK by bacterially expressing each protein and monitoring their interaction by their elution from a column.

Research Technician in Chemistry, Mount Holyoke College, Laboratory of Megan Nunez (2006-2008). Examine a bacterial predator/prey relationship on the nanoscale

Senior Thesis Student in Biochemistry, Mount Holyoke College, Laboratory of Megan Nunez (2005-2006). Explored the structural differences in the LPS of the predatory

bacteria *Bdellovibrio bacteriovorus* and LPS from several prey strains of *Escherichia coli*

Professional Memberships

Biophysical Society

American Chemical Society

Awards, Honors, and Grants Funded

National Defense Science and Engineering Graduate Fellow (2010-present)

NIH National Graduate Student Research Conference, participant (2012)

International Travel Award Recipient (2012)

Young Researcher (Participant), 61st Lindau Meeting of Nobel Laureates (2011)

Mary Lyon Scholar (2006)

Sigma Xi (2006)

Leadership, Service, and Outreach

Member, BioMed Advisory Committee, Brown University (2008-present). Work with the Dean of the Division of Biology and Medicine to plan career development and community building events for graduate students in the division.

President, Graduate Women in Science and Engineering (2010-2012). Work to build a community of female science and engineering graduate students and provide mentors for undergraduate women in the sciences. Merged with the Sarah Doyle Women's Center Graduate Women's Group in March 2012.

Head Teaching Assistant, Mount Holyoke College (2006). Worked with professors to develop grading schemes, developed a format for formal laboratory report submissions, and assisted other teaching assistants.

Chemistry Department Liaison, Mount Holyoke Chemistry Club, Mount Holyoke College (2005-2006). Worked with Chemistry Department in searches for new faculty and planned chemistry club activities.

Volunteer, Greenfield Animal Shelter (2004). Worked with full time shelter staff members to train and care for abused and neglected dogs and cats. Also worked with veterinary staff and helped find placement for animals after training was complete.

Papers

Catherine B. Volle and Sarah Delaney. (2012) “CAG/CTG repeats alter affinity for the histone core and positioning of DNA in the nucleosome” *Biochemistry* 59 9814-9825.

Sarah Delaney, Daniel A. Jarem, **Catherine B. Volle**, and Craig J. Yennie. (2012) “Chemical and biological consequences of oxidatively damaged guanine in DNA” *Free Radical Research* 46 420-441.

Catherine B. Volle, Daniel A. Jarem, and Sarah Delaney. (2012) “Trinucleotide repeat DNA alters structure to minimize the thermodynamic impact of 8-oxo-7,8-dihydroguanine” *Biochemistry* 51 52-62.

Catherine B. Volle, Megan A. Ferguson, Katherine E. Aidala, Eileen M. Spain, and Megan E. Núñez. (2008) “Spring Constants and Adhesive Properties of Native Bacterial Biofilm Cells Measured by Atomic Force Microscopy” *Colloids and Surfaces B: Biointerfaces* 67 32-40.

Catherine B. Volle, Megan A. Ferguson, Katherine E. Aidala, Eileen M. Spain, and Megan E. Núñez. (2008) “Quantitative Changes in the Elasticity and Adhesive Properties of Prey Cells During Predation by *Bdellovibrio bacteriovorus* 109J” *Langmuir* 24 8102-8110.

Megan A. Ferguson, Jacqueline Schmitt, Anil R. Sindhurakar, **Catherine B. Volle**, Megan E. Núñez, and Eileen M. Spain. (2008) “Rapid isolation of host-independent *Bdellovibrio bacteriovorus*” *Journal of Microbiological Methods* 73 279-281.

Selected Talks and Posters

Catherine B. Volle. “The Trouble with Triples: Elucidating the Behavior of Trinucleotide Repeat Sequences in Chromatin” (Talk) DNA Repair and Mutagenesis, MIT, February 12, 2013.

Catherine B. Volle and Sarah Delaney. “The Trouble with Triples: Elucidating the Behavior of Trinucleotide Repeat Sequences in Chromatin” (Poster) Biophysical Society Meeting, February 6, 2013.

Catherine B. Volle and Sarah Delaney. “The Trouble with Triples: Elucidating the Behavior of Trinucleotide Repeats in Chromatin” (Poster) National Graduate Student Research Conference, NIH, October 9, 2012.

Catherine B. Volle and Sarah Delaney. “The Trouble with Triples: Elucidating the Behavior of Trinucleotide Repeats in Chromatin” (Poster) Gordon Research Conference, Mutagenesis, August 23, 2012.

Catherine B. Volle and Sarah Delaney. “The Trouble with Triples: Elucidating the Behavior of Trinucleotide Repeats in Chromatin” (Poster) Gordon Research Conference, Chromatin Structure and Function, May 7-8, 2012.

Catherine B. Volle. “The Trouble with Triples: ROS-induced DNA Damage and Its Effect on Trinucleotide Repeat Sequences” (Talk) Providence Area Aging Research Forum, February 28, 2012.

Catherine B. Volle and Sarah Delaney. “Investigation of Trinucleotide Repeats in Nucleosome Core Particles” (Poster) Albany 2011: The 17th Conversation, June 17, 2011.

Catherine B. Volle and Sarah Delaney. “Investigation of Trinucleotide Repeats in Nucleosome Core Particles” (Poster) American Chemical Society Meeting, August 22, 2010.

Catherine B. Volle, Megan A. Ferguson, Katherine E. Aidala, Eileen M. Spain, and Megan E. Núñez. “Physical Properties of Native Biofilm Cells Explored by Atomic Force Microscopy” (Poster) Biophysical Society Meeting, February 10, 2009.

Catherine B. Volle and Megan E. Núñez. “Characterization of Lipopolysaccharide Monolayers from *Bdellovibrio* Predator and *E. coli* Prey” (Poster) American Society for Microbiology General Meeting, June 2, 2008.

Catherine B. Volle. “AFM investigations into the physical changes of biofilm cells during attack by *Bdellovibrio bacteriovorus*” (Talk) Invited lecturer, Harvard AFM Workshop, Asylum Research, February 12, 2008.

Catherine B. Volle and Megan E. Núñez. “Characterization of Lipopolysaccharide Monolayers from *Bdellovibrio* Predator and *E. coli* Prey” (Poster) American Chemical Society Annual Meeting, August 20, 2007.

Catherine B. Volle and Megan E. Núñez. “Characterization of Lipopolysaccharide Monolayers from *Bdellovibrio* Predator and *E. coli* Prey by Langmuir-Blodgett and Atomic Force Microscopy” (Poster) American Society for Microbiology General Meeting, May 24, 2007.

Megan A. Ferguson, **Catherine B. Volle**, Megan E. Núñez, and Eileen M. Spain. “Structure and function of a pigment produced by host-independent *Bdellovibrio bacteriovorus*” (Poster) American Society for Microbiology Annual Meeting, May 23, 2007.

Catherine B. Volle and Megan E. Núñez. “Characterization of Lipopolysaccharide Monolayers from *Bdellovibrio* Predator and *E. coli* Prey by Langmuir-Blodgett and Atomic Force Microscopy” (Poster) Biophysical Society Annual Meeting, March 6, 2007.

Acknowledgements

For my Friends and Family

Table of Contents

Signature Page.....	iii
Curriculum Vitae.....	iv
Acknowledgements.....	viii
Table of Contents.....	ix
List of Tables.....	xiv
List of Figures.....	xv
Chapter 1- Introduction: The Trouble with Triples.....	1
1.1 Repetitive DNA in the Genome.....	1
1.2 Huntington’s Disease.....	3
1.3 Fragile X Syndrome.....	4
1.4 Formation of Non-canonical Structure by TNRs.....	5
1.5 Expansion During DNA Replication.....	7
1.6 Expansion During Base Excision Repair.....	11
1.6.1 <i>Oxidative Damage in the Cell</i>	11
1.6.2 <i>The Oxidative Lesion 8-oxo-7,8-dihydroguanine</i>	13
1.6.3 <i>8oxo-G and TNR expansion</i>	15
1.7 Packaging of DNA in the Nucleus.....	18
1.8 Trinucleotide Repeats in Chromatin.....	21
1.9 Gap in Knowledge and Insights Provided by this Work.....	22
Chapter 2- Healthy-Length CAG/CTG and CGG/CCG Repeats Alter Affinity for the Histone Core and Positioning of DNA in the Nucleosome.....	25
2.1 Introduction.....	25

2.2 Experimental Methods.....	27
2.2.1 <i>Oligonucleotide synthesis and purification</i>	27
2.2.2 <i>Competitive nucleosome incorporation assay</i>	30
2.2.3 <i>Native polyacrylamide gel electrophoresis</i>	32
2.2.4 <i>DNase I, Exo III, and S1 nuclease reactions</i>	33
2.2.5 <i>Hydroxyl radical footprinting</i>	35
2.2.6 <i>Image processing</i>	36
2.3 Results and Discussion.....	37
2.3.1 <i>Substrate design</i>	37
2.3.2 <i>Healthy-length CAG/CTG repeats increase nucleosome formation</i>	39
2.3.3 <i>Nineteen CGG/CCG repeats have no effect on incorporation into nucleosomes; the addition of AGG/CCT interruptions does affect nucleosome incorporation</i>	44
2.3.4 <i>The genetic flanking sequence decreases nucleosome incorporation</i>	48
2.3.5 <i>DNase I and Exo III reveal the position of DNA in the nucleosome</i>	50
2.3.6 <i>Hydroxyl radical footprinting reveals that CAG/CTG repeats alter the rotational and translational positioning of DNA in a nucleosome</i>	56
2.3.7 <i>CGG/CCG repeats reposition DNA in the nucleosome; AGG/CCT interruptions further alter the positioning</i>	63
2.3.8 <i>CAG/CTG repeats possess a periodicity of ~10.7 bps per turn, regardless of flanking sequence</i>	71
2.3.9 <i>CGG/CCG repeats decrease the periodicity of DNA in the nucleosome, but an AGG/CCT interruption increases periodicity</i>	74
2.3.10 <i>The CCG/CGG repeat in the HTT flanking sequence “kinks” when incorporated into the nucleosome, but CGG/CCG repeats at the dyad axis do not</i>	75

2.4 Conclusions.....	79
Chapter 3- Trinucleotide Repeat DNA Alters Structure to Minimize the Thermodynamic Impact of 8-Oxo-7,8-dihydroguanine.....	25
3.1 Introduction.....	82
3.2 Experimental Methods.....	84
3.2.1 <i>Oligonucleotide synthesis and purification</i>	84
3.2.2 <i>Chemical probe analysis</i>	85
3.2.3 <i>Differential scanning calorimetry</i>	86
3.3 Results and Discussion.....	87
3.3.1 <i>Structural characterization of a TNR three-way junction</i>	87
3.3.2 <i>Influence of 8-oxoG on the structure of an even-numbered TNR sequence</i>	94
3.3.3 <i>Chemical probe analysis of odd-numbered, single-stranded TNR sequences lacking DNA damage</i>	99
3.3.4 <i>Structural characterization of ‘locked-loop’ hairpin standards</i>	102
3.3.5 <i>Influence of 8-oxoG on the structure of an odd-numbered, single-stranded TNR DNA</i>	104
3.3.6 <i>Calorimetric analysis of single-stranded TNR DNA lacking DNA damage</i>	107
3.3.7 <i>Analysis of damage-containing TNR structures by calorimetry</i>	108
3.4 Conclusions.....	113
Chapter 4- The Presence of a CAG Repeat Hairpin Decreases Affinity for the Histone Core and Modifies the Local Rotational Position of DNA in a Nucleosome.....	115
4.1 Introduction.....	115
4.2 Experimental Methods.....	117

4.2.1	<i>Oligonucleotide synthesis and purification</i>	117
4.2.2	<i>Oligonucleotide ligation and purification</i>	118
4.2.3	<i>Competitive nucleosome incorporation assay</i>	120
4.2.4	<i>Enzymatic probing reactions</i>	122
4.3	Results and Discussion.....	122
4.3.1	<i>CAG hairpins reduce the level of nucleosome formation</i>	122
4.3.2	<i>While the presence of 6 CAG/CTG repeats does not alter the translational positioning of DNA in a nucleosome, the presence of a (CAG)₁₀ hairpin does</i>	127
4.3.3	<i>The presence of a (CAG)₁₀ hairpin alters the rotational position of DNA in a nucleosome</i>	132
4.4	Conclusions.....	138
Chapter 5- Future Directions and Ongoing Research: Investigating the Effects of Oxidative Damage on DNA Packaging and the Ability of TNRs to Modify Di-Nucleosome Structure		140
5.1	Future Directions.....	140
5.2	Investigating the Effect of 8-oxoG on Nucleosome Formation.....	140
5.2.1	<i>DNA substrate preparation</i>	141
5.2.2	<i>Competitive nucleosome incorporation assay</i>	142
5.2.3	<i>Hydroxyl radical footprinting</i>	143
5.3	Assembling Di-nucleosomes to Investigate Higher Order Chromatin Structure.....	144
5.3.1	<i>Oligonucleotide synthesis and purification</i>	145
5.3.2	<i>Inserting the first S1 sequence into the pUC19 vector</i>	145
5.3.3	<i>Inserting the second S1 sequence into the pCBV_S1+ vector</i>	146
5.3.4	<i>Nucleosome exchange</i>	147

5.3.5 <i>Exonuclease mapping</i>	149
Chapter 6- Conclusions: Small Changes in Sequence, Structure, and Length can Have Large Effects on Nucleosome Formation.....	152
References.....	156
Appendix A: Protocols.....	192

List of Tables

Chapter 1

Table 1.1 Neurodegenerative disorders caused by microsatellite expansion.....	1
--	---

Chapter 2

Table 2.1: Sequences of DNA used in CAG/CTG and CGG/CCG nucleosome studies..	28
---	----

Table 2.2: Position of the dyad and translational shifts for CAG-containing strands of the S1 and HTT substrates.....	62
--	----

Table 2.3: Position of the dyad and translational shifts for CGG-containing strands of the S1 substrates.....	68
--	----

Table 2.4: Base pairs per helical turn determined using the CAG and CTG-containing strands of the S1 and HTT substrates.....	72
---	----

Table 2.5: Base pairs per helical turn determined using the CGG- and CCG-containing strands of the S1 substrates.....	75
--	----

Chapter 3

Table 3.1: DNA sequences used to investigate the behavior of 8-oxoG in CAG repeats	88
---	----

Table 3.2: DSC-Derived Thermodynamic Parameters.....	108
---	-----

List of Figures

Chapter 1

Figure 1.1 Hairpin formed by (CNG) ₁₀ TNRs, where N is any base.....	6
Figure 1.2 Structure of a three-way junction.....	7
Figure 1.3 Simple polymerase slippage can lead to small expansions or deletions.....	8
Figure 1.4 Replication fork stalling and reversal can lead to TNR expansion.....	10
Figure 1.5 The structures of G and 8-oxoG.....	14
Figure 1.6 Prevention of G to T transversion mutations caused by 8-oxoG.....	15
Figure 1.7 Different BER pathways.....	17
Figure 1.8 The toxic oxidation cycle.....	18
Figure 1.9 The nucleosome core particle.....	19

Chapter 2

Figure 2.1: Schematic representations of the DNA used to investigate CAG/CTG repeats in the nucleosome and its design.....	37
Figure 2.2: Schematic representations of the DNA used to investigate interrupted and uninterrupted CGG/CCG repeats.....	38
Figure 2.3: Representative competitive nucleosome incorporation reactions for the CAG/CTG repeat-containing substrates.....	41
Figure 2.4: ΔG of bending values calculated for the S1 and HTT substrates.....	42
Figure 2.5: Representative competitive nucleosome incorporation reactions for the CGG/CCG repeat-containing substrates.....	44
Figure 2.6: ΔG of bending values calculated for the S1 and FMR1 substrates.....	46
Figure 2.7: DNase I digestion of nucleosome samples containing CAG/CTG repeats with the CAG-containing strand radiolabeled.....	51
Figure 2.8: DNase I digestion of nucleosome samples containing CAG/CTG repeats with the CTG-containing strand radiolabeled.....	52

Figure 2.9: Exonuclease III digestion reveals weaker interactions between the ends of the CAG/CTG repeat-containing DNA and the histone core, where the CAG-containing strand is radiolabeled.....	53
Figure 2.10: Exonuclease III digestion reveals weaker interactions between the ends of the CAG/CTG repeat-containing DNA and the histone core, where the CAG-containing strand is radiolabeled.....	54
Figure 2.11: Exonuclease III digestion reveals weaker interactions between the ends of the CGG/CCG repeat-containing DNA and the histone core.....	55
Figure 2.12: Hydroxyl radical footprinting reveals the periodicity of the CAG/CTG repeat-containing DNA around the histone core, where the CAG-containing strand is radiolabeled.....	57
Figure 2.13: Hydroxyl radical footprinting reveals the periodicity of the CAG/CTG repeat-containing DNA around the histone core, where the CTG-containing strand is radiolabeled.....	58
Figure 2.14: Hydroxyl radical cleavage curves reveal the rotational setting of the CAG/CTG repeat-containing DNA around the histone core.....	60
Figure 2.15: Translational positioning curves reveal a change in the position of the dyad axis for both the S1_CAG and HTT series.....	62
Figure 2.16: Hydroxyl radical footprinting reveals the periodicity of the CGG/CCG repeat-containing S1 DNA around the histone core, where the CGG-containing strand is radiolabeled.....	64
Figure 2.17: Hydroxyl radical footprinting reveals the periodicity of the CGG/CCG repeat-containing S1 DNA around the histone core, where the CCG-containing strand is radiolabeled.....	65
Figure 2.18: Hydroxyl radical cleavage curves reveal the rotational setting of the CGG/CCG repeat-containing DNA around the histone core.....	66
Figure 2.19: Translational positioning curves reveal a change in the position of the dyad axis for the CGG/CCG repeat-containing S1 series.....	67
Figure 2.20: Hydroxyl radical footprinting reveals the periodicity of the CGG/CCG repeat-containing FMR1 DNA around the histone core, where the CGG-containing strand is radiolabeled.....	69
Figure 2.21: Hydroxyl radical footprinting reveals the periodicity of the CGG/CCG repeat-containing FMR1 DNA around the histone core, where the CGG-containing strand	

is radiolabeled.....	70
----------------------	----

Figure 2.22: S1 nuclease probing reveals that kinking occurs in the secondary CCG/CGG repeat of the HTT-CAG10 DNA when incorporated into a nucleosome.....	76
---	----

Figure 2.23: S1 nuclease digestion reveals no single stranded regions in the S1 or FMR1 DNA substrates.....	78
--	----

Chapter 3

Figure 3.1 Autoradiogram of DEPC reactions with CAG16-TWJ/CTG6-TWJ and CAG16-TWJ-8OX/CTG6-TWJ.....	90
---	----

Figure 3.2 Histograms displaying reactivity of TWJ substrates to DEPC.....	91
---	----

Figure 3.3 DEPC reactions with CAG10, CAG10-5OX, and CAG10-6OX.....	95
--	----

Figure 3.4 Reactivity of CAG10 substrates with DMS.....	96
--	----

Figure 3.5 Reactivity of CAG7, CAG11, and CAG15 with chemical probes.....	99
--	----

Figure 3.6 Histograms displaying reactivity of CAG7, CAG11, and CAG15 to DEPC and DMS.....	100
---	-----

Figure 3.7 Proposed structures for the locked-loop substrates.....	102
---	-----

Figure 3.8 Reactivity of the CAG7 locked-loop standards with DMS.....	103
--	-----

Figure 3.9 Reactivity of the CAG7 locked-loop standards with DEPC.....	104
---	-----

Figure 3.10 DEPC reactions with CAG11, CAG11-5OX, and CAG11-6OX.....	106
---	-----

Figure 3.11 Representative DSC thermograms of undamaged substrates.....	107
--	-----

Figure 3.12 Representative DSC thermograms of damaged substrates.....	109
--	-----

Chapter 4

Figure 4.1 Primary sequences of the individual DNA oligonucleotides and schematic representations of the assembly process used to create hairpin-containing DNA for nucleosome investigations.....	119
---	-----

Figure 4.2 Schematic representations of the duplex DNA substrates used to investigate the affect of hairpin formation on chromatin packaging.....	121
--	-----

Figure 4.3 Nucleosome incorporation ratios for the four S2 substrates.....	123
Figure 4.4 Exonuclease III digestion reveals the translation position of the S2 DNA around the histone core, where the S2_Comp strand is radiolabeled.....	128
Figure 4.5 Exonuclease III digestion reveals the translation position of the S2 DNA around the histone core, where the S2_C or S2_HP strand is radiolabeled.....	129
Figure 4.6 DNase I digestion reveals the rotational positioning of the S2 substrates in the nucleosome, where the S2_Comp strand is radiolabeled.....	132
Figure 4.7 DNase I digestion reveals the rotational positioning of the S2 substrates in the nucleosome, where the S2_C or S2_HP strand is radiolabeled.....	133
Figure 4.8 Alteration of rotational position in the nucleosome by the (CAG) ₁₀ hairpin.....	136
 Chapter 5	
Figure 5.1 Primary sequences of the individual DNA oligonucleotides and schematic representations of the assembly process used to create the single stranded substrates used to investigate the effects of the 8-oxoG (G ^{ox}) on nucleosome formation.....	141
Figure 5.2 Nucleosome incorporation ratios for the S1_C and S1_8oxoG substrates....	142
Figure 5.3 Hydroxyl radical footprinting reveals the position of the S1_C and S1_8oxoG DNA around the histone core.....	143
Figure 5.4 Primer design for construction of a di-nucleosome positioning sequence....	145
Figure 5.5 Plasmid map of pCBV_S1+.....	146
Figure 5.6 Plasmid map of pCBV_S1+S1.....	147
Figure 5.7 The PCR amplified S1+S1 sequence forms nucleosomes.....	148
Figure 5.8 Exonuclease III digestion reveals the translation position of the DNA around the histone cores, where the S1 forward primer is radiolabeled.....	150
Figure 5.9 Exonuclease III digestion reveals the translation position of the DNA around the histone cores, where the S1 reverse primer is radiolabeled.....	150

Chapter 1

Introduction: the Trouble with Triples

1.1 REPETITIVE DNA IN THE GENOME

Stretches of DNA containing repeating units of 1-6 nucleotides, termed microsatellites, are found throughout the genome.¹⁻⁵ Approximately 3% of the human genome is composed of microsatellite DNA. These regions are highly mutable, giving rise to both sequence and length polymorphisms. Over the course of evolution, changes in repetitive sequences within the genome drove cellular diversity.⁶ These types of changes in microsatellite regions, primarily of length, are useful in determining paternity, forensic analysis, and conservation biology.^{1,7-10} However, rapid changes in microsatellite length and/or sequence, usually occurring within a single generation, have also been linked to certain types of cancer¹¹⁻¹⁶ and a class of neurodegenerative disorders (Table 1.1).¹⁷⁻²¹

Table 1.1 Neurodegenerative disorders caused by microsatellite expansion. Adapted from 18.

Disease	Sequence	Location	Health y	Pre- mutation	Disease
Dentatorubral-pallidoluysian atrophy	CAG	<i>ATNI</i> (exon 5)	6–35	35–48	49–88
Huntington's Disease	CAG	<i>HTT</i> (exon 1)	6–29	29–37	38–180
Oculopharyngeal Muscular Dystrophy	GCN	<i>PABPN1</i> (exon 1)	10	12–17	>11
Spinocerebellar Ataxia Type 1	CAG	<i>ATXN1</i> (exon 8)	6–39	40	41–83
Spinocerebellar Ataxia Type 2	CAG	<i>ATXN2</i> (exon 1)	<31	31–32	32–200
Spinocerebellar Ataxia Type 3	CAG	<i>ATXN3</i> (exon 8)	12–40	41–85	52–86
Spinocerebellar Ataxia Type 6	CAG	<i>CACNA1A</i> (exon 47)	<18	19	20–33
Spinocerebellar Ataxia Type 7	CAG	<i>ATXN7</i> (exon 3)	4–17	28–33	>36 to >460
Spinocerebellar Ataxia Type 17	CAG	<i>TBP</i> (exon 3)	25–42	43–48	45–66
Spino and Bulbar Muscular Atrophy	CAG	<i>AR</i> (exon 1)	13–31	32–39	40

Myotonic Dystrophy Type 1	CTG	<i>DMPK</i> (3' UTR)	5–37	37–50	<50
Myotonic Dystrophy Type 2	CCTG	<i>CNBP</i> (intron 1)	<30	31–74	75–11,000
FRAAXE	GCC	<i>AFF2</i> (5' UTR)	4–39	40–200	>200
Friedreich's Ataxia	GAA	<i>FXN</i> (intron 1)	5–30	31–100	70–1,000
Fragile X Syndrome	CGG	<i>FMRI</i> (5' UTR)	6–50	55–200	200–4,000
Huntington's Disease Like 2	CTG	<i>JPH3</i> (exon 2A)	6–27	29–35	36–57
Spinocerebellar Ataxia Type 8	CTG	<i>ATXN8OS</i> (3' UTR)	15–34	34–89	89–250
Spinocerebellar Ataxia Type 10	ATTCT	<i>ATXN10</i> (intron 9)	10–29	29–400	400–4,500
Spinocerebellar Ataxia Type 12	CAG	<i>PPP2R2B</i> (5' UTR)	7–28	28–66	66–78

Expansion of repeat tracts associated with neurodegeneration primarily occurs during genetic transmission from parent to offspring, as well as during development of the progeny.¹⁷⁻²¹ In the genome, the length of the repeat tract falls into one of three categories: a genetically stable length found in healthy individuals, which will be transmitted from parent to offspring without undergoing expansion; an unstable, pre-mutation length that is prone to expansion; and a threshold length that, once reached, leads to manifestation of the disease state. Once the length of the repeat tract has crossed into the pre-mutation range, both the likelihood of further expansion events and size of the subsequent expansions increase. Furthermore, as the expansion is passed on to successive generations, the age of onset decreases and the disease severity increase, in a phenomenon termed genetic anticipation.

The changes in repeat number caused by genomic expansion can be substantial.¹⁷⁻²¹ When the repeat is located within a coding sequence, instability can occur when as few as ~30 repeats are present in the gene, yet the changes in repeat length are modest; ~10 repeats will be added per generational transmission. If the repeat is located in a non-

coding region, instability can also occur when as few as ~30 repeats are present, but expansion that occurs in a single generation can contain from 100 to as many as 10,000 repeats. In the following sections, we will focus on two of these neurodegenerative disorders, Huntington's Disease and Fragile X Syndrome.

1.2 HUNTINGTON'S DISEASE

Huntington's disease (HD) is a devastating neurodegenerative disorder caused by expansion of a CAG/CTG trinucleotide repeat (TNR) in exon 1 of the huntingtin gene (*HTT*), which codes for a series of poly-glutamines in the mature huntingtin protein.²²⁻²⁶ The number of repeats present in exon 1 falls into one of three categories: 5-35 repeats falls within the healthy range, 36-40 repeats falls within the pre-mutation range, and greater than 40 repeats falls within the disease range. Adult onset HD occurs when the repeat tract contains 40-50 CAG/CTG repeats, but a childhood form will manifest if the repeat number is greater than 70.^{23,27}

The 350 kDa huntingtin protein coded for by the *HTT* gene is found mainly in the cytoplasm of the cell; there is an identified nuclear export signal located at the C-terminal end of the protein.^{23,27} Huntingtin protein can be membrane-associated via palmitoylation at cysteine 214 and has also been found associated with vesicles, endosomal compartments, the endoplasmic reticulum, the Golgi apparatus, mitochondria, and microtubules. It is an extremely important protein within in the cell; mouse knockouts are embryonic lethal.²⁸ Huntingtin protein appears to play a role in diverse cellular functions including transcriptional regulation and axonal vesicle trafficking, and may also act as an anti-apoptosis factor.^{23,27} However, many cellular functions of the huntingtin protein are

not well described, owing to its promiscuous interactions with over 200 different binding partners.

It is generally accepted that a N-terminal fragment of the mature protein is responsible for disease manifestation, caused by a gain-of-function (GOF) in the mutant huntingtin protein. The observation of these fragments in the brains of HD patients and mice lead to the identification of several proteases that might be responsible for generating the exon 1 fragment.^{27,29-31} However, recent evidence suggests that the expansion of the CAG/CTG repeat tract may lead to aberrant splicing of the *HTT* mRNA in a CAG/CTG repeat length dependent manner.³² In these studies, mRNA splicing created a poly-adenylated exon 1 mRNA, which then undergoes translation, giving rise to an exon 1 protein. The exon 1 protein, formed either by protease cleavage or a truncated mRNA, then aggregates within the nucleus, which eventually leads to transcriptional dysregulation and neuronal cell death.^{23,27,33} The loss of brain tissue observed in the diseased state gives rise to the classic symptoms of HD: movement disorders, personality changes, and cognitive decline.

1.3 FRAGILE X SYNDROME

Like HD, Fragile X Syndrome (FXS) is caused by an expansion of a TNR. Here, a CGG/CCG repeat in the 5' untranslated region (UTR) of the Fragile X Mental Retardation 1 (*FMRI*) gene expands, leading to loss of the *FMRI* protein.³⁴⁻⁴¹ As with HD, the number of repeats present in *FMRI* falls into one of three categories: healthy individuals have 6-54 repeats, individuals carrying a pre-mutation have 55-200 repeats,

and FXS patients have over 200 repeats. Interestingly, healthy-length repeat tracts are interrupted by an AGG/CCT trinucleotide every 8-11 repeats.⁴²⁻⁴⁹

Even though these diseases are caused by essentially the same initiating event, the expansion of a TNR, HD and FXS manifest within an affected neuron in very different ways. While the expansion of the CAG/CTG repeats in HD produces a GOF mutant protein that can aggregate within the nucleus, expansion of the CGG/CCG repeat in FXS is a loss of function (LOF) mutation. In FXS, expanded CGG/CCG repeats form heterochromatin, silencing the *FMRI* gene, and leading to loss of the *FMRI* protein product, FMRP.^{38,50-52} FMRP is an RNA binding protein that negatively regulates the translation of mRNAs.⁵³⁻⁵⁵ The loss of FMRP leads to transcriptional dysregulation at the synapse, which causes changes in brain development. While it is unknown exactly what initiates the formation of heterochromatin, methylation of the CGG/CCG repeat region appears to be involved,^{38,50,51} allowing the formation of heterochromatin in the CGG/CCG repeat tract, which spreads along the gene, eventually compacting the promoter region.⁵⁶

1.4 FORMATION OF NON-CANONICAL SECONDARY STRUCTURE BY TNRs

TNRs possess an unusual capability: *in vitro*, TNRs can form several different non-canonical secondary structures such as hairpins, G-quadruplexes, and triplexes.^{21,57,58} Many different techniques have been applied to the study of non-canonical TNR structure including NMR, enzymatic and chemical probing, electrophoretic mobility, and thermal denaturation.⁵⁸⁻⁶⁹ All lines of experimentation have demonstrated the ability of both the

CAG and CTG repeats involved in HD and the CGG and CCG repeats involved in FXS to form hairpins. In particular, our lab has shown that CAG, CTG, and CGG repeats form hairpins in which the loop contains 4 bases and the stem contains G/C base pairs (bps) and either A•A, T•T, or G•G mismatches (Figure 1.1).^{59-61,67,68,70} Indeed, it is this capacity to form these various structures that allows expansion within the genome in a polymerase mediated manner (*vide infra*).^{18,71}

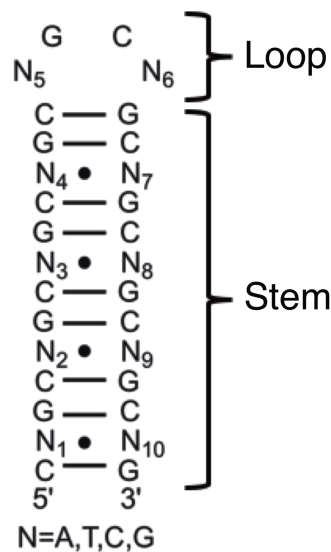


Figure 1.1 Hairpin formed by (CNG)₁₀ TNRs, where N is any base. These TNRs form hairpins that consist of a four base loop and a stem composed of C/G bps and N•N mismatches. The G in the loop is a hot spot for oxidative damage.

While most of the structural data available on TNR hairpins has been generated using single-stranded oligonucleotide models, TNRs do not exist as single-stranded oligonucleotides in the genome. TNR hairpins will form in the context of a duplex, creating a three-way junction in which the hairpin will be flanked on either side by well-matched duplex (Figure 1.2). The structure of these three-way junctions has been inferred from electrophoretic mobility, electron microscopy (EM), and differential scanning calorimetry (DSC), and combined with oligonucleotide data to produce reasonable

assumptions about the behavior of TNR hairpins in three-way junctions.⁷²⁻⁷⁹ It appears that the formation of hairpins within the three-way junction is dependent on the number of excess repeats; a three-way junction containing 6 or 8 additional CAG repeats appears to form an open bulge, but a three-way junction containing 15 CAG repeats appears to form a hairpin.⁷⁶⁻⁷⁹ However no direct structural characterization with enzymatic or chemical probes has taken place. Furthermore, many of these substrates are designed such that the repeat hairpin is immediately flanked by mixed-sequence duplex, forcing the hairpin to occupy a single position, which makes evaluation of the structure less complicated. Indeed, hairpins may form in this manner within the genome. However, it is equally likely that the TNR hairpin will not be flanked by mixed sequence but by base-paired TNRs. If the flanking sequence is composed of TNRs, the hairpin will have the ability to shift between many equivalent positions along the repeats in the complementary strand, creating a population of three-way junctions with the hairpin in various positions relative to the complementary repeats. Little is known about how these hairpins, once formed, might behave in the genome, though recent work using a zinc-finger nuclease has demonstrated their formation *in vivo*.⁸⁰

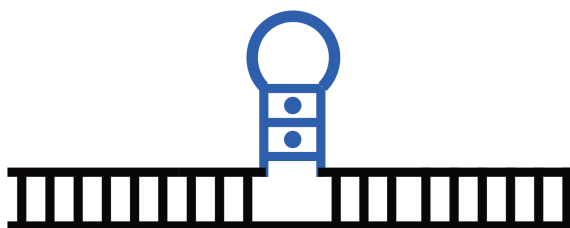


Figure 1.2 Schematic diagram of a three-way junction composed of a hairpin (blue) flanked on either side by well matched duplex (black).

1.5 EXPANSION DURING DNA REPLICATION

In actively dividing cells, DNA replication appears to play a role in repeat expansion through the phenomenon of polymerase slippage, in which the polymerase dissociates from the template strand and re-associates either upstream or downstream of its previous position (Figure 1.3).^{18,20,57,71,81} During TNR replication, expansion can occur when the DNA polymerase dissociates from the template strand and re-associates at an already replicated repeat upstream. This leads to an increase in repeat length after a second round of replication, if the excess DNA is not removed. If re-association of the polymerase occurs downstream of the replicated region, a deletion will occur. However, it is important to note that the increases or decreases in repeat number caused by simple polymerase slippage will be quite small, and does not account for the large change in repeat number observed with manifestation of these neurodegenerative disorders.

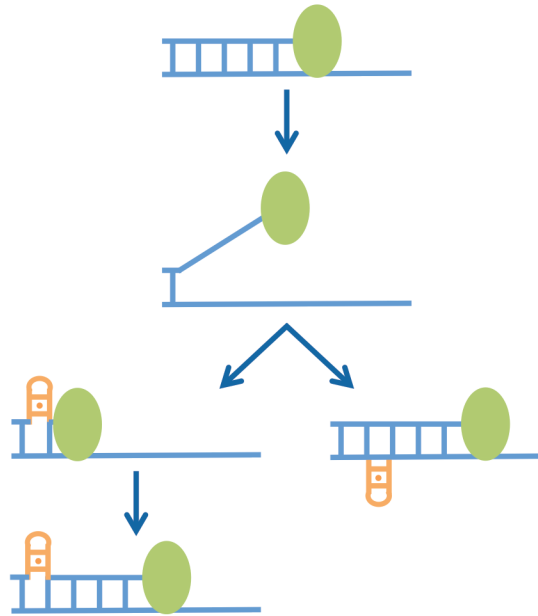


Figure 1.3 Simple polymerase slippage can lead to small expansions or deletions. As the polymerase (green) replicates the DNA, it can dissociate from the template strand. If the polymerase re-associates upstream (left) an expansion is created and if the polymerase re-associates downstream (right) a deletion is created after a second round of replication.

Lagging strand DNA synthesis is specifically implicated in replication dependent expansion by both the correlation between healthy repeat length and Okazaki fragment size as well as the altered instability of repeats in cell extracts that lack proteins crucial for Okazaki fragment processing, Flap Endonuclease 1 (FEN1)/RAD27, and other enzymes required for lagging strand synthesis.⁸²⁻⁸⁸ *In vitro* studies have demonstrated that secondary structure formation by TNRs can block FEN1 activity on Okazaki fragments, possibly allowing for persistence of these secondary structures in the genome.⁸⁹⁻⁹²

Interestingly, regions of the genome composed of TNRs stall polymerases during replication in a length dependent manner.^{20,57,93-95} The lagging strand appears to be under replicated at the stall site, indicating that the repeat blocks lagging strand synthesis.⁹⁵ Thus it was proposed that replication fork stalling and reversal could lead to expansion.^{20,57,93-95} As the leading strand polymerase replicates continuously through the repeat tract, the complementary repeat will be positioned in the Okazaki Initiation Zone (OIZ) of the lagging strand, a region that will remain transiently single-stranded (Figure 1.4). Thus it is possible for the OIZ to fold into a non-canonical structure and stall replication of the lagging strand. As leading and lagging strand syntheses are coordinated, leading strand synthesis stalls as well. The process of fork reversal can release stalled replication forks. The replication fork backs up, forming a four-way junction in which a single-stranded repeat tract is present off the 3' end of the leading strand. The single-stranded tail composed of repetitive DNA can then form a non-canonical secondary structure. To restart replication, the reversed fork is “flipped back” and the non-canonical secondary structure is incorporated into the genome, where it will lead to an increased repeat number after a second round of replication. If the repeat length is longer than that

of the OIZ, the process of replication stalling and fork reversal could happen several times within the repeat tract.

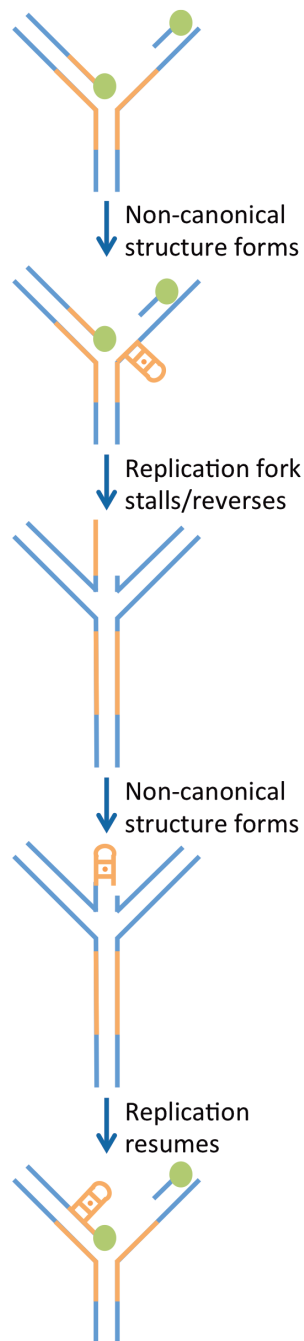


Figure 1.4 Replication fork stalling and reversal can lead to TNR expansion. The polymerase is shown in green, the TNR tract is shown in orange, and the non-repetitive flanking sequence is shown in blue.

Replication dependent expansion can be influenced by several factors. The location of the repeat tract relative to the origin of replication (ORI) is likely a crucial determinant of instability during DNA replication. Work in primate cells, using constructs that contain an SV40 ORI placed in different locations relative to the repeat tract, has shown that the bias toward expansion or deletion changes depending on the distance between the ORI and the repeat tract,⁹⁶ though analysis of ORI location in cells from individuals affected by TNR expansion indicates that the relationship might not be so clear.⁹⁷ The direction of replication, determining which repeat is placed in the lagging strand, is also an important factor in instability. In bacteria and yeast, several previous studies have demonstrated that TNR instability is modulated by the direction of replication.^{95,98-101} While in bacteria and yeast deletions are observed more commonly than expansions, experiments using the SV40 ORI in primate cells also demonstrated that instability is dependent upon replication direction, indicating that the differences in instability between single celled organisms and mammalian cells may be due to differences between replication fork dynamics or the length of the Okazaki fragment.⁹⁶

While these expansion pathways appear to be important in proliferative cells, expansion also occurs in the neurons where these disorders manifest.²⁴⁻²⁶ Neurons are fully differentiated and non-proliferative. What then accounts for continued TNR expansion in these cells?

1.6 EXPANSION DURING BASE EXCISION REPAIR

1.6.1 Oxidative damage in the cell

Damaging agents constantly bombard DNA in the genome.¹⁰² In particular, oxidative damage of DNA can occur from exposure to a wide variety of both endogenous and exogenous sources. The primary source of oxidative damage within the cell comes from byproducts formed during cellular respiration.^{103,104} 0.1-2% of oxygen that enters the respiratory pathway is converted into reactive oxygen species (ROS), including superoxide anion, radicals, and hydrogen peroxide (H₂O₂).¹⁰³⁻¹¹¹ Other electron transport chains in the endoplasmic reticulum and nuclear membrane may also generate ROS.¹¹² It is important to note that H₂O₂ generated in the cell, which cannot directly damage DNA, can be converted to highly reactive hydroxyl radicals (·OH) in the presence of metal ions; ·OH are the major cause of DNA damage inside the cell.^{113,114}

Respiration is not the only source of ROS generation in the mitochondria; several other mechanisms have been identified, including the Kreb's Cycle enzyme α -ketoglutarate dehydrogenase.^{103,104,115} ROS generation has also been observed in conjunction with oxidases, especially the peroxidases responsible for the oxidation of fatty-acids, which allows escape of H₂O₂ into the cytosol, where it can diffuse throughout the cell.^{110,116} Further ROS production is associated with the cellular processes of signaling,^{117,118} apoptosis,¹⁰⁶ and the inflammatory response,^{119,120} which is known to produce nitric oxide (·NO) and superoxide radicals. These chemicals can combine to form peroxynitrite (ONOO⁻), a potent oxidizing agent.¹²¹ Chronic viral infection has also been linked to high levels of ROS.¹²² While endogenous ROS is primarily responsible for oxidative damage to cellular DNA, extracellular factors such as ionizing radiation,¹²³⁻¹²⁶ gamma radiation,¹²⁷ UV light,^{128,129} and various drugs and chemicals such as bleomycin¹²⁴ and riboflavin¹³⁰⁻¹³² have also been shown to contribute to cellular levels of

oxidative damage, either by exerting a direct effect on the DNA, or by triggering processes that increase ROS production.

Of all the DNA bases, guanine (G) has the lowest redox potential, making it the most susceptible to oxidative damage within the cell.^{133,134} G will react with an assortment of oxidants to generate a guanine radical cation, the precursor to a wide range of modified bases.¹³⁵⁻¹³⁷ G exhibits an even greater vulnerability to oxidation if present in a series of guanines: a 5' G is more susceptible to oxidation if it is contained in a GG or GGG sequence.^{138,139} Furthermore, the radical cations formed during oxidative damage, arising from single electron abstraction from G, can migrate through the DNA and eventually become trapped at a GG or GGG sequence.¹⁴⁰⁻¹⁴² Thus the cell may exploit the increased ability of tandem guanines to oxidize: it has been proposed that GC rich domains located outside a gene's coding region could act as a trap for oxidative damage, sequestering any damage in the non-coding regions where it is less likely to be detrimental to the cell.^{143,144}

1.6.2 The oxidative lesion 8-oxo-7,8-dihydroguanine

While there are numerous products formed from guanine oxidation, the most common form of oxidatively damaged guanine found in the cell is 8-oxo-7,8-dihydroguanine (8-oxoG) (Figure 1.5A).^{127,145,146} It is a product formed from reaction of G with $\cdot\text{OH}$ (derived from H_2O_2 , ionizing radiation), one-electron oxidants (UV light, KBrO_3 , $\text{CO}_3^{\cdot-}$), singlet oxygen (cellular respiration), and ONOO^- (inflammatory response).^{128,133,137,147-149} The basal level of 8-oxoG is estimated at 1-2 per 10^6 guanine residues in nuclear DNA and 1-3 per 10^5 in mitochondrial DNA; up to 100,000 8-oxoG lesions could form each day per cell.^{102,133,150-154} Thus, as a major, if transient, component

of cellular DNA, it is important to understand the effect that the 8-oxoG lesion has on DNA behavior at a molecular and cellular level.

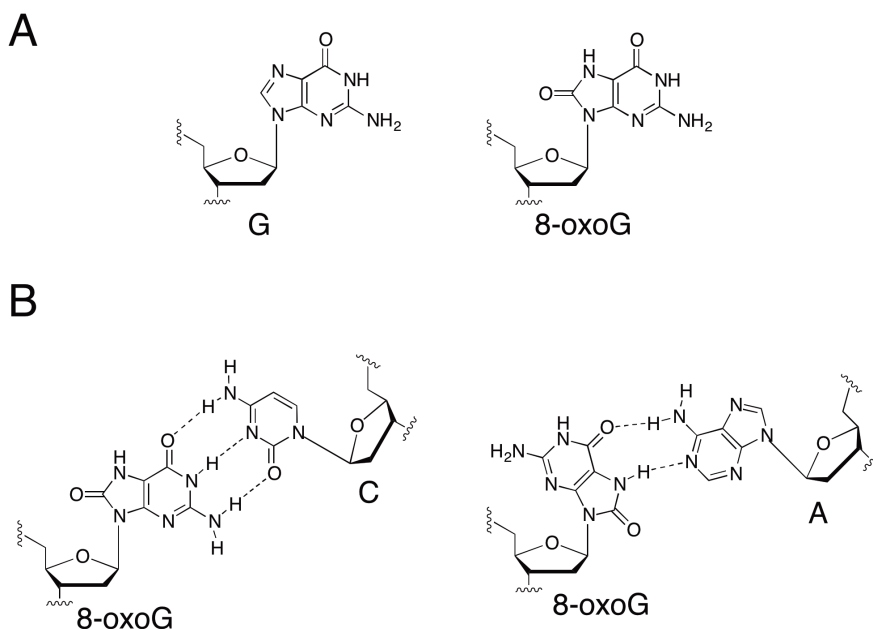


Figure 1.5 The structures of G and 8-oxoG (A) and the structure of 8-oxoG(anti):C(anti) and 8-oxoG(syn):A(anti) base pair (B). Figure adapted from 155.

8-oxoG is able to rotate around the glycosidic bond, converting from the canonical *anti* conformation to the *syn* conformation (Figure 1.5B).¹⁴⁷ When in the *syn* conformation, 8-oxoG can base pair with A via Hoogsteen bonding. Because of its ability to pair with A, 8-oxoG is mutagenic and can cause G to T transversion mutations if this mispair is not repaired prior to replication. To combat the 8-oxoG lesion, the cell has a repair network dedicated to recognizing and removing the oxidized base (Figure 1.6).¹⁵⁶ Repair begins with recognition of the 8-oxoG base paired to C by the glycosylase OGG1 (8-oxoguanine glycosylase 1). OGG1 catalyzes removal of the damaged base, leaving an abasic site opposite the C. Should OGG1 fail to recognize the 8-oxoG, a polymerase may insert an A opposite the lesion. The 8-oxoG/A bp is recognized by a second glycosylase,

MUTYH, which removes the A opposite the 8-oxoG, giving OGG1 a second chance at recognizing and removing the lesion.

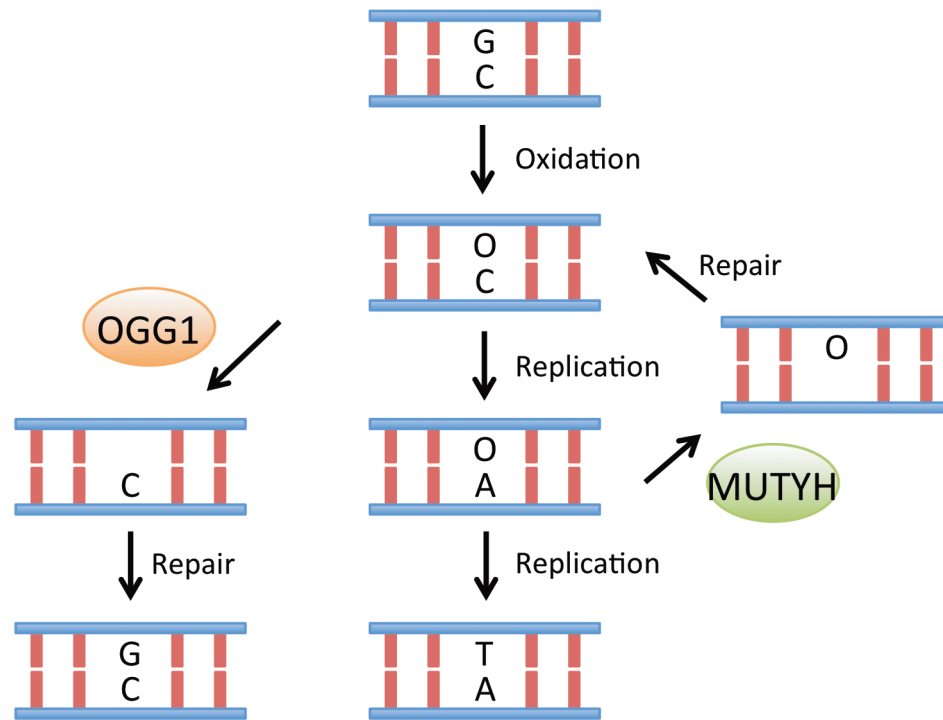


Figure 1.6 Prevention of G to T transversion mutations caused by 8-oxoG involves two glycosylases, OGG1 (orange) and MUTYH (green). OGG1 recognizes and removes the damaged base. If OGG1 doesn't catch the lesion before replication, MUTYH recognizes and removes an A paired with 8-oxoG. Figure adapted from 157.

1.6.3 8-oxoG and TNR expansion

A mouse model has greatly increased understanding of repeat expansion in adult-onset HD.^{158,159} Recently, it has been proposed that the base excision repair (BER) pathway initiated by the glycosylase OGG1 is involved in expansion. Both control and HD mice accumulate 8-oxoG lesions in the CAG repeat tract with age, but HD mice lacking this repair enzyme showed limited expansion of CAG repeats in the brain, compared to their control littermates. Furthermore, *in vitro* experiments that reconstituted BER with a (CAG)₂₀ construct showed repeat expansion while a non-repetitive construct did not expand. These initial experiments used the enzymes OGG1, APE1

(apurinic/apyrimidinic endonuclease 1), and pol β (DNA polymerase β), along with a ligase, which appeared to be sufficient to observe TNR expansion. Importantly, these experiments indicated pol β was adding multiple nucleotides, not just the G that was required for repair through short-patch (SP) BER, in which the lesion is replaced with a single nucleotide (Figure 1.7). Thus expansion may occur through the long-patch (LP) BER pathway, usually observed when a 5' dRp group, the product of APE1 cleavage, is refractory to the dRpase activity of pol β , leading to multi-nucleotide synthesis.^{160,161} This creates a 5' flap that must be removed before ligation can occur. Further studies included the enzyme FEN1, which, in addition to its role in processing Okazaki fragments, is responsible for cleaving the 5' flap in LP-BER.^{162,163} These experiments suggest that multi-nucleotide synthesis and strand displacement in the TNR region allows secondary structure formation in the flap, which blocks FEN1 access (Figure 1.7). While the majority of these experiments used a 5' dRp that is refractory to pol β , experiments with a naturally occurring 5' dRp group also showed expansion.^{159,162} Returning to the mouse model of HD, it appears that the stoichiometry of the BER enzymes also plays a role in expansion, determining whether or not the 5' flap is cleaved correctly.^{164,165} The stoichiometry of pol β and FEN1 in particular seem to explain why expansion occurs more readily in certain regions of the brain; in the cerebellum, which contains cells with high FEN1 activity, instability is limited, but the striatum, which contain cells with high pol β activity, is more susceptible to expansion.

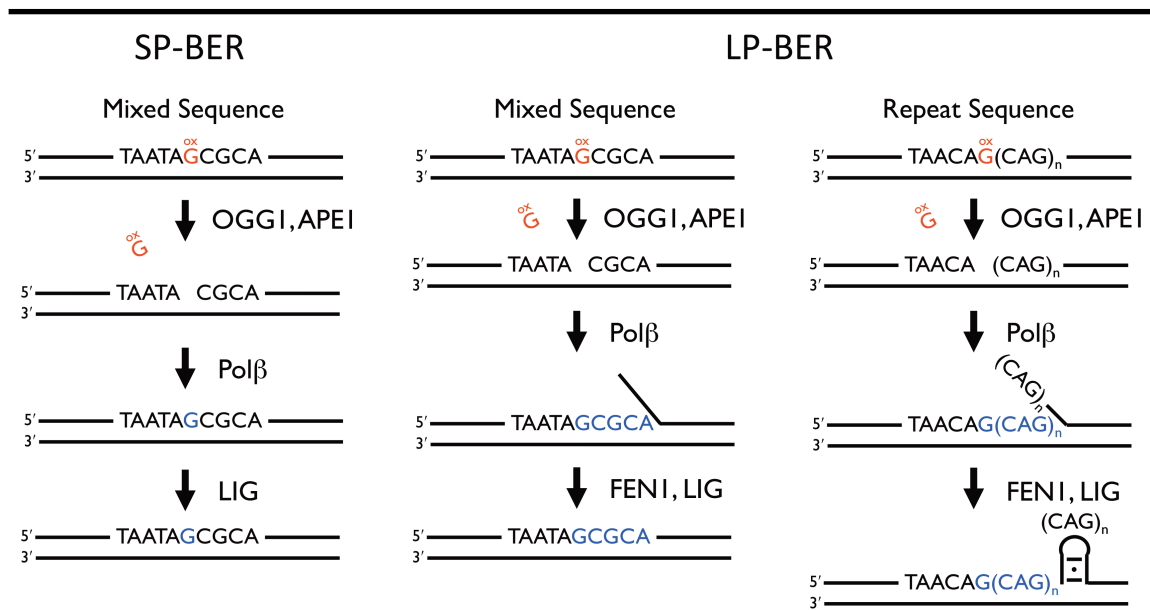


Figure 1.7 Different BER pathways. After recognition and removal of an 8-oxoG (red) by OGG1 and cleavage of the backbone by APE1, guanine insertion by pol β can proceed in two different manners. In SP-BER, a single nucleotide is inserted. During LP-BER multiple nucleotides are added, creating a 5' single stranded flap. If LP-BER occurs in a repeat sequence, the flap can form a hairpin, blocking FEN1 activity and allowing ligation of the hairpin into the repeat tract.

Further experiments performed with a CAG hairpin indicate that the G present in the loop of this structure is a hot spot for oxidative damage and that, once that damage had formed, it is repaired by OGG1 very slowly, indicating that the formation of 8-oxoG in the hairpin loop may persist in the genome.⁵⁹ Taken together, these experiments lead to the proposal of a “toxic oxidation cycle” that fuels the expansion of CAG repeats within the neuron (Figure 1.8).^{59,155,159,162} First, there is an initial oxidation event and formation of the 8-oxoG lesion within the repeat tract. This lesion is recognized and removed by OGG1, leaving an abasic site. This abasic site is converted into a strand break by APE1, and the strand break is recognized by pol β, which begins LP-BER, creating a single-stranded 5' flap. Because this flap is composed of repeats, it folds into a hairpin, making it resistant to cleavage by FEN1. Ligase then joins the hairpin-containing flap to the

newly repaired region. The newly ligated hairpin now contains a hot spot for oxidative damage, which will persist until the hairpin is reincorporated into duplex, and the cycle can begin again.

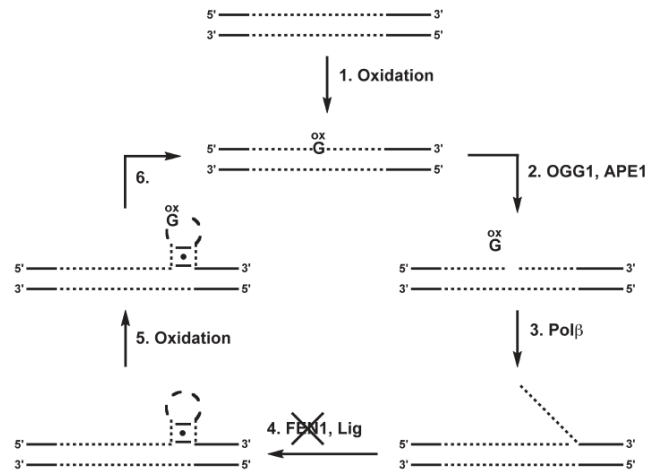


Figure 1.8 The toxic oxidation cycle. A guanine in the repeat DNA is oxidized, generating 8-oxoG (1). OGG1 and APE1 initiate BER by removing the oxidatively damaged base from duplex TNR DNA (dotted lines) and cleaving the backbone at the abasic site (2). The resulting gap is repaired via LP-BER, producing a TNR flap that is refractory to FEN1 (3). The flap, now containing a hairpin, is ligated, trapping the hairpin in the duplex (4). The resulting stem-loop hairpin is highly susceptible to DNA damage in the loop region (5) and the lack of efficient repair of this damage by OGG1 results in it being incorporated into the expanded duplex (6). The final product of the cycle now contains an 8-oxoG that will be recognized by OGG1, beginning the cycle anew and incrementally expanding the TNR tract of DNA. Adapted from 155.

1.7 PACKAGING OF DNA IN THE NUCLEUS

While much of the work elucidating the toxic oxidation cycle was carried out using duplex substrates, genomic DNA is packaged into chromatin, compressing the ~2 meters of DNA found in a single diploid cell into a structure that can fit within the nucleus.¹⁶⁶⁻¹⁶⁹ This compaction has profound effects on the regulation of various cellular processes, including DNA replication, transcription, and repair.

The first step in the compaction process is assembly of the histone core on the DNA, forming a nucleosome (Figure 1.9).^{166,167,169-174} The canonical histone core consists

of two each of histones H2A, H2B, H3, and H4. The histone proteins are obligate heterodimers *in vitro*; H3 dimerizes only with H4 and H2A dimerizes only with H2B. When H3 and H4 dimerize, creating the H3/H4 dimer, the H3/H4 dimer immediately forms a dimer with a second H3/H4 dimer, creating the (H3/H4)₂ tetramer. H2A and H2B also dimerize, forming the H2A/H2B dimer. The (H3/H4)₂ tetramer then forms the histone octamer with two H2A/H2B dimers. Each H2A/H2B dimer has two contact points with the (H3/H4)₂ tetramer; a four-helix bundle configuration connects H2B with H4 and the H2A docking domain contacts the other arm of the (H3/H4)₂ tetramer. Approximately 146 bps of DNA wrap around the histone octamer in a 1.65 turn negative superhelix. The DNA interacts tightly with the histone core along the length of the DNA, except for ~15-20 bases at either end of the DNA duplex where the interaction is far more transient. The macromolecular complex composed of both the histone core and the 146 bps of DNA is called the nucleosome core particle.

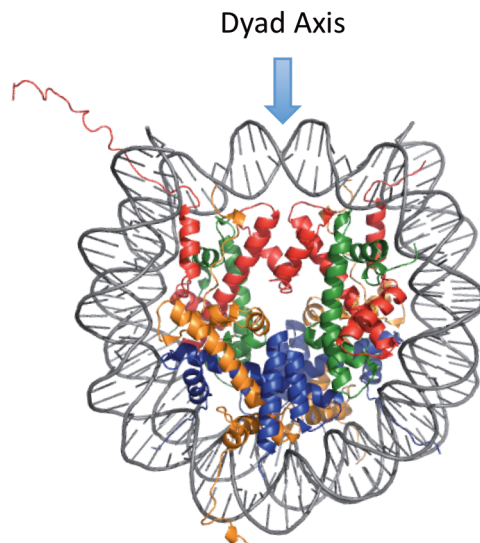


Figure 1.9 The nucleosome core particle. Composed of eight histone proteins and 146 bps of DNA, the NCP is the most basic unit of chromatin packaging. PDB 1AOI.

In the genome, nucleosome core particles are separated by varying lengths of linker DNA, forming the 10 nm chromatin fiber.^{166,168,169} Further compaction occurs with addition of the linker histones H1 and H5, which sit across the DNA at the dyad axis, and through interactions between the histone tails of neighboring nucleosomes, forming the “30 nm” fiber. While the “30 nm” fiber was once considered to have a uniform structure, new evidence has shown that, in the cell, the “30 nm” fiber likely adopts many different packing conformations.^{166,175} Scaffold proteins then assist with further compaction to form the metaphase chromosome.¹⁷⁶ During interphase, genes that are actively transcribed will be packaged as the more open euchromatin, while genes not being transcribed are condensed into the tightly packed heterochromatin.^{177,178}

In the cell, the nucleosome is assembled with help from several chaperone proteins.^{171-173,179} During DNA replication, newly synthesized H3/H4 dimer is bound by the chaperone Asf1, and the trimeric complex travels from the cytoplasm to the nucleus. Asf1 binds the H3/H4 dimer at the H3 interface involved in forming the (H3/H4)₂ tetramer, preventing two H3/H4 dimers from associating prematurely. Once in the nucleus, the H3/H4 dimer is passed off to another histone chaperone, CAF-1 or Rtt106, where the (H3/H4)₂ tetramer is formed. The tetramer is then deposited onto the DNA. Newly synthesized H2A/H2B dimers are bound by the chaperone Nap1 for transport to the nucleus, where the dimer is then deposited with the (H3/H4)₂ tetramer.

Nucleosome positioning is influenced by several different factors, including DNA primary sequence and the location of other DNA interacting proteins.^{169,180-182} Most importantly, nucleosome location is often correlated with the functional organization of the gene, as the presence of nucleosomes can inhibit the binding of transcriptional

activators and the initiation complex, and significantly slow the progress of polymerases. The location of a nucleosome within a stretch of DNA is defined by two parameters: the translational and rotational position. The translational position is defined by the 146 bps interacting with the histone core and describes the position of the DNA with regard to the boundaries of the nucleosome. The rotational position is defined by the regions of the duplex that interact with the histone core and describes the position of the DNA double helix with regard to the histone octamer.

Once formed, nucleosomes are not static structures. Evidence suggests that the nucleosome can, on its own, form several open states that involve disassembly of either the DNA or histone components, allowing access to regions of the DNA or histone folds that might otherwise be blocked.¹⁷⁴ Furthermore, covalent modifications of the histone tails can alter the chromatin architecture, as can ATP-dependent chromatin remodelers, which alter the organization of a gene by moving, ejecting, or restructuring nucleosomes.^{172,177,178,183} These dynamic modifications allow controlled access to specific regions of the genome, and help define the functional organization of a gene.

1.8 TRINUCLEOTIDE REPEATS IN CHROMATIN

In vitro and *in vivo* studies have shown that TNRs are found both in tightly packed, transcriptionally-silent heterochromatin and the more loosely packed, transcriptionally-active euchromatin.⁵² Healthy-length repeats associated with HD and FXS are found in euchromatin, allowing transcription through the associated genes.^{56,164,184-186} However, once expansion has occurred, disease length FXS repeats are found in heterochromatin, and the associated gene is silenced through heterochromatin

spreading.^{56,185,186} Interestingly, HD repeats are found in association with histones containing euchromatin markers regardless of whether they are stable or disease length.^{56,164} Previous studies *in vitro* have shown that disease-length CAG/CTG repeats are preferential locations for nucleosome formation and that, as the length of the repeat tract increases, the ability of CAG/CTG repeats to incorporate into a nucleosome also increases.¹⁸⁷⁻¹⁸⁹ However, disease-length CGG/CCG repeats have been shown to be much poorer incorporators.¹⁹⁰⁻¹⁹²

1.9 GAP IN KNOWLEDGE AND INSIGHTS PROVIDED BY THIS WORK

Despite the limited investigation into the behavior of TNRs associated with HD and FXS in the packaged genome, there are several lines of evidence that suggest understanding changes in the functional organization of genes within HD and FXS cells could be important in understanding expansion and disease progression. As demonstrated by the molecular pathogenicity of FXS, the formation of heterochromatin plays a key role in disease manifestation.^{38,50-52} Recent evidence has also suggests that expression levels of the histone variant macroH2A are correlated with the progression of HD, making macroH2A a possible biomarker for disease progression and therapeutic treatment.¹⁹³ Additionally, histone deacetylase (HDAC) inhibitors show promise as a treatment for a range of neurodegenerative disorders.^{23,27,33,194,195} Finally, as experiments have demonstrated that modifications to a TNR tract can extensively alter nucleosome positioning in a yeast mini-chromosome,^{196,197} we must investigate the behavior of the repeat tract within the packaged genome in order to fully understand both the expansion

process and how expansion influences the functional organization of the TNR-containing gene.

While others have primarily investigated the behavior of disease-length repeats in chromatin, it is equally important to consider the behavior of healthy-length repeats. To understand the expansion process, we must investigate how the functional organization of the repeat tract changes as it expands, and that can only be accomplished by understanding the behavior of the healthy-length repeat tract. Furthermore, many of the previous experiments use DNA that is not only removed from its genetic context, but is placed within flanking sequences without regard for the effect of flanking sequence on nucleosome formation. Many experiments use series of DNA substrates with inconsistent DNA flanking sequences, complicating interpretation of the results. Here, we use DNA substrates designed for systematic evaluation of the behavior of healthy length repeats in the nucleosome core particle, creating substrates designed to investigate the innate behavior of the repeat tract, regardless of genetic context. Furthermore, we also created substrates designed to investigate the ability of the gene flanking sequence to modulate the innate behavior of the repeat tract. Finally, no previous investigation has determined what effect the presence of a hairpin has on the behavior of DNA in a nucleosome. Indeed, little structural information exists about the behavior of a TNR hairpin in a three-way junction, a substrate that must first be characterized before any nucleosome experiments can take place. We find that small changes to the repeat tract, in length, primary sequence, or secondary structure, can alter the behavior of DNA in the nucleosome, potentially altering the functional organization of the gene. Furthermore, we provide evidence that healthy- and disease-length repeats do not exhibit the same

behavior in a packaged context, and illuminate important differences between the two classifications.

Chapter 2

Healthy-Length CAG/CTG and CGG/CCG Repeats Alter Affinity for the Histone Core and Positioning of DNA in the Nucleosome¹

2.1 INTRODUCTION

As discussed in the previous chapter, trinucleotide repeats (TNRs) are genetically unstable.^{17-20,198} They occur throughout the genome and their expansion or contraction, which is proposed to be mediated by the ability of repeats to form non-canonical secondary structures during DNA replication, transcription, and/or repair, can affect varied cellular processes such as gene expression, mRNA processing, and protein folding.^{17-20,198} Furthermore, TNR expansion has been linked to a class of neurodegenerative disorders, including Huntington's Disease (HD) and Fragile X Syndrome (FXS).^{22,34-41,199}

In the genome, the length of the repeat tract falls into one of three categories: a short, genetically stable repeat length found in healthy individuals, an intermediate, unstable repeat length that is prone to expansion, and a threshold repeat length that, once reached, leads to manifestation of the disease state.^{17-20,198} This threshold number of repeats varies by disease. Onset of HD occurs when there are more than 40 CAG/CTG repeats in exon 1 of the huntingtin (*HTT*) gene, whereas healthy individuals have between 5 and 36 repeats.²² Onset of FXS occurs when there are greater than 200 CGG/CCG repeats in the 5' untranslated region (UTR) of the Fragile X Mental Retardation 1 (*FMRI*) gene.^{37,38,41,42,50-52} Healthy individuals have between 5-55 repeats. Interestingly, sequencing analysis of healthy *FMRI* alleles has shown that AGG/CCT interruptions are

¹ Adapted from Volle, C. B. and Delaney, S. (2012) *Biochemistry* **51** 9814-9825

present every 9-10 repeats, with (CGG)₉₋₁₀AGG(CGG)₉AGG(CGG)₉ being the most common allele.⁴²⁻⁴⁸ The importance of these interruptions is underscored by the observation that uninterrupted CGG/CCG tracts containing 34-59 repeats can expand to disease-length, yet similar repeat tracts with AGG/CCT interruptions are stably transmitted to offspring.^{43,45,47,49} Thus, AGG/CCT interruptions appear to play an important role in suppressing expansion of CGG/CCG TNRs.

In vivo studies have shown that TNRs are found in both tightly packed, transcriptionally-silent heterochromatin and the more loosely packed, transcriptionally-active euchromatin.^{52,56} Interestingly, HD repeats are found in association with histones containing euchromatin markers regardless of whether they are healthy- or disease-length.¹⁶⁴ However, the transition from euchromatin to heterochromatin is an important step in the pathogenesis of FXS.^{54,56,200}

The most basic unit of packaging in chromatin is the nucleosome, which is comprised of ~146 base pairs (bps) of DNA wrapped around a histone octamer.^{170,201-204} Previous studies have shown that disease-length CAG/CTG repeats are preferential locations for nucleosome formation¹⁸⁸ and that as the length of the repeat tract increases across the disease range, the ability of the DNA to incorporate into a nucleosome also increases.¹⁸⁷ However, disease-length CGG/CCG repeats show little affinity for the histone core.¹⁹¹ Notably, previous work has focused mainly on the incorporation of disease-length TNRs into nucleosomes; much less information is available on how repeat tracts found in healthy individuals incorporate.^{56,187-192,205} However, in order to better understand the mechanism by which expansion occurs, it is important to consider the behavior of both healthy- and disease-length repeat tracts in a nucleosome. Furthermore,

the DNA used in previous nucleosome incorporation studies was often constructed from genomic clones containing human repeat tracts; these DNA substrates often vary in their overall length, providing opportunities for multiple translational positions, and/or are removed from their genetic flanking sequence. For these reasons, we synthesized 146 bp sequences containing either CAG/CTG or CGG/CCG repeat tracts with lengths corresponding to those found in healthy individuals. In one series, a nucleosome positioning sequence flanks the repeats, while the other series uses the genetic flanking sequence. We assessed the ability of these sequences to incorporate and to position in nucleosomes. These results describe how healthy-length TNRs behave in chromatin and how this behavior compares to that of disease-length repeat tracts. An understanding of how these sequences behave in nucleosomes defines how the functional organization of the *HTT* and *FMRI* genes, in particular the CAG/CTG or CGG/CCG repeats respectively, can influence the accessibility of the DNA, progression of RNA and DNA polymerases, and occurrence of genomic expansion.

2.2 EXPERIMENTAL PROCEDURES

2.2.1 Oligonucleotide synthesis and purification

Oligonucleotides were synthesized using standard phosphoramidite chemistry on a BioAutomation DNA/RNA synthesizer (sequences can be found Table 2.1). The oligonucleotides were synthesized dimethoxytrityl (DMT)-on and were deprotected in NH_4OH at 55 °C for at least 36 h. One round of HPLC purification was carried out at 90 °C using a Polymer Labs Reverse Phase column (4.6 × 250 mm) using 100 mM triethyl ammonium acetate (TEAA) in 99% acetonitrile:1% water (solvent A) and 100 mM

TEAA in 1% acetonitrile:99% water (solvent B) as the mobile phases (gradient: solvent A was increased from 5% to 25% over 25 min; 1 mL/min). Following removal of the 5'-DMT group by incubation in 80% glacial acetic acid for 12 min at room temperature, the oligonucleotides were dried, resuspended in formamide, and purified by 8% denaturing polyacrylamide gel electrophoresis (PAGE) (0.4 mm thick, 45 cm long). After electrophoresis at 65 W for 4 h, DNA was visualized by UV shadowing. Bands containing the 146-mer were excised from the gel, crushed and soaked overnight at 37 °C in 10 mM Tris, 1 mM EDTA, pH 7.5, and the supernatant collected. The DNA was then lyophilized, resuspended in 1 mL of water and desalted using Sephadex G-50. The desalted DNA was then subjected to a second round of HPLC purification (gradient: solvent A was increased from 0% to 25% over 40 min; 1 mL/min). The pUC19 fragment was PCR amplified from the plasmid pUC19 using the primers forward: 5'-CCATTCGCCATTCAGGCTGCGCAA-3' and reverse: 5'-GCGGATAACAATTTACACAGGAA-3'.

Table 2.1: Sequences of DNA used in CAG/CTG and CGG/CCG nucleosome studies

Name	Sequence (5'- 3')	Complement (5'- 3')
S1	S1a: GTAAATTAATAGTTGAAGTTGTAGT AAATGTTAATGTAGATCTGTTGTTCCGA TATTACCAAAACCTTCACCTAATAGCGC TAGTACACAGGAGAAGGACATGAACAT GAACCTAATGAACACAACAATAATGT AAGTGCCCATG	S1b: CATGGGCACTTACATTATTTGTTGTG TTCATTAGGTTTCATGTTTCATGTCCTTCTC CTGTGTACTAGCGCTATTAGGTGAAGGT TTGGTAATATCGGAACAACAGATCTAC ATTAACATTTACTACAACCTCAACTATT AATTTAC
pUC19	CCATTCGCCATTCAGGCTGCGCAACTGT TGGGAAGGGCGATCGGTGCGGGCCTCTT CGCTATTACGCCAGCTGGCGAAAGGGG GATGTGCTGCAAGGCGATTAAAGTTGGGT AACGCCAGGGTTTTCCAGTCACGACGT TGTAACACGACGGCCAGTGAATTCGAGC TCGGTACCCGGGGATCCTCTAGAGTCGA CCTGCAGGCATGCAAGCTTGGCGTAATC ATGGTCATAGCTGTTTCTGTGTGAAAT TGTTATCCGC	GCGGATAACAATTTACACAGGAAACA GCTATGACCATGATTACGCCAAGCTTGC ATGCCTGCAGGTCGACTCTAGAGGATCC CCGGGTACCGAGCTCGAATTCAGTGCC GTCGTTTTACAACGTCGTGACTGGGAAA ACCCTGGCGTTACCCAACTTAATCGCCT TGCAGCACATCCCCCTTCGCCAGCTGG CGTAATAGCGAAGAGGCCCGCACCGAT CGCCCTCCCAACAGTTGCGCAGCCTGA ATGGCGAATGG
S1- CAG10	GTAAATTAATAGTTGAAGTTGTAGTAAA TGTTAATGTAGATCTGTTGTTCCGATATT	CATGGGCACTTACATTATTTGTTGTGTTC ATTAGGTTTCATGTTTCATGTCCTTCTC-

	ACCA-(CAG) ₁₀ - GAGAAGGACATGAACATGAACCTAATG AACACAACAAATAATGTAAGTGCCCATG	(CTG) ₁₀ - GTAATATCGGAACAACAGATCTACATTA ACATTTACTACAACCTCAACTATTAATTT AC CATGGGCACTTACATTATTTGTTGTGTT ATTAGGTTTCAT-(CTG) ₂₀ - AACAGATCTACATTAACATTTACTACAA CTTCAACTATTAATTTAC
S1- CAG20	GTAAATTAATAGTTGAAGTTGTAGTAAA TGTTAATGTAGATCTGTT-(CAG) ₂₀ - ATGAACCTAATGAACACAACAAATAAT GTAAGTGCCCATG	
S1- CAG30	GTAAATTAATAGTTGAAGTTGTAGTAAA TGT-(CAG) ₃₀ - ACAACAAATAATGTAAGTGCCCATG	CATGGGCACTTACATTATTTGTTGT- (CTG) ₃₀ - ACATTTACTACAACCTCAACTATTAATTT AC
HTT- CAG10	GACCGCCATGGCGACCCCTGGAAAAGCTC ATCAACCGGTTTCGAGTCCCTCAAGTCCT TC-(CAG) ₁₀ - CAAGCACCGCCACCGCCGCCGCCGCCGC CGCCGCCTCCTCAGCTTCCTCAGCCTCCT C	GAGGAGGCTGAGGAAGCTGAGGAGGCG GCGGCGGCGGCGGCGGCGGTTGGCGGTG CTTG-(CTG) ₁₀ - GAAGGACTTGAGGGACTCGAACCGGTT GATGAGCTTTTCCAGGGTCGCCATGGCG GTC
HTT- CAG20	CCTGGAAAAGCTCATCAACCGGTTTCGAG TCCCTCAAGTCCTTC-(CAG) ₂₀ - CAAGCACCGCCACCGCCGCCGCCGCCGC CGCCGCCTCCTCAGC	GCTGAGGAGGCGGCGGCGGCGGCGGCG GCGGTGGCGGTGCTTG-(CTG) ₂₀ - GAAGGACTTGAGGGACTCGAACCGGTT GATGAGCTTTTCCAGG
HTT- CAG30	CAACCGGTTTCGAGTCCCTCAAGTCCTTC- (CAG) ₃₀ - CAAGCACCGCCACCGCCGCCGCCGCCGC	GCGGCGGCGGCGGCGGCGGTGGCGGTGCTT G-(CTG) ₃₀ - GAAGGACTTGAGGGACTCGAACCGGTT G
S1- CGG19	GTAAATTAATAGTTGAAGTTGTAGTA AATGTTAATGTAGATCTGTT(CGG) ₁₉ A ACATGAACCTAATGAACACAACAAAT AATGTAAGTGCCCATG	CATGGGCACTTACATTATTTGTTGTGT TCATTAGGTTTCATGTT(CCG) ₁₉ AACAGA TCTACATTAACATTTACTACAACCTTCA ACTATTAATTTAC
S1- 1AGGa	GTAAATTAATAGTTGAAGTTGTAGTA AATGTTAATGTAGATCTGTT(CGG) ₄ AG G(CGG) ₁₄ AACATGAACCTAATGAACAC AACAAATAATGTAAGTGCCCATG	CATGGGCACTTACATTATTTGTTGTGT TCATTAGGTTTCATGTT(CCG) ₁₄ CCT(CC G) ₄ AACAGATCTACATTAACATTTACTA CAACTTCAACTATTAATTTAC
S1- 1AGGb	GTAAATTAATAGTTGAAGTTGTAGTA AATGTTAATGTAGATCTGTT(CGG) ₈ AG G(CGG) ₁₀ AACATGAACCTAATGAACAC AACAAATAATGTAAGTGCCCATG	CATGGGCACTTACATTATTTGTTGTGT TCATTAGGTTTCATGTT(CCG) ₁₀ CCT(CC G) ₈ AACAGATCTACATTAACATTTACTA CAACTTCAACTATTAATTTAC
S1- 2AGG	GTAAATTAATAGTTGAAGTTGTAGTA AATGTTAATGTAGATCTGTT(CGG) ₄ AG G(CGG) ₉ AGG(CGG) ₄ AACATGAACCTA ATGAACACAACAAATAATGTAAGTGC CCATG	CATGGGCACTTACATTATTTGTTGTGT TCATTAGGTTTCATGTT(CCG) ₄ CCT(CC G) ₉ CCT(CCG) ₄ AACAGATCTACATTAAC ATTTACTACAACCTCAACTATTAATTT AC
FMR1- CGG19	CGGCGGCGGTGACGGAGGCGCCGCTG CCAGGGGGCGTGCGGCAGCG(CGG) ₁₉ C TGGGCCTCGAGCGCCCGCAGCCCACT TCTCGGGGGCGGGCTC	GAGCCCGCCCCCGAGAGGTGGGCTGC GGGCGCTCGAGGCCAG(CCG) ₁₉ CGCT GCCGCACGCCCCCTGGCAGCGGCGCC TCCGTACCCGCCGCCG
FMR1-	CGGCGGCGGTGACGGAGGCGCCGCTG	GAGCCCGCCCCCGAGAGGTGGGCTGC

1AGGa	CCAGGGGGCGTGCGGCAGCG(CGG) <u>4A</u> <u>GG(CGG)</u> ₁₄ CTGGGCCTCGAGCGCCCGC AGCCACCTCTCGGGGGCGGGCTC	GGGCGCTCGAGGCCAG(CCG) <u>14CCT</u> (CCG) ₄ CGCTGCCGCACGCCCCCTGGCA GCGGCGCCTCCGTCACCGCCGCCG
FMR1- 1AGGb	CGGCGGCGGTGACGGAGGCGCCGCTG CCAGGGGGCGTGCGGCAGCG(CGG) <u>8A</u> <u>GG(CGG)</u> ₁₀ CTGGGCCTCGAGCGCCCGC AGCCACCTCTCGGGGGCGGGCTC	GAGCCCGCCCCGAGAGGTGGGCTGC GGGCGCTCGAGGCCAG(CCG) <u>10CCT</u> (CCG) ₈ CGCTGCCGCACGCCCCCTGGCA GCGGCGCCTCCGTCACCGCCGCCG
FMR1- 2AGG	CGGCGGCGGTGACGGAGGCGCCGCTGC CAGGGGGCGTGCGGCAGCG(CGG) <u>4AGG</u> (CGG) ₉ AGG(CGG) ₄ CTGGGCCTCGAGCGC CCGCAGCCACCTCTCGGGGGCGGGCTC	GAGCCCGCCCCGAGAGGTGGGCTGCG GGCGCTCGAGGCCAG(CCG) <u>4CCT(CCG)</u> <u>9CCT(CCG)</u> ₄ CGCTGCCGCACGCCCCCTGG CAGCGGCGCCTCCGTCACCGCCGCCG

2.2.2 Competitive nucleosome incorporation assay

Oligonucleotides were 5'-³²P end-labeled using T4 polynucleotide kinase (New England Biolabs) following the manufacturer's protocol. Radiolabeled DNA, along with a 1.5-fold excess of its complement, was resuspended to a final volume of 50 μ L in Tris buffer (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.5), heated at 95 °C for 5 min and cooled at a rate of 1 °C/min to 25 °C, forming duplex.

Nucleosomes were isolated from chicken erythrocytes following a previously published protocol.²⁰⁶ Briefly, erythrocytes were collected by centrifugation and washed copiously with Solution X (15 mM Tris, 60 mM KCl, 15 mM NaCl, 0.15 mM spermine, 0.5 mM spermidine, 0.34 M sucrose, 2 mM EDTA, 0.5 mM EGTA, 0.2 mM PMSF, 1 mM benzamidine, 15 mM 2-mercaptoethanol, pH 7.5; the last 3 components were added immediately prior to use). The cellular membranes were lysed by addition of 0.1% Nonidet p40 to Solution X followed by gentle resuspension. The nuclei were then collected by centrifugation. The nuclei were again washed with Solution X to remove hemoglobin. A final wash was performed by resuspending the nuclei in Solution N (15 mM Tris, 60 mM KCl, 15 mM NaCl, 0.15 mM spermine, 0.5 mM spermidine, 0.34 M

sucrose, 0.2 mM PMSF, 15 mM 2-mercaptoethanol, pH 7.5; the last 2 components were added immediately prior to use).

The chromatin was digested by addition of CaCl_2 to a final concentration of 1 mM and incubated with 200 U/mL micrococcal nuclease (New England Biolabs). The digestion proceeded for 7 min at 37 °C before quenching with EDTA (final concentration of 2 mM). Nuclei were then collected by centrifugation and lysed by addition of Solution R (10 mM Tris, 0.2 mM EDTA, 0.2 mM PMSF, pH 8.0; PMSF was added immediately prior to use). The chromatin was then allowed to shake overnight at 4 °C, bringing the digested chromatin into solution.

The digested chromatin was collected by centrifugation, concentrated to a volume of no more than 4 mL using Amicon concentrators with a 3 kDa molecular weight cut-off, and the NaCl concentration adjusted to 0.65 M by drop wise addition of 4 M NaCl. This solution was applied to a Sepharose 4B column (1.5 cm $\times$ 75 cm) equilibrated with Solution S (20 mM Tris, 0.2 mM EDTA, 0.63 M NaCl, 0.2 mM PMSF, pH 7.5; PMSF was added immediately prior to use). Using a peristaltic pump, digested chromatin was eluted by flowing Solution S at $\sim$ 0.5 mL/min. 25 fractions, containing 5 mL each, were collected and assessed by SDS-PAGE to determine in which fractions the nucleosomes eluted, using the known molecular weights of the individual histone proteins.²⁰⁷ Samples containing only the core histone proteins were pooled, concentrated, subjected to a second digestion with micrococcal nuclease to ensure nucleosomes were present as mono-nucleosomes, and re-purified using the same procedure. Nucleosomes were dialyzed against Solution D (10 mM Tris, 0.2 mM EDTA, 0.2 mM PMSF, pH 7.5; PMSF was added immediately prior to use), concentrated to $\sim$ 1 mL, the concentration

determined by Bradford Assay. The final concentration was adjusted to 1 mg/mL with solution D.

Four incorporation reactions were prepared per DNA substrate; three reactions contained nucleosomes while one reaction lacked nucleosomes.²⁰⁸ The reaction lacking nucleosomes serves as a control for migration of the DNA duplex during native PAGE. Each 50 μ L incorporation reaction consisted of 100 ng of radiolabeled 146 bp duplex, 2.5 μ g calf-thymus DNA, 25 μ g isolated nucleosomes, 1 M NaCl, and 10X exchange buffer (EB) (1X EB; 20 mM Tris, 1 mM EDTA, 1 mM 2-mercaptoethanol, pH 8.0). For reactions that lacked nucleosomes, an equal volume of Solution D was substituted. The initial incorporation reaction was incubated for 1 h at room temperature. Next, 17 μ L of 1X EB was added to each reaction and incubated for 1 h at room temperature. The addition of 16.5 μ L of 1X EB followed by 1 h incubation at room temperature was repeated twice, doubling the reaction volume. Then 100 μ L 1X EB was added to the reaction and incubated for 15 min at room temperature; this process was repeated twice more. After a final addition of 100 μ L 1X EB, bringing the final reaction volume to 500 μ L, the reaction was allowed to incubate for 1 h at 37 °C. Each reaction was then transferred to an Amicon Micro-concentrator with a 3 kDa molecular weight cut-off and centrifuged at 4 °C for 30 min at 14,000 $\times$ g, reducing the final volume to ~30 μ L. These incorporation reactions were completed at least three times for each DNA sequence for a total of at least nine incorporations per DNA.

2.2.3 Native polyacrylamide gel electrophoresis

Thirty μ L samples from the competitive nucleosome incorporation assay were mixed with 6 μ L non-denaturing dye (15% ficoll, 0.25% bromophenol blue (w:v), 0.25%

xylene cyanol (w:v)), and 10 μ L were applied to a 5% native polyacrylamide gel (4 mm thick, 45 cm long). Samples were electrophoresed at 100 V for 12 h and the products were visualized by phosphorimagery, using a BioRad FX or BioRad Pharos scanner and the accompanying Quantity One software. These conditions provided clear resolution between single-stranded, duplex, and nucleosome samples. The ratio of incorporated DNA (DNA in a nucleosome) to free DNA was determined by comparing the number of counts in each band.

2.2.4 DNase I, Exo III, and S1 nuclease reactions

For enzymatic probing, DNA was exchanged onto the nucleosomes in a similar manner to that described for the competitive nucleosome incorporation assays, with a few alterations. First, no competitor DNA was added to the reaction and the amount of radiolabeled duplex was increased to 10 pmol. Second, the exchange buffer was altered (1X EB; 5 mM Tris, pH 7.5). Finally, the incubation time was shortened; while the first 1 h incubation remained, the subsequent 1 h incubations were changed to 45 min, and 15 min incubations changed to 10 min. This procedure produced samples in which ~90-95% of the DNA incorporated into the nucleosome.

After concentration of the exchanged nucleosomes, enzymatic probing reactions were prepared, combining 0.4 units DNase I (New England Biolabs), 50 units Exo III (New England Biolabs), or 42.5 units S1 nuclease (Promega) with 5 μ L of the appropriate buffer for the enzyme to be used: DNase I buffer (100 mM Tris, 25 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.6), NEB1 buffer (100 mM Bis Tris propane-HCl, 100 mM $MgCl_2$, 10 mM DTT, pH 7.0), or S1 nuclease buffer (500 mM sodium acetate, 2.8 M NaCl, 45 mM $ZnSO_4$, pH 4.5). The reaction volume was adjusted to 50 μ L with

exchanged, concentrated nucleosomes. The samples were incubated at 37 °C for 5 min and then quenched by the addition of 6 µL of 0.1 M EDTA and 6 µL of 50% glycerol. Samples were then loaded onto a 5% native polyacrylamide gel (8 mm thick, 16 cm long), and electrophoresed at 4 °C, 150 V for 3.5 h. The gel was used to expose a phosphorplate for approximately 25 min. The plate was imaged using a BioRad Pharos scanner, and the image used to identify the position of the DNA incorporated into nucleosomes. These bands were excised, crushed, and soaked overnight at 37 °C in 10 mM Tris, 1 mM EDTA, pH 7.5. After centrifugation to pellet the remaining gel, the supernatant was concentrated in a 15 mL volume Amicon concentrator with a 3 kDa molecular weight cut-off. Centrifugation occurred for 50 min at 4 °C, 3220 × g leaving ~0.5 mL of recovered nucleosomes. SDS was added to 0.01%, and samples were heated at 90 °C for 10 min. To preserve the duplex structure, this step was omitted for S1 nuclease reactions carried out after the incorporated DNA had been separated from the histone core. Samples were then extracted twice with 400 µL of phenol:chloroform:isoamyl alcohol, ethanol precipitated, and the resulting pellet dried under vacuum.

Reactions were also performed on free duplex. In the Exo III and S1 nuclease reactions the amount of enzyme added was the same as the nucleosome reactions. However, the enzyme concentration of DNase I optimized for the nucleosome reactions resulted in over-cleavage of the free duplex and so a 1:10 dilution was used. Duplex was annealed as described previously and 5 pmol was used to adjust the reaction volume to 50 µL. In all enzymatic reactions of free duplex, the incubation time was shortened to 2 min and the reactions were quenched by addition of 6 µL 0.1 M EDTA and 6 µL of 1 mg/mL

calf-thymus DNA. The DNA was immediately ethanol precipitated and dried under vacuum. A standard Maxam-Gilbert A/G reaction performed on either strand of S1 (S1a or S1b, Table 2.1) was used to create a marker.

After the DNA pellets were dry, the number of counts was determined in a Perkin-Elmer scintillation counter and diluted with formamide to 5,000 counts per microliter. Three microliters of these samples were applied to an 8% denaturing polyacrylamide gel (0.4 mm thick, 65 cm long), and electrophoresed at 65 W for 2.5 h. The gel was dried and used to expose a phosphorplate for 96 h.

2.2.5 Hydroxyl radical footprinting

Nucleosome preparation was the same for the hydroxyl radical reactions as it was for the enzymatic probing. To generate hydroxyl radicals, 7.5 μL each of 1 mM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$:2 mM EDTA, 10 mM sodium ascorbate, and 0.6% H_2O_2 was added to 52.5 μL of exchanged nucleosomes, bringing the total reaction volume to 75 μL . The solution was incubated for 15 min at room temperature. The reactions were quenched by addition of 8 μL of 50% glycerol and 8 μL of 0.1 M EDTA and samples were immediately applied to a native polyacrylamide gel as described above. The sample preparation process was the same as for the enzymatic probing experiments from this point forward. Free duplex reactions were carried out under the same reaction conditions as the nucleosome samples except they were incubated for 5 min at room temperature, quenched with 8 μL of 50% glycerol and 8 μL of 1 mg/mL calf-thymus DNA, and ethanol precipitated. From this point forward sample preparation followed the same procedure as the enzymatic probing samples. Each sample was electrophoresed in three separate gels, one of which was run in the same manner as described above. However, to

better visualize reactivity at the 5' or 3' end of the DNA, samples were also electrophoresed for 1.75 h (to visualize the 5' end) or 3.5 h (to visualize the 3' end).

2.2.6 Image processing

Phosphorplates were imaged using a BioRad Pharos scanner and the associated Quantity One software. Histograms were generated using the Quantity One software, but all other image processing was done using BioRad's Image Lab software. First, the gel image was aligned and the lanes identified. Lane width and position was adjusted and the position of each band manually determined. Using the accompanying Maxam-Gilbert sequencing reactions, each band was assigned by nucleotide and the corresponding intensity determined. Intensity was graphed as a function of nucleotide, smoothed by applying a moving average with a window of 5, and that data imported into Origin 8.2 (OriginLab Corporation). The data was used to create a graph, which was baseline subtracted and used for curve fitting experiments.

In order to determine the periodicity of the DNA around the histone core, a sine wave was fit to the hydroxyl radical footprinting data, the phase of which is equal to the best-fit period of the DNA.²⁰³ Using the non-linear curve fitting function in the Origin software, a best-guess sine wave, with a phase of ~10, was fit to the data. From the initial guess the software refined the fit until the Chi-squared was optimized. The sine wave was applied to the entire length of the DNA or to certain sections, such as the repeat tract or flanking sequences.

The Origin software was also used to determine the location of maxima and minima reactivity to the hydroxyl radical, which was used to construct translational positioning maps, revealing the location of the dyad axis.²⁰⁹ The nucleotide position of

each maximum was determined and the number of nucleotides between two consecutive maxima ($y = \max_n - \max_{n-1}$) was plotted against the average value of the maxima ($x = [\max_n + \max_{n-1}] / 2$). The same process was applied to the minima. The points determined for both the maxima and minima were smoothed with a moving average window of 2 or 3 and graphed.

2.3 RESULTS AND DISCUSSION

2.3.1 Substrate design

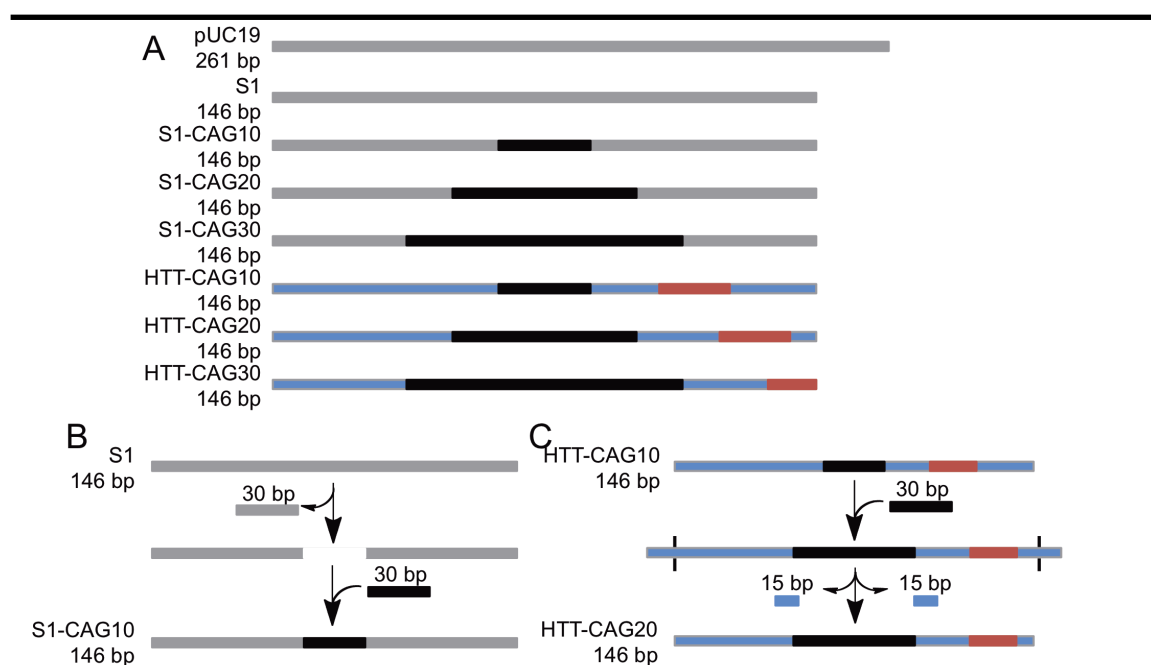


Figure 2.1: (A) Schematic representations of the DNA used to investigate CAG/CTG repeats in the nucleosome and its design. S1 and pUC19 serve as controls and contain no repeat tracts. S1-CAG10, S1-CAG20, and S1-CAG30 contain 10, 20, and 30 CAG/CTG repeats (black), respectively, within the S1 flanking sequence (gray). HTT-CAG10, HTT-CAG20, and HTT-CAG30 also contain 10, 20, and 30 CAG/CTG repeats (black), respectively; however these substrates contain the *HTT* flanking sequence (NG_009378.1, blue). The CCG/CGG repeats of the *HTT* flanking sequence are shown in dark red. (B) Schematic design of the S1 series of DNA in which the CAG/CTG repeats replace the center of the 146 bp S1 control. (C) Schematic design of the HTT series of DNA in which the size of the repeat tract was increased, and the flanking sequence was pushed outwards and extra sequence trimmed from the ends to maintain a 146 bp length. Notably, all the 146 bp DNA sequences were prepared directly by phosphoramidite chemistry, and the schematics in panels B and C represent our design strategy rather than experimental preparation.

In the first series of DNA substrates, those containing healthy-length CAG/CTG repeats, we started with the 146 bp control duplex S1, a strong nucleosome positioning sequence with a single rotational and translational setting, which is derived from the pD89 sequence (Figure 2.1A).^{209,210} The repeat-containing DNA was designed by replacing the center of the S1 duplex with 10, 20, or 30 CAG/CTG repeats to generate S1-CAG10, S1-CAG20, and S1-CAG30, respectively, while maintaining the overall 146 bp length (Figure 2.1B). While the length of the DNA substrates might restrict nucleosome formation, we chose to limit the length to 146 bp, the length of DNA wrapped around a single histone core, to minimize the probability that these substrates would form a heterogeneous population of nucleosomes.

To assess the ability of healthy-length CGG/CCG repeats to incorporate into nucleosomes, we designed a 146 bp substrate containing 19 CGG/CCG repeats centered within S1 (S1-CGG19) (Figure 2.2). We then substituted specific CGG/CCG repeats for

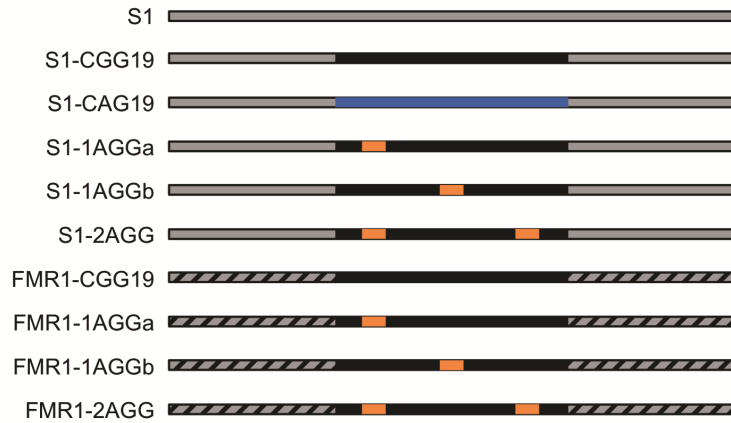


Figure 2.2: Schematic representations of the DNA used to investigate interrupted and uninterrupted CGG/CCG repeats. S1 serves as a control and contains no triplet repeat tracts. S1-CGG19 contains 19 CGG/CCG repeats (black) within the S1 flanking sequence (gray). S1-CAG19 contains 19 CAG/CTG repeats (blue) within the S1 flanking sequence (gray). S1-1AGGa, S1-1AGGb, and S1-2AGG contain AGG/CCT interruptions (orange) within the repeat tract. FMR1-CGG19, FMR1-1AGGa, FMR1-1AGGb, and FMR1-2AGG also contain CGG/CCG repeats (black) and AGG/CCT interruptions (orange), however they contain the *FMR1* flanking sequence (NG_007529.1, dashed).

AGG/CCT interruptions, creating substrates with either one (S1-1AGGa, S1-1AGGb) or two interruptions (S1-2AGG). We chose 19 CGG/CCG repeats as well as the placement of the interruptions from a previously characterized set of interrupted and uninterrupted (CGG)₁₉ oligonucleotides, which were designed using biological precedence.⁶⁰

While the S1 sequence has been used in previous nucleosome investigations, and therefore serves as a good control, repeats are not flanked by the S1 sequence in the genome. Thus we designed substrates containing the same CAG/CTG repeat tract lengths, but changed the flanking sequence to that of the *HTT* gene, creating HTT-CAG10, HTT-CAG20, and HTT-CAG30. These substrates were designed to mimic an expansion event; the TNR region was expanded and the *HTT* flanking sequence was pushed outwards, allowing us to determine what effect of the *HTT* flanking sequence has on packaging of healthy-length CAG/CTG repeats. The ends of the duplex were trimmed to maintain a 146 bp length (Figure 2.1C). We also created substrates containing the same interrupted and uninterrupted CGG/CCG repeats now containing the *FMRI* gene flanking sequence. (Figure 2.2)

2.3.2 Healthy-length CAG/CTG repeats increase nucleosome formation

To understand the behavior of healthy-length CAG/CTG repeats in chromatin, we first performed competitive nucleosome incorporation assays, determining the extent to which a particular duplex incorporates into a nucleosome.²⁰⁸ In this assay, the DNA of interest was radiolabeled and exchanged onto the histone core of nucleosomes isolated from chicken erythrocytes. Unlabeled competitor DNA, in this case calf-thymus DNA, acts as an internal control and, as it was in gross excess over the radiolabeled DNA, approximates an equilibrium distribution of the DNA incorporated into the nucleosome.

Following the incorporation assay, native PAGE was used to separate radiolabeled duplex incorporated into nucleosomes from free, unincorporated duplex. As a mixed-sequence control, we also used a 261 bp fragment of pUC19 and examined its incorporation into a nucleosome.

We find that the S1 duplex incorporates into a nucleosome with the same efficiency as the pUC19 duplex, each with an incorporation ratio of ~ 2 , indicating that there are two radiolabeled DNA duplexes incorporated into a nucleosome for every one radiolabeled duplex that remains as free DNA (Figure 2.3A,D). These results indicate that, while the S1 duplex was chosen because of its positioning ability, it incorporates with the same efficiency as the mixed-sequence pUC19 DNA. Notably, since some radiolabeled DNA remains free and is not incorporated into a nucleosome, the S1 duplex provides a suitable baseline from which there is the potential to observe either increased or decreased incorporation upon introduction of a CAG/CTG sequence.

Interestingly, by native PAGE we observe two distinct bands for pUC19 duplex incorporated into nucleosomes, indicating that the pUC19 fragment forms at least two species of nucleosomes with different translational positions, in which the translational position is defined as the position of the histone core relative to the ends of the DNA duplex (Figure 2.3A). Although in previous studies^{190,191,205} multiple, distinct species are not observed for nucleosomes formed from this pUC19 fragment, it is possible that our PAGE conditions allowed for resolution of multiple species.

When the center of the S1 duplex is replaced by CAG/CTG repeats there is a significant increase in the ratio of DNA found in a nucleosome relative to free DNA (Figure 2.3B,D). Even the addition of only ten CAG/CTG repeats (S1-CAG10) provides

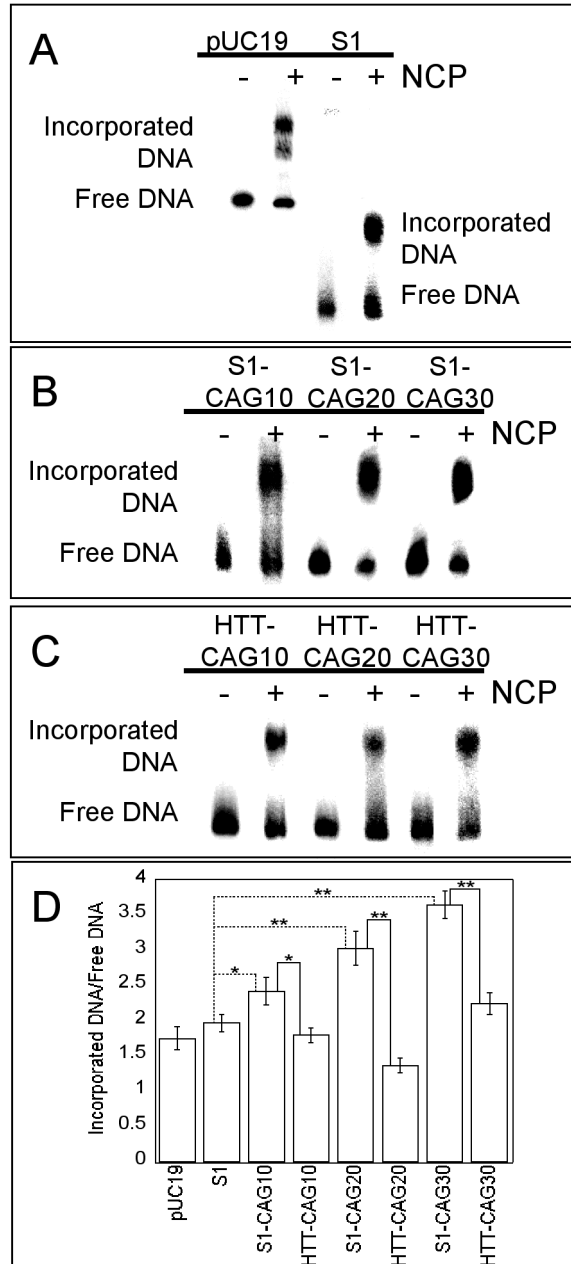


Figure 2.3: Representative competitive nucleosome incorporation reactions for the CAG/CTG repeat-containing substrates. An incorporation performed in the absence of chicken nucleosomes (NCP) (-NCP) and an incorporation performed in the presence of NCP (+NCP) are shown. The incorporation assays with pUC19 and S1 controls are shown in panel A, the incorporation assays with S1 DNA containing 10, 20 or 30 CAG/CTG repeats are shown in panel B, and the incorporation assays with HTT DNA are shown in panel C. The ratio of incorporated DNA to free DNA is shown in panel D. The error bars represent the standard error for each substrate. * $p \leq 0.05$ and ** $p \leq 0.005$ by Student's T-Test.

a small but significant increase in nucleosome incorporation compared to the S1 control; the ratio increases from 2 to 2.4. Addition of another ten CAG/CTG repeats (S1-CAG20) further increases the incorporation ratio to 3. The largest CAG/CTG repeat tract studied here, within S1-CAG30, has the highest incorporation ratio of the investigated substrates at 3.7. As the length of the repeat tract is the only difference between the S1 substrates, the increase in incorporation can be attributed to the increase in the number of CAG/CTG repeats.

The increased incorporation ratio for DNA containing CAG/CTG repeats relative to mixed-sequence controls has traditionally been ascribed to the flexibility induced by the AG/CT dinucleotide, which lowers the free energy (ΔG) of duplex bending, indicating a more flexible duplex overall.^{190,211-213} Furthermore, CAG/CTG duplex has been shown to be an inherently flexible molecule based on electrophoretic mobility and cyclization kinetics.²¹⁴ It is thought that more flexible DNA is better able to bend around

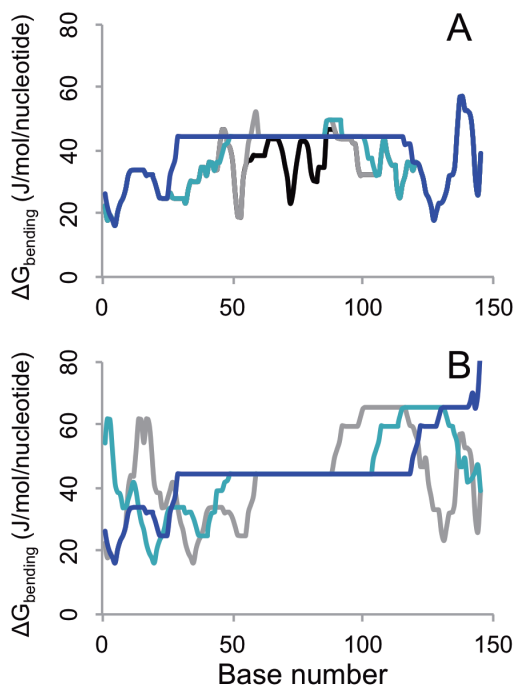


Figure 2.4: ΔG of bending values calculated for the S1 and HTT substrates. Using the equation described by Satchwell and Travers,²¹² and as applied by Metzenberg,²¹¹ the ΔG of bending was calculated across each substrate, using a 6-nucleotide window. S1 (black), S1-CAG10 (gray), S1-CAG20 (green), and S1-CAG30 (blue) are shown in panel A. HTT-CAG10 (gray), HTT-CAG20 (green), and HTT-CAG30 (blue) are shown in panel B.

the histone core, leading to higher incorporation ratios.²¹⁵ However, based on analysis of the ΔG of bending for each substrate, and thus an analysis of the flexibility of the substrate, it appears that the ΔG of bending for the CAG/CTG repeats is in the same range as that of the S1 sequence it replaces (Figure 2.4A). Thus, there must be other factors not accounted for in the ΔG of bending that make CAG/CTG repeats such robust incorporators, such as the ability to compress the minor groove and accommodate the unique constraints imposed by bending around the histone core.

The trends we observe in the S1 series are consistent with previously reported results for longer, disease-length CAG/CTG repeats, in which the presence of the repeats was found to facilitate nucleosome formation. By direct observation of nucleosome formation by electron microscopy (EM), repeat tracts containing up to 250 CAG/CTG repeats have been shown to be preferential sites for nucleosome formation.¹⁸⁸ Furthermore, repeat tracts containing either 75 or 130 CAG/CTG repeats have been shown to significantly increase the ratio of incorporated DNA to free DNA.¹⁸⁷ Importantly, because of the large number of repeats, the flanking sequence does not contribute to the incorporation in these experiments, where nucleosome formation occurs almost exclusively within the repeat tract.

Interestingly, it has been reported that there is negligible difference in the incorporation efficiency between DNA containing 10 CAG/CTG repeats and DNA containing either 55 or 62 CAG/CTG repeats.¹⁸⁹ However, the 55 CAG/CTG repeat sequence forms nucleosomes with several translational positions, many of which contain either the 3' or 5' flanking sequence at the same position as the nucleosomes formed with

10 CAG/CTG repeats, indicating that the flanking sequence and not the repeat may be dominating the incorporation process.

2.3.3 Nineteen CGG/CCG repeats have no effect on incorporation into nucleosomes; the addition of AGG/CCT interruptions does affect nucleosome incorporation

When 19 CGG/CCG repeats are centered in the S1 sequence (S1-CGG19) there is no significant change in the ratio of incorporated DNA to free DNA (Figure 2.5A,C), relative to the S1 control. This contrasts with the observed increase for S1-CAG19, where 19 CAG/CTG repeats are centered in the S1 sequence. To our knowledge, this is the first direct comparison of these triplet repeat sequences where the number of repeats, flanking sequence, and overall DNA length are identical. Therefore, the 1.8 fold difference in nucleosome incorporation observed for S1-CGG19 and S1-CAG19 is derived exclusively from the identity of the triplet repeat sequence and provides a quantitative measure of the differing ability of these repeats to form nucleosomes under these conditions.

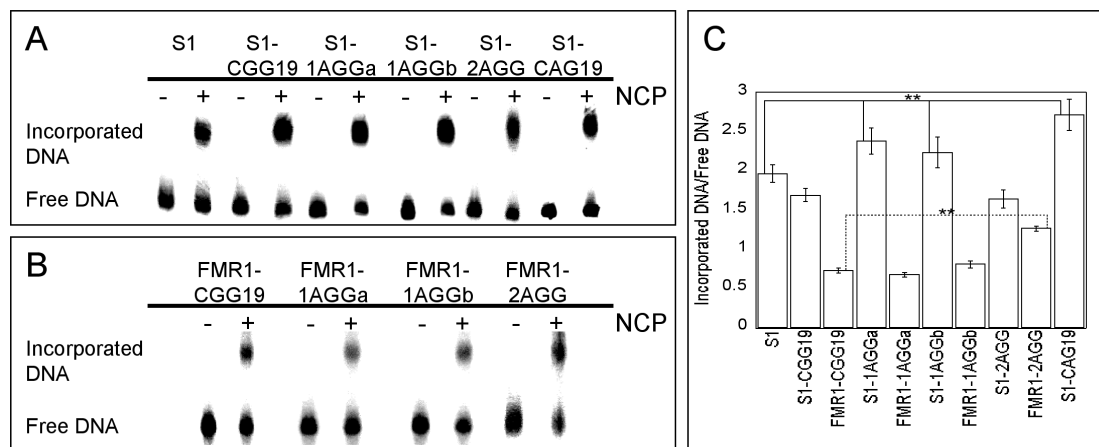


Figure 2.5: Representative competitive nucleosome incorporation reactions for the CGG/CCG repeat-containing substrates. The incorporation assays with the S1 series of CGG/CCG substrates are shown in panel A and the incorporation assays with FMR1 series of substrates are shown in panel B. The ratio of incorporated DNA to free DNA is shown in panel C. The error bars represent the standard error for each substrate. ** $p \leq 0.005$ by Student's T-Test.

In previous experiments designed to observe the location of nucleosome formation on a closed, circular plasmid by electron microscopy, 76 CGG/CCG repeats were found to reduce nucleosome formation, meaning that the region of the plasmid containing the CGG/CCG repeats was populated with a nucleosome less frequently than other regions of the plasmid.¹⁹¹ In other experiments, 10 and 24 CGG/CCG repeats placed in the center of a pUC19 fragment incorporated into a nucleosome less readily than a control that lacked the repeats. However, it is notable that these experiments were performed using DNA substrates that did not contain uniform flanking sequences, opening the possibility that the primary sequence of the flanking DNA could be affecting the observed results. Furthermore, it is also possible that rather than the CGG/CCG repeats themselves decreasing nucleosome incorporation, the introduction of the repeats disrupted a region of the pUC19 DNA sequence that promotes nucleosome incorporation. In other work, experiments performed with 6 CGG/CCG repeats in the histone H4 gene, found that CGG/CCG repeats can favorably contribute to nucleosome incorporation, though these experiments were also performed without regard for flanking sequence.¹⁸⁹ Perhaps our experiments with 19 repeats represent a middle ground where the length of the CGG/CCG repeat tract is too long to facilitate nucleosome incorporation but is not long enough to destabilize interaction with the histone core.

To investigate the consequences of interruptions on nucleosome incorporation, we introduced AGG/CCT interruptions into the S1-CGG19 duplex (Figure 2.2). When a single CGG/CCG is replaced with AGG/CCT in either the fifth or ninth repeat (S1-1AGGa and S1-1AGGb, respectively) there is a ~1.5-fold increase in the incorporation ratio as compared to S1-CGG19 (Figure 2.5A,C). This result indicates that the presence

of just one AGG/CCT interruption significantly impacts the ability of the DNA to interact with the histone core.

It is noteworthy that this observed increase in nucleosome incorporation is due to a change in 1 out of 146 bps; to introduce the AGG/CCT interruption, a C/G bp was changed to A/T. Based on analysis of the free energy (ΔG) of DNA bending for these (Figure 2.6A-D) and other substrates,^{190,211-213} it has been suggested that AGG/CCT interruptions represent a localized decrease in ΔG within the *FMR1* repeat tract, thereby decreasing the energetic barrier for CGG/CCG repeat bending around the histone core, making the incorporation of the interrupted sequences into a nucleosome slightly more

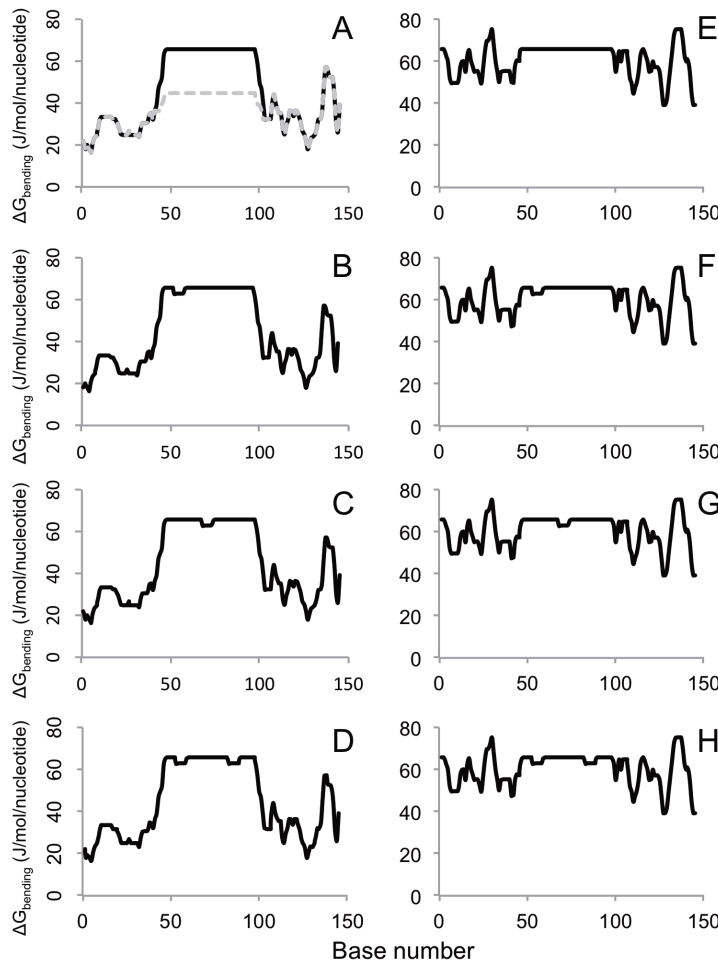


Figure 2.6: ΔG of bending values calculated for the S1 and FMR1 substrates. Using the equation described by Satchwell and Travers,²¹² and as applied by Metzenberg,²¹¹ the ΔG of bending was calculated across each substrate, using a 6-nucleotide window. S1-CGG19, S1-1AGGa, S1-1AGGb, and S1-2AGG (black) are shown in panels A-D respectively. S1-CAG19 (dashed gray) is also shown in panel A. FMR1-CGG19, FMR1-1AGGa, FMR1-1AGGb, and FMR1-2AGG (black) are shown in panels E-H respectively.

energetically favorable. However, as observed for the CAG/CTG repeats, it is also likely that other factors pertaining to nucleosome formation play a role as well, but are not accounted for in these calculations.

Interestingly, when an AGG/CCT interruption is included at both the fifth and fifteenth repeats (S1-2AGG), the level of incorporation is comparable to the uninterrupted S1-CGG19 (Figure 2.5A,C). We had anticipated that DNA containing two interruptions would incorporate more readily than S1-CGG19, as observed for the sequences with a single interruption. We speculate that the lack of observable benefit to incorporation with the S1-2AGG substrate is due to the position of the interruptions within the repeat tract. In the S1-2AGG substrate, the interruptions are positioned equidistant from the ends of the repeat tract, with ten intervening CGG/CCG repeats located in the center of the DNA. Because of the positioning effects of the S1 flanking sequence, this configuration places an interruption at either edge of a region within the nucleosome spanning the dyad axis where the DNA is locally underwound.^{170,203,204} As discussed above, uninterrupted CGG/CCG repeats possess a relatively large ΔG of bending compared to the S1 flanking sequence, and therefore have an energetic barrier that limits the ability to bend around the histone core.²¹¹ Thus, because of the positioning effect of the S1 flanking sequence, each interruption provides a localized decrease in ΔG at either end of this region, but the intervening CGG/CCG repeats may not be able to readily accommodate the underwinding required for nucleosome formation, leading to lower incorporation. Notably, the results obtained with one and two AGG/CCT interruptions underscores the importance of both the location of the interruption within the repeat tract and the location of DNA around the histone core.

The relationship between nucleosome incorporation and the presence of AGG/CCT interruptions in CGG/CCG repeat tracts was investigated previously.¹⁹⁰ Using DNA fragments derived from human *FMRI* repeat tracts, when 3 or 2 interruptions were present in tracts of 39 or 53 CGG/CCG repeats respectively, the interruptions did not influence nucleosome incorporation on histones isolated from HeLa cells. However, the interruptions did increase formation of hyperacetylated nucleosomes. The histones used in our experiments are likely comparable in acetylation to those isolated from HeLa cells, as they were both isolated from biological sources, but the absolute measure of epigenetic modifications is unknown. Interestingly, when ATG/CAT interruptions were added to a CAG/CTG repeat tract, the incorporation ratio decreased.¹⁹⁰

2.3.4 The genetic flanking sequence decreases nucleosome incorporation

While it is important to understand the innate influence of CAG/CTG and CGG/CCG repeats on nucleosome incorporation, it is equally important to consider the effect of the genetic flanking sequence. To this end, we first repeated the competitive nucleosome incorporation using the HTT series of DNA; substrates designed with the same number of CAG/CTG repeats now flanked by the *HTT* gene sequence (Figure 2.1). The HTT substrates all had significantly lower incorporation ratios than their corresponding S1 counterparts; the ratios for HTT-CAG10, HTT-CAG20 and HTT-CAG30 are 1.8, 1.2, and 2.2 respectively (Figure 2.3C,D). The significant decrease observed for each substrate in the HTT series indicates that the *HTT* flanking sequence possesses an inherent negative effect on nucleosome incorporation relative to the corresponding S1 flanking sequence, offsetting the benefit from the CAG/CTG repeats.

It is important to note that the *HTT* series contains not only the primary CAG/CTG repeats but also a secondary CCG/CGG repeat in the flanking sequence (Figure 2.1A), a TNR that has decreased nucleosome incorporation ratios in various contexts,^{190-192,205,216} including the results described above. Compared to the CAG/CTG repeats, CCG/CGG repeats have a much higher ΔG of bending, which helps to increase the overall ΔG of bending for the *HTT* flanking sequence compared to S1 (Figure 2.4B). The lack of sequential A/T bps within the relatively G/C rich *HTT* flanking sequence, which are known to be important in facilitating bending around the histone core, also likely contributes to the observed decrease in incorporation.¹⁸¹

As with the CAG/CTG repeats, it is equally important to understand the effect of interruptions when the CGG/CCG repeats are flanked by the biologically relevant *FMR1* gene sequence. Thus we designed DNA substrates that have the same repeat and interruption configuration as those in the S1 series, but contain the *FMR1* flanking sequence (Figure 2.2). In each substrate, both those with and without interruptions, the *FMR1* flanking sequence decreases the observed level of incorporation, relative to the same repeat flanked by the S1 sequence (Figure 2.5B,C). This decrease in incorporation is likely due at least in part to an increase in the ΔG of bending for the *FMR1* flanking sequence relative to the S1 flanking sequence (Figure 2.6E-H). Interestingly, while a single AGG/CCT interruption yielded an increase in incorporation of the S1 substrates, there is no significant change when a single AGG/CCT interruption is present in the *FMR1* substrates. In contrast, while the two AGG/CCT interruptions in S1-2AGG produced no change in the amount of incorporation relative to the S1-CGG19, there is a significant increase in the incorporation observed in the *FMR1*-2AGG substrate relative

to the FMR1-CGG19 substrate. Thus, not only is the number and placement of the AGG/CCT interruptions an important factor in determining the ability of CGG/CCG repeats to incorporate into a nucleosome, the flanking sequence is also important as it can further modulate the impact of the interruptions.

In the cells of FXS patients, disease-length CGG/CCG repeats are associated with heterochromatic markers, indicating that these repeats are contained within tightly packed heterochromatin.^{50-52,56,200,217} Heterochromatin in the CGG/CCG repeat region spreads along the *FMR1* gene, condensing the *FMR1* promoter and silencing the gene. Packaging in heterochromatin, and thus gene silencing, appears to occur only with very long, uninterrupted repeats. Based on nucleosome incorporation assays, this silencing does not come from the innate ability of CGG/CCG repeats or the *FMR1* flanking sequence to form nucleosomes; both the CGG/CCG repeats and *FMR1* flanking sequence decrease the amount of DNA incorporated into a nucleosome, relative to their respective controls. Therefore, it is most likely that heterochromatic markers associated with the expanded CGG/CCG repeats silence the *FMR1* gene.^{52,190,192,218} However, the experiments described here shed light on the changes in packaging that occur as the genomic CGG/CCG repeats expand. As demonstrated with the S1-CGG19 substrate, short, uninterrupted repeats do not change the ability of DNA to incorporate into a nucleosome. The lack of ability to increase incorporation will leave shorter genomic CGG/CCG repeats available for proper progression of the transcription machinery within the cell, and while short CGG/CCG repeats have been shown to promote polymerase stalling in a bacterial plasmid,⁹⁵ they do not promote polymerase stalling in eukaryotic cells.^{94,197}

2.3.5 DNase I and Exo III reveal the position of DNA in the nucleosome

To better understand the interaction of these healthy-length repeats with the histone core, we performed enzymatic probing on both free and incorporated duplex. We began by investigating the CAG/CTG repeats using DNase I, an endonuclease that creates single-strand nicks.²¹⁹⁻²²² In separate experiments in which either the CAG-containing (Figure 2.7) or CTG-containing (Figure 2.8) strand is radiolabeled, both free duplex and nucleosome samples were subjected to digestion by DNase I, and the products were resolved by denaturing PAGE. While DNase I does not possess strict sequence dependency, it does cleave preferentially at some sequences, mainly depending on minor groove width^{223,224} and DNA flexibility.^{225,226} Indeed, we observe unequal reactivity along all of the free duplex samples.

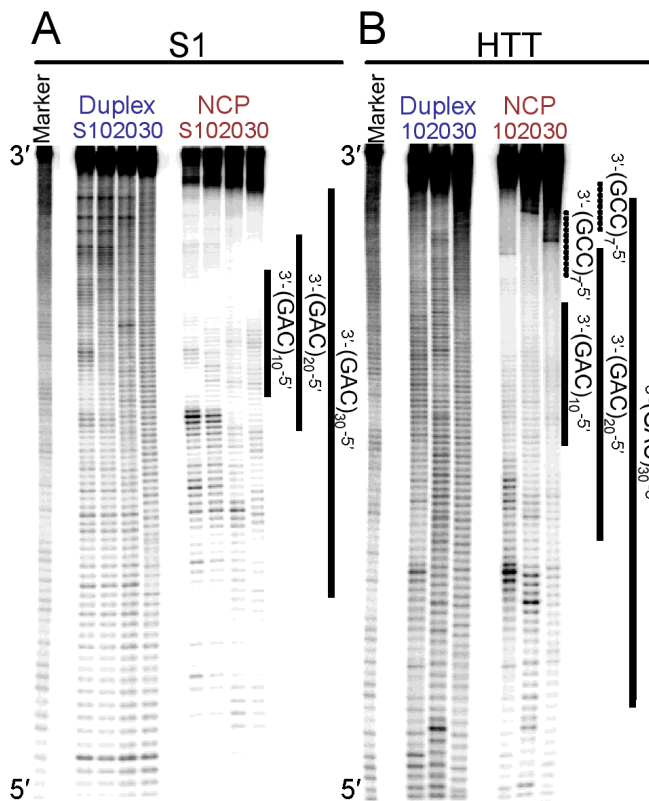


Figure 2.7: DNase I digestion reveals successful incorporation of the CAG/CTG repeat-containing DNA around the histone core, as evidenced by the altered reaction pattern in the nucleosome samples as compared to the duplex. DNase I cleavage of the CAG-containing strands of the S1 substrates are shown in panel A, while cleavage of the CAG-containing strands of the HTT substrates are shown in panel B. Reactions were performed on both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. The number of repeats is indicated above each lane, with S indicating the S1 control. The marker lane consists of the Maxam-Gilbert A/G sequencing reaction performed on S1a. The location of the repeat tracts is indicated on the right-hand side of the gel.

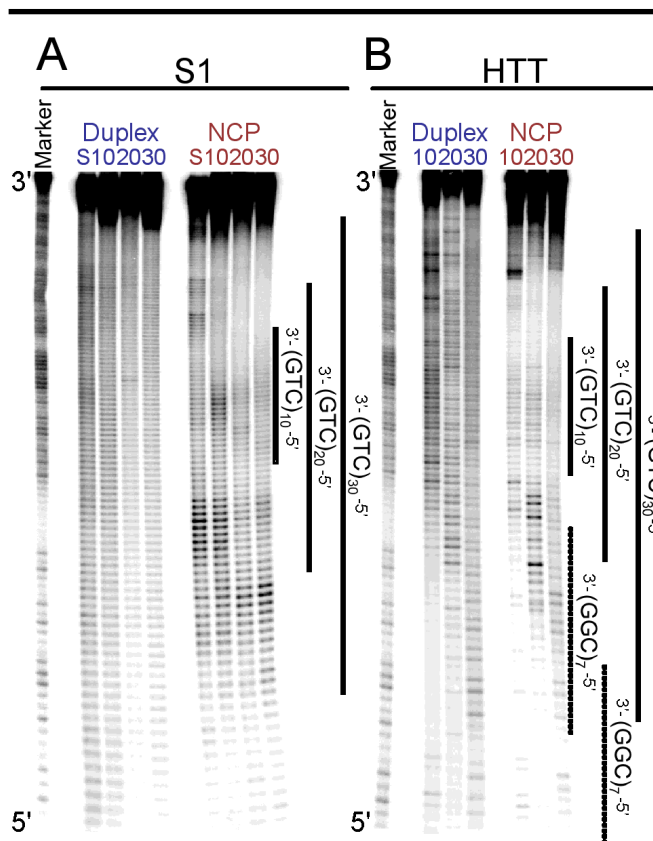


Figure 2.8: DNase I digestion reveals successful incorporation of the CAG/CTG repeat-containing DNA around the histone core, as evidenced by the altered reaction pattern in the nucleosome samples as compared to the duplex. DNase I cleavage of the CTG-containing strands of the S1 substrates are shown in panel A, while cleavage of the CTG-containing strands of the HTT substrates are shown in panel B. Reactions were performed on both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. The number of repeats is indicated above each lane, with S indicating the S1 control. The marker lane consists of the Maxam-Gilbert A/G sequencing reaction performed on S1b. The location of the repeat tracts is indicated on the right-hand side of the gel.

When the DNA is incorporated into a nucleosome, its cleavage by DNase I changes, revealing a pattern of regions with high reactivity toward DNase I followed by regions with little or no reactivity toward DNase I. We observe this reactivity for both the S1 and HTT series incorporated into nucleosomes (Figure 2.7, Figure 2.8). Where the DNA backbone faces out into solution, DNase I can cleave, but DNA facing into the histone core will be inaccessible to the enzyme.²¹⁹⁻²²² The DNA at the end of the duplex is more readily cleaved than the DNA in the center, producing dark bands near the top of each lane, due to decreased interactions between the extremes of the DNA duplex and the histone core.²²⁷

Exo III is a 3' to 5' exonuclease, which digests duplex DNA.²²⁸⁻²³⁰ When the enzyme encounters DNA that is interacting with the histone core it pauses, which results

in intense strand cleavage. These locations of intense strand cleavage map the translational position of the DNA. As with the DNase I reactions, in separate experiments in which either the CAG-containing (Figure 2.9) or CTG-containing (Figure 2.10) strand was radiolabeled, both free duplex and nucleosome substrates were subjected to partial digestion by Exo III and the products were resolved by denaturing PAGE. Like DNase I, Exo III has some sequence specificity,²³¹ indicated by the uneven reaction along the free duplex samples for both the S1 and HTT series.

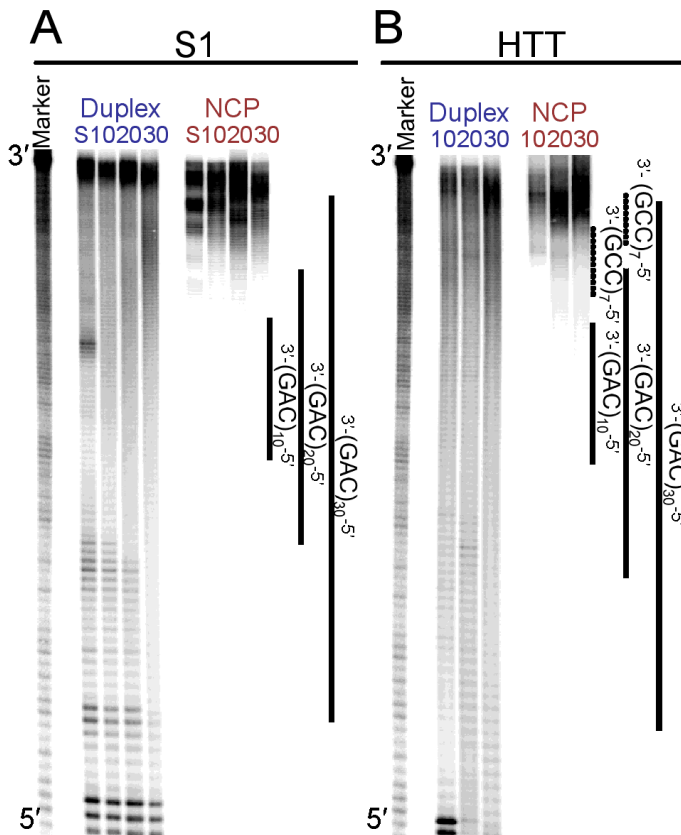
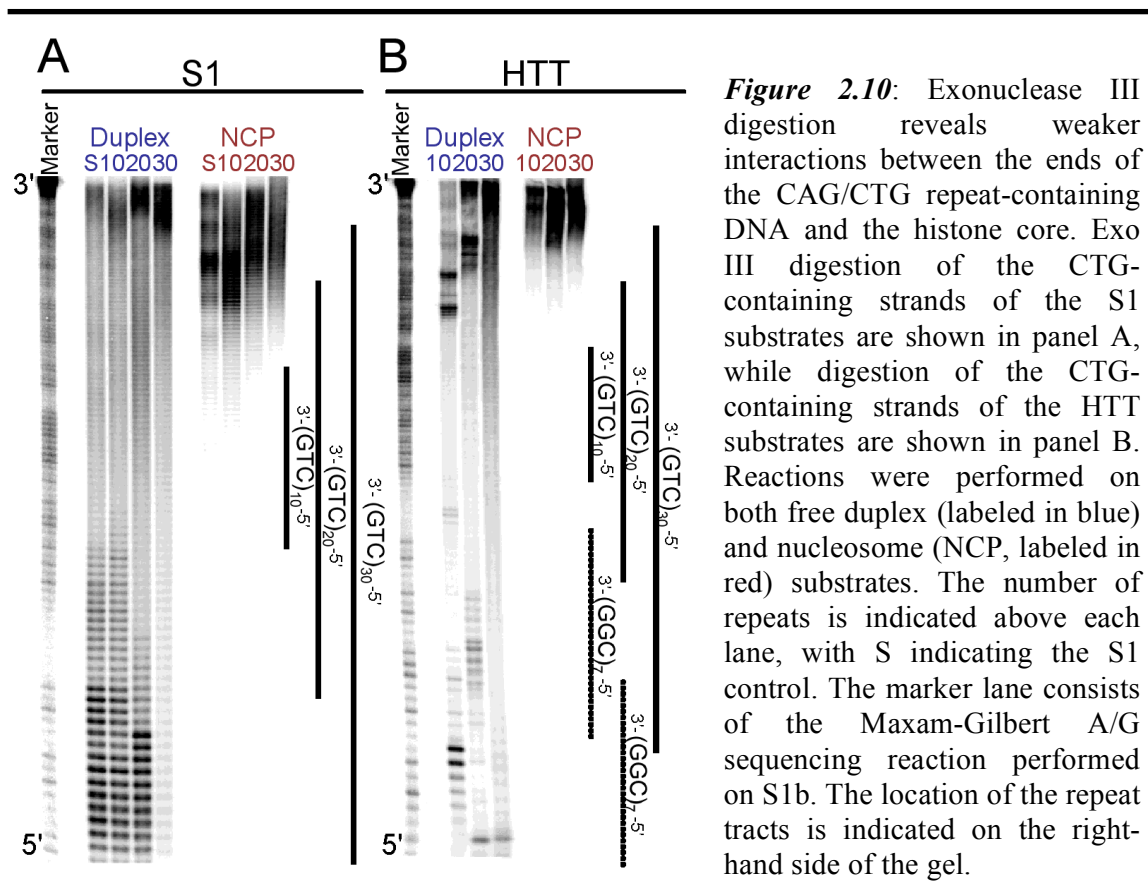


Figure 2.9: Exonuclease III digestion reveals weaker interactions between the ends of the CAG/CTG repeat-containing DNA and the histone core. Exo III digestion of the CAG-containing strands of the S1 substrates are shown in panel A, while digestion of the CAG-containing strands of the HTT substrates are shown in panel B. Reactions were performed on both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. The number of repeats is indicated above each lane, with S indicating the S1 control. The marker lane consists of the Maxam-Gilbert A/G sequencing reaction performed on S1a. The location of the repeat tracts is indicated on the right-hand side of the gel.



There is a clear change in the Exo III digestion pattern when the duplex is incorporated into a nucleosome (Figure 2.9, Figure 2.10). Enzyme reactivity in the nucleosome samples is limited to the ends of the DNA. In the S1 series, there are 3-4 highly reactive sites observed regardless of the strand radiolabeled (Figure 2.9A, Figure 2.10A). While the presence of multiple reactive sites could indicate that the nucleosomal S1 series is forming multiple translational positions, it is more likely that, due to salt-induced shielding of the extremes of the DNA duplex, the interactions between the DNA and the histone core are weaker at the ends, allowing some Exo III to push the DNA off the histone core and continue its digestion.²²⁷ Furthermore, the S1 control sequence has been shown previously to adopt a single translational position under similar conditions.^{209,210}

There are only slight differences observed for Exo III digestion of the HTT substrates, indicating that the three HTT DNAs form nucleosomes with similar translational positions (Figure 2.9B, Figure 2.10B). There is only one main reactive site in each of the HTT substrates, indicating that the *HTT* flanking sequence is in contact with the histone core.

We applied the same Exo III digestion technique to investigate the interaction between the CGG/CCG repeat-containing DNA and the histone core. When the

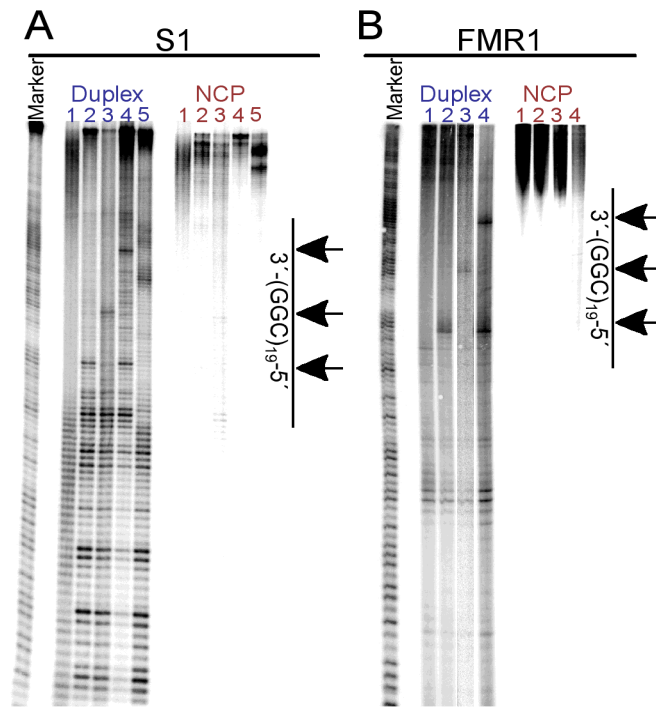


Figure 2.11: Exonuclease III digestion reveals weaker interactions between the ends of the CGG/CCG repeat-containing DNA and the histone core. Exo III digestion of the CGG-containing strands of the S1 substrates are shown in panel A, while digestion of the CGG-containing strands of the FMR1 substrates are shown in panel B. Reactions were performed on both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. In panel A, lanes 1 contain the S1 control, lanes 2 contain S1-CGG19, lanes 3 contain S1-1AGGa, lanes 4 contain S1-1AGGb, and lanes 5 contain S1-2AGG. In panel B, lanes 1 contain FMR1-CGG19, lanes 2 contain FMR1-1AGGa, lanes 3 contain FMR1-1AGGb, and lanes 4 contain FMR1-2AGG. The marker lane consists of the Maxam-Gilbert A/G sequencing reaction performed on S1a. The location of the repeat tract and the A of the interruptions (arrows) are indicated on the right-hand side of the gel.

CGG/CCG repeats, both with and without AGG/CCT interruptions, are contained within the S1 positioning sequence, there is clear reactivity along the length of each free duplex, evidenced by the appearance of discrete bands in the denaturing gel (Figure 2.11A, duplex). However, once the DNA is incorporated into a nucleosome, there is a change in the reactivity; namely strand cleavage only arises from Exo III digestion of the DNA at the end of the duplex (Figure 2.11A, NCP). The same change in reactivity is present in the FMR1 substrates as well (Figure 2.11B). It is interesting to note that, while reactivity in both the S1 and FMR1 nucleosomal substrates is restricted to the same region, the reactivity within that region is different. In the S1 substrates there are discrete, high intensity regions corresponding to the Exo III pause sites. However there are no distinct pause sites present in the FMR1 substrates; the entire region is equally reactive. The difference in reactivity at the ends of the DNA indicates that the FMR1 DNA may not be uniformly positioned (*vide infra*).

2.3.6 Hydroxyl radical footprinting reveals that CAG/CTG repeats alter the rotational and translational positioning of DNA in a nucleosome

While the enzymatic probing with DNase I and Exo III reveals the general position of DNA in the nucleosome, the results can be complicated by the sequence preferences associated with each enzyme. A more sensitive method of determining the position of the DNA in a nucleosome is hydroxyl radical footprinting. The hydroxyl radical reacts preferentially with the C5' hydrogen atom of the sugar-phosphate backbone, located in the minor groove, inducing cleavage at every nucleotide. Regions of most intense strand cleavage indicate where the minor groove is solvent exposed.^{203,209,221,232,233}

To assess the position of each strand of the duplex with respect to the histone core, reactions were preformed in separate experiments where either the CAG-containing

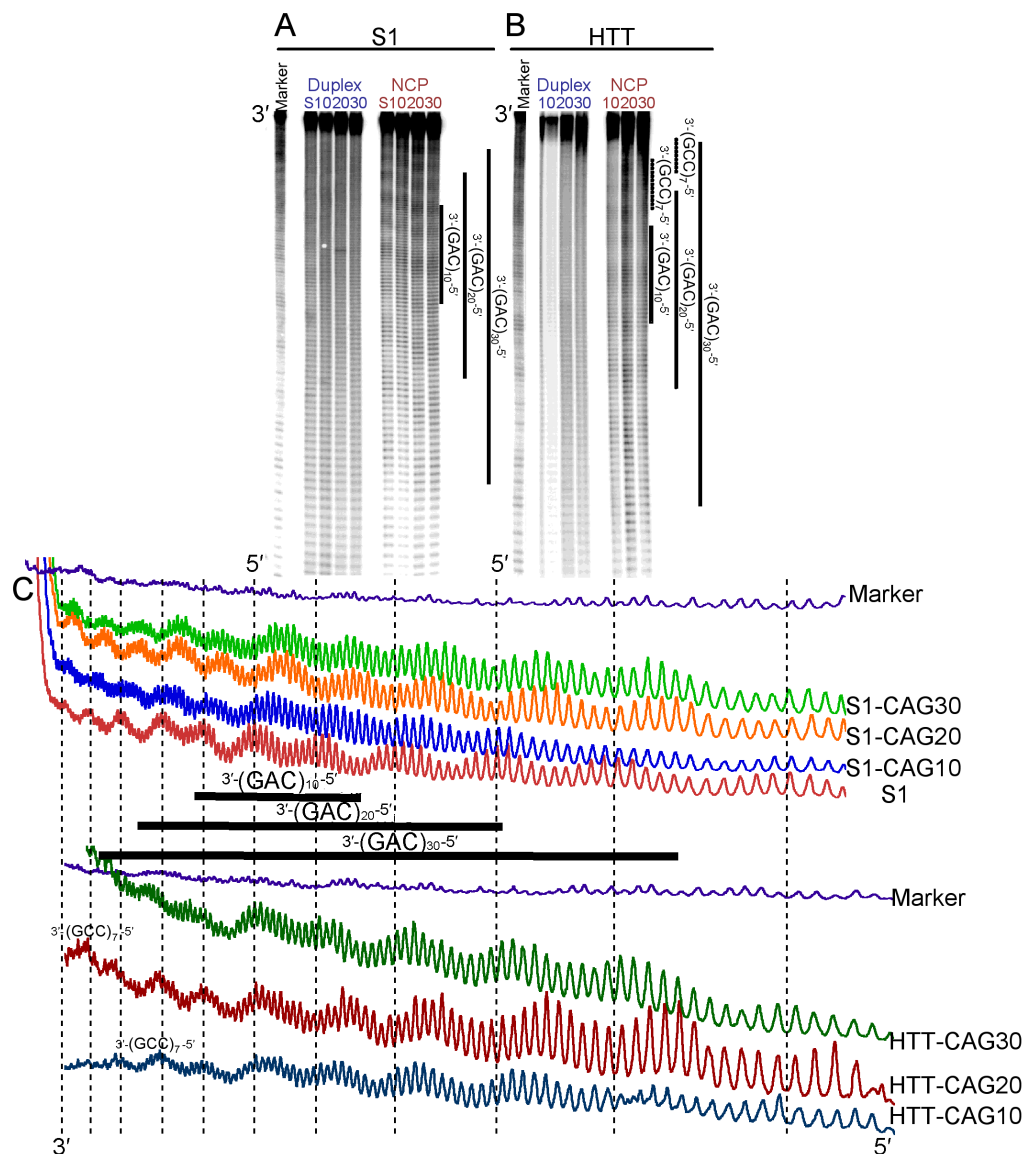


Figure 2.12: Hydroxyl radical footprinting reveals the periodicity of the CAG/CTG repeat-containing DNA around the histone core. Reactions were performed on both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. Reactivity toward hydroxyl radical of the CAG-containing strands of the S1 substrates are shown in panel A, while reactivity of the CAG-containing strands of the HTT substrates are shown in panel B. The number of repeats is indicated above each lane, with S indicating the S1 control. The marker lane consists of the Maxam-Gilbert A/G sequencing reaction performed on S1a. The location of the repeat tracts is indicated on the right-hand side of the gel. Panel C contains histograms generated from the gels presented in A and B. The dashed lines indicate the maxima of the S1 substrate.

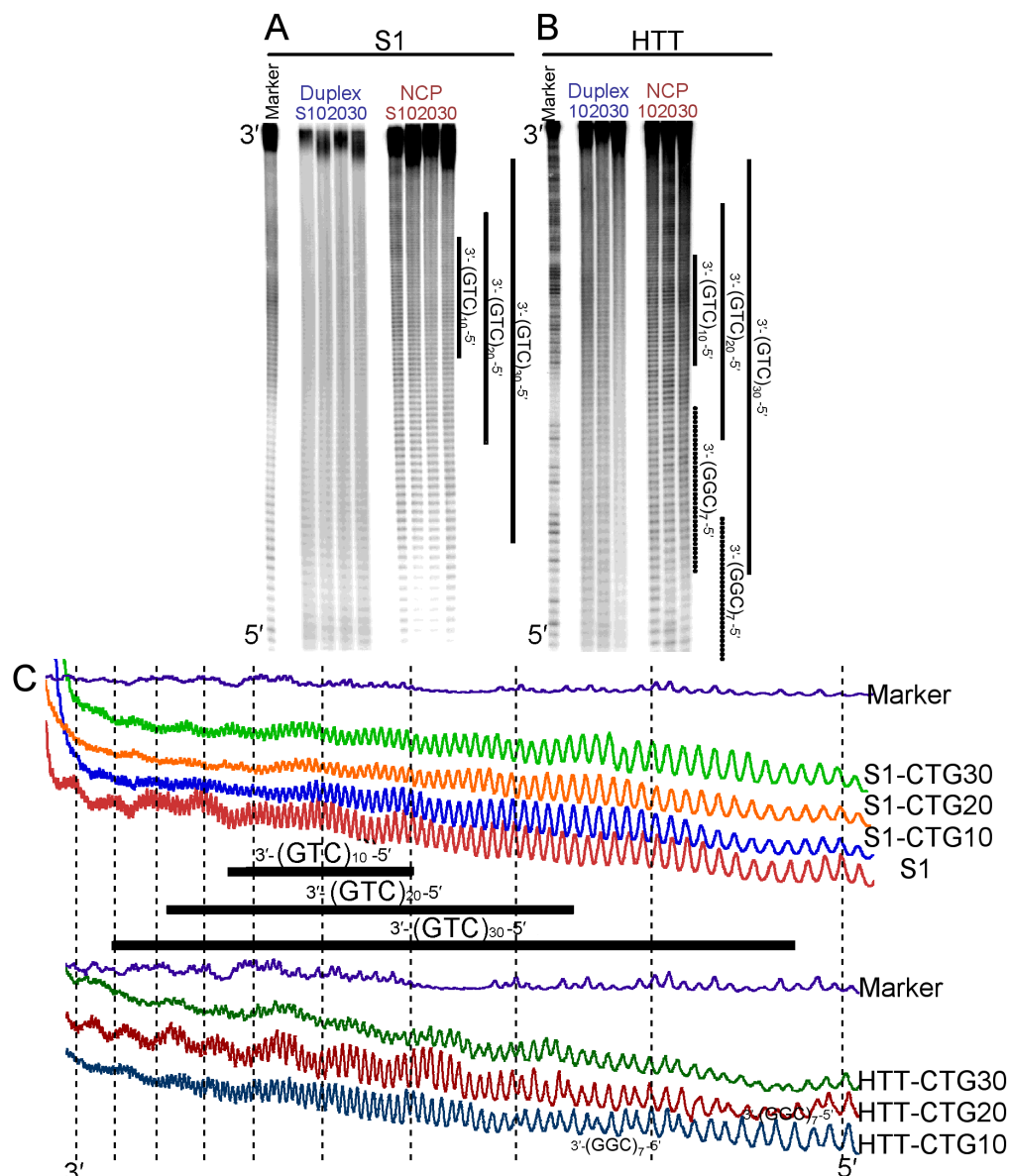


Figure 2.13: Hydroxyl radical footprinting reveals the periodicity of the CAG/CTG repeat-containing DNA around the histone core. Reactions were performed on both free duplex (labeled in blue) and nucleosomes (NCP, labeled in red) substrates. Reactivity toward hydroxyl radical of the CTG-containing strands of the S1 substrates are shown in panel A, while reactivity of the CTG-containing strands of the HTT substrates are shown in panel B. The number of repeats is indicated above each lane, with S indicating the S1 control. The marker lane consists of the Maxam-Gilbert A/G sequencing reaction performed on S1b. The location of the repeat tracts is indicated on the right-hand side of the gel. Panel C contains histograms generated from the gels presented in A and B. The dashed lines indicate the maxima of the S1 substrate.

(Figure 2.12) or CTG-containing (Figure 2.13) strand was radiolabeled. The products of the reactions were resolved by denaturing PAGE, in which 100-120 nucleotides were

observed by standard electrophoresis procedures. The PAGE images were used to create histograms for each sample (Figure 2.12C, Figure 2.13C).

When the free duplex is exposed to hydroxyl radicals, the reaction produces a nearly even pattern of reactivity along the DNA, with a slight increase in reactivity at purines (Figure 2.12A,B, Figure 2.13A,B).^{147,209,221,233} Once the DNA is incorporated into a nucleosome, we observe an oscillating pattern of high and low reactivity towards the hydroxyl radical. Areas of increased reactivity indicate regions where the minor groove faces away from the histone core and is accessible to the hydroxyl radical, whereas areas of reduced reactivity indicate regions where the minor groove faces in towards the histone core and is protected from cleavage by hydroxyl radicals.

Analysis of the hydroxyl radical footprinting reactions on the S1 series reveals the inherent positioning ability of the CAG/CTG repeat tract. The pattern of reactivity is clearly different amongst the S1 series in both the gels and the resulting histograms (Figure 2.12A, Figure 2.13A,C). The vertical, dashed lines in Figure 2.12C indicate the position of each maxima for the S1 substrate, from which the offset of the maxima and minima of other substrates is determined. Using the S1 control as a baseline, we observe that the maxima of S1-CAG20 and S1-CAG30 are shifted by ~5 nucleotides; where a maxima occurs in S1, the S1-CAG20 and S1-CAG30 have minima. While the offset between the S1 and the S1-CAG10 substrate is not as pronounced, the nucleotide with maximum reactivity is offset by ~1-2 nucleotides in the repeat region (Figure 2.12C). However, the shift appears to be localized at the repeat tract, as the extremes of the flanking sequence share maxima and minima with the S1 substrate.

By overlaying the footprinting patterns obtained for the complementary strands, it is possible to identify regions where the minor groove faces out into solution, allowing analysis of the rotational position of the DNA (Figure 2.14).²⁰⁹ By comparing which regions of the minor groove face out into solution, we can determine differences in the rotational position between the different substrates. Consistent with the maxima and minima observed in the histograms, the rotational position of the S1 and S1-CAG10 substrates are quite similar, but differ from the rotational position of the S1-CAG20 and S1-CAG30 substrates. This result indicates that for CAG/CTG repeat tracts longer than 20, the repeats override the S1 flanking sequence to define the rotational position (Figure 2.14A-D).

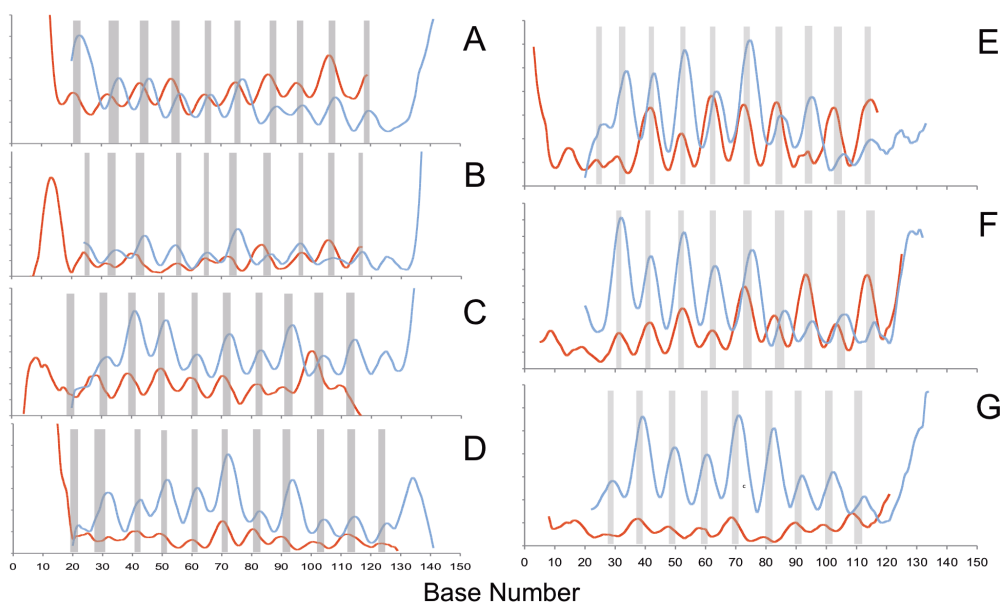


Figure 2.14: Hydroxyl radical cleavage curves reveal the rotational setting of the CAG/CTG repeat-containing DNA around the histone core. The gray bars indicate areas where the minor groove is solvent exposed. The S1 series of substrates, S1 (A), S1-CAG10 (B), S1-CAG20 (C), and S1-CAG30 (D) are displayed on the left, while the HTT series, HTT-CAG10 (E), HTT-CAG20 (F), and HTT-CAG30 (G), are displayed on the right. S1a or the CAG-containing strand is shown in red and S1b or the CTG-containing strand is shown in blue.

While the HTT series possesses some small changes in the maxima and minima present in the gel and resulting histograms, they are much more uniformly aligned than the substrates within the S1 series (Figure 2.12B,C, Figure 2.13B,C). Upon closer examination of the footprinting patterns for the complementary strands, the HTT-CAG10 and HTT-CAG20 share a rotational position, which is different than the rotational position occupied by HTT-CAG30 (Figure 2.14E-G). In these constructs, it is likely that the secondary CCG/CGG repeat, which is reduced to only two of the seven repeats in HTT-CAG30, is playing some role in determining the rotational position of the HTT substrates. The S1-CAG30 and HTT-CAG30 have the same rotational position, indicating that the repeats and not the flanking sequence are determining the rotational position in these substrates. However, unlike the S1 substrates, 20 CAG/CTG repeats is not enough to overcome the influence of the *HTT* flanking sequence; it takes 30 CAG/CTG repeats.

We also examined the periodicity associated with each maxima and minima in the hydroxyl radical footprinting data. By graphing these values, we generated the translational positioning curves shown in Figure 2.15, allowing us to identify extremes in the periodicity with respect to the DNA backbone.²⁰⁹ The position of the maximum in the curve defines the dyad axis, which is given in Table 2.2. The dyad axis of each repeat-containing substrate is located well within the repeat tract, consistent with previous observations that CAG/CTG repeats of various lengths preferentially assemble at the dyad axis.¹⁸⁷⁻¹⁸⁹ Upon addition of 10 CAG/CTG repeats to the S1 substrate, there is a shift in the position of the dyad axis of 2 bps. An additional 10 and 20 repeats shifts the dyad another 2 and 3 bps, respectively. However, when the repeats are placed within the

HTT flanking sequence, the position of dyad axis remains constant at position 71, a further example of the ability of the *HTT* flanking sequence to modulate the behavior of the CAG/CTG repeat.

Table 2.2: Position of the dyad and translational shifts for CAG-containing strands of the S1 and HTT substrates

Substrate	Dyad position	Shift (nucleotides) ^a
S1	73	-
S1-CAG10	71	-2
S1-CAG20	69	-4
S1-CAG30	68	-5
HTT-CAG10	71	-2
HTT-CAG20	71	-2
HTT-CAG30	71	-2

^a With respect to the S1 control.

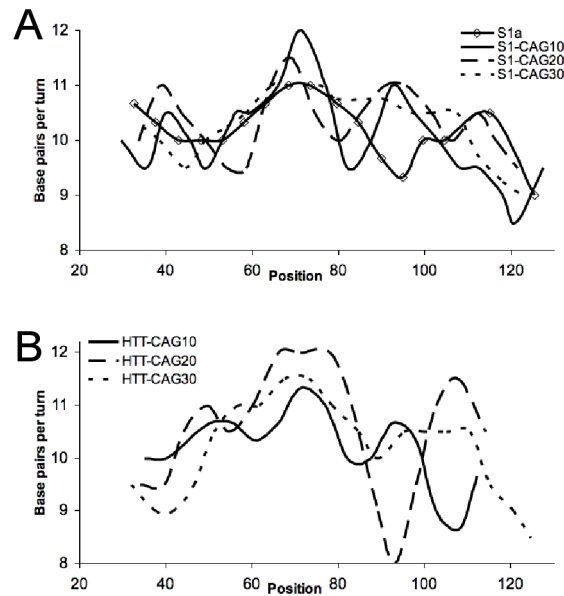


Figure 2.15: Translational positioning curves reveal a change in the position of the dyad axis for both the S1 and HTT series. The locations of the maxima and minima are then graphed with respect to their periodicity. The curve reveals the variation in periodicity around the histone core, and the highest peak indicates the location of the dyad axis. In the S1 series (A), the translational positioning curves reveal 2-5 bp offsets in the location of the dyad axis. In comparison, the HTT series (B) all have the same dyad axis placement.

It is important consider that the length of the DNA substrates used here is limited to the length of DNA associated with a single nucleosome. This length minimizes the possibility that the DNA would form a heterogeneous population of nucleosomes with multiple translational positions, which would confound our analysis. Within the cell, the histone core will have access to a longer DNA substrate with which to form a nucleosome, and while a nucleosome will likely form at the CAG/CTG repeat, given the affinity of the CAG/CTG repeat tract for the histone core, the nucleosomes formed with longer DNA substrates may have multiple DNA positions. However, it is likely that even with longer DNA substrates, CAG/CTG repeats will still alter DNA positioning, though the alterations may be more complex.

2.3.7 CGG/CCG repeats reposition DNA in the nucleosome; AGG/CCT interruptions further alter the positioning

To further understand how the AGG/CCT interruptions modify the behavior of the CGG/CCG repeat tract in the nucleosome, we used hydroxyl radical footprinting to determine the position of the DNA relative to the histone core. First, we used this technique to analyze the ability of 19 CGG/CCG repeats to reposition the S1 DNA in a nucleosome (Figure 2.16, Figure 2.17). Separate experiments were performed with either the CGG-containing (Figure 2.16) or CCG-containing (Figure 2.17) strand radiolabeled and the products are resolved by denaturing PAGE. As described above, exposure of free duplex to hydroxyl radicals results in cleavage along the entire DNA substrate in a relatively even manner, observed as the appearance of discrete bands in the gel (Figure 2.16A, Figure 2.17A; duplex). However, once the DNA is incorporated into a nucleosome, the cleavage pattern changes to the characteristic pattern of oscillating high

and low reactivity associated with nucleosome incorporation (Figure 2.16A, Figure 2.17A; NCP). The amount of cleavage at each nucleotide was determined and used to construct the plots shown in Figure 2.16B and Figure 2.17B, allowing comparison of the overall DNA position between the various DNA substrates. The characteristic, oscillating

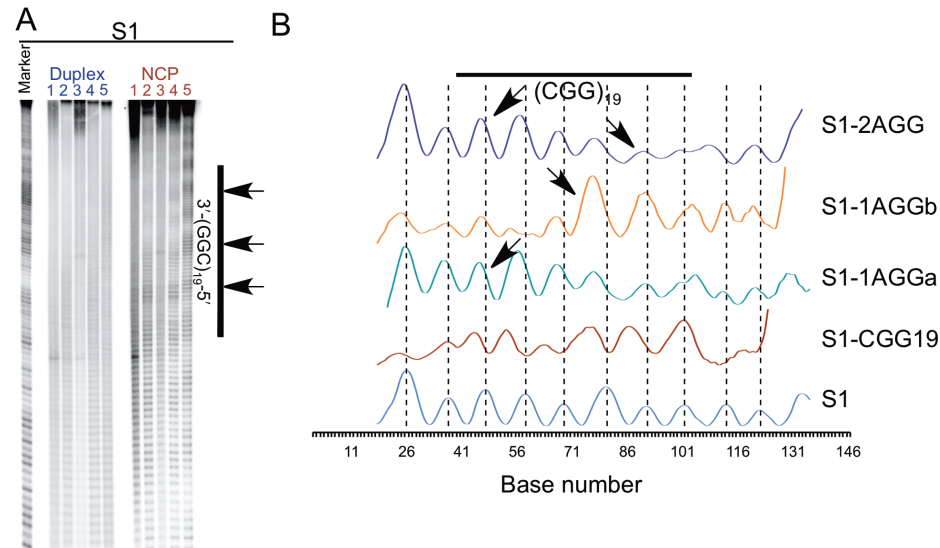


Figure 2.16: Hydroxyl radical footprinting reveals the periodicity of the CGG/CCG repeat-containing S1 DNA around the histone core. Radiolabeled samples, both free duplex (labeled in blue) and nucleosomes (NCP, labeled in red) substrates, were reacted with hydroxyl radicals. Reactivity for the CGG-containing strands of the S1 substrates are shown in panel A. Reactions were performed on both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. In panel A, lanes 1 contain the S1 control, lanes 2 contain S1-CGG19, lanes 3 contain S1-1AGGa, lanes 4 contain S1-1AGGb, and lanes 5 contain S1-2AGG. The marker lane consists of the Maxam-Gilbert A/G sequencing reaction performed on S1a. The location of the repeat tract and interruptions are indicated on the right-hand side of the gel. Panel B shows the reactivity at each nucleotide, generated from the gel presented in A. The dashed lines indicate the maxima of the S1 substrate and the arrows indicate the position of the A of the AGG interruptions.

pattern produced by reaction of the S1 nucleosome substrates with the hydroxyl radical indicates a uniform population of nucleosomes, where the DNA occupies a single rotational and translational position. The vertical, dashed lines in Figure 2.16B and

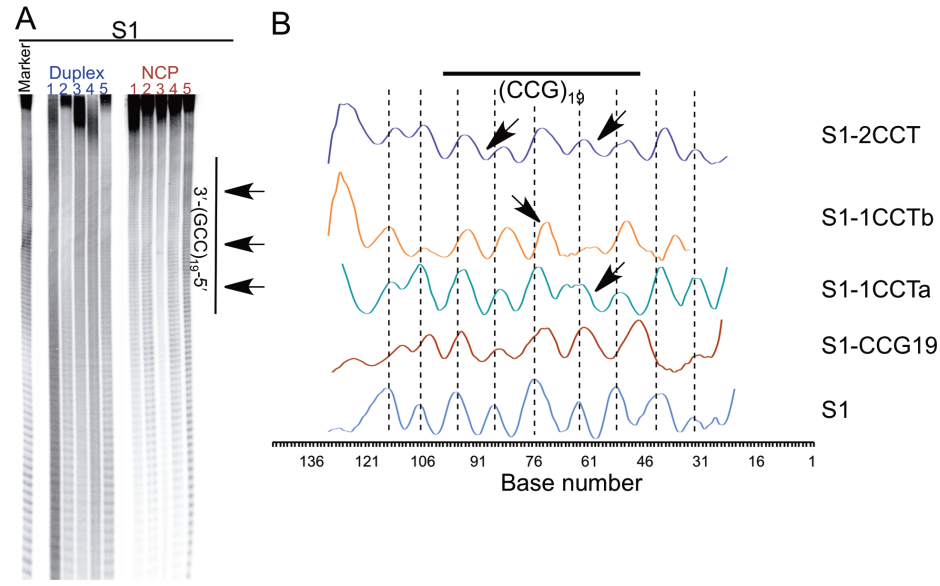


Figure 2.17: Hydroxyl radical footprinting reveals the periodicity of the CGG/CCG repeat-containing S1 DNA around the histone core. Radiolabeled samples, both free duplex (labeled in blue) and nucleosomes (NCP, labeled in red) substrates, were reacted with hydroxyl radicals. Reactivity for the CCG-containing strands of the S1 substrates are shown in panel A. Reactions were performed on both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. In panel A, lanes 1 contain the S1 control, lanes 2 contain S1-CCG19, lanes 3 contain S1-1CCTa, lanes 4 contain S1-1CCTb, and lanes 5 contain S1-2CCT. The marker lane consists of the Maxam-Gilbert A/G sequencing reaction performed on S1b. The location of the repeat tract and interruptions are indicated on the right-hand side of the gel. Panel B contains histograms generated from the gel presented in A. The dashed lines indicate the maxima of the S1 substrate and the arrows indicate the position of the CCT interruptions.

Figure 2.17B indicate the maxima for the S1 control. It is from these lines that the offset in maxima and minima of the other DNA substrates can be determined. When 19 CGG/CCG repeats are located in the center of the S1 DNA, there is a shift of ~4 bases in the position of the maxima and minima relative to the S1 substrate, particularly in the repeat region, indicating a difference in how the DNA is positioned in the nucleosome. Thus, not only can healthy-length CAG/CTG repeats reposition DNA in a nucleosome, but healthy-length CGG/CCG repeats can reposition DNA in a nucleosome as well. When an AGG/CCT interruption replaces a CGG/CCG repeat, the maxima and minima

are again offset, signifying that the DNA again positions differently, now occupying a different overall position than either the S1 or S1-CGG19 nucleosomes.

To examine the rotational positioning of each DNA substrate, the locations where the minor groove faces out into solution were determined and used to compare the rotational position between the S1 substrates (Figure 2.18).²⁰⁹ Upon addition of 19 CGG/CCG repeats to the S1 DNA, there is a shift in the rotational positioning of the DNA in the nucleosome. Interestingly, upon the addition of the AGG/CCT interruptions,

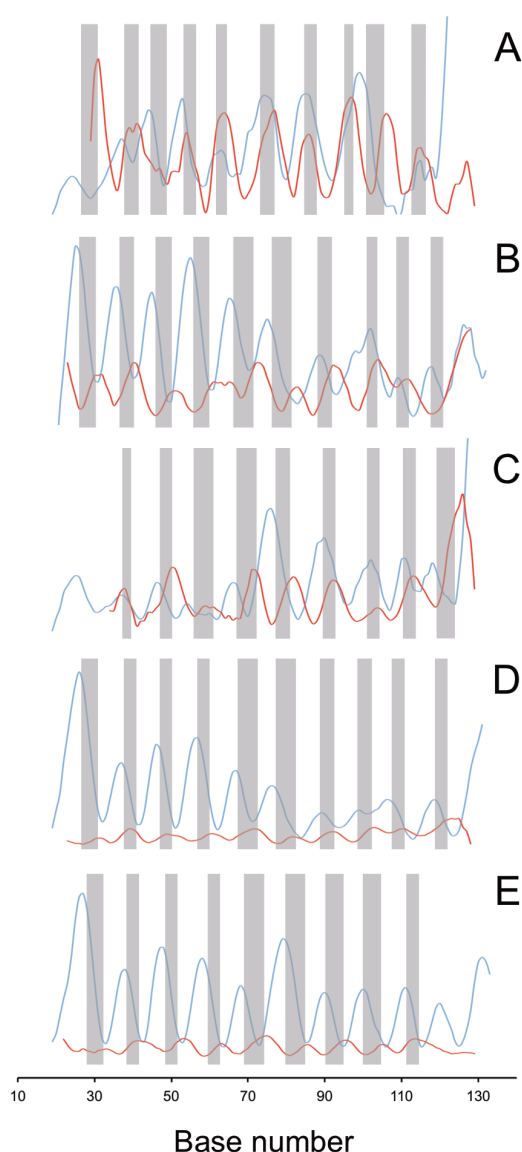


Figure 2.18: Hydroxyl radical cleavage curves reveal the rotational setting of the CGG/CCG repeat-containing DNA around the histone core. The gray bars indicate areas where the minor groove is solvent exposed. Only the S1 series of substrates, S1-CGG19 (A), S1-1AGGa (B), S1-1AGGb (C), S1-2AGG (D), and the S1 control (E) are displayed, as the FMR1 substrates occupy many rotational positions. The S1a or CGG-containing strand is shown in blue and the S1b or CCG-containing strand is shown in red.

the S1-1AGGa, S1-1AGGb, and S1-2AGG DNA all adopt the same rotational position, different from both the S1 control and the S1-CGG19 substrates. Thus, not only does the presence of the CGG/CCG repeat alter the rotational positioning of DNA in the nucleosome, but the addition of an AGG/CCT interruption, a change of 1 or 2 bps out of 146, further alters the rotational positioning.

Next, the translational positioning (Figure 2.19) and the nucleotide positioned at the dyad axis were determined for each S1 substrate (Table 2.3).²⁰⁹ When the CGG/CCG

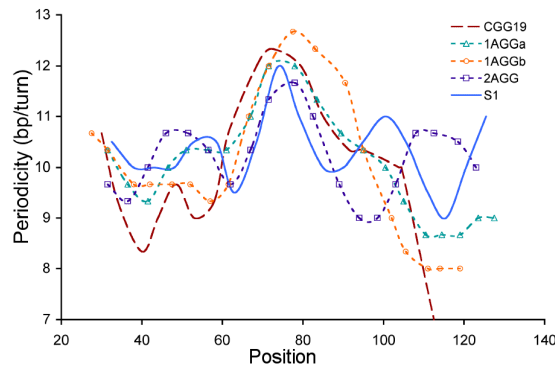


Figure 2.19: Translational positioning curves reveal a change in the position of the dyad axis for the CGG/CCG repeat-containing S1 series. The locations of the maxima and minima are then graphed with respect to their periodicity. The curve reveals the variation in periodicity around the histone core, and the highest peak indicates the location of the dyad axis.

repeat tract replaces the center base pairs of the S1 sequence, there is a 2 bp shift in the position of the dyad axis. Upon examination of the translational positioning curves, it is apparent that each of the interrupted substrates, S1-1AGGa, S1-1AGGb, and S1-2AGG, have a different nucleotide positioned at the dyad axis than S1-CGG19. Thus, as we observed with the rotational positioning, the CGG/CCG repeat alters the translational positioning relative to the S1 control, and the addition of an interruption further alters the translational positioning.

Table 2.3: Position of the dyad and translational shifts for CGG-containing strands of the S1 substrates

Substrate	Dyad position	Translational shift (base pairs) ^a
S1	74	-
S1-CGG19	72	-2
S1-1AGGa	74.75	+0.75
S1-1AGGb	77.5	+3.5
S1-2AGG	78	+4

^a With respect to the S1 control.

The ability of AGG/CCT interruptions to alter the positioning of the CGG/CCG repeat may contribute to the “gray zone” between healthy and pre-mutation lengths, where uninterrupted repeats classified as healthy-length can expand, but longer, interrupted repeats do not.^{43,45,47,49} Perhaps by modulating either the ability of the repeat tract to form a nucleosome or by altering its rotational and translational position around the histone core, the interruptions offer some protection from expansion. Furthermore, as a single interruption has an innate ability to reposition the repeat tract in a nucleosome, and since the loss of a single interruption has been linked with expansion, it is possible that the loss of the AGG/CCT interruption and its associated positioning effect benefits the expansion process.

In contrast to the hydroxyl radical footprint observed for the S1 substrates, the FMR1 substrates produce rather ambiguous results (Figure 2.20 and Figure 2.21). While reaction of the free duplex DNA with hydroxyl radicals produces even reactivity along the DNA substrate, just as observed in the S1 substrates, the oscillation between high and low reactivity when the FMR1 DNA is incorporated into a nucleosome is quite different from that observed for the S1 substrates (compare Figure 2.16B with Figure 2.20B). This cleavage pattern indicates that the FMR1 substrates form heterogeneous populations of

nucleosomes where the DNA adopts several slightly different positions around the histone core, leading to overlapping, oscillating patterns. The formation of a heterogeneous population by the FMR1 substrates was unexpected, considering we designed our DNA substrates with only 146 bps to minimize the possibility of just such an occurrence. However, CGG/CCG repeats in a larger fragment of the human *FMR1* gene have previously been shown to form multiple translational positions.¹⁹²

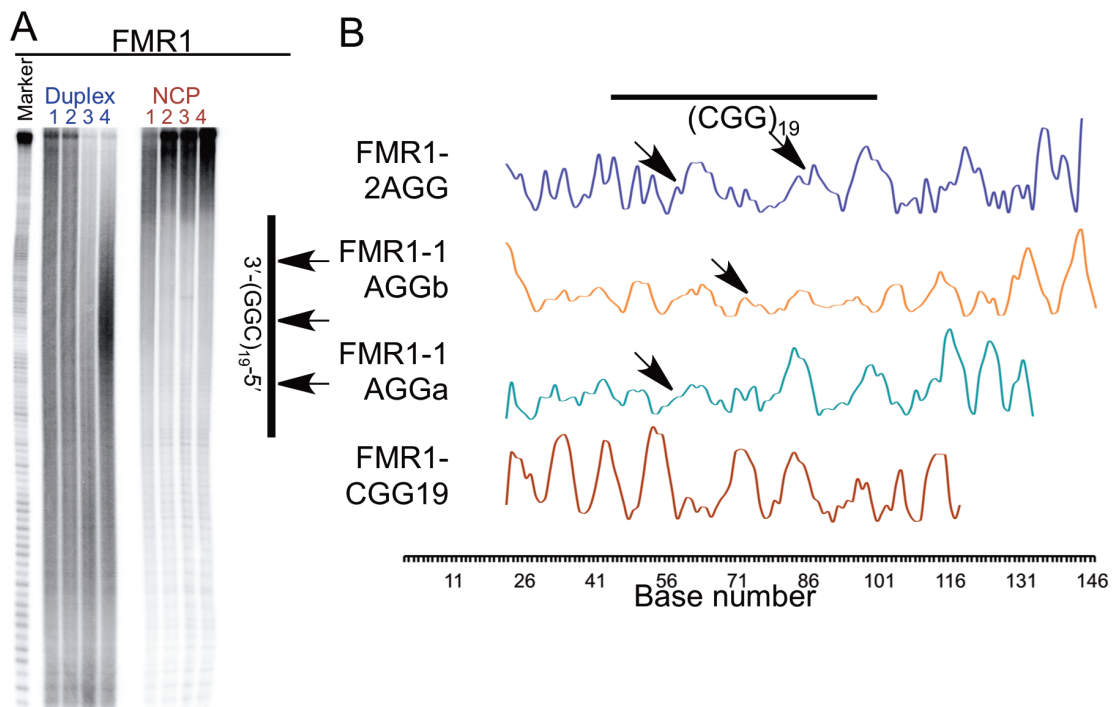


Figure 2.20: Hydroxyl radical footprinting reveals the periodicity of the CGG/CCG repeat-containing FMR1 DNA around the histone core. Radiolabeled samples, both free duplex (labeled in blue) and nucleosomes (NCP, labeled in red) substrates, were reacted with hydroxyl radicals. Reactivity toward hydroxyl radical for the CGG-containing strands of the FMR1 substrates are shown in panel A. Reactions were performed on both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. In panel A, lanes 1 contain FMR1-CGG19, lanes 2 contain FMR1-1AGGa, lanes 3 contain FMR1-1AGGb, and lanes 4 contain FMR1-2AGG. The marker lane consists of the Maxam-Gilbert A/G sequencing reaction performed on S1a. The location of the repeat tract and interruptions are indicated on the right-hand side of the gel. Panel B shows the reactivity at each nucleotide, generated from the gel presented in A. The arrows indicate the position of the A in the AGG interruptions.

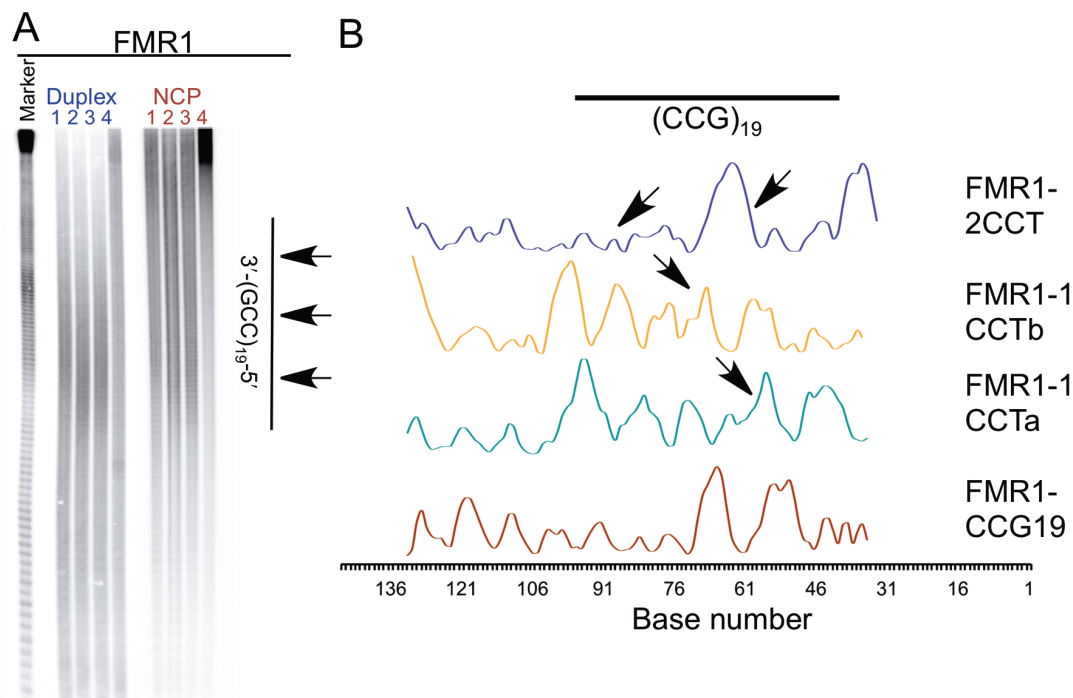


Figure 2.21: Hydroxyl radical footprinting reveals the periodicity of the CGG/CCG repeat-containing FMR1 DNA around the histone core. Radiolabeled samples, both free duplex (labeled in blue) and nucleosomes (NCP, labeled in red) substrates, were reacted with hydroxyl radicals. Reactivity toward hydroxyl radical for the CCG-containing strands of the FMR1 substrates are shown in panel A. Reactions were performed on both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. In panel A, lanes 1 contain FMR1-CCG19, lanes 2 contain FMR1-1CCTa, lanes 3 contain FMR1-1CCTb, and lanes 4 contain FMR1-2CCT. The marker lane consists of the Maxam-Gilbert A/G sequencing reaction performed on S1b. The location of the repeat tract and interruptions are indicated on the right-hand side of the gel. Panel B contains histograms generated from the gel presented in A. The arrows indicate the position of the CCT interruptions.

Unfortunately, analysis of the hydroxyl radical footprinting data requires a homogenous population, thus we cannot draw any specific conclusions about how the *FMR1* gene sequence might modulate the position of the interrupted and uninterrupted CGG/CCG repeat tracts, only that, unlike the CAG/CTG repeats, the CGG/CCG repeats have no inherent positioning ability.

It is important to consider that the AGG/CCT interruptions are likely altering the rotational and translational positions of the FMR1 substrates, just as they do in the S1 substrates, but because of the formation of multiple species of nucleosomes, those changes cannot be detected by this technique. However, the ability of the FMR1 substrates to adopt multiple positions relative to the histone core indicates that the repeat tract is less limited in its positioning options as compared to the S1 substrates. This may explain why the FMR1-2AGG substrate has the highest incorporation of the FMR1 series; because the FMR1 flanking sequence cannot dictate the position of the DNA as effectively as the S1 flanking sequence, the DNA is better able to balance the competing positioning forces of the repeat tract, the AGG/CCT interruptions, and the FMR1 flanking sequence, leading to higher levels of incorporation.

2.3.8 CAG/CTG repeats possess a periodicity of ~10.7 bps per turn, regardless of flanking sequence

Upon fitting a sine wave to the hydroxyl radical footprinting data generated for the S1 series of CAG/CTG repeat-containing substrates, we observe a general increase in the average periodicity for the entire DNA as the repeat tract expands. The S1 substrate has a periodicity of 10.35 ± 0.02 bps/turn (Table 2.4), consistent with previously reported values for this substrate, as well as similar DNA incorporated into nucleosomes.^{170,203,209} For S1-CAG10 the periodicity increases to 10.44 ± 0.04 bps/turn, a change of ~0.1 bps/turn, which represents a global underwinding of the duplex by ~1 bp. The periodicity of the S1-CAG20 DNA further increases to 10.61 ± 0.02 bps/turn, or a global underwinding of ~2 bps relative to the S1 substrate. However, this appears to be the maximum change in periodicity associated with the addition of CAG/CTG repeats since

S1-CAG30 has a periodicity comparable to S1-CAG20, consistent with our proposal that the CAG/CTG repeats have fully dominated the influence of the S1 flanking sequence in the S1-CAG20 and S1-CAG30 substrates.

Table 2.4: Base pairs per helical turn determined using the CAG and CTG-containing strands of the S1 and HTT substrates

Substrate	Total	Repeat	Flanking (5' end)	Flanking (3' end)
S1a	10.35 ± 0.02	N/A	N/A	N/A
S1b	10.34 ± 0.02	N/A	N/A	N/A
S1-CAG10	10.44 ± 0.04	10.71 ± 0.02	10.34 ± 0.04	10.30 ± 0.04
S1-CTG10	10.44 ± 0.02	10.70 ± 0.03	10.30 ± 0.02	10.32 ± 0.09
S1-CAG20	10.61 ± 0.02	10.70 ± 0.02	10.34 ± 0.04	10.30 ± 0.07
S1-CTG20	10.58 ± 0.05	10.69 ± 0.02	10.30 ± 0.02	10.30 ± 0.08
S1-CAG30	10.62 ± 0.03	10.66 ± 0.03	10.40 ± 0.1	10.30 ± 0.06
S1-CTG30	10.64 ± 0.02	10.71 ± 0.03	10.30 ± 0.02	10.23 ± 0.03
HTT-CAG10	10.55 ± 0.04	10.69 ± 0.03	10.54 ± 0.08	10.01 ± 0.01
HTT-CTG10	10.51 ± 0.03	10.69 ± 0.02	9.63 ± 0.04	10.54 ± 0.02
HTT-CAG20	10.52 ± 0.02	10.67 ± 0.03	10.52 ± 0.09	10.51 ± 0.03
HTT-CTG20	10.51 ± 0.02	10.68 ± 0.03	10.50 ± 0.03	10.51 ± 0.09
HTT-CAG30	10.71 ± 0.05	10.70 ± 0.04	10.55 ± 0.3	10.50 ± 0.03
HTT-CTG30	10.70 ± 0.03	10.68 ± 0.03	10.48 ± 0.1	10.51 ± 0.03

The periodicity of the HTT substrates is different from the S1 series. HTT-CAG10 and HTT-CAG20 have a periodicity of ~ 10.55 bps/turn, but the periodicity of the HTT-CAG30 substrate is larger at 10.71 ± 0.05 bps/turn. This translates to a global underwinding of ~ 2 bps for HTT-CAG10 and HTT-CAG20, and ~ 3 bps for HTT-CAG30, compared to S1. The relative periodicities correlate with the trends observed for the rotational position of each substrate; the substrates where the CAG/CTG repeat tract is the determinant of rotational position (S1-CAG20, S1-CAG30, HTT-CAG30) all have a similar periodicity.

When a sine wave is fit to the hydroxyl radical footprinting data corresponding to specific regions of the DNA substrates, an interesting trend is observed. The periodicity of all the CAG/CTG repeat regions is ~ 10.7 bps/turn (Table 2.4). Based on both the chemical and enzymatic probing, each CAG/CTG repeat tract resides at the ~ 30 bp

region spanning the dyad axis, which is known to have a naturally larger periodicity than other regions of DNA in the nucleosome.^{170,203,209} But as the repeat tract continues to expand off this region, the periodicity associated with the repeat tract remains constant at ~10.7. Taken together, the shared periodicity of the repeat regions from both the S1 and HTT substrates indicate that the preferred periodicity of the CAG/CTG repeat tract in a nucleosome, under these conditions, is ~10.7 bps/turn. Furthermore, the increased periodicity of the CAG/CTG repeat tracts relative to the flanking sequence provides additional insight as to why CAG/CTG repeats are such good incorporators and prefer to associate with the region spanning the dyad axis; a preferred periodicity congruent with that of the dyad axis allows them to facilitate nucleosome formation at this region. It is likely that the ability to adopt this specific periodicity is one of the factors not accounted for in the ΔG of bending that makes CAG/CTG repeats such robust incorporators.

In the genome, CAG/CTG repeats have been found in heterochromatin and euchromatin.^{52,56,164,234-236} Both healthy and disease-length repeats associated with HD are found in the more open, transcriptionally active euchromatin. In contrast, while shorter CAG/CTG tracts associated with another neurodegenerative disorder, myotonic dystrophy (DM1), are present in regions of euchromatin, the expanded, disease-length repeats are associated with several heterochromatic markers. Notably, disease-length DM1 repeat tracts can contain up to 1800 CAG/CTG repeats, which would span ~27 nucleosomes, compared to the single nucleosome that healthy-length repeats would occupy. The innate incorporation ability of CAG/CTG repeats, along with a defined periodicity compatible with the underwinding needed for interaction with the region around the dyad axis, is consistent with the shift from euchromatin to heterochromatin

associated with repeat expansion in DM1, though it is likely that other cellular factors play a role as well.⁵⁶ Conversely, the CAG/CTG repeats associated with HD expand over a much shorter range, with the longest expanded repeat tracts containing ~100 repeats, which would span less than 2 nucleosomes.²² Thus, while a nucleosome is more likely to position within the disease-length *HTT* repeat tract, the gene remains as euchromatin and transcription can proceed normally. However, the same gene containing a healthy-length repeat tract must balance the competing incorporation and positioning influences of the repeat tract and flanking sequence, likely leading to a difference in functional organization, and thus different levels of DNA accessibility, which could affect genomic expansion.

2.3.9 CGG/CCG repeats decrease the periodicity of DNA in the nucleosome, but an AGG/CCT interruption increases periodicity

To determine the periodicity of the CGG/CCG repeat-containing S1 substrates, a sine wave was fit to each curve presented in Figure 2.16B. The periodicities are given in Table 2.5. Consistent with the analysis described above, the periodicity of the S1 control is 10.37 ± 0.01 bp/turn. Upon inclusion of the 19 CGG/CCG repeats, the periodicity decreases to 10.14 ± 0.05 bp/turn, representing a global duplex overwinding of ~2 bp relative to the S1 substrate. However, when the AGG/CCT interruptions are present, the periodicity returns to ~10.4 bp/turn. The same pattern is observed when the periodicity is determined specifically for the repeat region; the repeat region of the S1-CGG19 substrate has a lower periodicity than the repeat region of any of the substrates containing interruptions.

Table 2.5: Base pairs per helical turn determined using the CGG- and CCG-containing strands of the S1 substrates

Substrate	Total	Repeat Region
S1a	10.37 ± 0.01	N/A
S1b	10.37 ± 0.02	N/A
S1-CGG19	10.14 ± 0.05	10.38 ± 0.03
S1-CCG19	10.23 ± 0.03	10.42 ± 0.02
S1-1AGGa	10.36 ± 0.02	10.56 ± 0.03
S1-1CCTa	10.37 ± 0.02	10.52 ± 0.02
S1-1AGGb	10.40 ± 0.04	10.55 ± 0.04
S1-1CCTb	10.41 ± 0.03	10.52 ± 0.02
S1-2AGG	10.40 ± 0.02	10.49 ± 0.03
S1-2CCT	10.41 ± 0.02	10.52 ± 0.02

The periodicity of both the uninterrupted and interrupted healthy-length CGG/CCG repeat tracts is less than that of healthy-length CAG/CTG repeats, which is consistently ~ 10.7 bp/turn. However, both repeats are located in the same area of the nucleosome: the region surrounding the dyad axis, an area that generally adopts a periodicity of ~ 10.7 bp/turn, best for bending around that region of the histone core. The difference in periodicity between the CGG/CCG and CAG/CTG repeats provides an additional rationale for the differences observed in DNA incorporation for these two TNR sequences. While the CAG/CTG repeats adopt a periodicity well suited for bending around the dyad axis, uninterrupted CGG/CCG repeat tracts cannot adopt such a relatively large periodicity. However, once an interruption is present, the repeat tract can now adopt a periodicity closer to the 10.7 bp/turn preferred for bending around the dyad axis, leading to higher levels of incorporation.

2.3.10 The CCG/CGG repeat in the HTT flanking sequence “kinks” when incorporated into the nucleosome, but CGG/CCG repeats at the dyad axis do not

S1 nuclease, which recognizes single-stranded regions of DNA or RNA has been used previously to characterize the non-canonical secondary structures formed by CAG or CTG repeat oligonucleotides.^{63,237} To assess the ability of CAG/CTG repeats to form

non-canonical structures in the nucleosome, we incubated both free duplex and nucleosomal substrates from the HTT series with S1 nuclease (Figure 2.22A,B). We found no evidence of non-canonical structure formation in any of the CAG/CTG repeat regions, as indicated by a lack of strand cleavage. However, there is reactivity in the secondary CCG repeat tract of the incorporated HTT-CAG10 substrate (Figure 2.22A, boxed region), indicating that there is a region of unpaired DNA in the nucleosome. We refer to this region as “kinked”. While we expect that the complementary region in the

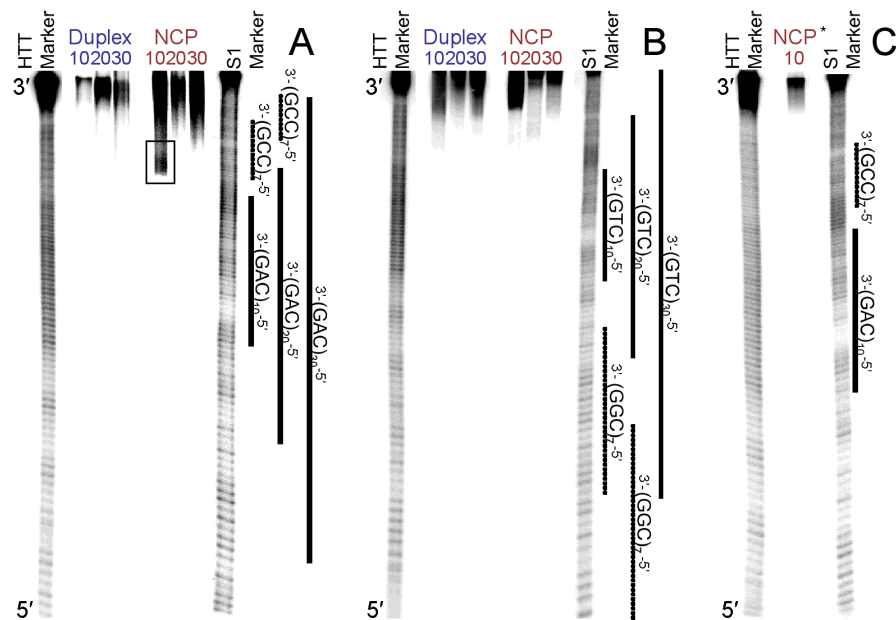


Figure 2.22: S1 nuclease probing reveals that kinking occurs in the secondary CCG/CGG repeat of the HTT-CAG10 DNA when incorporated into a nucleosome. The HTT series of substrates, with either the CAG-containing strand radiolabeled (A) or the CTG-containing strand radiolabeled (B) was incubated with S1 nuclease, a probe for single-stranded DNA. Reactions were performed with both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. A region of reactivity (panel A, boxed region) is observed for HTT-CAG10 incorporated into a nucleosome (A), but once that DNA is removed from a nucleosomal context, the reactivity is no longer observed (C, NCP* indicates that the DNA was incorporated but was removed from the NCP before probing). The S1 nuclease reactions were electrophoresed along with an S1 marker and an HTT marker, consisting of the Maxam-Gilbert A/G sequencing reaction performed on S1a or the CAG-containing strand of HTT-CAG10.

CGG repeat is also kinked, no cleavage by S1 nuclease was observed (Figure 2.22B). However, based on the hydroxyl radical footprinting, the CGG repeats complementary to the kinked region are facing in towards the histone core, making them inaccessible to the nuclease. Thus, the complementary strand may be kinked as well, but is undetectable by S1 nuclease.

As a control, we incorporated the HTT-CAG10 substrate into a nucleosome, removed the proteins by performing a phenol:chloroform:isoamyl alcohol extraction and treated the freed duplex DNA with S1 nuclease. Upon removal of the DNA from the constraints of the histone core, no cleavage of DNA by the nuclease was observed, indicating that kinking is a reversible process (Figure 2.22C).

The region in which this kink occurs is distinguished by its overwinding. The CCG/CGG repeat in the HTT-CAG10 duplex is located in an area of the nucleosome that is locally overwound; the center of the (CCG)₇ repeat falls in a trough on the translational positioning curve for the HTT-CAG10 nucleosome (Figure 2.15B). Furthermore, the entire flanking region possesses a global overwinding; the periodicity of the CCG repeat containing flanking sequence is much lower than the periodicities determined for any of the other flanking sequences (Table 2.4). It has previously been shown that super-helical location (SHL) 1.5 and 4.5 have a distinct bending of the DNA, but the kink does not form at these locations.¹⁷⁰ The formation of the kink is likely due in combination to the high ΔG of bending for the CCG/CGG repeats and the overwinding necessary for bending around the histone core.^{190,211-213} Kinks have been observed in previous experiments with superhelically-stressed DNA and is considered a way of relieving some of this stress.²³⁸⁻²⁴⁰ We observe that once the CCG/CGG repeat is moved to a different

location in the nucleosome, kinks no longer form. In HTT-CAG20 the CCG/CGG repeats are shifted close to the end of the duplex such that it is likely more favorable to simply disassociate from the histone core than to kink the DNA. This is reflected in the Exo III footprint of HTT-CAG20 nucleosomes; Exo III shows strong digestion through the CCG/CGG repeat tract (Figure 2.9B). In HTT-CAG30 part of the CCG/CGG repeat is removed and no cleavage by S1 nuclease is observed. It is unknown whether cellular conditions will allow for kinking, yet this process of relieving the stress of bending around the histone core could have interesting consequences within the cell.

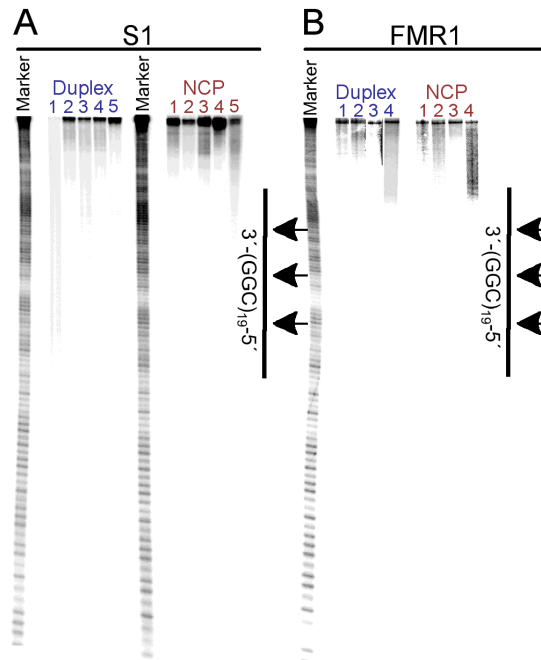


Figure 2.23: S1 nuclease digestion reveals no single stranded regions in the S1 or FMR1 DNA substrates. Radiolabeled DNA substrates were incubated with S1 nuclease, a probe for single-stranded DNA. Reactions were performed with both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. Reaction with the S1 substrates is shown in panel A, while reaction with the FMR1 substrates are shown in panel B. In panel A, lanes 1 contain the S1 control, lanes 2 contain S1-CGG19, lanes 3 contain S1-1AGGa, lanes 4 contain S1-1AGGb, and lanes 5 contain S1-2AGG. In panel B, lanes 1 contain FMR1-CGG19, lanes 2 contain FMR1-1AGGa, lanes 3 contain FMR1-1AGGb, and lanes 4 contain FMR1-2AGG. In both panels, the marker lane consists of the Maxam-Gilbert A/G sequencing reaction performed on S1a. The location of the repeat tract and interruptions (arrows) are indicated on the right-hand side of the gel.

We performed the same experiments using the S1- and FMR1-CGG DNA to investigate whether CGG/CCG repeats also form kinks in this context (Figure 2.23). We find no areas of increased reactivity to S1 nuclease in nucleosome samples containing these CGG/CCG repeats compared to the free duplex samples, indicating that they behave as canonical duplex when incorporated into the nucleosome. Notably, the CCG/CGG repeats in the HTT-CAG10 substrate were located towards the 3' end of the DNA, which placed the CCG/CGG repeats in a region of the nucleosome where the DNA was locally overwound. However, the CGG/CCG repeats in these substrates are located at the center of the DNA, which positions the CGG/CCG repeats in the region spanning the dyad axis where they are locally underwound. The difference in position within the DNA substrate places the repeats in different environments within the nucleosome, providing a probable explanation for the difference in behavior.

2.4 CONCLUSIONS

While CAG/CTG repeats are naturally strong nucleosome incorporation and positioning sequences, the effect of healthy-length repeats is moderated by the *HTT* flanking sequence. When placed within the S1 sequence, the CAG/CTG repeats are able to significantly increase incorporation with each addition of 10 repeats. However, when the repeats are flanked by the *HTT* gene sequence, each substrate showed a significant decrease in nucleosome incorporation compared to the corresponding S1 substrate. Twenty CAG/CTG repeats are sufficient to override the preferred positioning of the S1 flanking sequence and adopt the orientation relative to the histone core most favorable for the CAG/CTG repeat tract. However, in the *HTT* gene sequence, 30 CAG/CTG repeats

are required to override the positioning effect of the *HTT* flanking sequence and adopt the most favorable position for the repeat tract. Interestingly, the only feature of the CAG/CTG repeats that is not influenced by the flanking sequence is their periodicity in the nucleosome; the periodicity is ~10.7 bps/turn in each substrate. Thus, while disease-length CAG/CTG repeats will likely interact with the histone core in a manner dictated solely by the CAG/CTG repeat, CAG/CTG repeat tracts found in healthy individuals must balance the incorporation and positioning preferences of both the repeat and the flanking sequence, leading to different levels of nucleosome occupancy and forms of positioning between healthy- and disease-length repeat tracts, the consequences of which could alter the propensity for genomic expansion and manifestation of the disease state within the packaged genome.

Furthermore, we find that while 19 CGG/CCG repeats themselves have no impact on incorporation of the 146 bp S1 duplex, a single AGG/CCT interruption significantly increases incorporation into the nucleosome. When the S1 flanking sequence is replaced by the *FMRI* gene sequence, the changes induced by the addition of AGG/CCT interruptions are much smaller, primarily due to the overall barrier to incorporation presented by the *FMRI* flanking sequence. Not only does the presence of interruptions alter the ability of the repeat tract to form a nucleosome, but once incorporated, also alters the rotational and translational positioning of the repeat tract around the histone core.

It has been proposed that AGG/CCT interruptions play a protective role within the CGG/CCG repeat tract by destabilizing non-canonical secondary structures.^{60,241} The work presented here provides insight into an additional protective function; the presence

of AGG/CCT interruptions may alter the behavior of the repeat tract in the packaged genome. The ability to alter the behavior of the repeat tract in the nucleosome may partially explain the difference in expansion ability between an interrupted and uninterrupted repeat tract whose length falls within the so called “gray zone” of healthy-length repeats, and why the loss of a single interruption can have such dire consequences. Combining their ability to destabilize non-canonical secondary structures and their ability to alter the behavior of the repeat tract within the packaged genome, the AGG/CCT interruptions clearly play an important role in maintenance of the CGG/CCG repeat tract in the *FMRI* gene.

Taken together, these results highlight the importance of understanding how healthy-length repeats behave in a nucleosome. In both HD and FXS, the healthy-length repeats behave differently than their disease-length counterparts when interacting with the histone core. Furthermore, small changes in the primary sequence of the repeat tract can lead to large consequences for both nucleosome incorporation and DNA positioning. Without understanding the changes that take place within the packaged repeat, we cannot fully understand any changes in the functional organization of these genes or the mechanisms of expansion within the packaged genome.

Significance: these results demonstrate the differences between healthy and disease length repeats in the packaged genome. By understanding these differences it might be possible to design drugs that could either target the appropriate trans-acting factors to reverse any changes in functional organization associated with the disease state or directly target the expanded repeats with the intended purpose of modify the chromatin organization, allowing proper progression of the various polymerases.

Chapter 3

Trinucleotide Repeat DNA Alters Structure to Minimize the Thermodynamic

Impact of 8-Oxo-7,8-dihydroguanine²

3.1 INTRODUCTION

Cellular DNA is constantly bombarded with reactive oxygen species (ROS) from both intra- and extracellular sources.¹⁰² Of the four DNA nucleobases, guanine (G) has the lowest redox potential¹³³ and is most readily oxidized by ROS. A commonly observed form of oxidatively damaged DNA is 8-oxo-7,8-dihydroguanine (8-oxoG) (Figure 1.4).¹⁴⁷ Interestingly, 8-oxoG is able to rotate about the glycosidic bond to convert from the canonical *anti* conformation to the *syn* conformation. When in the *syn* conformation 8-oxoG can base pair with A via Hoogsteen bonding. Because of its ability to pair with A, 8-oxoG is mutagenic and can cause G to T transversion mutations if this mispair is not repaired prior to replication. In order to maintain genetic integrity, 8-oxoG is removed from DNA by the repair protein 8-oxo-7,8-dihydroguanine glycosylase (OGG1), which recognizes the modified G and catalyzes cleavage of the *N*-glycosidic linkage to the sugar-phosphate backbone (Figure 1.5).¹⁵⁶

Surprisingly, while repair of DNA damage is usually beneficial, removal of 8-oxoG by OGG1 has been implicated in the expansion of CAG/CTG trinucleotide repeat (TNR) sequences linked to Huntington's Disease (HD).^{159,162,164} CAG/CTG repeats in exon 1 of the huntingtin gene are genetically unstable and their expansion provides the molecular basis for HD.²² Using a mouse model of HD, expansion at CAG/CTG repeat tracts was abrogated in mice lacking OGG1.¹⁵⁹ Subsequent work *in vitro* lead to the

² Adapted from Volle, C. B. and Delaney, S. (2012) *Biochemistry* **51** 52-62

proposal of a “toxic oxidation cycle” fueling the expansion of CAG/CTG repeats (Figure 1.7).^{159,162} As part of this cycle, OGG1 removes 8-oxoG from duplex DNA and polymerase β performs strand-displacement synthesis, incorporating multiple nucleotides as part of a long-patch base excision repair (LP-BER) event. The formation of a non-canonical structure by the displaced flap, a CAG hairpin, renders the DNA refractory to cleavage by flap endonuclease 1. Ligase incorporates the hairpin into the DNA, forming a three-way junction containing the CAG hairpin flanked on either side by duplex. Additionally, we have shown that the hairpin formed by the displaced flap is hyper-susceptible to oxidation, and that OGG1 does not efficiently remove 8-oxoG from the hairpin loop.^{59,242} We proposed that this combination of susceptibility to oxidation and inefficient repair would exacerbate TNR expansion during LP-BER.²⁴²

While we have previously shown that a TNR hairpin is more susceptible than duplex to oxidative damage, the effect of this damage on the structure and stability of the hairpin DNA has not been defined, either as an oligonucleotide or as a three-way junction. Indeed, no direct structural characterization of three-way junctions containing a CAG hairpin has taken place. Before we can investigate the effect of oxidative damage on hairpin formation or the effect of hairpin formation on other cellular processes, such as nucleosome incorporation and positioning, we must first understand the behavior of the hairpin itself within its biologically relevant DNA context; the three-way junction.

Given the known propensity for damage to accumulate in the non-canonical structure formed by TNR sequences, it is necessary to define the structural and thermodynamic impact of this damage. Furthermore, in these experiments we provide a structural analysis of TNR DNA in a three-way junction, with and without the 8-oxoG

lesion. We find that TNR DNA has the ability to modulate its structure to accommodate 8-oxoG, both in the context of a three-way junction and in single-stranded oligonucleotides. Furthermore, we show that this structural change minimizes the thermodynamic impact of the DNA lesion.

3.2 EXPERIMENTAL PROCEDURES

3.2.1 Oligonucleotide synthesis and purification

Oligonucleotides were synthesized using standard phosphoramidite chemistry on a BioAutomation DNA/RNA synthesizer. The oligonucleotides containing 8-oxoG were synthesized trityl-off (on instrument removal of the 5'-dimethoxytrityl group) and the oligonucleotides were deprotected in 1M β -mercaptoethanol in NH_4OH at 55 °C for 24 h. Two rounds of HPLC purification were then carried out using a Dionex DNAPac PA100 anion-exchange column (4 × 250 mm) using 10% acetonitrile (solvent A) and 0.8 M ammonium chloride in 10% acetonitrile (solvent B) as the mobile phases (gradient: solvent B was increased from 55% to 75% over 5 min, then 75% to 90% over 15 min, and 90% to 100% over 5 min; 1 mL/min).

For the synthesis of oligonucleotides lacking 8-oxoG, the 5'-trityl group was retained to facilitate purification by HPLC. Purification of these oligonucleotides was performed on a Dynamax Microsorb C18 column (10 × 250 mm) using acetonitrile (solvent A) and 30 mM ammonium acetate (solvent B) as the mobile phases (gradient: solvent A was increased from 5% to 25% over 25 min; 3.5 mL/min). Following removal of the 5'-trityl group by incubation in 80% glacial acetic acid for 12 min at room

temperature, the oligonucleotides were subjected to a second round of HPLC purification (gradient: solvent A was increased from 0% to 15% over 35 min; 3.5 mL/min). The longer sequences used to construct the three-way junctions were purified as described above, but an additional gel purification step was performed after HPLC to ensure removal of longer failure sequences. Samples were suspended in denaturing loading buffer (80% formamide, 10 mM EDTA, 1% xylene cyanol), applied to a 14% denaturing polyacrylamide gel, and electrophoresed at 80 W until the dye had migrated 8 in. The DNA was then excised and eluted from the gel by crushing and soaking. The supernatant, containing the purified DNA, was collected and desalted using Sephadex G50 fine resin.

2.2.2 Chemical probe analysis

Oligonucleotides were 5'-³²P end-labeled using T4 polynucleotide kinase (New England Biolabs) following the manufacturer's protocol. 10 pmol of labeled DNA were suspended in 100 µL of Tris buffer (20 mM Tris-HCl, 70 mM NaCl, 10 mM EDTA, pH 7.5), heated to 90 °C for 5 min and slowly cooled to room temperature. Samples were then divided to 2 pmol per aliquot, mixed with 5 µg calf thymus DNA and brought to 100 µL total volume before being incubated with either 1% DEPC (v/v) or 0.05% DMS in ethanol (v/v) for 5-30 min at 37 °C. Reactions were quenched by addition of 20 µL stop solution (1.5 M sodium acetate, 1 M β-mercaptoethanol, 250 µg/mL tRNA, pH 7.0), ethanol precipitated, and the pellet dried *in vacuo*. Samples were then treated with 10% piperidine (v/v) for 30 min at 90 °C before drying *in vacuo*. Samples were resuspended in denaturing loading buffer (80% formamide, 10 mM EDTA, 1% xylene cyanol, 1% bromophenol blue), incubated at 90 °C for 3 min before loading onto a 16% denaturing polyacrylamide gel (0.4 mm thick, 45 mm long). Samples were electrophoresed at 80 W

until 4 in separated the xylene cyanol and bromophenol blue dyes. The products were visualized by phosphorimagery, using a BioRad FX scanner and the accompanying Quantity One software. Histograms were generated by drawing a line through the lane corresponding to the 15 min time point and were background subtracted using the “rolling ball” technique with the radius set to 25.

Three-way junction constructs were formed by annealing 10 pmol of radiolabeled CAG16-TWJ or CAG16-8-TWJ with a 1.5-fold excess of the complementary sequence CTG6-TWJ. DEPC reactions on the three-way junction constructs were performed using 2.5% DEPC (v/v) and incubated for 5-30 min at 37 °C. Sample work-up was the same as above except that samples were run on a 14% gel.

3.2.3 *Differential scanning calorimetry*

Calorimetry was performed using a TA Instruments NanoDSC III. Oligonucleotides were suspended in 700 μ L of 20 mM sodium phosphate, 70 mM NaCl, pH 7.5. All samples were de-gassed *in vacuo* for 12 min at 25 °C before analysis. Data were obtained by monitoring the excess power required to maintain the same temperature in both the sample cell and a reference cell filled with de-gassed buffer. Each cycle of the experiment consists of a forward scan, in which the temperature is raised from 10 °C to 95 °C at a rate of 1 °C/min, a reverse scan, in which the temperature is decreased from 95 °C to 10 °C at 1 °C/min, and an equilibration period of 10 min between each forward and reverse scan. The resulting thermograms display the excess heat capacity (C_p) as a function of temperature. Ten thermograms, consisting of both forward and reverse scans, were obtained for each DNA substrate. The thermograms were normalized for concentration, background corrected using a buffer-buffer scan, and a polynomial

baseline correction was performed using the NanoAnalyze software after fitting the linear portions of the pre- and post-transitional baseline. ΔH was determined by integrating the area under the excess heat capacity curve. ΔS was determined by plotting C_p/T against the temperature in K and integrating under the resulting curve. ΔG was calculated at 37 °C using the values obtained for ΔH and ΔS and the equation $\Delta G = \Delta H - T\Delta S$ where T is in K.

3.3 RESULTS AND DISCUSSION

3.3.1 Structural characterization of a TNR three-way junction

A DNA three-way junction mimics a genomic repeat tract after an expansion has occurred, where there are more CAG repeats present than complementary CTG repeats. To determine the influence of 8-oxoG on these post-expansion TNR sequences, we used a chemical probe of nucleobase solvent accessibility to investigate the structure of three-way junctions by revealing the presence of loops, bulges, or overhangs. These three-way junctions contain 16 CAG repeats, six of which base pair with a CTG repeat in the complementary strand, and the remaining ten are predicted to form a stem-loop hairpin. We have shown previously that a single strand containing 10 CAG repeats adopts a stem-loop hairpin in which the loop contains four nucleobases and G/C base pairs and A•A mismatches comprise the stem (Figure 1.1).^{59,67,68} Here, the hairpin is flanked by duplex in order to form a three-way junction. Furthermore, the CAG/CTG repeats are bordered by a well-matched duplex comprised of a mixed sequence.

The first three-way junction lacks 8-oxoG and was assembled using the sequences CAG16-TWJ and CTG6-TWJ (Table 3.1). In the second three-way junction, G8 in

CAG16-TWJ is site-specifically replaced by 8-oxoG, where G8 refers to the guanine in the eighth CAG repeat when counting from the 5'-end of the TNR sequence (CAG16-TWJ-8OX). Due to the repetitive nature of the sequence in these three-way junctions, it is expected that the CAG stem-loop can position itself in several ways relative to the 6 complementary CTG repeats.^{243,244} Indeed, such behavior was confirmed by structural analysis using the chemical probe diethyl pyrocarbonate (DEPC), a small molecule that reveals the solvent accessibility of purines, with preferential modification of adenines.²⁴⁵ Upon incubation with DEPC, the CAG16-TWJ/CTG6-TWJ substrate shows hyper-reactivity at A6 through A12 (Figure 3.1A, Figure 3.2A), though interestingly, we do not observe equal reactivity at these adenines, indicating that the CAG hairpin forms at six of the seven possible positions along the complement, but that each of these structures does not form in equal amounts. Based on the level of reactivity at A11 and A12, which exhibit the highest degree of modification by DEPC, the top structure in Figure 3.1A is most likely predominant within the population. The composition of the flanking mixed-sequence duplex regions may skew the formation of hairpins; the two G/C base pairs flanking the first CAG repeat may limit the dynamics of the system and prevent the first CAG repeat from being part of the stem.

Table 3.1: DNA sequences used to investigate the behavior of 8-oxoG in CAG repeats

Name	Nucleotide Sequence
CAG16-TWJ	5'-(CCTTAGCTCC)(CAG) ₁₆ (TAAGGGGAAC)-3'
CAG16-8OX-TWJ	5'-CCTTAGCTCC)(CAG) ₇ (CAG ^{ox})(CAG) ₈ (TAAGGGGAAC)-3'
CTG6-TWJ	5'-(GTTCCCCTTA)(CTG) ₆ (GGAGCTAAGG)-3'
CAG7	5'-(CAG) ₇ -3'
CAG10	5'-(CAG) ₁₀ -3'
CAG11	5'-(CAG) ₁₁ -3'
CAG15	5'-(CAG) ₁₅ -3'
CAG7-3'Overhang	5'-(CAG) ₂ (CTTTTG)(CAG) ₃ -3'
CAG7-5'Overhang	5'-(CAG) ₃ (CTTTTG)(CAG) ₂ -3'
CAG7-Blunt Stem	5'-(CAG) ₂ (CTTTTTTG)(CAG) ₂ -3'
CAG10-5OX	5'-(CAG) ₄ (CAG ^{ox})(CAG) ₅ -3'
CAG10-6OX	5'-(CAG) ₅ (CAG ^{ox})(CAG) ₄ -3'
CAG11-5OX	5'-(CAG) ₄ (CAG ^{ox})(CAG) ₆ -3'
CAG11-6OX	5'-(CAG) ₅ (CAG ^{ox})(CAG) ₅ -3'

G^{ox} represents 8-oxo-7,8-dihydroguanine.

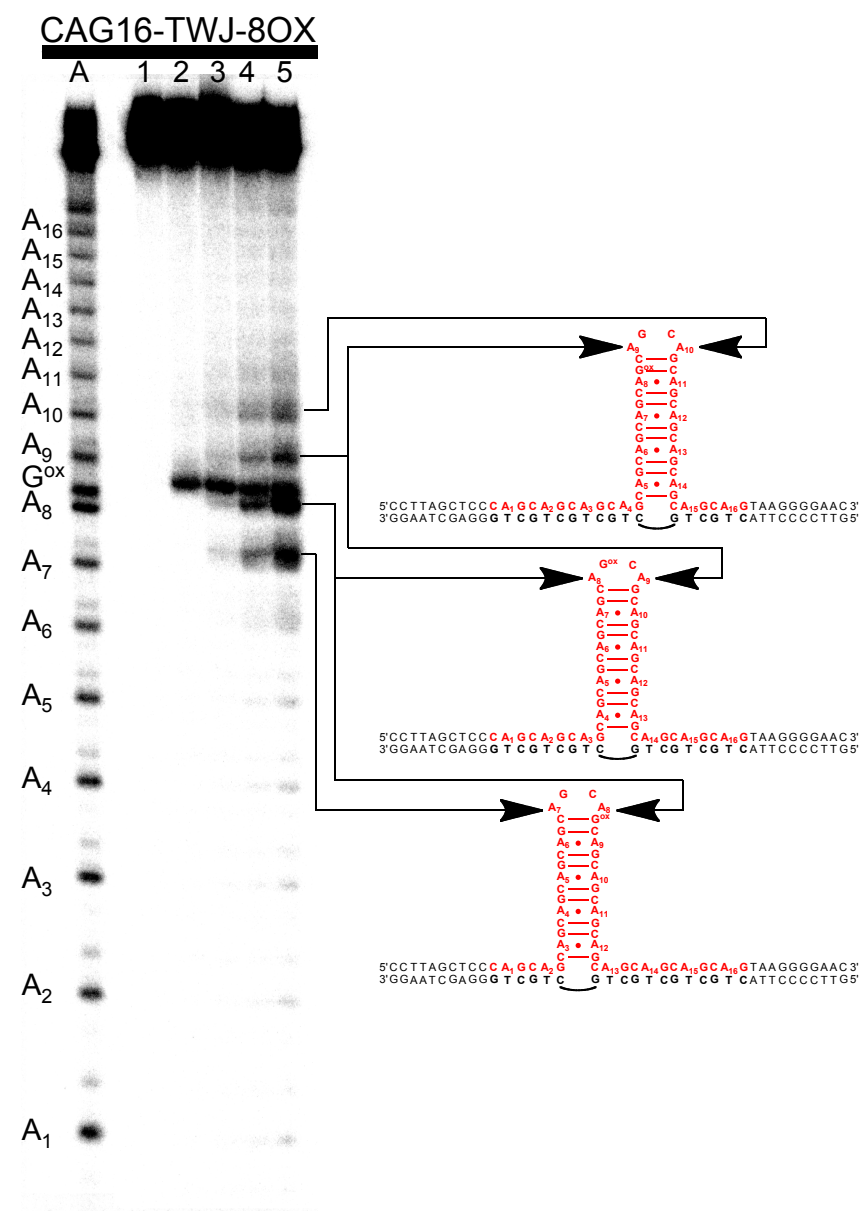


Figure 3.1 Autoradiogram of DEPC reactions with (A) CAG16-TWJ/CTG6-TWJ and (B) CAG16-TWJ-8OX/CTG6-TWJ. Lanes 1 are DNA alone, lanes 2 are piperidine-treated DNA, and lanes 3-5 are DNA treated with 2.5% (v/v) DEPC at 37 °C for 5, 15 or 30 min respectively, followed by treatment with piperidine. Lanes A are a marker created by treatment of the single-stranded CAG16-TWJ or CAG16-TWJ-8OX substrate with 1% DEPC at 37 °C for 15 min. The proposed structures derived from the reactivity of each substrate are shown to the right of the autoradiogram.

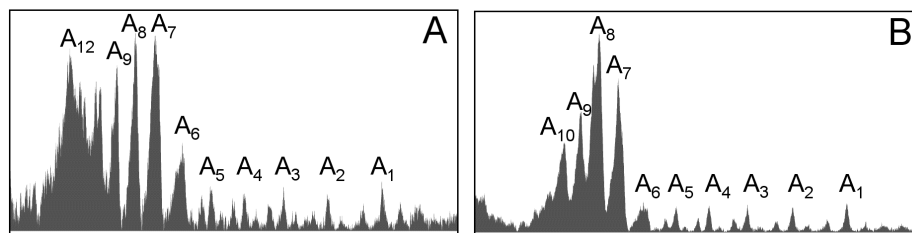


Figure 3.2 Histograms displaying reactivity of TWJ substrates to DEPC. Histograms for CAG16-TWJ (A) or CAG16-TWJ-8OX (B) are derived from lanes five of Figure 1, in which the substrate was incubated with 1% (v/v) DEPC at 37 °C for 30 m, followed by treatment with piperidine.

Notably, based on reactivity towards DEPC, we propose that the CAG repeats form a stem-loop hairpin as part of the three-way junction, and not an open bulge in which there is no intra-strand hydrogen bonding between the CAG repeats. In a bulge, every adenine in the CAG repeat region would be unpaired and solvent accessible and thus highly reactive toward DEPC. Furthermore, since a bulge would be able to shift its position relative to the CTG repeats in the complementary sequence, A2 through A16 would be equally accessible and modifiable by DEPC. Due to the lack of intense, ubiquitous DEPC reactivity at these adenines, we conclude that the CAG repeats in the three-way junction form a stem-loop hairpin, and not a bulge. This contrasts with previous EM data which shows these structures as open bulges; however EM is performed in low salt conditions which may influence the structure formation.

Work using the fluorescent adenine analogue 2-aminopurine (2-AP), which exhibits increased fluorescence with increasing solvent exposure, has shown that a hairpin, and not an open bulge, is formed by 15 CAG repeats when part of a three-way junction.⁷⁹ Interestingly, 8 CAG repeats adopt an open bulge, not a structured stem-loop, in the context of a three-way junction, but form a structured stem-loop when the TNR sequence is not part of a three-way junction.^{78,246} Degtyareva and co-workers proposed that in the

shorter tract of 8 CAG repeats, folding into a stem-loop is disrupted by A•A mismatches in the stem and by base pairing in the flanking duplex regions.⁷⁸ Indeed, we propose here that the flanking duplex regions influence the likelihood of formation of the CAG hairpin at various positions along the complementary CTG repeats.

In order to determine the extent to which 8-oxoG alters the structure of the three-way junction, we replaced a G in the CAG stem-loop with 8-oxoG. Importantly, when 8-oxoG is introduced into the 8th repeat, in the center of the CAG repeats, the observed hyper-reactivity to DEPC is restricted to adenines A7 through A10 (Figure 3.1B, Figure 3.2B). This pattern of reactivity corresponds to the formation of the stem-loop at three of the seven potential positions along the CTG complement. These results can be compared to those obtained in the absence of 8-oxoG, in which A6 through A12 were hyper-reactive to DEPC and the CAG stem-loop formed at six of the seven potential positions along the CTG complement. Importantly, the loss of hyper-reactivity at certain adenines indicates that 8-oxoG is modulating the location where the hairpin forms within the repeat tract.

Recently the behavior of shorter CAG repeat bulges was defined in a similar three-way junction substrate using DSC.²⁴³ CAG repeats were allowed to pair with CTG repeats in the complementary strand, giving rise to substrates capable of adopting 1, 3, or 5 different structures, depending on the number of CAG and CTG repeats present in the substrate. The small CAG repeat bulges were able to freely roll across the CTG complement in a phenomenon termed “rolimerization”, giving rise to several energetically equivalent structures. However, the population could be biased towards an individual structure by the presence of an abasic site. In a more constrained system

designed to investigate the behavior of the bases located at the three-way junction itself, the bases were determined to be highly dynamic.²⁴⁴ Even when the position of the hairpin arm is locked in place, the bases in the junction have several different options for base pairing. Furthermore, the dynamics associated with the three-way junction were demonstrated to affect the repair outcome.

Piperidine treatment of the three-way junction containing 8-oxoG results in a small amount of strand cleavage at the 8-oxoG (~2%) (Figure 3.1B, Figure 3.2B), however the 8-oxoG lesion is known to be refractory to piperidine.^{247,248} Therefore, the strand cleavage observed in piperidine-treated DNA controls does not arise from 8-oxoG. This piperidine sensitivity is likely derived from the oxidation of 8-oxoG during deprotection to form hyperoxidized guanine lesions, some of which are sensitive to cleavage by piperidine.²⁴⁹ It is not uncommon for such impurities to be present in synthesized oligonucleotides containing 8-oxoG, despite the presence of β -mercaptoethanol during deprotection to reduce the propensity for further oxidation of 8-oxoG. Using analysis by mass spectrometry a 10% impurity has been reported previously.²⁴⁸ Here, we observe ~2% of the 8-oxoG substrates reacting with piperidine. However, because the DEPC reactions are performed under single-hit conditions with only ~5% of the total DNA reacting with the chemical probe, strand cleavage of hyperoxidized substrates would appear as only 0.1% reactivity of the total DNA substrate and would not be above background. The percentage of strand cleavage at 8-oxoG is consistent in all the piperidine treated, lesion-containing substrates, and is not affected by the addition of the chemical probe. This observation indicates that the strand cleavage observed at 8-oxoG is due to piperidine and not reaction with the chemical probe. For this

reason the amount of strand cleavage at 8-oxoG does not reflect its solution accessibility and cannot be used in our structural analysis. Indeed, our structural predications are derived from the reactivity towards DEPC of the adenines and guanines in the substrates and not reactivity of 8-oxoG.

When considering the amount of reactivity to DEPC at each adenine in CAG16-TWJ-8OX/CTG6-TWJ, these structures seem to position the 8-oxoG either directly in the loop or in the base pairs nearest the loop, eliminating the structures where the lesion is positioned further down the stem (Figure 3.1B). We hypothesize that positioning the lesion at or near the loop within the three-way junction could decrease or eliminate the enthalpic penalty associated with 8-oxoG/C base pairing, as it is known that 8-oxoG base paired with C in a duplex confers an enthalpic destabilization to the structure.²⁵⁰⁻²⁵² To better understand the behavior of the three-way junction, it is advantageous to study the behavior of the stem-loop component alone; this approach will allow us to better address the combination of structural and energetic influences 8-oxoG has on TNR DNA.

3.3.2 Influence of 8-oxoG on the structure of an even-numbered TNR sequence

Since we observed that the structure of the CAG16-TWJ-8OX/CTG6-TWJ sequence is modulated in order to position 8-oxoG in or near the loop, we wondered if 8-oxoG would have a similar effect on the TNR sequence when removed from the three-way junction. To investigate the inherent ability of 8-oxoG to impact the structure of TNR DNA outside the context of a three-way junction, we performed chemical probe experiments with CAG10 (Table 3.1), as this mimics the hairpin of the three-way junction.

When incubated with DEPC, the CAG10 substrate shows reactivity at every A, with increased reactivity at A5 and A6, indicating their presence in a loop (Figure 3.3A).

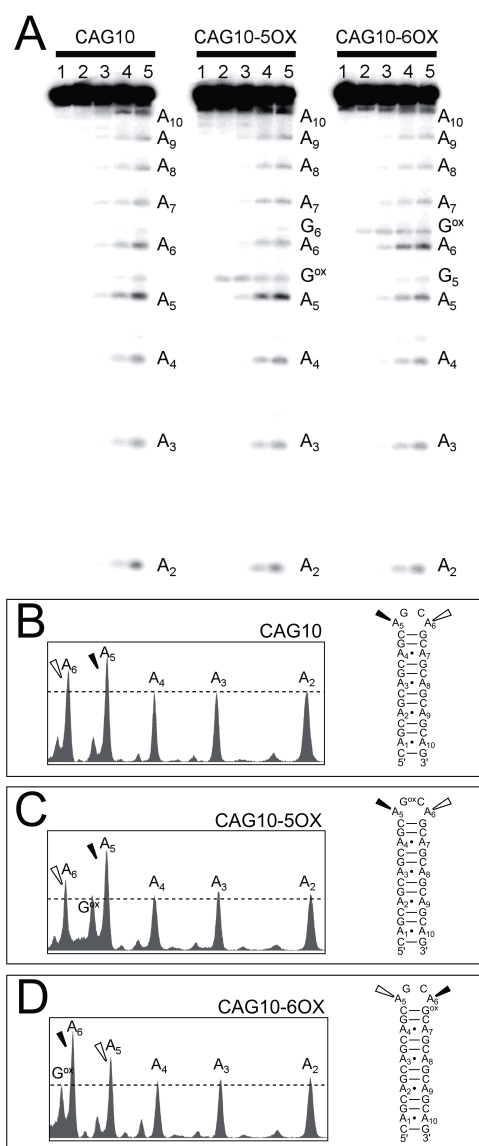


Figure 3.3 (A) Autoradiogram of DEPC reactions with CAG10, CAG10-5OX, and CAG10-6OX. Lanes 1 are DNA alone, lanes 2 are piperidine-treated DNA, and lanes 3-5 are DNA treated with 1% (v/v) DEPC at 37 °C for 5, 15 or 30 min respectively, followed by treatment with piperidine. Histograms of DEPC reactivity, derived from lanes 4, with (B) CAG10, (C) CAG10-5OX, and (D) CAG10-6OX. The dashed line in each histogram demarks the level of reactivity of adenines in the stem region, which serves as a baseline to establish hyper-reactivity of the nucleobases in the loop. Proposed structures for CAG10, CAG10-5OX, and CAG10-6OX are shown next to the corresponding histogram. The closed wedge indicates hyper-reactivity to DEPC and the open wedge indicates moderate hyper-reactivity to DEPC.

A histogram displaying a graphical representation of the reactivity of CAG10 is provided in Figure 3.3B. Furthermore, when incubated with the chemical probe dimethyl sulfate (DMS), which reveals the solvent accessibility of guanines,²⁵³ reactivity is observed at every G with hyper-reactivity at G5 (Figure 3.4A,B), consistent with G5 being positioned in a loop. This pattern of reactivity to DEPC and DMS indicate the formation of a four nucleobase loop that includes A5, G5, C6, and A6 and a blunt-ended stem (Figure 3.3B). Interestingly, the two reactive adenines in the loop, A5 and A6, are not modified to the same extent by DEPC. This difference in reactivity may derive from differences in stacking interactions with the adjacent nucleobases, as we proposed in earlier work.⁶⁷ The

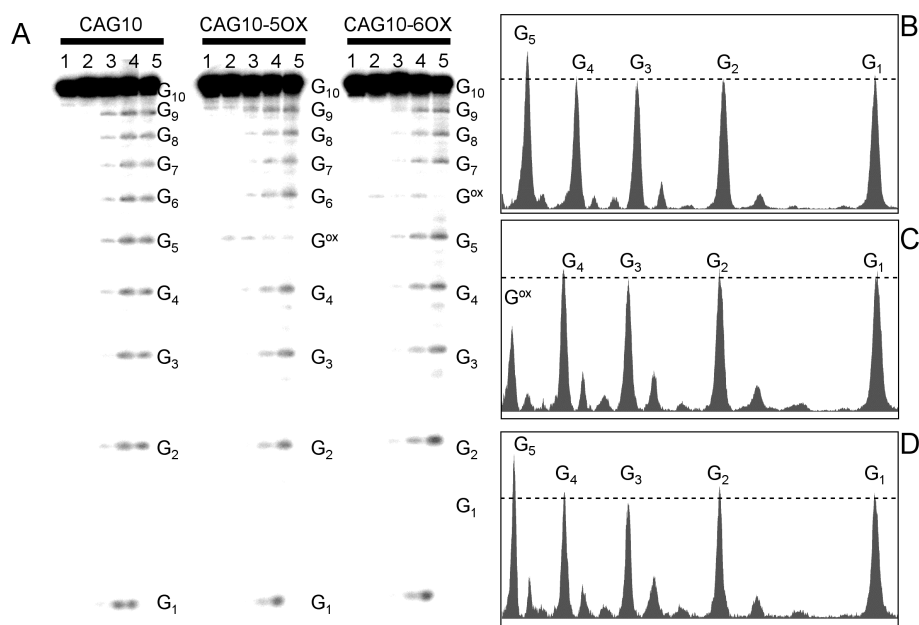


Figure 3.4 Reactivity of CAG10 substrates with DMS. (A) Autoradiogram of DMS reactions with CAG10, CAG10-5OX, and CAG10-6OX. Lanes 1 are DNA alone, lanes 2 are piperidine-treated DNA, and lanes 3-5 are DNA treated with 0.05% (v:v) at 37 °C for 5, 15 or 30 min respectively, followed by treatment with piperidine. Histograms derived from lanes 4 are shown in (B) CAG10, (C) CAG10-5OX, and (D) CAG10-6OX. The dashed line demarks the level of reactivity of guanines in the stem region, which serves as a baseline to establish hyper-reactivity of nucleobases.

folding of CAG10 into a hairpin is consistent with the structures observed for several other TNRs, including those comprised of CAG, CTG, CGG, and CCG repeats, which have been shown to form hairpins *in vitro*.^{58-60,62-68,72,78,79,242,246,254,255}

It is important to note that while hyper-reactivity to chemical probes was observed at particular adenines and guanines, some level of reactivity was observed at every A and G in CAG10, consistent with the established reactivity patterns of DEPC and DMS. DEPC is known to modify unpaired adenines in loops, bulges, and overhangs, but not adenines in well-matched A-T base pairs.^{256,257} However, the adenines in the stem of the TNR hairpins are part of A•A mismatches, which are known to have increased dynamics relative to an A-T base pair,²⁵⁸ and therefore we would expect these adenines to be modified by DEPC, although to a lesser extent than the unpaired adenines in the loop region. DMS is known to modify guanines in G/C base pairs,²⁵³ and we observed modification of such guanines in the stem region; however unpaired guanines, such as those in the loop, are modified to a greater extent due to increased accessibility to DMS. Interestingly, we observe only slight modification by DEPC of the adenines in the stem of the three-way junctions, and only for the longest incubation time; the flanking duplex regions may minimize the dynamics of the stem, leading to less modification by DEPC.

In order to determine the extent to which 8-oxoG influences the structure of CAG10, we site-specifically replaced G5 or G6 with 8-oxoG to form CAG10-5OX and CAG10-6OX, respectively. When an 8-oxoG is present in the CAG10 TNR sequence, the reactivity to DEPC and DMS indicate that the overall hairpin structure does not change, and that the position of the four nucleobase loop and the presence of a blunt-ended stem are maintained (Figure 3.3C,D, Figure 3.4C,D). However, while the overall hairpin

structure is maintained, there are subtle differences in the DEPC reactivity profile when 8-oxoG is present; notably, we observe an increase in the reactivity of the adenine 5' to the 8-oxoG. In fact, of all the adenines in CAG10-5OX, A5 exhibits the highest reactivity to DEPC (Figure 3.3C), while A6 of the CAG10-6OX substrate exhibits the highest reactivity to DEPC (Figure 3.3D). The observation of differing levels of solvent accessibility near an 8-oxoG is consistent with the recent work of Singh and co-workers, which demonstrated that 8-oxoG has a disruptive effect on a base pair 5' of an 8-oxoG/C base pair.²⁵² Our data indicate that 8-oxoG can exert a similar effect within a single-stranded DNA loop.

To confirm the presence of a blunt-ended stem in CAG10-5OX and CAG10-6OX, we performed DMS reactions to monitor the reactivity of G1 and G2. In particular, if the 8-oxoG in CAG10-6OX were to be repositioned into the loop, a two repeat overhang would form at the 5' end, leading to greater reactivity to DMS of the guanines in those repeats. We observed no increase in DMS reactivity at G1 or G2, consistent with a blunt-ended stem (Figure 3.4C,D). Thus, all three CAG10 substrates, regardless of the presence of 8-oxoG, form hairpins with a four nucleobase loop and a blunt-ended stem (Figure 3.3B-D). Therefore, unlike the effects exhibited by 8-oxoG in the three-way junction, the presence of 8-oxoG in CAG10 results only in minor changes to the solution accessibility of the adjacent nucleobase and does not dramatically alter the overall structure adopted by the sequence.

Recent work from our laboratory has indicated that single strands containing an odd number of CTG repeats, the complement of CAG, form populations composed of two structures, both with four nucleobase loops and either a 3' or a 5' overhang.⁶⁸ We

hypothesized that if odd-numbered CAG repeats exhibit the same behavior, they may have the ability to reposition and better accommodate the lesion, which could provide a more accurate model of TNR behavior within a three-way junction. Therefore, we performed chemical probing experiments on CAG sequences with an odd number of repeats.

3.3.3 Chemical probe analysis of odd-numbered, single-stranded TNR sequences lacking DNA damage

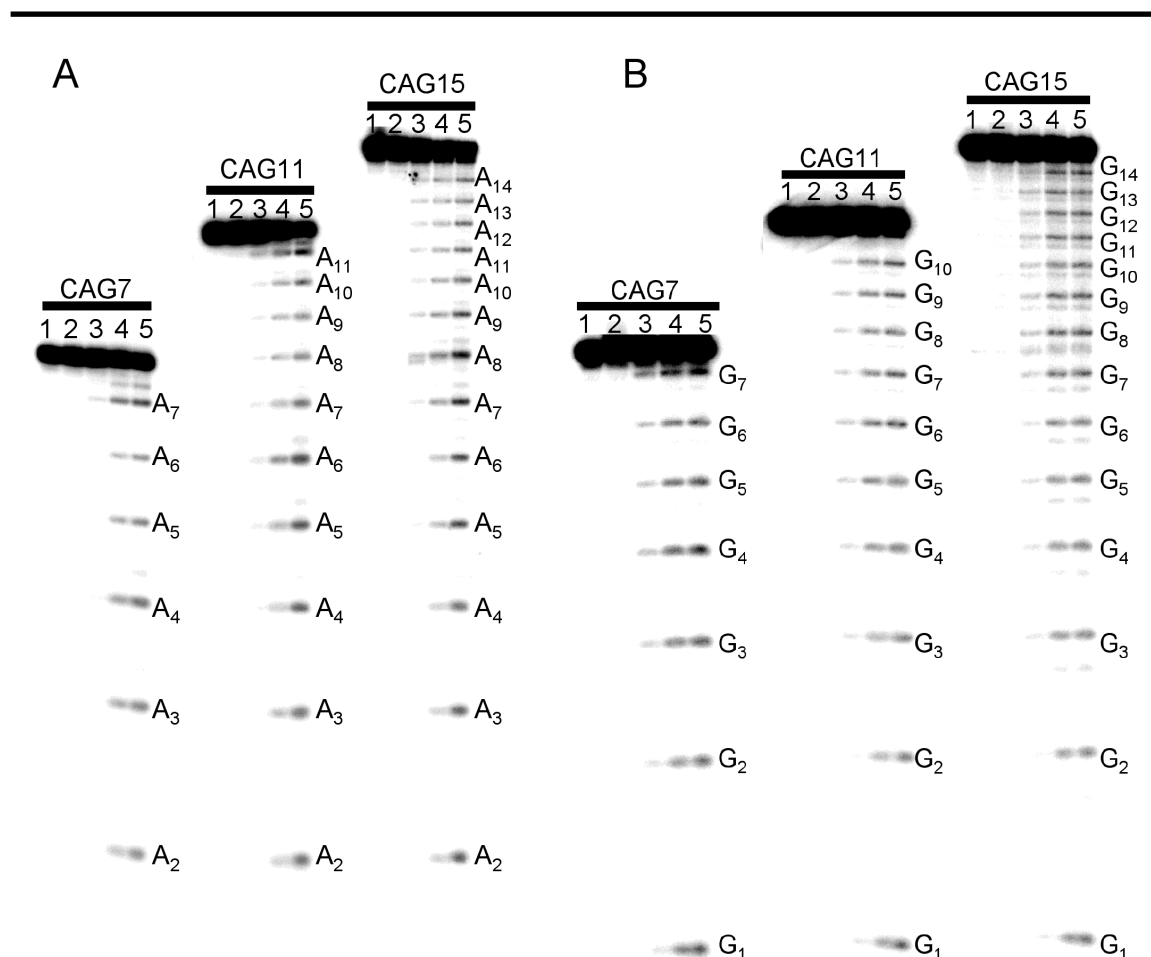


Figure 3.5 Reactivity of CAG7, CAG11, and CAG15 with chemical probes. (A) Reactivity of CAG7, CAG11, and CAG15 with DEPC. Lanes 1 are DNA alone, lanes 2 are piperidine-treated DNA, and lanes 3-5 are DNA treated with 1% (v/v) DEPC at 37 °C for 5, 15 or 30 min respectively, followed by treatment with piperidine. (B) Reactivity of CAG7, CAG11, and CAG15 with DMS. Lanes 1 are DNA alone, lanes 2 are piperidine-treated DNA, and lanes 3-5 are DNA treated with 0.05% (v:v) DMS.

When CAG7, CAG11, and CAG15 were incubated with DEPC, reactivity at every A in the sequence was observed, but hyper-reactivity was observed at adenines in the center of each sequence (Figure 3.5). Histograms displaying a graphical representation of the reactivity of each sequence are provided in Figure 3.6A and reveal that the adenines hyper-reactive towards DEPC are as follows: A3, A4, A5, and A7 in CAG7; A5, A6, and A7 in CAG11; and A7, A8, and A9 in CAG15. When each single-stranded substrate was incubated with DMS, reactivity was observed at each G, with hyper-reactivity observed at the guanines in the center and at the ends of each sequence: G1, G3, G4, and G7 in CAG7; G1, G5, and G6 in CAG11; and G1, G7, and G8 in CAG15.

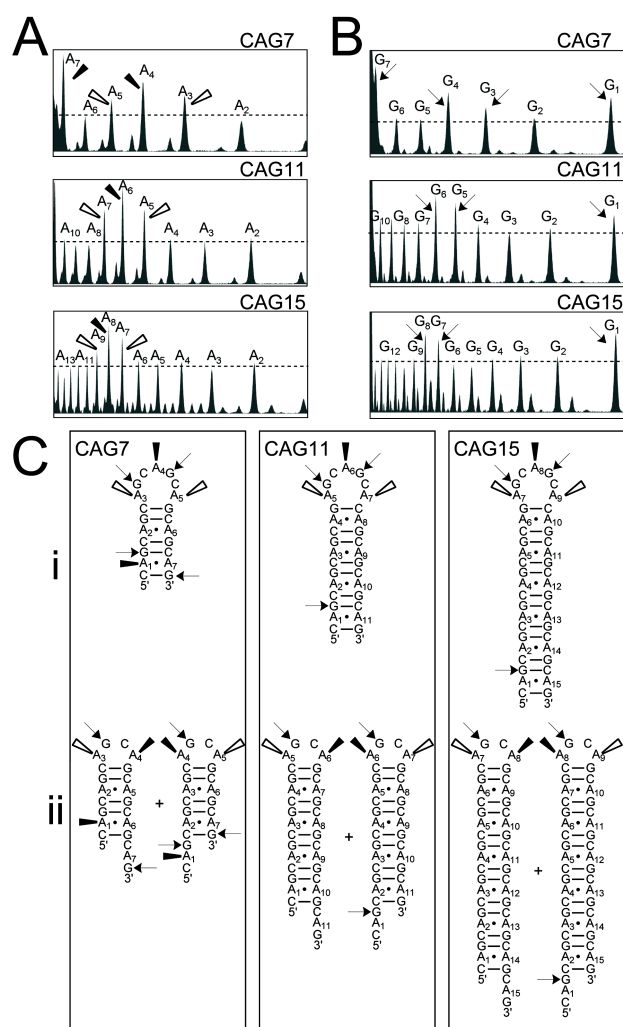


Figure 3.6 Histograms displaying reactivity of CAG7, CAG11, and CAG15 to (A) DEPC and (B) DMS. Histograms are derived from lanes 4 of the autoradiograms shown in Supporting Information Figure S3, in which the DNA was treated with the chemical probe at 37 °C for 15 min. The dashed line demarks the level of reactivity of adenines or guanines in the stem region, which serves as a baseline to establish hyper-reactivity of nucleobases in the loop. The closed wedge indicates the greatest degree of hyper-reactivity to DEPC, the open wedge indicates moderate hyper-reactivity to DEPC, and the arrow indicates hyper-reactivity to DMS. (C) Two potential scenarios for the structures of CAG7, CAG11, and CAG15 derived from their reactivity to DEPC and DMS. Structures containing a seven nucleobase loop and blunt-ended stem are shown in C(i) while a population of two stem-loops each containing a four nucleobase loop and either a 5' or 3' overhang is shown in C(ii).

CAG15 (Figure 3.5, Figure 3.6B). The results obtained in the chemical probing experiments indicate the formation of hairpins by CAG7, CAG11, and CAG15, with the most highly modified adenines and guanines in either the loop region or at the base of the stem (Figure 3.6C). The remaining adenines and guanines in the sequences participate in A•A mismatches and G/C base pairs in the stem. However, there are two scenarios which could produce the chemical reactivity profiles we observed for CAG7, CAG11, and CAG15 and both would be consistent with a hairpin(s): (1) a homogeneous population of a single stem-loop with a 7 nucleobase loop and blunt-ended stem (Figure 3.6C i) or (2) a population consisting of two different stem-loop structures, each with loops containing 4 nucleobases along with either a 5' or 3' overhang of one CAG repeat (Figure 3.6C ii).

Relative to the other guanines in the stem, the increased reactivity toward DMS observed at G1 in CAG7, CAG11, and CAG15 argues for the presence of a 5' overhang, as G1 would be positioned in the stem in scenario (i), and would not be expected to have increased reactivity relative to G2. Since neither the hairpin with the 5' nor the 3' overhang should be favored, we would expect both species to be present; therefore, in addition to the 5' overhang indicated by hyper-reactivity to DMS at G1, we would also expect to observe a 3' overhang. A 3' overhang would be indicated by hyper-reactivity to DMS at G7, G11, and G15 for CAG7, CAG11, and CAG15, respectively. Indeed, hyper-reactivity at G7 of CAG7 was observed (Figure 3.6B), which is consistent with CAG7 forming a population of two stem-loop hairpins. However, due to the longer length of CAG11 and CAG15, we did not obtain sufficient electrophoretic resolution of the unmodified DNA from the fragment corresponding to cleavage induced by DMS at their respective 3' guanines; this prevented direct observation of a 3' overhang species for

these substrates. Note that we used DMS, and not DEPC, to confirm the presence of overhangs in CAG7, CAG11, and CAG15; this is because an A1 fragment is not retained on the gel.

3.3.4 Structural characterization of ‘locked-loop’ hairpin standards

To confirm that CAG7, CAG11, and CAG15 form a population of two hairpin structures as depicted in Figure 3.6C ii, we mimicked both the blunt-ended and overhang stem-loop hairpins using ‘locked-loop’ standards. We chose to derive the locked-loop standards from CAG7 as it possesses the same pattern of reactivity to chemical probes as the longer CAG11 and CAG15 substrates, but also allows for electrophoretic resolution of unmodified DNA from DNA that is modified at the terminal 3' repeat. In these standards, the nucleobases in the loop of the proposed CAG7 structure in either scenario (i) or (ii) are replaced by thymines, forcing the stem to adopt a 3' overhang (CAG7-3'Overhang), a 5' overhang (CAG7-5'Overhang), or a blunt-ended stem (CAG7-Blunt

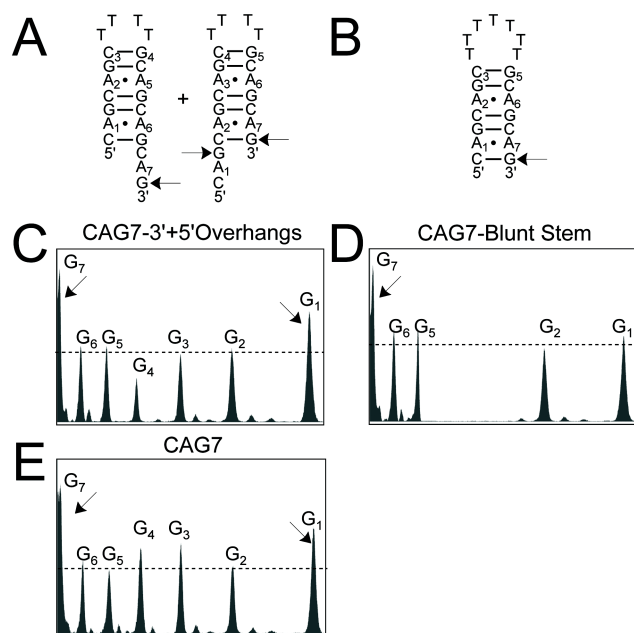


Figure 3.7 (A) Proposed structures for CAG7-3'Overhang (left) and CAG7-5'Overhang (right) and (B) CAG7-Blunt Stem. Histograms displaying reactivity to DMS for (C) CAG7-3'+5'Overhangs, (D) CAG7-Blunt Stem, and (E) CAG7. Histograms are derived from lane 4 of the autoradiograms shown in Figure S4 and S5 where the DNA was treated with 0.05% (v/v) DMS at 37 °C for 15 minutes. The arrow indicates hyper-reactivity to DMS.

Stem) (Figure 3.7A, B). CAG7-3'Overhang and CAG7-5'Overhang were combined in equal amounts to replicate the proposed population of structures with overhangs (CAG7-3'+5'Overhangs).

When incubated with DMS, both G1 and G7 of CAG7-3'+5'Overhangs display hyper-reactivity compared to the guanines in the neighboring repeats, consistent with the increased reactivity expected for the solvent-exposed overhangs (Figure 3.7C). Interestingly, while G7 of CAG7-Blunt Stem is hyper-reactive to DMS, G1 reacts at the same level as G2 (Figure 3.7D). For this blunt-ended substrate, the hyper-reactivity observed at G7 is likely due to end fraying of the stem, since this nucleobase is part of the terminal base pair. The reactions with DMS (Figure 3.7, Figure 3.8) and DEPC (Figure 3.9) confirmed the expected folding of all locked-loop standards.

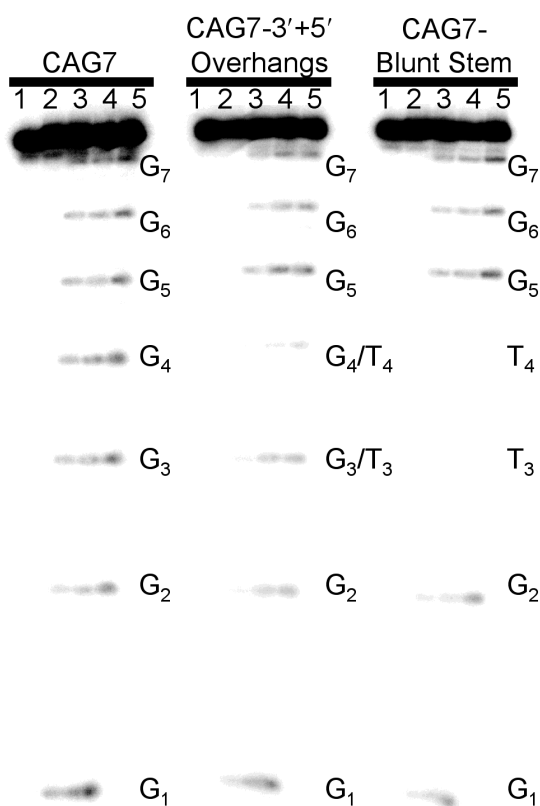


Figure 3.8 Reactivity of the CAG7 locked-loop standards with DMS. Autoradiogram of DMS reactions with (left) CAG7, (center) CAG7-3'+5'Overhangs, and (right) CAG7-Blunt Stem. Lanes 1 are DNA alone, lanes 2 are piperidine-treated DNA, and lanes 3-5 are DNA treated with 0.05% (v:v) DMS at 37 °C for 5, 15 or 30 min respectively, followed by treatment with piperidine.

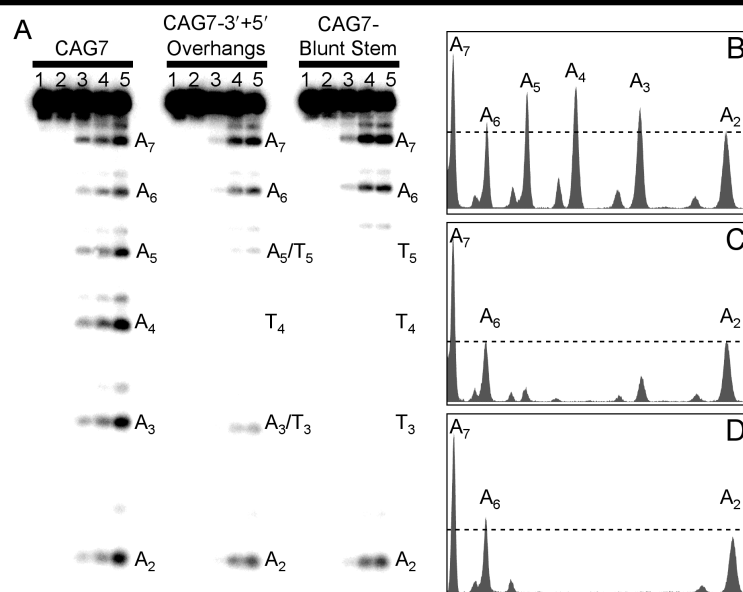


Figure 3.9 Reactivity of the CAG7 locked-loop standards with DEPC. (A) Autoradiogram of DEPC reactions with CAG7, CAG7-3'+5'Overhangs, and CAG7-Blunt Stem. Lanes 1 are DNA alone, lanes 2 are piperidine-treated DNA, and lanes 3-5 are DNA treated with 1% (v/v) DEPC at 37 °C for 5, 15 or 30 min respectively, followed by treatment with piperidine. Histograms derived from lanes 4 are shown in (B) CAG7, (C) CAG7-3'+5'Overhangs, and (D) CAG7-Blunt Stem. The dashed line in demarks the level of reactivity of adenines in the stem region, which serves as a baseline to establish hyper-reactivity of nucleobases.

We then compared the reactivity of G1 and G7 toward DMS in the locked-loop CAG7 standards to the reactivity of the same nucleobases in CAG7. Since G1 and G7 of CAG7 display hyper-reactivity to DMS similar to that of the CAG7-3'+5'Overhangs standard (Figure 3.7C,E), and not that of CAG7-Blunt Stem (Figure 3.7B), we conclude that CAG7 and the other odd numbered repeat sequences, CAG11 and CAG15, form populations of two stem-loop hairpins with four nucleobase loops and either a 3' or 5' overhang (Figure 3.6C ii). Using enzymatic probes, other laboratories have proposed the presence of overhangs in structures formed by 15 CAG repeats.⁶³

3.3.5 Influence of 8-oxoG on the structure of an odd-numbered, single-stranded TNR DNA

The results discussed above demonstrate the adaptability of odd numbered CAG repeats compared to CAG10. Thus, to better address the impact of 8-oxoG on the structure of TNR DNA and provide a better mimic for the hairpin arm of the three-way junction, we performed chemical probe experiments with odd numbered substrates that contain 11 CAG repeats and in which a single G was site-specifically replaced by 8-oxoG. We chose CAG11 as the model substrate since it has the same number of base pairs and mismatches in the stem as CAG10 and the hairpin arm of the three-way junction.

The first substrate contains 8-oxoG in the fifth repeat (CAG11-5OX), and the second substrate contains 8-oxoG in the sixth repeat (CAG11-6OX) (Table 3.1). As described above, upon incubation with DEPC, CAG11 that lacks DNA damage shows the highest degree of reactivity at A5, A6, and A7 (Figure 3.10B). The CAG11-5OX substrate has increased reactivity toward DEPC at A5 and A6 when compared to the flanking adenines (Figure 3.10C). This result indicates that A5 and A6 are part of the loop, while the remaining adenines are in A•A mismatches in the stem. In order for these particular adenines to be present in the loop, one CAG repeat must overhang at the 3' end of the stem. In contrast to CAG11-5OX, in which A5 and A6 were most susceptible to modification by DEPC, CAG11-6OX displays increased reactivity at A6 and A7 (Figure 3.10D). This result indicates that CAG11-6OX also forms a stem-loop structure with a four nucleobase loop, but unlike CAG11-5OX, the loop includes A6 and A7. In this case one CAG repeat must overhang the 5' end of the stem. As observed for CAG10, the two hyper-reactive adenines in CAG11-5OX and CAG11-6OX are modified by DEPC to different extents, which could be an effect of different stacking interactions with the

adjacent bases. Notably, in both CAG11-5OX and CAG11-6OX the 8-oxoG is in the loop region.

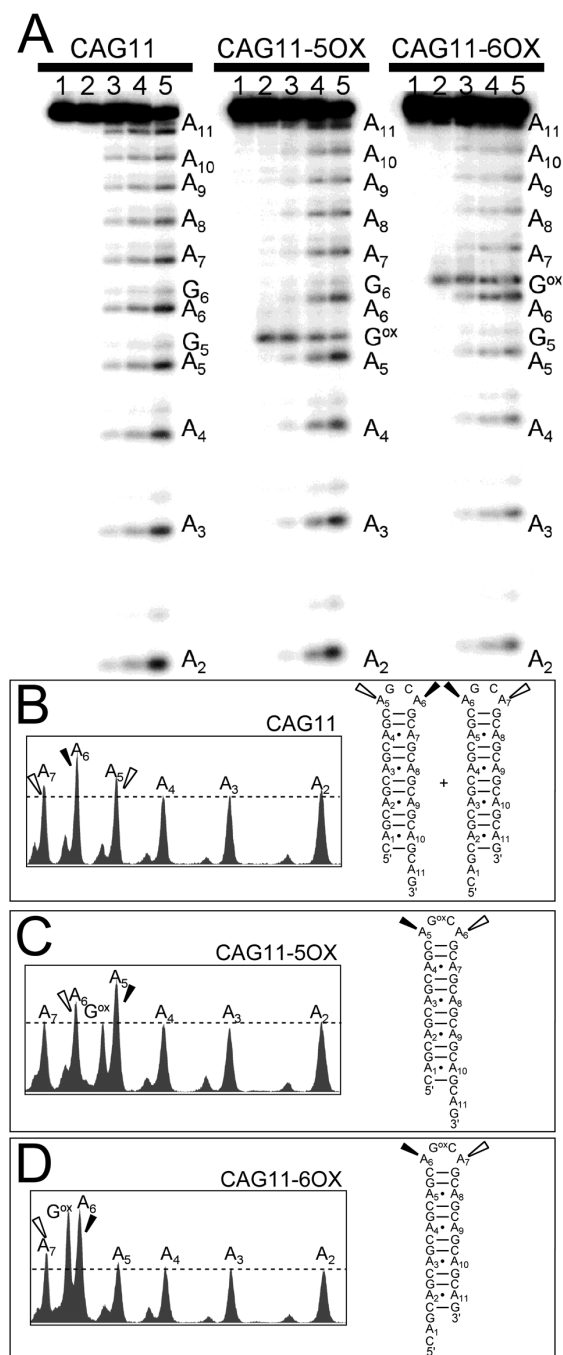


Figure 3.10 (A) Autoradiogram of DEPC reactions with CAG11, CAG11-5OX, and CAG11-6OX. Lanes 1 are DNA alone, lanes 2 are piperidine-treated DNA, and lanes 3-5 are DNA treated with 1% (v/v) DEPC at 37 °C for 5, 15 or 30 min respectively, followed by treatment with piperidine. Histograms of DEPC reactivity, derived from lanes 4, with (B) CAG11, (C) CAG11-5OX, and (D) CAG11-6OX. The dashed line in each histogram demarks the level of reactivity of adenines in the stem region, which serves as a baseline to establish hyper-reactivity of the nucleobases in the loop. Proposed structures for CAG11, CAG11-5OX, and CAG11-6OX are shown next to the corresponding histogram. The closed wedge indicates the greatest degree of hyper-reactivity to DEPC and the open wedge indicates moderate hyper-reactivity to DEPC.

We draw two conclusions from the chemical probing experiments of the CAG11 substrates. First, stem-loop TNR DNA containing 8-oxoG adopts a structure different

from TNR DNA lacking the damage. The presence of the 8-oxoG imposes a structural constraint that results in formation of a *single* hairpin, as compared to the CAG11 control that is a mixed population containing *two* species of hairpins. Second, the position of the 8-oxoG within CAG11 determines the structure. Indeed, the hairpin adapts in order to position 8-oxoG in the loop.

3.3.6 Calorimetric analysis of single-stranded TNR DNA lacking DNA damage

To characterize further the structures adopted by the CAG7, CAG10, CAG11, and CAG15 sequences lacking 8-oxoG, we employed DSC to elucidate the thermal and thermodynamic parameters associated with the transition from the hairpin to the unstructured state. Representative thermograms for CAG7, CAG10, CAG11, and CAG15 are shown in Figure 3.11. A DSC thermogram displays excess heat capacity as a function of temperature and the area under the curve is a direct measure of the enthalpy (ΔH) associated with the structural transition.²⁵⁹ Entropy (ΔS), free energy (ΔG), and melting temperature (T_m) can be obtained as described previously.²⁵⁹

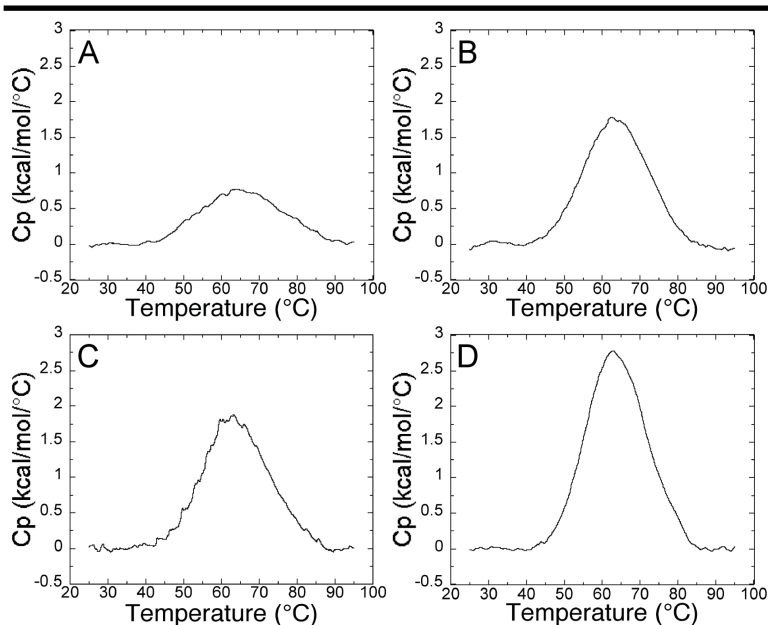


Figure 3.11
Representative thermograms, both background corrected using a buffer-buffer scan and baseline corrected using a polynomial fit, for (A) CAG7 at 33.1 μM , (B) CAG10 at 30 μM , (C) CAG11 at 16.8 μM , and (D) CAG15 at 40 μM . All samples were suspended in buffer containing 20 mM sodium phosphate, 70 mM NaCl, pH 7.5.

While there is an increase in the T_m between CAG7 and CAG10, and between CAG7 and CAG11, there is no significant difference in T_m between CAG10, CAG11, and CAG15 (Table 3.2). Previous work with CAG repeats, ranging in length from 6 to 25 repeats, has shown that, while there may be some fluctuations in T_m , there is no linear relationship between T_m and CAG repeat length.^{61,63}

Table 3.2: DSC-derived thermodynamic parameters

Substrate	T_m^a (°C)	$\Delta H^{a,b}$ (kcal/mol)	$\Delta S^{a,b}$ (cal/mol/K)	$\Delta G^{a,c}$ (kcal/mol)	$\Delta\Delta H$ (kcal/mol)	$\Delta\Delta S$ (cal/mol/K)	$\Delta\Delta G$ (kcal/mol)
CAG7	61.2 ± 1.3	23.7 ± 1.8	71.3 ± 5.2	1.72 ± 0.18	-	-	-
CAG10	63.0 ± 0.5	33.8 ± 1.3	100.7 ± 4.0	2.63 ± 0.09	-10.1 ± 2.2 ^d	-29.4 ± 6.6 ^d	-0.9 ± 0.2 ^d
CAG11	63.3 ± 0.1	34.0 ± 2.0	101.1 ± 6.0	2.67 ± 0.15	-10.3 ± 2.7 ^d	-29.8 ± 7.9 ^d	-1.0 ± 0.2 ^d
CAG15	63.5 ± 0.7	51.4 ± 2.0	152.8 ± 6.1	4.00 ± 0.15	-27.7 ± 2.7 ^d	-81.5 ± 8.1 ^d	-2.3 ± 0.2 ^d
CAG10-5OX	60.2 ± 0.7	33.4 ± 1.1	100.1 ± 3.2	2.34 ± 0.16	N/A ^{e,f}	N/A ^{e,f}	N/A ^{e,f}
CAG10-6OX	61.1 ± 0.5	29.8 ± 0.9	89.1 ± 2.5	2.15 ± 0.10	4.0 ± 1.8 ^e	11.6 ± 4.7 ^e	0.5 ± 0.1 ^e
CAG11-5OX	59.5 ± 0.1	34.6 ± 2.7	104.1 ± 8.0	2.36 ± 0.18	N/A ^{e,f}	N/A ^{e,f}	N/A ^{e,f}
CAG11-6OX	59.5 ± 0.1	34.9 ± 2.6	104.9 ± 7.9	2.37 ± 0.19	N/A ^{e,f}	N/A ^{e,f}	N/A ^{e,f}

^aDNA in 20 mM sodium phosphate, 70 mM NaCl, pH 7.5. The error represents the standard deviation derived from a minimum of three experiments. ^bValues derived by integration of the area under the excess heat capacity curve. ^cValues at 37 °C. ^dWith respect to CAG7. ^eWith respect to the equivalent substrate lacking 8-oxoG. ^fNot statistically different.

In contrast to the lack of linear relationship for T_m , we observed increasing ΔH values with respect to the length of the repeat sequence (Table 3.2), indicating that as repeat length increases, so too does the enthalpic stability of the structure. With respect to nucleic acids, both hydrogen bonding and base stacking interactions are reflected in the value of ΔH .²⁵⁹ Thus, the increase in ΔH observed here is rationalized by the increase in number of hydrogen bonds and base stacking interactions as the length of the stem increases. However, this increase in enthalpic stabilization as a function of TNR length is tempered by an increase in the entropic contributions. Nevertheless, ΔG increases with repeat length, indicating that the thermodynamic stability of the hairpins increases with length. These thermodynamic data are consistent with previously reported, DSC-derived values for similarly sized CAG repeats that lack DNA damage.^{63,65}

3.3.7 Analysis of damage-containing TNR structures by calorimetry

All of the 8-oxoG-containing sequences, composed of either 10 or 11 CAG repeats, show a decrease in T_m of ~ 2 -4 °C compared to their undamaged control substrates, regardless of whether 8-oxoG is positioned in the loop or the loop-closing base pair (Table 3.2). It is known that 8-oxoG is thermally destabilizing when base paired to C in a duplex,^{250-252,260,261} although this is an example where 8-oxoG decreases T_m when present outside the context of a well-matched duplex. The decrease in T_m is likely caused by a modest perturbation of the base stacking interactions in the loop when 8-oxoG is present, as supported by our chemical probing experiments.

With respect to the thermodynamic impact of 8-oxoG, we first considered its effects on the blunt-end CAG10 hairpin. Compared to the undamaged CAG10 control, we observed no difference in the ΔH associated with the transition from the hairpin to the unstructured sequence for CAG10-5OX, where 8-oxoG is positioned in the loop. When compared to the value for CAG10, there is a small but significant decrease in the ΔH for

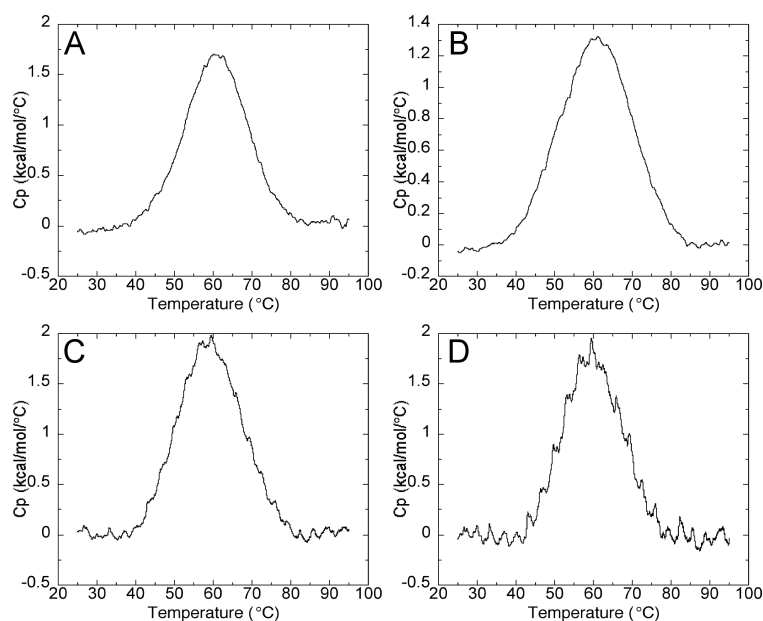


Figure 3.12
Representative DSC thermograms, both background corrected using a buffer-buffer scan and baseline corrected using a polynomial fit, for (A) CAG10-5OX at 50.6 μ M, (B) CAG10-6OX at 45.8 μ M, (C) CAG11-5OX at 23.4 μ M, and (D) CAG11-6OX at 12.9 μ M. All samples are in 20 mM sodium phosphate, 70 mM NaCl, pH 7.5.

CAG10-6OX where the 8-oxoG is positioned in the loop-closing base pair (Table 3.2) (Figure 3.12). However, it is important to note that the $\Delta\Delta H$ of 4 kcal/mol observed for CAG10-6OX is much smaller than the values reported for 8-oxoG base paired to C in a duplex, where $\Delta\Delta H$ values ranging from 7.3 to 52.9 kcal/mol have been reported depending on sequence context and length of the duplex.²⁵⁰⁻²⁵²

As observed for ΔH , ΔS is the same for CAG10 and CAG10-5OX, but there is a decrease in ΔS for CAG10-6OX compared to CAG10, indicating that the hairpin formed by CAG10-6OX is more disordered than its counterparts. These changes in ΔH and ΔS lead to a small but significant decrease in ΔG compared to CAG10, indicating that the structure formed by CAG10-6OX is thermodynamically less stable than the structures formed by either CAG10 or CAG10-5OX.

For CAG11 we observed that 8-oxoG does not influence ΔH . The $\Delta\Delta H$ comparing the melting transitions of CAG11-5OX and CAG11-6OX with the CAG11 control are nearly zero (Table 3.2). The proposed structures for CAG11, CAG11-5OX, and CAG11-6OX have the same number of base pairs and mismatches in the stem and, therefore, the same number of hydrogen bonds (Figure 3.10). Thus, the hydrogen bonding pattern is not expected to alter ΔH of the structures. 8-oxoG could potentially alter base stacking in the loop but given the similar ΔH values for the CAG11 substrates the contribution of base stacking to enthalpic stability must be similar for all substrates. Therefore, while our chemical probe data indicate that 8-oxoG alters base stacking in the loop, our thermodynamic data suggest that any alteration of stacking interactions caused by 8-oxoG in the loop of CAG11-5OX and CAG11-6OX is not sufficient to significantly alter the enthalpic stability of the hairpin. Since it is known that 8-oxoG is enthalpically

destabilizing when base paired with C in duplex,²⁵⁰⁻²⁵² 8-oxoG is likely positioned in the loop of these hairpin structures to avoid the enthalpic penalty associated with base stacking and/or hydrogen bonding that would occur if 8-oxoG were in the stem region.

Here, we find that the positioning of 8-oxoG in the loop of TNR DNA decreases the T_m but does not affect ΔH . Another example where a modified nucleobase alters thermal but not enthalpic stability is 2-AP in a (CAG)₆ hairpin.²⁶² When 2-AP was positioned in the loop, there was an increase in T_m but there was no change in ΔH . Lee and co-workers proposed that the increase in thermal stability is a result of favorable base stacking interactions of 2-AP.

As observed for ΔH , the ΔS describing the transition from the structured to the unstructured state is the same for all three CAG11 substrates. It is reasonable to expect that the proposed structures would have similar degrees of disorder as each structure has the same number of G/C base pairs, A•A mismatches, and unpaired nucleobases. Thus, the similar ΔS values observed here indicate that 8-oxoG in the loop makes no significant contribution to the entropy of the system. Lastly, since 8-oxoG does not influence ΔH or ΔS of the stem-loop structure adopted by CAG11, the introduction of 8-oxoG to CAG11 does not affect the thermodynamic stability of the stem-loop hairpin.

Recent work done in the context of TNR Ω -DNA focused on investigating the thermodynamics of the tetrahydrofuran (THF) abasic site analogue, both alone and in combination with 8-oxoG, can broaden our understanding of the impact that 8-oxoG has on the structure of CAG repeat DNA.⁷⁵⁻⁷⁷ The Ω -DNA is a three-way junction that contains an open bulge formed by 6 CAG repeats and is flanked on both sides by mixed-sequence duplex. Based on a calorimetric analysis, Völker and co-workers concluded

that, because the Ω -DNA accommodates the THF abasic site without altering the thermodynamic parameters, the THF abasic site can alter the distribution of the structural microstates occupied by the Ω -DNA.⁷⁵ In other work, the same authors proposed allosteric coupling between a 8-oxoG and/or a THF abasic site in the Ω -DNA.^{76,77} This finding again demonstrates the ability of these lesions to alter the microstates adopted by the Ω -DNA. Besides an abasic site, here we provide structural and thermodynamic data that supports the notion that the CAG repeats can also adapt their structure to accommodate the oxidatively damaged nucleobase 8-oxoG.

Our results demonstrate that when 8-oxoG replaces G5 or G6 of CAG10, there is no overall change in the hairpin structure. In contrast, when 8-oxoG replaces G5 or G6 of CAG11, the hairpin structure changes to accommodate 8-oxoG in the loop, an adjustment that we propose is enthalpically driven. Why then does CAG10-6OX not position the 8-oxoG in the loop, and create a 5' overhang, if a position in the loop is favored for 8-oxoG? If an overhang were to form, two base pairs in the stem would be lost, resulting in a decrease in $\Delta H \sim 10$ kcal/mol, using nearest neighbor calculations.²⁶³ The thermodynamic consequence arising from the loss of those base pairs would be more detrimental to the stability of the hairpin than the benefit that would come from positioning the 8-oxoG in the loop (+ 4 kcal/mol). Thus the positioning of 8-oxoG in the TNR structure represents a thermodynamic balancing act; when there is the possibility of positioning 8-oxoG in the loop while still maintaining the base pairs and stacking of the stem, there is a significant favorability for 8-oxoG to occupy that position.

Reactions performed with DEPC showed the three-way junction CAG16-TWJ-8OX/CTG6-TWJ forms several unequivalent structures; the 8-oxoG occupies different

positions within the hairpin arm. Although DSC data obtained for the three-way junction could potentially yield thermodynamic parameters of the entire population, they cannot be used to determine the thermodynamic parameters of the individual structures. Importantly, we can use the thermodynamic information obtained for the CAG10 and CAG11 sequences to better understand the behavior of the hairpin arm of the three-way junction. Because substitution of 8-oxoG at either G5 or G6 of CAG11 leads to formation of a single structure, we are able to determine thermodynamic parameters and evaluate the impact of the lesion on the TNR DNA by comparing to the unmodified substrate. In contrast to the enthalpic destabilization observed in duplex DNA, we have shown that 8-oxoG positioned in the loop of either a CAG10 or CAG11 hairpin conveys no enthalpic penalty to the structure, and an 8-oxoG positioned in the loop closing base pair, as observed in CAG10-6OX, shows only a minor decrease in enthalpy, suggesting that positions in or near the loop are enthalpically favorable. It is of note that the structures lost upon addition of 8-oxoG to CAG16-TWJ/CTG6-TWJ are those that position the lesion furthest from the loop, where the lesion is expected to have the largest enthalpically destabilizing effect.

3.4 CONCLUSIONS

Analysis via chemical probing revealed the ability of 8-oxoG to alter the population of structures formed by a three-way junction containing 16 CAG repeats. To better understand the innate ability of CAG repeat sequences to adapt to the presence of 8-oxoG, we performed the same chemical probe analysis on CAG10, which is identical in sequence to the hairpin arm of the three-way junction, as well as two CAG10 substrates

containing 8-oxoG. We observed that 8-oxoG did not alter the structure of CAG10, and positioned 8-oxoG in either the loop or loop-closing base pair depending on the location of the lesion within the sequence. However, when 8-oxoG was introduced into CAG11, we observed a structural modification; 8-oxoG was positioned in the loop of both the CAG11-5OX and CAG11-6OX structures. The ability of the CAG11 sequence to adapt its structure to accommodate the 8-oxoG lesion is likely due to the enthalpic penalty associated with the formation of an 8-oxoG/C base pair in the stem region.

The formation of hairpins has been proposed to play a key role in mediating expansion by ligation of a structured flap into the repeat tract during LP-BER.^{59,159,162,164,242} If 8-oxoG is present in the repeat tract as these structures form in the genome, they may form such that the 8-oxoG is positioned at or near the loop. Furthermore, if 8-oxoG forms within a hairpin, the structure may rearrange to position the 8-oxoG in the loop. Additionally, if oxidation occurs after formation of the hairpin, the 8-oxoG will likely remain in the loop rather than re-arranging to form an 8-oxoG/C bp in the stem. Importantly, our previous work has shown that 8-oxoG is inefficiently repaired when positioned in the loop of a hairpin,⁵⁹ thus 8-oxoG would persist and allow another round of the “toxic oxidation cycle” to occur. These results underscore the importance of understanding repeat structure and how the interplay with DNA damage can influence expansion of TNR regions in the genome.

Significance: these results clearly indicate that once an 8-oxoG has formed in the loop, it will remain in that location rather than forming a base pair in either the stem or the flanking duplex. Thus it may be possible to exploit differences between the lesion in a base pair and in the loop to target hairpins and prevent expansion.

Chapter 4

The Presence of a CAG Repeat Hairpin Decreases Affinity for the Histone Core and Modifies the Local Rotational Position of DNA in a Nucleosome

4.1 INTRODUCTION

As described in the previous chapters, the number of trinucleotide repeats (TNRs) associated with various neurodegenerative disorders can be highly variable.^{17-20,198} Furthermore, the number of repeats associated with a given gene is not fixed within an organism; the number varies based on age and cell type. Many TNRs have the potential to undergo both genomic expansions and contractions; processes controlled by various factors.

Most TNR expansion events occur in a polymerase-mediated fashion, through the formation of non-canonical secondary structures such as hairpins. A common cause of expansion in somatic cells is the formation of non-canonical structure during DNA replication.^{71,95,264-266} Expansion of the TNR tract can occur during replication fork stalling and subsequent reversal (Figure 1.3).^{20,57,93-95} As DNA replication reaches the repeat tract, the repeat on the lagging strand will be positioned in the transiently single stranded Okazaki initiation zone (OIZ), allowing the possible formation of secondary structure, which causes the replication fork to stall. The fork then reverses, forming a four-way junction in which a single stranded repeat tract is present off the 3' end of the leading strand. This repeat can also form a non-canonical secondary structure, which will be incorporated into the genome once fork progression resumes. If the excess DNA is not removed, the expansion is fixed in the genome after a second round of replication.

Recently, hairpin formation in the lagging strand during DNA replication of CAG and CTG repeats has been characterized *in vivo* using zinc-finger nucleases.⁸⁰

While a DNA replication dependent mechanism provides an explanation for expansion in proliferative cells, many of the cells where these neurodegenerative disorders manifest are non-proliferative. Expansion of the CAG/CTG repeats in the *HTT* gene is readily observed in the neurons of both HD patients and HD mice,^{24-26,159,164,267} yet these terminally differentiated neurons are not undergoing DNA replication and cell division. Thus, another metabolic process must be responsible for the expansions that can accumulate over time in non-proliferative cells. In HD, these expansions arise from repair of oxidative damage within the repeat tract.^{155,159,162,164,242,268} During base excision repair (BER) of the oxidative lesion 8-oxo-7,8-dihydroguanine (8-oxoG) located within the repeat tract, polymerase β can perform a multi-nucleotide addition creating a 5' single stranded flap, in a process called long-patch base excision repair (LP-BER). Since the 5' flap is composed of repeats, it can fold into a hairpin, making it refractory to cleavage by flap endonuclease 1. The hairpin is then ligated into the newly repaired strand, incorporating the expansion into the genome. While expansion likely occurs during repair of several different TNRs, the only repeats currently confirmed to expand during LP-BER are CAG¹⁵⁹ and CGG²⁶⁸ repeats.

Importantly, while both of these processes involve the formation of hairpin structures, little is known about how those structures respond to nucleosome packaging. As chromatin compaction is the next step in the cell cycle after DNA replication, it is important to understand how any hairpins formed during replication might behave within a nucleosome. Furthermore, hairpins are predicted to form as chromatin undergoes

subsequent levels of condensation, making their interaction with the histone core a potentially important aspect of correct genomic compaction.²⁶⁹ Additionally, previous research indicates that LP-BER can occur within a nucleosome, though at reduced efficiency.²⁷⁰⁻²⁷⁴ Should LP-BER mediated expansion occur in a repeat tract that is packaged in a nucleosome, what would be the effect of hairpin formation on nucleosome structure?

Here we investigate the ability of a 246 bp DNA duplex, with and without a (CAG)₁₀ hairpin, to form nucleosomes. Furthermore, we examine how the presence of a hairpin might alter the position of the DNA once incorporated. These results will help define the behavior of CAG hairpins formed within the genome and how they might influence the functional organization of a repeat tract.

4.2 EXPERIMENTAL PROCEEDURES

4.2.1 Oligonucleotide synthesis and purification

The oligonucleotides used to construct the substrates used in these experiments were synthesized by standard phosphoramidite chemistry as described previously. The oligonucleotides were synthesized dimethoxytrityl (DMT)-on and were deprotected in NH₄OH at 55 °C for 24-36 h. In the first round of HPLC purification, the sample was applied to a Polymer Labs Reverse Phase column (4.6 × 250 mm) and elution was carried out at 90 °C using 100 mM triethyl ammonium acetate (TEAA) in 99% acetonitrile:1% water (solvent A) and 100 mM TEAA in 1% acetonitrile:99% water (solvent B) as the mobile phases (gradient: solvent A was increased from 5% to 25% over 25 min; 1 mL/min). Following removal of the 5'-DMT group by incubation in 80% glacial acetic

acid for 12 min at room temperature, the oligonucleotides were dried, and then subjected to a second round of HPLC purification (gradient: solvent A was increased from 0% to 25% over 40 min; 1 mL/min).

4.2.2 Oligonucleotide ligation and purification

After purification of all 33 individual oligonucleotides (Figure 4.1), ligations were performed to generate the single stranded components of the S2 substrates. First, each piece of the S2 substrate was phosphorylated at the 5'-end using T4 polynucleotide kinase (PNK, New England Biolabs). The piece containing the eventual 5' end of the ligation product (S2_1, S2_6) was phosphorylated with ^{32}P , following the manufacturer's instructions. The remaining pieces involved in the ligation (S2_2, S2_3_C, S2_3_HP, S2_CAG6_3_C, S2_CAG6_3_HP, S2_4, S2_5, S2_7, S2_8, S2_CTG6_7, S2_CTG6_8, S2_9, S2_CTG6_10) were phosphorylated under the same conditions, but the reaction was run in T4 DNA ligase buffer (50 mM Tris-HCl, 10 mM MgCl_2 , 1 mM ATP, 10 mM DTT, pH 7.5), to provide ATP for the phosphorylation. The phosphorylated DNA, in a 1.1:1 molar ratio of phosphorylated DNA to the appropriate scaffolds, was re-suspended to a final volume of 100 μL in Tris buffer (10 mM Tris-HCl, 100 mM NaCl, pH 7.5), heated at 95 $^{\circ}\text{C}$ for 5 min and slowly cooled to 25 $^{\circ}\text{C}$ at a rate of 1 $^{\circ}\text{C}/\text{min}$ to form duplex.

After annealing was complete, the volume was increased to 170 μL with water and 20 μL fresh 10X ligase buffer was added. 10 μL of T4 DNA ligase was added to the annealed DNA, and the reaction was allowed to proceed at 16 $^{\circ}\text{C}$, overnight. The ligase

1: TTCGAACACTGTTACTCCTAATTTTAGAAGGCTTCAAGACGAAGTAGTA
2: GTAAATTAATAGTTGAAGTTGTAGTAAATGTTAATGTAGATCTGTTGTTCCGATATTA
3_C: CCAAACCTTCACCTAATAGCGCTAGTACA
4_HP: CCAAACCTTCACCTCAGCAGCAGCAGCAGCAGCAGCAGCAGAATAGCGCTAGTACA
5: CAGGAGAAGGACATGAACATGAACCTAATGAACACACAACAAATATGTAAGTGCCCATG
5: ATAGTGTGCGATCCTCACCAAGAATTATATCGTCCGCAAACCTTAACTCA
ScaffoldA: CCTTCTCCTGTGTACTAGCGCTATTAGGTGAAGGTTTGGTAATATCGGA
ScaffoldB: ACTACAACCTCAACTATTAATTTACTACTACTCTCGTCTTGAAGACCTGG
ScaffoldC: ATTCTTGGTGAGAGTCGCACACTCATATGGGCATCATACATATTTGTTGT



6: TGAGTTAAAGTTTTCGGGACGATATAAATCTTGGTGAGGATGCGACACTAT
7: CATGGGCACATTACATATATTTGTTGTTCATTAGGTTTCATGTTTCATGTCCTTCTCTGTGTACTAGCGCTATT
8: AGGTGAAGGTTTGGTGAATATCGGAACAACAGATCTACATTAACATTTCTACTACAATTCACATATTAATTTAC
9: TACTACTTCGTCCTTGAAGACCTTTAAAATAGAGGTAACAGTTTCGAA
ScaffoldD: TCCGATATTACCAAAACCTTCACCTAATAGCGGTAGTATACACAGGAGAAGG
ScaffoldE: AGAAGGTTCTCAAGACGAAGTAGTAGTAATTAATAGTTGAAGTTGTAGT
ScaffoldF: ACAACAAATAATGTAAAGTGCCCATGATAGTGTGCGCATCTTCAACGAAT



1: TTCGAACACTGTTACTCCTAATTTTAGAAGTCTTCAAGACGAAGTAGTA
2: GTAAATTAATAGTTGAAGTTGTAGTAAATGTTAATGTAGATCTGTTGTTCCGATATTA
3_CAG6_C: CCAAAACAGCAGCAGCAGCAGCAGCAGTACA
3_CAG6_HP: CCAAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTACA
4: CAGGAGAAGGACATGAACATGAACCTAATGAACACACAAATAATGTAAGTGCCCATG
5: ATAGTGTGCGCATCCTCACCAAGAATTATATCGTCCGCAAACTTTAACTCA
ScaffoldA_CTG6: CCTTCTCTGTGTACTAGCGCTATTAGGTGAAGGTTTTGGTAATATCGGA
ScaffoldB: ACTACAACTTCAACTATTAACTTACTACTACTTCGCTTGAAGACCTCT
ScaffoldC: ATCTCTGGGTAGGATGCGACATTACTTATGGGCATCTACATTATTTGTTGT



6: TGAGTTAAAGTTTGGCGGACGATATAATTCTTGGTGAGGATGCGACACTAT
7_CTF6: CATGGGCACTTACATTAATTGTTGTGTTCATTAGGTTCAATGTTATGTCCTTCTCTG
10_CTF6: TGTACTCTGCTGCTGCTGCTGCTGTTTGG
8_CTF6: TAATATCGGAACAACAGATCATACATTAACATTACTACAACCTAACTATTAATTAC
9: TACTACTTCGTCTTGAAGACCTTCTAAAATTAGGAGTAACAGTGTTCTGAA
ScaffoldD_CAG6: TCCGATATTACCAAAACCTTCACCTAATAGCGCTAGTACACAGGAGAAGG
ScaffoldE: AGAAGGCTTCAAGCAGAGTAGTAGTAATTAATAGTCTGAAGTTGAGG
ScaffoldF: ACAACAAATAATGTAAAGTGCCCATGATAGTGTCGATGCTACCAAGAAT



119

was heat inactivated by incubation at 65 °C for 20 minutes, and then the reaction was dried under vacuum.

The entire ligation reaction was then applied to a 6% denaturing polyacrylamide gel and electrophoresed for 3 h at 65 W (0.4 mm thick, 45 mm long). The full-length ligated product was then excised from the gel, crushed, soaked overnight in buffer (10 mM Tris, 1 mM EDTA, pH 7.5), centrifuged, and the supernatant collected. The pelleted gel was checked for radioactivity, and resuspended for another overnight soak. The second supernatant was also collected and pooled with the first. Samples were then lyophilized for ~24 h, resuspended in 2 mL of 10 mM Tris buffer, and placed in dialysis tubing (MWC 8 kDa). These samples were dialyzed against 2 L of 10 mM Tris for at least 24 hours, with one change of buffer.

After dialysis, samples were transferred to eppendorf tubes and dried *in vacuo*. Samples were resuspended in 48 µL of NEB buffer 3 (100mM NaCl, 50mM Tris-HCl, 10mM MgCl₂, 1mM DTT, pH 7.9) and 2 µL of Calf Intestinal Phosphatase (CIP). Samples were de-phosphorylated overnight at 37 °C. The enzyme was deactivated by incubation at 90 °C for 10 min. The de-phosphorylated samples were then cleaned using Qiagen QIAquick PCR purification kit. Treatment with CIP removed ~90% of the radiolabel from each DNA substrate. After elution from the QIAquick spin column, the total amount of ligated product was determined using a NanoDrop spectrophotometer.

4.2.3 Competitive nucleosome incorporation assay

Ligated oligonucleotides were 5'-³²P end-labeled using T4 PNK following the manufacturer's protocol. Radiolabeled DNA, along with a 1.5-fold excess of its complement, was resuspended to a final volume of 60 µL in Tris buffer (10 mM Tris-

HCl, 100 mM NaCl, pH 7.5), heated at 95 °C for 5 min and slowly cooled to 25 °C at a rate of 1 °C/min to form duplex (Figure 4.2).

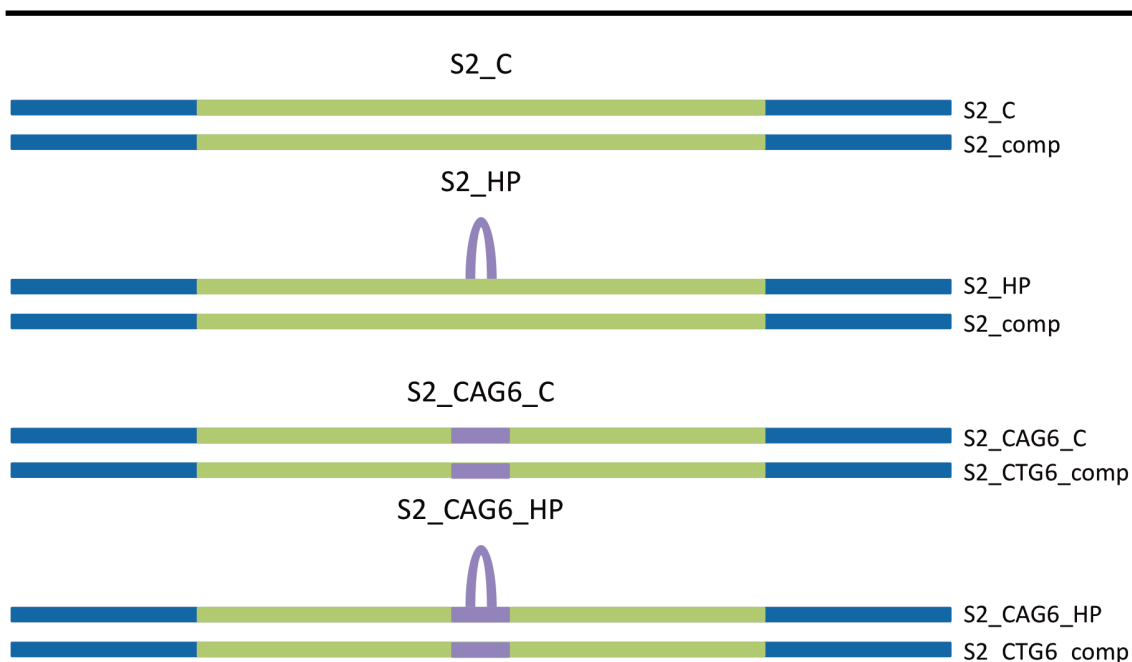


Figure 4.2 Schematic representations of the duplex DNA substrates used to investigate the affect of hairpin formation on chromatin packaging. Each of the ligated constructs was annealed with its complement to form the duplex substrates used in the competitive nucleosome exchange assays and the enzymatic footprinting. The location of the S1 positioning sequence is shown in green and CAG/CTG repeats are shown in purple.

Competitive nucleosome incorporation reactions were performed as described previously, except that the exchange buffer was changed to 5 mM Tris, pH 7.5, and due to the low incorporation of the hairpin containing DNA, the amount of competitor calf-thymus DNA was reduced to 1 µg. 30 µL of the final incorporation reaction was mixed with 6 µL non-denaturing dye (15% ficoll, 0.25% bromophenol blue (w:v), 0.25% xylene cyanol (w:v)), and 15 µL was applied to a 5% native polyacrylamide gel (8 mm thick, 16 cm long). The samples were electrophoresed at 150 V for 4 hours in 4 °C and the products were visualized by phosphorimagery. The ratio of incorporated DNA to free DNA was determined by comparing the number of counts in each band.

4.2.4 Enzymatic probing reactions

Exo III and DNase I reactions were carried out as described in the previous chapter, with a few small changes. Briefly, the histone core was exchanged onto the S2 DNA duplexes in exchange reactions containing 25 µg of NCPs for the S2_C and S2_CAG6_C DNA substrates and 75 µg of NCPs for the S2_HP and S2_CAG6_HP DNA substrates. The samples were incubated with either 50 units of Exo III (New England Biolabs) or 0.4 units of DNase I (New England Biolabs) for 5 min at 37 °C. The reaction was quenched by the addition of 0.1 M EDTA and the samples were immediately applied to a native polyacrylamide gel. After recovering the DNA, 6000 counts per sample were electrophoresed at 65 W for 2.5 h in a 6% denaturing polyacrylamide gel (0.4 mm thick, 65 cm long). After drying, the gel was used to expose a phosphor plate for 120 h. Reactions were also performed on free duplex, as described previously.

4.3 RESULTS AND DISCUSSION

4.3.1 CAG hairpins reduce the level of nucleosome formation

To determine the effect of hairpin formation of chromatin assembly, we first performed competitive nucleosome incorporation assays, in which radiolabeled substrate and unlabeled competitor DNA are exchanged onto nucleosomes isolated from chicken erythrocytes.^{208,275} The amount of radiolabeled substrate incorporated into a nucleosome is then determined and compared to the amount that remains as free duplex. These values, called the incorporation ratio, can be compared across different samples and provide information on how well a given substrate forms a nucleosome.

The first set of substrates, S2_C and S2_HP, contain the nucleosome positioning sequence S1^{209,210} in the center of the 246 bp construct (Figure 4.2). Like the S1 sequence, the S2 sequence is also derived from the pD89 sequence, which comes from the *GLN3* promoter. Based on previous characterization of the S1 positioning sequence shown in chapter 2, the (CAG)₁₀ hairpin was placed in the center of the duplex, where it was predicted to extend away from the histone core in a nucleosome formed at the S1 positioning sequence. Interestingly, when S2_C is used in the competitive nucleosome exchange assay, nucleosomes form at two positions along the duplex, evidenced by the appearance of two distinct higher order products; the slower migrating species is labeled C_S and the faster migrating species C_F (Figure 4.3). The electrophoretic mobility of

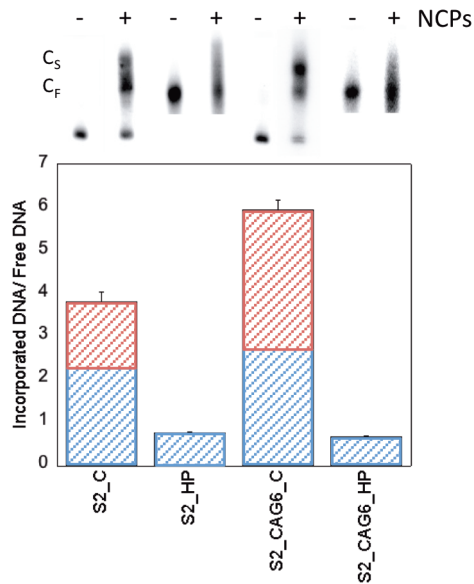


Figure 4.3 Nucleosome incorporation ratios for the four S2 substrates. The incorporation ratio for each substrate is shown in the chart with a representative incorporation performed in the absence (–NCPs) and presence of NCP (+NCPs) shown above. The amount of DNA associated with the faster migrating band (bottom, C_F) is shown in blue and the amount of DNA associated with the slower migrating band (top, C_S) is shown in red.

nucleosomes is determined by the location of the histone core relative to the center of the associated DNA; when the histone core is located toward the center of the DNA, the nucleosome has a slower electrophoretic mobility, while a histone core located toward either end of the DNA yields a nucleosome with a faster electrophoretic mobility.²⁷⁶ Thus, S2_C forms nucleosomes with at least two different translational positions.

While S2_C is composed entirely of well-matched base pairs, S2_HP contains a (CAG)₁₀ hairpin in the center of the construct, forming a three-way junction (Figure 4.2). When the S2_HP substrate is used in a competitive nucleosome incorporation assay, there is a dramatic decrease in the amount of radiolabeled substrate incorporated into a nucleosome (Figure 4.3). While two clear NCP bands appear in the S2_C competitive nucleosome incorporation assay, the NCPs formed with the S2_HP DNA barely form a single distinct band; they run almost as a tail off the duplex band. However, this region of radioactivity is not present in the assay performed in the absence of nucleosomes, thus this region of slower migration must be due to the interaction of the S2_HP DNA with a histone core.

We have previously investigated the behavior of a (CAG)₁₀ hairpin in a similar context.⁷⁰ The previous construct, described in chapter 3, was formed by annealing an oligonucleotide containing a (CAG)₁₆ with an oligonucleotide containing a (CTG)₆, forming a three-way junction where 6 of the CAG repeats are in well matched base pairs and the remaining 10 CAG repeats form a hairpin. Due to the repetitive nature of the DNA, the CAG hairpin can form at many positions along the complementary CTG repeat, providing some flexibility in the location of the hairpin within the three-way junction. Thus, we were interested in determining what effect that added flexibility would

have on the ability of the hairpin containing DNA to form a nucleosome. Starting from the S2 sequence, we replaced the center-most bases with 6 CAG/CTG repeats, creating the S2_CAG6_C substrate (Figure 4.2). We then added an additional 10 CAG repeats to the CAG-containing stand, creating the S2_CAG6_HP substrate.

When the S2_CAG6_C substrate was used in the competitive nucleosome incorporation assay, there is a significant increase in the amount of incorporation observed compared to S2_C (Figure 4.3). This increase is due to the presence of the 6 CAG/CTG repeats, the only difference between the S2_C and S2_CAG6_C substrates. We have previously observed increased incorporation ratios for 146 bp duplexes containing as few as 10 CAG/CTG repeats, and the increase in incorporation ratio observed here is likely due to similar forces.²⁷⁵ Interestingly, the majority of the observed increase in nucleosome formation seems to come specifically from formation of the slower-migrating species, C_s, the species that contains the histone core at the center of the DNA. This species likely places the CAG/CTG repeats at the dyad axis, a region of the nucleosome for which CAG/CTG repeats possess a high affinity, as discussed in chapter 2. Thus it seems quite likely that the increase in nucleosome formation in the S2_CAG6_C substrate is primarily due to the affinity of the CAG/CTG repeats for the dyad axis.

Despite the observed increase in incorporation ratio for the S2_CAG6_C substrate due to the presence of the short CAG/CTG repeat tract, there is no increase in the incorporation ratio for the S2_CAG6_HP substrate compared to the S2_HP substrate (Figure 4.3). Even though the CAG/CTG repeats that benefit nucleosome formation in S2_CAG6_C DNA are also present in the S2_CAG6_HP substrate, the presence of the

hairpin must destabilize the nucleosome complex to such an extent that the addition of the CAG/CTG repeats provides no benefit to nucleosome formation. Thus we can conclude that, even given the presence of the beneficial CAG/CTG repeats and the ability of the (CAG)₁₀ hairpin to adopt multiple positions relative to the 6 CTG repeats, the presence of the hairpin itself remains extremely disruptive to nucleosome formation. As CAG/CTG repeats prefer to associate with the histone core,²⁷⁵ formation of a CAG hairpin within the repeat tract may translate to the unpacking of CAG/CTG repeats that would normally be packaged in a nucleosome, altering the functional organization of the repeat tract.

While no previous work has specifically investigated the incorporation of hairpin containing DNA into nucleosomes, analysis of nucleosome formation in the genome by another non-canonical DNA structure, the G-quadruplex, has been performed.²⁷⁷⁻²⁷⁹ A G-quadruplex is a stacked series of rings, called a G-quartet, composed of four guanines each forming four hydrogen bonds with its neighboring guanines. G-quadruplexes are proposed to form throughout the genome, primarily at telomeres and gene promoters.²⁷⁹ Genome-wide computational analyses of several sets of genomic nucleosome occupancy data indicate that regions conducive to forming G-quadruplexes are mostly found outside of nucleosomes, likely because the G-quadruplex would destabilize the interaction between the DNA and histone core.^{277,278} Thus, it is likely that hairpins will exhibit the same behavior within the packaged genome and would be much less likely to form a nucleosome, consistent with the competitive nucleosome incorporation data presented here.

4.3.2 While the presence of 6 CAG/CTG repeats does not alter the translational positioning of DNA in a nucleosome, the presence of a (CAG)₁₀ hairpin does

While the migration of nucleosome samples through a native gel provides general information about the position of the histone core, Exo III can provide more detail about the translation positioning of the DNA. As Exo III digests the DNA, it pauses when it encounters DNA interacting with the histone core, leading to an area of intense strand cleavage.²²⁸⁻²³⁰ As expected, when Exo III digests free duplex, there is reactivity along the entire duplex (Figure 4.4, 4.5). However, there is a clear change in reactivity once the free duplex is incorporated into a nucleosome.

When the S2_C DNA is incorporated into a nucleosome, it forms a population of at least two nucleosomes with different translational positions (Figure 4.3). However, as the Exo III digests away the DNA flanking the histone core, the two species of nucleosomes become homogenized, and what were previously two distinct populations of nucleosomes begin to migrate together in the native gel. Thus, the sample applied to the denaturing gel represents a combination of Exo III reaction on both species of nucleosome. However, using both the migration of each undigested species in the native gel and the location of Exo III pause sites, we can assign the position of the histone core in both species.

The slower-migrating species of S2_C nucleosome (S2_C_S) has a histone core positioned in the center of the DNA, and the pattern of Exo III pause sites is consistent with the histone core occupying the S1 positioning sequence, located in the middle of the S2_C duplex, the center of which is indicated by a black bar (Figure 4.4A). Based on the

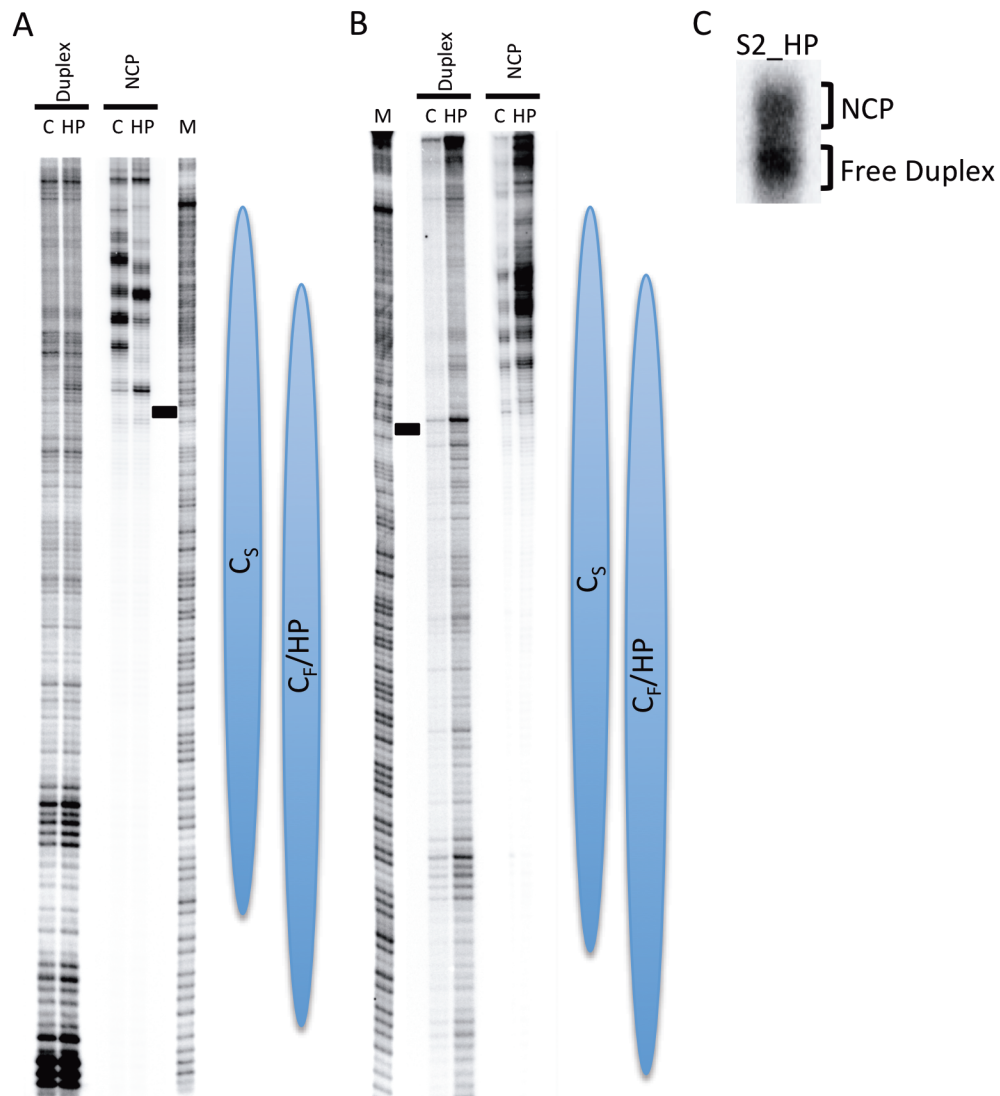


Figure 4.4 Exonuclease III digestion reveals the translation position of the DNA around the histone core. Exo III digestion of the S2_C (C) and S2_HP (HP) substrates is shown in panel A, while digestion of the S2_CAG6_C (C) and S2_CAG6_HP (HP) substrates is shown in panel B. In these reactions, the complementary strand of the duplex is radiolabeled. Reactions were performed on both free duplex and nucleosome (NCP) samples. The marker lane (M) consists of the Maxam-Gilbert A/G sequencing reaction performed on single stranded S2_C, and the dark line next to the marker lane indicates the position of the hairpin. The blue ovals indicate the position of the histone core for each substrate. Panel C shows the migration of the S2_HP NCP and free duplex in the native gel after Exo III treatment.

migration of the faster-migrating species of S2_C nucleosome (S2_C_F), the histone core is positioned toward one end of the DNA, but it is not clear from the migration which end contains the nucleosome. However, there appears to be significant digestion from the 3'

end to almost the center of the DNA duplex in the Exo III-treated S2_C sample applied to the denaturing gel in Figure 4.4. In order to observe Exo III digestion this far into the duplex, the 3' end of the DNA must be readily accessible to the enzyme in some

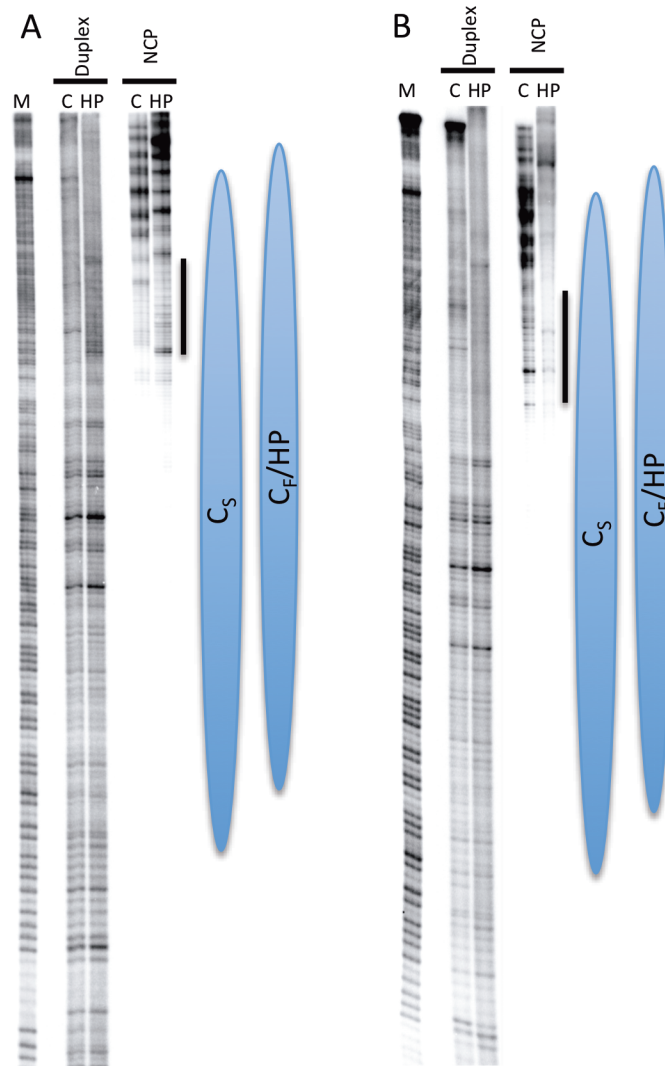


Figure 4.5 Exonuclease III digestion reveals the translation position of the DNA around the histone core. Exo III digestion of the S2_C (C) and S2_HP (HP) substrates is shown in panel A, while digestion of the S2_CAG6_C (C) and S2_CAG6_HP (HP) substrates is shown in panel B. In these reactions, the C or HP strand of the duplex is radiolabeled. Reactions were performed on both free duplex and nucleosome (NCP) samples. The marker lane (M) consists of the Maxam-Gilbert A/G sequencing reaction performed on single stranded S2_C, and the dark line next to the marker lane indicates the position of the hairpin. The blue ovals indicate the position of the histone core for each substrate.

population of nucleosomes. Thus, it is likely the S2_C_F species which produces the extended digestion pattern, indicating that the histone core is positioned near the radiolabel, which in these substrates is located on the 5'-end of the complementary strand (Figure 4.4A). The assigned locations of the histone core for the S2_C_S and S2_C_F nucleosomes are consistent with the Exo III digestion observed when the S2_C strand is radiolabeled (Figure 4.5A).

As observed for the incorporated S2_C DNA, the S2_CAG6_C DNA also forms two species of nucleosomes, and like the S2_C_S and S2_C_F species, they cannot be distinguished in the native gel after Exo III treatment. However, using the same analysis applied to the S2_C nucleosome species, we can determine the position of the histone core in both the slower migrating S2_CAG6_C_S nucleosome and the faster migrating S2_CAG6_C_F nucleosome. The histone core associated with the S2_CAG6_C_S species is positioned in the center of the DNA, and again the Exo III digestion suggests it is located at the S1 positioning sequence (Figure 4.4B). Furthermore, by placing the CAG/CTG repeats at the dyad axis, this positioning best explains the increase in incorporation ratio between the S2_C and S2_CAG6_C substrates.

When the complementary strand is radiolabeled, the Exo III treated S2_CAG6_C sample also shows extensive digestion from the 3' end, as observed for the S2_C substrate (Figure 4.4). The digestion is consistent with a histone core positioned near the radiolabel, at the same location as observed for the S2_C_F substrate. The proposed nucleosome positions for S2_CAG6_C_S and S2_CAG6_C_F derived from Exo III digestion of samples in which the complementary strand is radiolabeled are consistent with the Exo III digestion of samples where the S2_CAG6_C strand is radiolabeled

(Figure 4.5B). Thus, the two species of nucleosome formed by the S2_CAG6_C DNA occupy the same positions as the nucleosomes formed by the S2_C DNA, indicating that the presence of the 6 CAG/CTG repeats does not affect the translational positioning of either nucleosome species.

While it was difficult to resolve S2_HP and S2_CAG6_HP DNA incorporated into a nucleosome from DNA that remained as free duplex during the competitive nucleosome incorporation assay, by increasing the amount of donor nucleosomes and removing the competitor DNA, we were able to better distinguish between the two bands (Figure 4.4C). Interestingly, the S2_HP and S2_CAG6_HP nucleosome species also display extended Exo III digestion from the 3' end when the complementary strand is radiolabeled, similar to Exo III digestion assigned to the S2_C_S and S2_CAG6_C_S nucleosomes (Figure 4.4). Here it is beneficial to initially consider digestion of samples where the complementary strand is radiolabeled, due to the difference in length between the S2_C and S2_HP strand, which can make interpretation of the Exo III digestion pattern more complex. The location of the hairpin is indicated by the presence of a black bar, marking the bases that flank the hairpin arm in the three-way junction. Because of the similarities in the amount and location of Exo III digestion, it is likely that the S2_HP and S2_CAG6_HP DNA forms nucleosomes at the same location found in the S2_C_F and S2_CAG6_C_F nucleosomes, positioning the hairpin at the 5' end of the 146 bps forming the nucleosome, near the edge of the ~20 bps that have only transient contact with the histone core, suggesting that the hairpin containing DNA can form a nucleosome, but the translational position must be such that the hairpin is not in tight contact with the histone core. Thus, even though we designed these substrates to position the hairpin at the dyad

axis, with the hairpin facing away from the histone core, formation of a nucleosome at that location is very much disfavored.

In a small population of S2_HP and S2_CAG6_HP nucleosomes, there is additional digestion of the duplex, with a pause site occurring just 3' to the (CAG)₁₀ hairpin, the location of which is indicated by the black bar in Figure 4.4. The pause site is less pronounced in the S2_CAG6_HP nucleosomes; this lack of clear reactivity may be due either to the difference in sequence context or to the ability of the hairpin to adopt multiple positions relative to the (CTG)₆ in the complementary strand. When the S2_HP or S2_CAG6_HP strand is radiolabeled there is also evidence that some population of nucleosomes give rise to additional digestion at the 5' edge of the hairpin, the location of which is indicated by the black line in Figure 4.5. The increased ability of Exo III to digest this sub-population of the S2_HP and S2_CAG6_HP nucleosomes indicates that the interaction between the region of the DNA containing the hairpin and the histone core is not always stable and may in fact induce additional unwrapping of the DNA from the histone core. These nucleosomes would be far more dynamic than those formed by simple duplex DNA, providing a possible explanation for why the incorporation ratio is low for the hairpin containing substrates, and for why these nucleosomes do not migrate as a clear band in the competitive nucleosome exchange assay.

4.3.3 The presence of a (CAG)₁₀ hairpin alters the rotational position of DNA in a nucleosome

While digestion with Exo III allows us to determine the translational positioning of the DNA in a nucleosome, we can also use the enzyme DNase I to investigate the rotational position of the DNA in a nucleosome.²¹⁹⁻²²² While DNase I does not possess

any strict sequence preference, it does prefer some stretches of DNA to others, based mainly on minor groove width^{223,224} and DNA flexibility.^{225,226} Thus, when free duplex is treated with DNase I, we do not observe equal reactivity along the entire duplex (Figure 4.6, Figure 4.7, Duplex). However, as observed with the Exo III reactions, when the duplex is incorporated into a nucleosome there is a clear change in reactivity: upon incorporation there is an oscillation to the reactivity in which regions with high levels of cleavage are followed by regions with little or no reactivity (Figure 4.6, Figure 4.7, NCP). As discussed previously, where the DNA backbone faces out into solution, the DNase I can recognize and cleave the DNA, but where the DNA backbone faces toward the histone core, the DNA is inaccessible to the enzyme, thus mapping the rotational position of the DNA.

Unlike treatment with Exo III, when the S2_C nucleosomes are treated with DNase I, it is possible to resolve the two separate species of nucleosome formed by the DNA substrate. This is because the two enzymes have a different mechanism of action; the Exo III digests the DNA from the 3' end, removing multiple nucleotides in succession, whereas the DNase I cleaves the DNA backbone, inducing only a single strand break. The presence of a single strand break does not alter the migration of the nucleosome in a native gel, and thus the two species can be analyzed separately. When the S2_C nucleosomes are treated with DNase I it becomes apparent that the S2_C_S and S2_C_F species have slightly different rotational positions, as judged from the regions of the DNA that are highly reactive to the enzyme (Figure 4.6A, compare the pattern of reactivity in NCP lane C_S and C_F). The differences in rotational position are also apparent

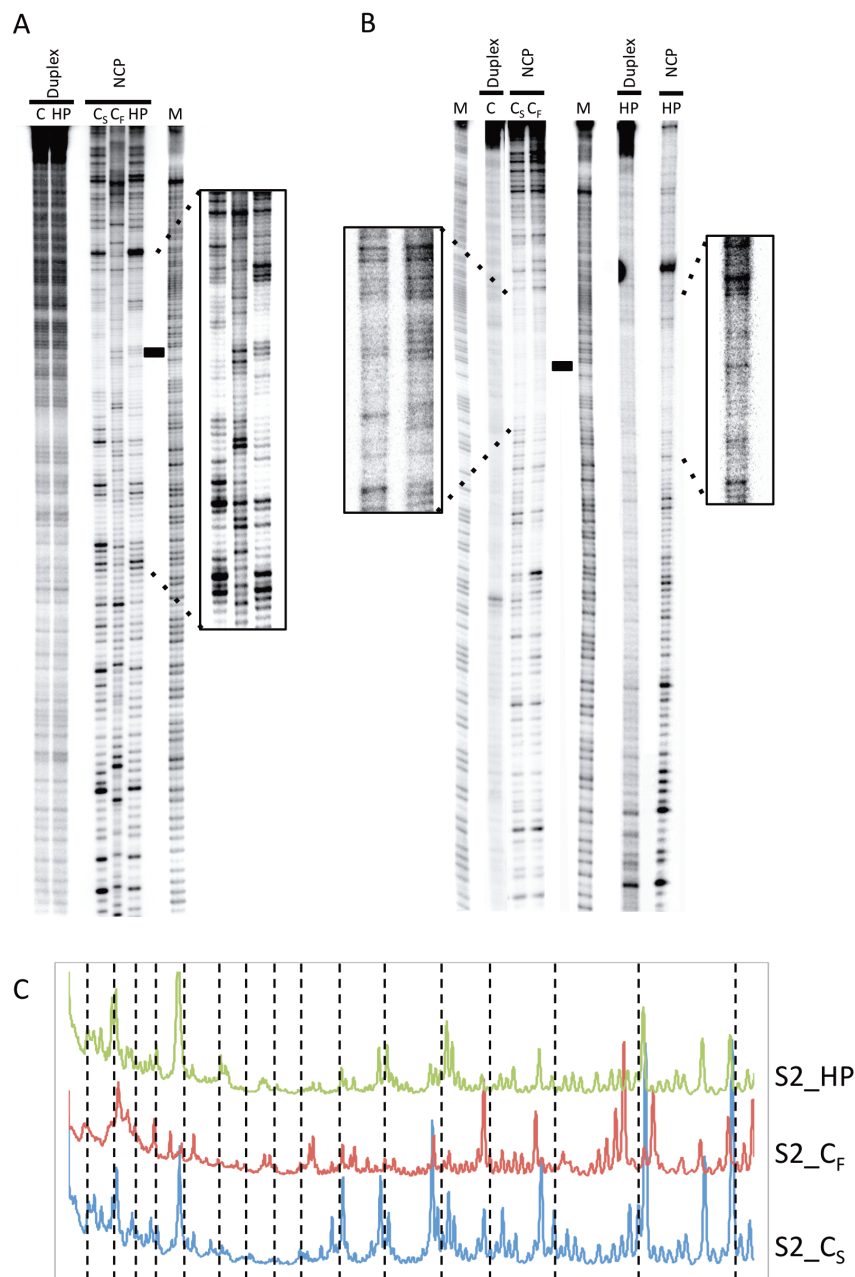


Figure 4.6 DNase I digestion reveals the rotational positioning of the S2 substrates in the nucleosome. DNase I cleavage of the S2_C and S2_HP substrates is shown in panel A, while digestion of the S2_CAG6_C and S2_CAG6_HP substrates is shown in panel B. The data shown in panel B are from two separate images. Reactions were performed on both free duplex and nucleosome (NCP) substrates. In these reactions, the complementary strand of the duplex is radiolabeled. The inserts consist of a magnified view of the region surrounding the hairpin. While the image is taken from the same data, the contrast has been increased to highlight the cleavage pattern. The marker lane (M) consists of the Maxam-Gilbert A/G sequencing reaction performed on single stranded S2_C, and the dark line next to the marker lane indicates the position of the hairpin.

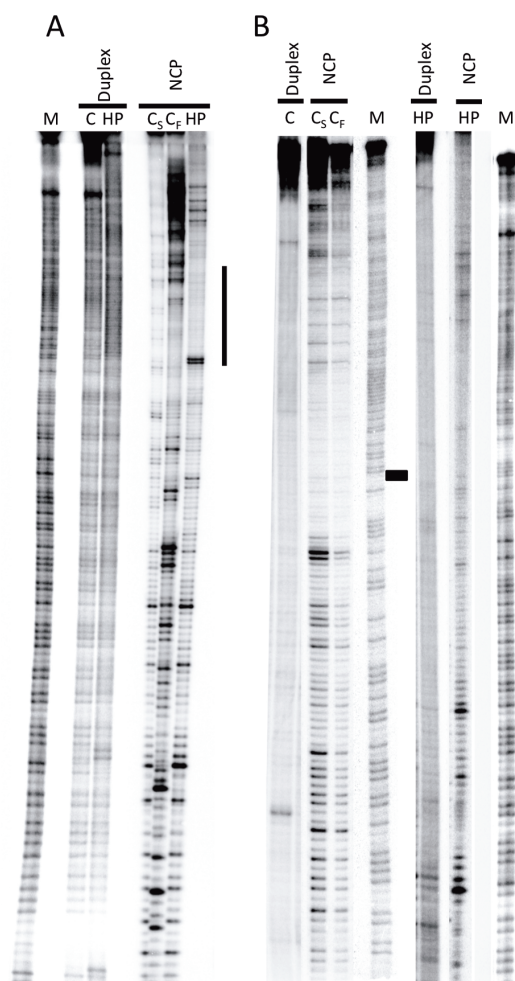


Figure 4.7 DNase I digestion reveals the rotational positioning of the S2 substrates in the nucleosome. DNase I cleavage of the S2_C and S2_HP substrates is shown in panel A, while digestion of the S2_CAG6_C and S2_CAG6_HP substrates is shown in panel B. The data shown in panel B are from two separate images. Reactions were performed on both free duplex and nucleosome (NCP) substrates. In these reactions, the C or HP strand of the duplex is radiolabeled. The inserts consist of a magnified view of the region surrounding the hairpin. While the image is taken from the same data, the contrast has been increased to highlight the cleavage pattern. The marker lane (M) consists of the Maxam-Gilbert A/G sequencing reaction performed on single stranded S2_C, and the dark line next to the marker lane indicates the position of the hairpin.

in the histograms shown in Figure 4.6C, which are derived from the gel shown in Figure 4.6A. The dashed lines indicate where the DNA backbone of the S2_C_S substrate exhibits the highest amount of cleavage by DNase I, marking the regions where the DNA backbone faces out into solution. Shifts in the rotational position can be observed by

comparing the maxima of DNase I reactivity of the other substrates to these dashed lines. Interestingly, the S2_HP nucleosome appears to possess an overall rotational position more similar to the S2_C_S nucleosome, even though they do not share the same translational position. However, upon examination of the region directly surrounding the hairpin, it becomes apparent that the areas of the DNA duplex with heightened reactivity to DNase I are slightly different in all three nucleosome species, indicating that the (CAG)₁₀ hairpin locally alters the rotational position of the DNA in the S2_HP nucleosome (Figure 4.6A NCP, magnified insert). The changes in rotational position are modeled in Figure 4.8. The position of the base on the complementary strand in the base pair immediately 5' to the hairpin is shown as a black line, relative to the histone core. In substrate the base is facing out into solution. Importantly, the S2_HP nucleosome possesses a high level of reactivity with the DNase I in the complementary strand slightly

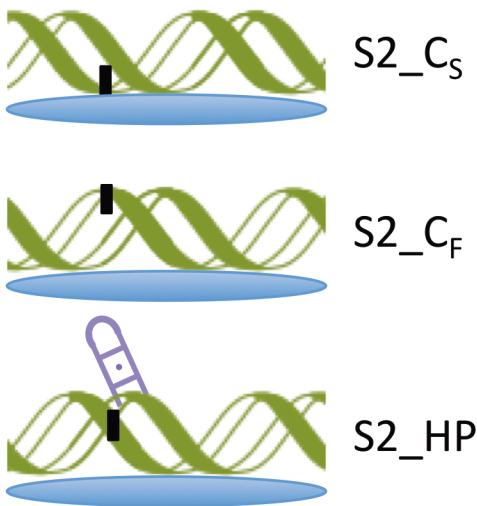


Figure 4.8 Alteration of rotational position in the nucleosome by the (CAG)₁₀ hairpin. The black line indicates the base on the complementary strand (S2_comp) in the base pair immediately 5' to the hairpin. The position of this base relative to the histone core (blue) is different in each substrate. Importantly, the rotational position of the S2_HP substrate places the hairpin (purple) in a location where it faces away from the histone core.

3' to the hairpin (Figure 4.6A, magnified insert), indicating that the hairpin is facing straight out from the histone core, in what is likely the most favorable position (Figure 4.8). Interestingly, this position could make the loop of the hairpin readily accessible to oxidants, and possibly initiate or continue the toxic oxidation cycle.

Previous research has used well-matched double-stranded flaps and hairpins placed in the *Xenopus* 5S positioning element as barriers to chromatin remodeling enzymes such as SWI/SNF and Mi-2.^{280,281} The hairpin substrates were placed near the center of the DNA incorporated into the nucleosome, placing the hairpin at the dyad axis, and designed such that the hairpin faced away from the histone core upon assembly. DNase I and/or hydroxyl radical footprinting were used to investigate the rotational position of the DNA in the nucleosome. These hairpins showed no ability to alter the rotational position of the DNA. Thus the observed alteration of rotational positioning by the (CAG)₁₀ hairpin may be a feature of CAG repeat hairpins, or a hairpin within the S2 DNA context.

When the S2_CAG6_C and S2_CAG6_HP nucleosomes are reacted with DNase I, the same oscillating pattern of reactivity occurs. However, the DNase I does not react strongly with CAG/CTG repeats, making the region surrounding the hairpin difficult to interpret. Indeed, due to the poor reactivity we are unable to determine whether the hairpin locally alters the rotational position in this context. Yet based on the reactivity of DNase I with the non-CAG/CTG repeat DNA, it is possible to determine that the S2_CAG6_C_S, S2_CAG6_C_F, and S2_CAG6_HP nucleosomes have the same overall rotational position, though they do not necessarily share the same translational position (Figure 4.6B, compare lanes NCP, C_S, C_F, and HP). This may be due to the positioning

effect of the 6 CAG/CTG repeats. We have previously demonstrated that as few as 10 CAG/CTG repeats positioned at the dyad axis can exert an effect over the positioning of DNA in a nucleosome, and it is possible that the 6 CAG/CTG repeats are exerting the same effect in this system.²⁷⁵

5.4 CONCLUSIONS

In this work we demonstrate how destabilizing a hairpin can be to nucleosome assembly. When a (CAG)₁₀ hairpin is included in the S2 sequence, there is a dramatic decrease in the ability of the DNA to form a nucleosome. Even if an additional 6 CAG/CTG repeats replace the DNA flanking the hairpin, a change which provides an increase in nucleosome formation for the well-matched duplex control, the amount of nucleosome formation by the hairpin containing substrate is still quite low. Using enzymatic probing to characterize the structure of the nucleosomes formed, we find that the addition of a hairpin to the S2_C or S2_CAG6_C substrates alters the translational positioning of the DNA; while the S2_C and S2_CAG6_C substrates form nucleosomes with two different translational positions, the hairpin-containing substrates occupy only a single translational position, eliminating the translational position where the hairpin is placed at the dyad axis. The translational position of the hairpin-containing substrates indicates that the hairpin prefers to occupy a location where it has transient interaction with the histone core, even though the alternative is to occupy a favorable positioning sequence. Furthermore, the hairpin is able to alter the local rotational position of the DNA such that the hairpin faces out into solution. While it seems likely that within the genome, DNA containing a hairpin will behave much like G-quadruplexes, excluding the

histone core from the region, should a CAG hairpin form while the DNA is already in a nucleosome, as might occur during LP-BER in a CAG/CTG repeat tract, the newly formed CAG hairpin may be able to alter the position of the DNA within the nucleosome, altering the functional organization of the repeat tract. Furthermore, since the hairpin will likely face out into solution, the histone core will provide little protection from oxidative damage and it is possible that the loop will be readily oxidized as part of the toxic oxidation cycle. Thus the formation of a hairpin within the CAG/CTG repeat tract could have a significant effect on both expansion and the functional organization of the *HTT* gene.

Significance: these results define the behavior of CAG repeat hairpins within the packaged genome. Furthermore, as these hairpins will be accessible within the packaged genome, either because they are positioned between nucleosomes during DNA packaging or face away from the histone core if formed by LP-BER, it might be possible to target them for removal.

Chapter 5

Future Directions and Ongoing Research: Investigating the Effects of Oxidative Damage on DNA Packaging and the Ability of TNRs to Modify Di-Nucleosome Structure

5.1 FUTURE DIRECTIONS

The results presented in the previous chapters provide a strong foundation for further exploration of TNRs within biological contexts. However, there are many more facets of both the expansion process and nucleosome packaging that have yet to be explored. We conducted the initial experiments on nucleosome formation using DNA substrates limited to 146 bps, the minimal length necessary to form a nucleosome, to ensure formation of a uniform population. However, it will be beneficial to conduct further experiments using longer DNA substrates, such as the S2 sequence used in chapter 4. Furthermore, in all of the substrates investigated here, the repeat was placed in the center of the DNA, providing placement at the dyad axis of the nucleosome. It will now be important to investigate the behavior of repeat tracts positioned off the dyad axis. Additionally, while we have so far focused on the DNA component of the nucleosome, it will also be beneficial to consider the histone components along with other epigenetic factors, such as DNA methylation. In the following sections, we describe new areas of research currently underway.

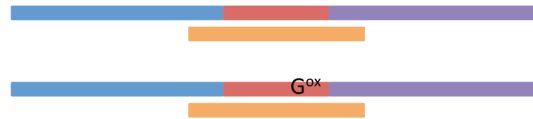
5.2 INVESTIGATING THE EFFECT OF 8-OXOG ON NUCLEOSOME FORMATION

Previous research has shown that the repair of the oxidative lesion 8-oxo-7,8-dihydroguanine (8-oxoG) within the CAG repeat tract is an initiating event for expansion within the cell.^{159,162,164} Some research is available on how 8-oxoG^{272,282} and other lesions^{270,271,273,274} repaired by the base excision repair (BER) pathway are repaired in a packaged context, but no information is available on how the lesion itself might alter the interactions between the histone core and the DNA. Thus, we have begun performing experiments meant to elucidate how the presence of an 8-oxoG lesion might alter the behavior of DNA in a nucleosome.

5.2.1 DNA substrate preparation

S1_C, S1_8oxoG

1:GTAAATTAATAGTTGAAGTTGTAGTAAATGTTAATGTAGATCTGTTGTTCCGATATTA
 2_C:CCAAAACCTTCACCTAATAGCGCTAGTACA
 2_8oxoG:CCAAAACCTTCACCTAATAG^{ox}CGCTAGTACA
 3:CAGGAGAAGGACATGAACATGAACCTAATGAACACAACAAATAATGTAAGTGCCCATG
 ScaffoldA:ACTACAACCTTCAACTATTAATTTACTACTACTTCGTCCTTGAAGACCTTCT



S1_comp

4:CATGGGCACTTACATTATTTGTTGTGTTTCATTAGGTTTCATGTTTCATGTCCTTCTCCTGTGTACTAGCGCTATT
 5:AGGTGAAGGTTTTGGTAATATCGGAACAACAGATCTACATTAAACATTTACTACAACCTTCAACTATTAATTTAC
 ScaffoldB:AGAAGGTCTTCAAGACGAAGTAGTAGTAAATTAATAGTTGAAGTTGTAGT



Figure 5.1 Primary sequences of the individual DNA oligonucleotides and schematic representations of the assembly process used to create the single stranded substrates used to investigate the effects of the 8-oxoG (G^{ox}) on nucleosome formation. Each oligonucleotide is color-coded; the color of the primary sequence matches the color in the schematic.

The oligonucleotides used to create the 146 bp S1 positioning sequence are the same as those used in creating the S2 DNA, described in chapter 4 (Figure 5.1). Each sequence was synthesized and purified as described previously (Chapter 4). To include

the 8-oxoG lesion in the final 146-mer, the most 5' G in S1_2 was replaced with the 8-oxoG lesion. This substrate was synthesized and purified as described previously (Chapter 3). These sequences were ligated as described in chapter 4 to create the single stranded S1_C, S1_8oxoG, and S1_comp strands (Figure 5.1). Each S1 substrate was constructed by annealing the appropriate oligonucleotides with DNA scaffolds and ligating the individual pieces.

S1_C and S1_8oxoG were each radiolabeled and annealed with an excess of the S1_comp strand. These DNA samples were then used for competitive nucleosome exchanges and hydroxyl radical footprinting as described previously. To observe reactivity to the hydroxyl radical on the complementary strand, S1_comp was radiolabeled and annealed with an excess of either the S1_C or S1_8oxoG strand.

5.2.2 Competitive nucleosome incorporation assay

Competitive nucleosome incorporations were performed as described in chapter 2. When the incorporation ratios for each substrate were compared, it is apparent that the 8-oxoG decreases the affinity of the S1 DNA for the histone core (Figure 5.2). Based on previous characterization of the S1_C DNA, the 8-oxoG lesion is predicted to be in a

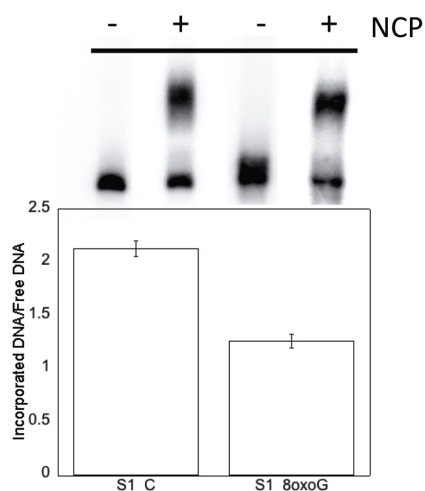


Figure 5.2 Nucleosome incorporation ratios for the S1_C and S1_8oxoG substrates. The incorporation ratio for each substrate is shown in the chart with a representative incorporation performed in the absence (–NCPs) and presence of NCP (+NCPs) shown above.

region of the DNA facing away from the histone core. However, without knowing how the presence of the 8-oxoG might alter the position of the DNA in the nucleosome, it is difficult to determine to what this decrease might be due.

5.2.3 Hydroxyl radical footprinting

Hydroxyl radical footprinting was carried out as described previously, except the incubation time for free duplex was shortened to 2 min and the incubation time for NCPs shortened to 5 min. Preliminary evidence suggests that the 8-oxoG has a localized, disruptive effect on the position of the S1 DNA in a nucleosome, as judged from the deviations in the local maxima and minima of reactivity with the hydroxyl radical between the two substrates. This change in reactivity indicates that the lesion not only lowers the affinity for the histone core, in this context an 8-oxoG lesion appears to alter the positioning of DNA in a nucleosome, though these experiments have not yet been repeated (Figure 5.3). Furthermore, the hydroxyl radical footprinting data indicates that the lesion is located in a region that faces away from the histone core. We predicted that

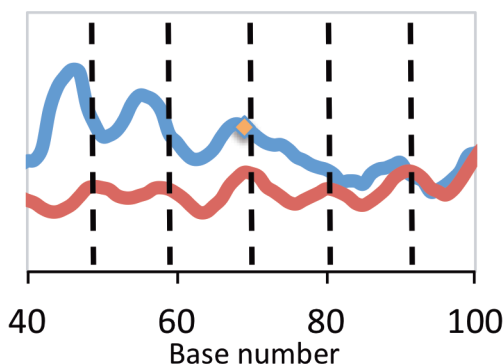


Figure 5.3 Hydroxyl radical footprinting reveals the position of the S1 DNA around the histone core. Reactions were performed on S1_C (red) and S1_8oxoG (blue) DNA substrates incorporated into nucleosomes. The location of the 8-oxoG is indicated with the orange diamond. The dashed lines indicate the maxima of the S1 substrate.

this position was likely to be the least disruptive to nucleosome formation, due to the lack of contact between the lesion and the histone core. However, the lesion is located near the H3 tail, which may account for changes in affinity for and positioning around the histone core. In addition to repeating the hydroxyl radical footprinting on the substrate described here, it will also be beneficial to investigate the results of lesion incorporation at other positions relative to the histone core. Luckily, the next two guanines in the S1_C sequence are located 2 and 6 bases away, providing a location where the lesion should be positioned midway between the histone core and the environment and a location where the lesion should face in toward the histone core. Further experiments should also include substrates containing CAG/CTG repeats, to determine how the lesion might affect the functional organization of the repeat tract.

5.3 ASSEMBLING DI-NUCLEOSOMES TO INVESTIGATE HIGHER ORDER CHROMATIN STRUCTURES

While it is important to understand the behavior of TNRs in a single nucleosome, the genes that contain the repeat tracts will not be occupied by only a single nucleosome. While the healthy-length repeat tract may occupy a single nucleosome, the entire gene will occupy many nucleosomes and will be further packaged into higher order chromatin structures. Furthermore, modification of a repeat tract in a yeast mini-chromosome has induced changes in nucleosome location that are not restricted to the region containing the repeat.^{196,197} To determine how packaging of the repeat tract might affect higher order chromatin formation, we have begun development of a plasmid-based system that will allow us to generate di-nucleosomes, and eventually multi-nucleosome arrays.

5.3.1 Oligonucleotide synthesis and purification

The S1 duplex and the respective primers was synthesized and purified as described in chapter 2. In order to insert two S1 sequences into the pUC19 vector, forward and reverse primers containing the appropriate restriction sites were designed for either end of the S1 duplex (Figure 5.4). In addition, to provide adequate distance between the two S1 sequences, 22 nucleotides were added to the BamHI 1 and BamHI 2 primers, so that the S1 sequences would be separated by 50 nucleotides, with the BamHI restriction site in the center.

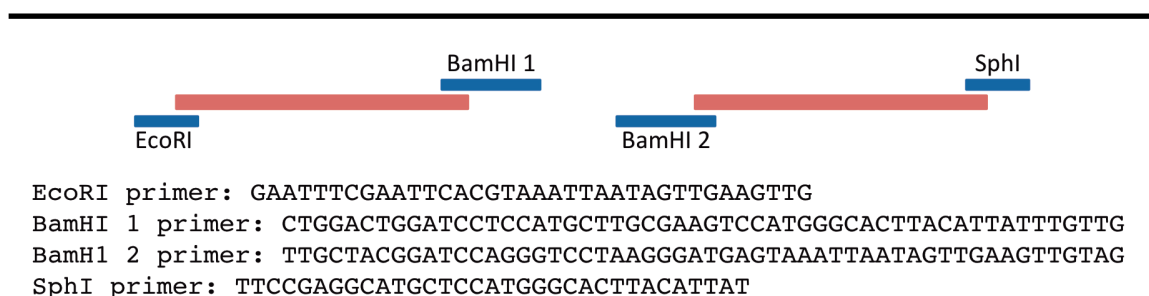


Figure 5.4 Primer design for construction of a di-nucleosome positioning sequence. Using the S1 positioning sequence (red), we constructed two sets of primers for PCR amplification (blue). The EcoRI and BamHI 1 primers were used in the first step of construction, inserting a S1 sequence into the pUC19 vector. The BamHI 2 and SphI primers were used to insert a second S1 sequence flanking the first. In order to provide sufficient space between the two S1 sequences, each BamHI primer contacted and additional 22 nucleotides of random sequence.

5.3.2 Inserting the first S1 sequence into the pUC19 vector

To create a di-nucleosome positioning sequence, S1 was first PCR amplified using the EcoRI and BamHI 1 primers. After PCR clean up with a Qiagen QIAquick PCR purification kit, both the S1+ (S1 plus the additional 22 nucleotides added to the BamHI 1 primer) insert and the pUC19 vector were digested with EcoRI and BamHI (NEB) overnight at 37 °C in NEB buffer 3 (100mM NaCl, 50mM Tris-HCl, 10mM MgCl₂, 1mM DTT, pH 7.9). After heat inactivation at 65 °C, the insert and vector were cleaned again to remove the restriction enzymes. The vector was then de-phosphorylated with calf

intestinal phosphatase (CIP, NEB) at 37 °C for 2 hours. After cleaning the vector to remove the CIP, the vector and insert were mixed at a 1:3 molar ratio and ligated overnight at 16 °C. 5 µL of the ligation was transformed into 50 µL of DH5-α cells, plated on LB plates containing 100 µg/mL ampicillin, and incubated overnight at 37 °C. Individual colonies were selected for overnight growth in 5 mL liquid LB plus 100 µg/mL ampicillin. These cultures were used in a QIAprep Mini-Prep Kit to isolate the vector now containing the S1+ insert. To ensure that the S1+ insert had been properly incorporated into the vector backbone, a diagnostic PCR was performed using the EcoRI and BamHI 1 primers. Constructs that have successfully incorporated the S1+ insert are now labeled pCBV_S1+ (Figure 5.5).

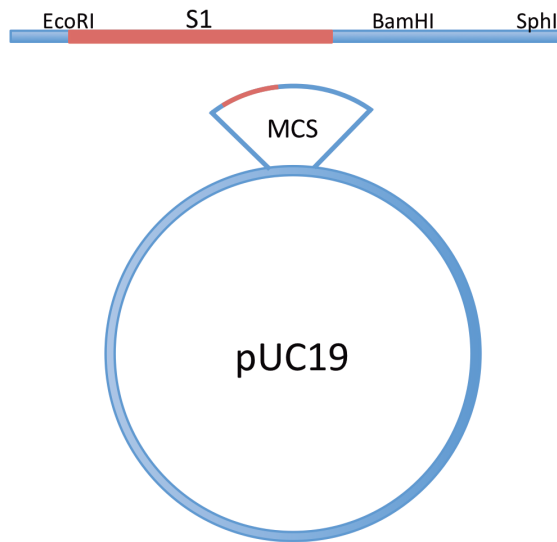


Figure 5.5 Plasmid map of pCBV_S1+. The S1 sequence (red) is inserted between the EcoRI and BamHI restriction sites within the multiple cloning site (MCS) of the pUC19 vector (blue).

5.3.3 Inserting the second S1 sequence into the pCBV_S1+ vector

First, S1 was PCR amplified using the BamHI 2 and SphI primers. After PCR cleanup, the insert and pCBV_S1+ construct were digested with BamHI and SphI at 37

°C overnight in NEB buffer 4 (50mM Potassium Acetate, 20mM Tris-acetate, 10mM Magnesium Acetate, 1mM DTT, pH 7.9). From this point forward, the dephosphorylation, ligation, and transformation were carried out in the same manner as for the construction of the pCBV_S1+ construct. After isolating the plasmid from the overnight culture, diagnostic PCR was performed using the EcoRI and SphI primers. However, since these primers will bind with both the S1 sequences individually, diagnostic PCR was also performed with primers specific to the region of the pUC19 backbone directly flanking the S1+S1 insert: forward primer 5'-GTAAAACGACGGCCAGTGAATTC-3' and reverse primer 5'-GATTACGCCAAGCTTGCATGC-3'. Constructs containing the full S1+S1 insert were named pCBV_S1+S1 (Figure 5.6).

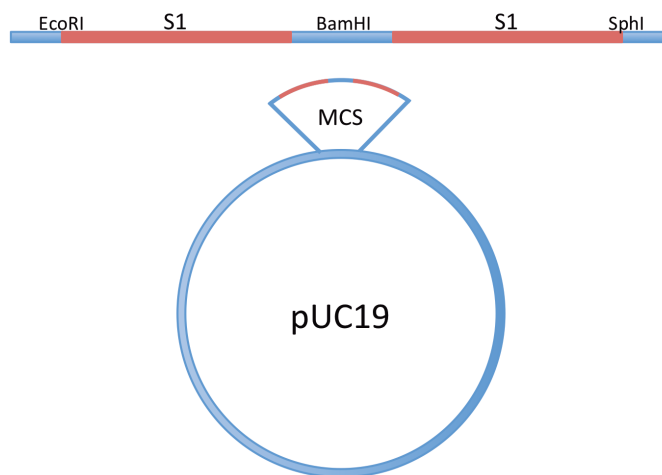


Figure 5.6 Plasmid map of pCBV_S1+S1. The second S1 sequence (red) is inserted between the BamHI and SphI restriction sites within the multiple cloning site (MCS) of the pUC19 vector (blue).

5.3.4 Nucleosome exchange

To generate a radiolabeled DNA substrate, either the pUC19 forward or reverse primer was 5' radiolabeled using ^{32}P following the manufacturers instructions. After PCR

amplification, the amplified DNA was electrophoresed in a 1% agarose gel. Bands containing the S1+S1 DNA were extracted from the gel using a Qiagen QIAquick Gel Extraction Kit. Samples were then quantitated using a NanoDrop spectrophotometer.

A 50 μ L incorporation reaction consisting of 100 ng of radiolabeled S1+S1 duplex, 25 μ g isolated nucleosomes, 1 M NaCl, and 10X exchange buffer (EB) (1X EB; 5 mM Tris, pH 7.5) and an equivalent reaction lacking NCPs were incubated at room temperature. Donor nucleosomes were isolated as described previously and the exchange reaction followed the same procedure described in chapter 2. After the exchange was complete, samples were run in a 4% native polyacrylamide gel at 150 V for 4 hours at 4 °C. The gel was then dried and used to expose a phosphorplate overnight. After exposure, the plate was then imaged using a BioRad Pharos phosphorimager.

In order to determine if the S1+S1 substrate was indeed capable of forming dinucleosomes, we performed the salt-dialysis exchange reaction under saturating conditions, such that there would be ample histone cores available for each S1 sequence. Upon analysis by native PAGE, the reaction performed in the absence of donor nucleosomes runs as a single band, indicating both the purity of the PCR amplified sample and the migration of the free duplex. In the presence of donor chromatin, almost all of the S1+S1 substrate was found in a slower migrating species, indicating that the S1+S1 substrate forms nucleosomes (Figure 5.7).

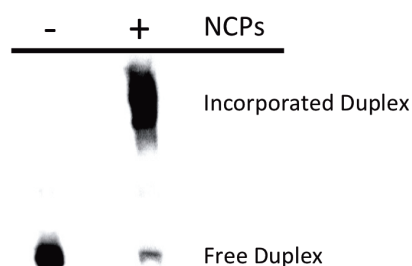


Figure 5.7 The PCR amplified S1+S1 sequence forms nucleosomes. An incorporation performed in the absence of chicken nucleosomes (NCP) (-NCP) and an incorporation performed in the presence of NCP (+NCP) are shown.

However, we cannot determine whether the S1+S1 substrate forms mono- or dinucleosomes from analysis of the native gel. We can only determine that the S1+S1 sequence is interacting with the core histones, but not how many histone cores are present on the substrate. Previous research using other positioning sequences suggests that the presence of a saturating amount of histone cores is sufficient to drive formation of multinucleosome complexes.²⁸³⁻²⁸⁵ However, to further investigate the composition of the incorporated S1+S1 substrate, it is best to evaluate the exchanged DNA with enzymatic probes.

5.3.5 Exonuclease mapping

To determine if both nucleosome-positioning sequences are occupied in the S1+S1 incorporated DNA, Exonuclease III (Exo III) was used to map the position of the interacting nucleosomes. Nucleosomes were exchanged onto 500 ng of S1+S1 DNA in the manner described above. After the exchange was complete, an Exo III reaction was carried out as described in chapters 2 and 4. Following electrophoresis in a 4% native polyacrylamide gel, nucleosome samples were excised. After recovering the DNA, 1000 counts per sample were loaded into a 4% denaturing polyacrylamide gel (0.4 mm thick, 65 cm long). After electrophoreses at 65 W for 2.5 h, the gel was dried and used to expose a phosphorplate for 120 h.

When Exo III encounters DNA interacting with the histone core, it pauses, leading to areas of increased strand cleavage. These areas of intense cleavage can be used to mark the position of the histone core relative to the end of the DNA duplex. When the S1+S1 free duplex is incubated with Exo III, there is unequal reactivity along the entire DNA substrate (Figure 5.8, - NCPs). However, once the DNA is incorporated into a

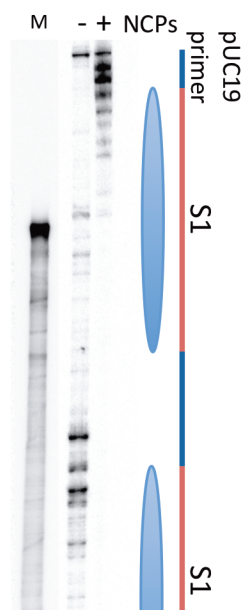


Figure 5.8 Exonuclease III digestion reveals the translation position of the DNA around the histone cores. Reactions were performed on both free duplex (-NCPs) and nucleosome (+NCPs) samples. The marker lane (M) consists of single stranded S2_HP, a 276 nucleotide long sequence discussed in chapter 4. The blue ovals located to the right of the image indicate the position of the histone core and the red lines indicate the location of the S1 positioning sequences.

nucleosome, there is a change in the reactivity pattern. Where the Exo III digestion runs the full length of the free duplex, the digestion is limited to the 3' end of the incorporated DNA (Figure 5.8, + NCPs). The Exo III digests through the pUC19 primer and into the first S1 sequence. If the digestion pattern of just the S1 sequence is compared with previous Exo III experiments performed on S1 in a mono-nucleosome (Figure 2.5, NCP, lane S), we find the reactivity is nearly identical, indicating that this S1 sequence contains a well-positioned nucleosome. When these reactions were carried out with the complementary strand radiolabeled (Figure 5.9), we observe the same change in Exo III

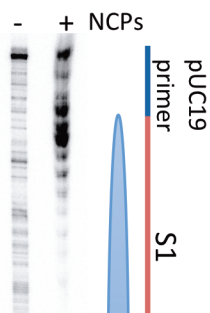


Figure 5.9 Exonuclease III digestion reveals the translation position of the DNA around the histone cores. Reactions were performed on both free duplex (-NCPs) and nucleosome (+NCPs) samples. The blue oval located to the right of the image indicates the position of the histone core and the red line indicates the location of the S1 positioning sequence.

reactivity, indicating that both S1 sequences contain well-positioned nucleosomes and thus we have successfully created di-nucleosomes. We can now use these techniques to construct DNA containing a wide variety of both healthy- and disease-length repeats.

Chapter 6

Conclusions: Small Changes in Sequence, Structure, and Length can Have Large Effects on Nucleosome Formation

In the work presented in the previous chapters, we have investigated the behavior of several TNRs in different biological contexts, particularly that of the nucleosome core particle. In order to investigate the behavior of healthy-length CAG/CTG repeats associated with HD within the packaged genome, we designed 146 bp DNA duplexes containing various lengths of CAG/CTG repeats. Our results demonstrate that, like their disease-length counterparts, healthy-length repeats have an increased affinity for the histone core, though that affinity is moderated by the presence of the *HTT* flanking sequence. Furthermore, healthy-length CAG/CTG repeats alter the position of DNA in a nucleosome in a length-dependent manner. As the repeats expand across the healthy range, both the translational and rotational positioning change until the repeat tract is 20-30 repeats in length. At this length the CAG/CTG repeats dictate the position of the DNA in the nucleosome, regardless of the flanking sequence. Interestingly, the periodicity of the CAG/CTG repeats was consistently ~10.7 bp/turn regardless of length or flanking sequence. This periodicity matches the preferred periodicity adopted by DNA around the ~30 bp region surrounding the dyad axis, a difficult region of the histone core for the DNA to accommodate, which may explain why CAG/CTG repeats are preferentially located at this region. These results highlight the changes in DNA packaging that take place as the repeat tract expands and importantly illustrate a key difference between healthy- and disease-length repeats; while healthy-length repeats must balance the competing effects of the CAG/CTG repeats and the flanking sequence on both affinity for

the histone core and DNA position in the nucleosome, disease-length CAG/CTG repeats are the dominant force in both nucleosome formation and positioning, regardless of the flanking sequence, as the disease-length repeat tract is itself able to occupy the DNA in closest contact with the histone core, leaving the flanking sequence in the regions that only transiently interact. Further expansion within the disease range, increasing the repeat number over 49, will produce nucleosomes that are fully occupied by CAG/CTG repeats.

We next investigated the behavior of healthy-length CGG/CCG repeats in the nucleosome. Here we were interested in a particular feature of the healthy-length repeat tract associated with FXS; healthy-length repeats contain AGG/CCT interruptions every 9-10 repeats and an interrupted repeat tract is less likely to expand than an uninterrupted repeat tract.⁴²⁻⁴⁸ We first determined that, unlike their disease-length counterparts, which decrease the levels of nucleosome formation, 19 CGG/CCG repeats had little effect on nucleosome incorporation and the presence of a single interruption actually increased affinity for the histone core. Furthermore, not only did the presence of the uninterrupted 19 CGG/CCG repeats alter the rotational and translational position of the DNA, but the addition of the AGG/CCT interruptions further altered the position of DNA in a nucleosome. When we measured the periodicity of the interrupted and uninterrupted CGG/CCG repeats, we found that the presence of an interruption increases the periodicity of the repeat tract, indicating that interrupted CGG/CCG repeat tracts are better able to accommodate bending around the histone core, particularly at the dyad axis. We again demonstrated the effects of flanking sequence on the behavior of the repeat tract; not only does the *FMRI* flanking sequence decrease affinity for the histone core, but it also allowed the formation of a heterogeneous population of nucleosomes. Interestingly, these

results may provide insight into the protective role played by the AGG/CCT interruptions; the presence of an interruption may alter the interaction between the repeat tract and the histone core in the packaged genome, which could partially explain the difference in expansion ability between an interrupted and uninterrupted repeat and why the loss of a single interruption can have such immediate consequences.

Before we could investigate the effect of hairpin formation on chromatin packaging, we needed to characterize the behavior of a CAG hairpin in the DNA alone. Building from oligonucleotide systems previously characterized in the lab,^{59,242} we demonstrated that the 8-oxoG lesion is able to restrict the location of the CAG hairpin arm in a three-way junction, such that the lesion is positioned at or near the loop, where repair would be inefficient. We then demonstrated that the ability to alter the structure of the three-way junction is driven by the thermodynamic properties of the 8-oxoG lesion within the CAG hairpin. Furthermore, these results indicate that during the toxic oxidation cycle, upon oxidation of the G in the hairpin loop, the resulting 8-oxoG will remain within the loop rather than shifting to a location in the hairpin stem or flanking duplex, where the lesion might be more easily repaired.

Finally, we addressed the behavior of the three-way junction in the packaged genome. Using a slightly longer DNA construct, we determined that the presence of a (CAG)₁₀ hairpin significantly decreases the affinity of the DNA duplex for the histone core, regardless of whether the hairpin is in a fixed position or can move relative to the complementary strand. However, using enzymatic footprinting, we determined that the hairpin has the ability to alter both the translational position of the DNA as well as locally alter the rotational position of the DNA around the histone core, which could have

profound effects on the functional organization of the repeat tract if a hairpin was to form on a packaged CAG repeat during the toxic oxidation cycle.

Thus we have established a systematic and thorough exploration of the behavior of several healthy-length TNRs in various contexts, both packaged and unpackaged and both hairpin and canonical duplex. These results describe the behavior of healthy length repeats within the packaged genome, revealing that even relatively small changes in length, sequence context, and structure can have profound effects on the functional organization of genes containing TNRs. Importantly, these results also highlight the differences in genomic packaging between healthy- and disease-length repeats, providing a reminder that, in order to understand how the expansion process occurs within the packaged genome, we cannot assume that healthy- and disease-length repeats behave in the same manner; there will likely be significant differences between the two classifications in both the affinity for the histone core and the position of the DNA in a nucleosome. By performing experiments using TNRs incorporated into nucleosomes, both as single NCPs and higher order chromatin structures, we can investigate the behavior of TNRs in the packaged genome with a high level of control, looking for small changes that might influence the expansion of the repeat tract. Using the NCP as our substrate, we can better understand every facet of expansion and how it occurs the packaged genome.

REFERENCES

- (1) Bhargava, A., and Fuentes, F. F. (2009) Mutational dynamics of microsatellites, *Mol. Biotechnol.* 44, 250–266.
- (2) Kelkar, Y. D., Tyekucheva, S., Chiaromonte, F., and Makova, K. D. (2008) The genome-wide determinants of human and chimpanzee microsatellite evolution, *Genome Res.* 18, 30–38.
- (3) Pumpernik, D., Oblak, B., and Borstnik, B. (2008) Replication slippage versus point mutation rates in short tandem repeats of the human genome, *Mol. Genet. Genomics* 279, 53–61.
- (4) Madsen, B. E., Villesen, P., and Wiuf, C. (2008) Short tandem repeats in human exons: a target for disease mutations, *BMC Genomics* 9, 410.
- (5) Tóth, G., Gáspári, Z., and Jurka, J. (2000) Microsatellites in different eukaryotic genomes: survey and analysis, *Genome Res.* 10, 967–981.
- (6) Slattery, J. P., Murphy, W. J., and O'Brien, S. J. (2000) Patterns of diversity among SINE elements isolated from three Y-chromosome genes in carnivores, *Mol. Biol. Evol.* 17, 825–829.
- (7) Alaeddini, R., Walsh, S. J., and Abbas, A. (2010) Forensic implications of genetic analyses from degraded DNA: A review, *Forensic Sci. Int. Genet.* 4, 148–157.
- (8) Li, D., Cui, H., Wang, C., Ling, S., Huang, Z., and Zhang, H. (2011) A fast and effective method to perform paternity testing for Wolong giant pandas, *Chin. Sci. Bull.* 56, 2559–2564.
- (9) Alacs, E. A., Georges, A., FitzSimmons, N. N., and Robertson, J. (2009) DNA

- detective: a review of molecular approaches to wildlife forensics, *Forensic Sci. Med. Pathol.* 6, 180–194.
- (10) McClure, M., Sonstegard, T., Wiggans, G., and Van Tassell, C. P. (2012) Imputation of microsatellite alleles from dense SNP genotypes for parental verification, *Front. Genet.* 3, 140.
 - (11) Boland, C. R., and Goel, A. (2010) Microsatellite instability in colorectal cancer, *YGAST* 138, 2073–2087.e3.
 - (12) Fujiwara, Y., Chi, D. D., Wang, H., Keleman, P., Morton, D. L., Turner, R., and Hoon, D. S. (1999) Plasma DNA microsatellites as tumor-specific markers and indicators of tumor progression in melanoma patients, *Cancer Res.* 59, 1567–1571.
 - (13) Shah, S. N., Hile, S. E., and Eckert, K. A. (2010) Defective mismatch repair, microsatellite mutation bias, and variability in clinical cancer phenotypes, *Cancer Res.* 70, 431–435.
 - (14) Popat, S. (2004) Systematic review of microsatellite instability and colorectal cancer prognosis, *J. Clin. Oncol.* 23, 609–618.
 - (15) Chen, X. Q., Stroun, M., Magnenat, J. L., Nicod, L. P., Kurt, A. M., Lyautey, J., Lederrey, C., and Anker, P. (1996) Microsatellite alterations in plasma DNA of small cell lung cancer patients, *Nat. Med.* 2, 1033–1035.
 - (16) Nawroz, H., Koch, W., Anker, P., Stroun, M., and Sidransky, D. (1996) Microsatellite alterations in serum DNA of head and neck cancer patients, *Nat. Med.* 2, 1035–1037.
 - (17) Gatchel, J. R., and Zoghbi, H. Y. (2005) Diseases of unstable repeat expansion:

- mechanisms and common principles, *Nat. Rev. Genet.* 6, 743–755.
- (18) McMurray, C. T. (2010) Mechanisms of trinucleotide repeat instability during human development, *Nat. Rev. Genet.* 11, 786–799.
 - (19) López Castel, A., Cleary, J. D., and Pearson, C. E. (2010) Repeat instability as the basis for human diseases and as a potential target for therapy, *Nat. Rev. Mol. Cell. Biol.* 11, 165–170.
 - (20) Mirkin, S. M. (2007) Expandable DNA repeats and human disease, *Nature* 447, 932–940.
 - (21) Wells, R. D. (2007) Non-B DNA conformations, mutagenesis and disease, *Trends Biochem. Sci.* 32, 271–278.
 - (22) The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes, *Cell* 72, 971–983.
 - (23) Imarisio, S., Carmichael, J., Korolchuk, V., Chen, C.-W., Saiki, S., Rose, C., Krishna, G., Davies, J. E., Ttofi, E., Underwood, B. R., and Rubinsztein, D. C. (2008) Huntington's disease: from pathology and genetics to potential therapies, *Biochem. J.* 412, 191–209.
 - (24) Kennedy, L., Evans, E., Chen, C.-M., Craven, L., Detloff, P. J., Ennis, M., and Shelbourne, P. F. (2003) Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis, *Hum. Mol. Genet.* 12, 3359–3367.
 - (25) Telenius, H., Kremer, B., Goldberg, Y. P., Theilmann, J., Andrew, S. E., Zeisler, J., Adam, S., Greenberg, C., Ives, E. J., and Clarke, L. A. (1994)

Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm, *Nat. Genet.* 6, 409–414.

- (26) Gonitell, R., Moffitt, H., Sathasivam, K., Woodman, B., Detloff, P. J., Faull, R. L. M., and Bates, G. P. (2008) DNA instability in postmitotic neurons, *Proc. Natl. Acad. Sci. USA* 105, 3467–3472.
- (27) Ross, C. A., and Tabrizi, S. J. (2011) Huntington's disease: from molecular pathogenesis to clinical treatment, *Lancet Neurol.* 10, 83–98.
- (28) Dragatsis, I., Efstratiadis, A., and Zeitlin, S. (1998) Mouse mutant embryos lacking huntingtin are rescued from lethality by wild-type extraembryonic tissues, *Development* 125, 1529–1539.
- (29) DiFiglia, M., Sapp, E., Chase, K. O., Davies, S. W., Bates, G. P., Vonsattel, J. P., and Aronin, N. (1997) Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain, *Science* 277, 1990–1993.
- (30) Landles, C., Sathasivam, K., Weiss, A., Woodman, B., Moffitt, H., Finkbeiner, S., Sun, B., Gafni, J., Ellerby, L. M., Trotter, Y., Richards, W. G., Osmand, A., Paganetti, P., and Bates, G. P. (2010) Proteolysis of mutant huntingtin produces an exon 1 fragment that accumulates as an aggregated protein in neuronal nuclei in Huntington disease, *J. Biol. Chem.* 285, 8808–8823.
- (31) Schilling, G., Klevytska, A., Tebbenkamp, A. T. N., Juenemann, K., Cooper, J., Gonzales, V., Slunt, H., Poirer, M., Ross, C. A., and Borchelt, D. R. (2007) Characterization of huntingtin pathologic fragments in human Huntington disease, transgenic mice, and cell models, *J. Neuropathol. Exp. Neurol.* 66, 313–320.

- (32) Sathasivam, K., Neueder, A., Gipson, T. A., Landles, C., Benjamin, A. C., Bondulich, M. K., Smith, D. L., Faull, R. L. M., Roos, R. A. C., Howland, D., Detloff, P. J., Housman, D. E., and Bates, G. P. (2013) Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease, *Proc. Natl. Acad. Sci. USA*.
- (33) Sadri-Vakili, G., and Cha, J.-H. J. (2006) Mechanisms of disease: Histone modifications in Huntington's disease, *Nat. Clin. Pract. Neurol.* 2, 330–338.
- (34) Fu, Y. H., Pizzuti, A., Fenwick, R. G., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., and de Jong, P. (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy, *Science* 255, 1256–1258.
- (35) Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barceló, J., and O'Hoy, K. (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene, *Science* 255, 1253–1255.
- (36) Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., and Zhang, F. P. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome, *Cell* 65, 905–914.
- (37) Fu, Y. H., Kuhl, D. P., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J., Holden, J. J., Fenwick, R. G., and Warren, S. T. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability:

resolution of the Sherman paradox, *Cell* 67, 1047–1058.

- (38) Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M., and Mandel, J. (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome, *Science* 252, 1097–1102.
- (39) Dietrich, A., Kioschis, P., Monaco, A. P., Gross, B., Korn, B., Williams, S. V., Sheer, D., Heitz, D., Oberle, I., and Toniolo, D. (1991) Molecular cloning and analysis of the fragile X region in man, *Nucleic Acids Res.* 19, 2567–2572.
- (40) Heitz, D., Rousseau, F., Devys, D., Saccone, S., Abderrahim, H., Le Paslier, D., Cohen, D., Vincent, A., Toniolo, D., and Valle, Della, G. (1991) Isolation of sequences that span the fragile X and identification of a fragile X-related CpG island, *Science* 251, 1236–1239.
- (41) Kremer, 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) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)_n, *Science* 252, 1711–1714.
- (42) Kunst, C. B., and Warren, S. T. (1994) Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles, *Cell* 77, 853–861.
- (43) Eichler, E. E., Holden, J. J., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., Richards, C. S., Ward, P. A., and Nelson, D. L. (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene, *Nat. Genet.* 8, 88–94.
- (44) Zhong, N., Yang, W., Dobkin, C., and Brown, W. T. (1995) Fragile X gene

- instability: anchoring AGGs and linked microsatellites, *Am. J. Hum. Genet.* 57, 351–361.
- (45) Zhong, N., Ju, W., Pietrofesa, J., Wang, D., Dobkin, C., and Brown, W. T. (1996) Fragile X “gray zone” alleles: AGG patterns, expansion risks, and associated haplotypes, *Am. J. Med. Genet.* 64, 261–265.
 - (46) Nolin, S. L., Lewis, F. A., Ye, L. L., Houck, G. E., Glicksman, A. E., Limprasert, P., Li, S. Y., Zhong, N., Ashley, A. E., Feingold, E., Sherman, S. L., and Brown, W. T. (1996) Familial transmission of the FMR1 CGG repeat, *Am. J. Hum. Genet.* 59, 1252–1261.
 - (47) Nolin, S. L., Brown, W. T., Glicksman, A., Houck, G. E., Gargano, A. D., Sullivan, A., Biancalana, V., Brøndum-Nielsen, K., Hjalgrim, H., Holinski-Feder, E., Kooy, F., Longshore, J., Macpherson, J., Mandel, J.-L., Matthijs, G., Rousseau, F., Steinbach, P., Väisänen, M.-L., Koskull, von, H., and Sherman, S. L. (2003) Expansion of the fragile X CGG repeat in females with premutation or intermediate alleles, *Am. J. Hum. Genet.* 72, 454–464.
 - (48) Dombrowski, C., Lévesque, S., Morel, M. L., Rouillard, P., Morgan, K., and Rousseau, F. (2002) Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles, *Hum. Mol. Genet.* 11, 371–378.
 - (49) Fernandez-Carvajal, I., Posadas, B. L., Pan, R., Raske, C., Hagerman, P. J., and Tassone, F. (2010) Expansion of an FMR1 grey-zone allele to a full mutation in two generations, *J. Mol. Diagn.* 11, 306–310.
 - (50) Sutcliffe, J. S., Nelson, D. L., Zhang, F., Pieretti, M., Caskey, C. T., Saxe, D.,

- and Warren, S. T. (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome, *Hum. Mol. Genet.* 1, 397–400.
- (51) Luo, S., Robinson, J. C., Reiss, A. L., and Migeon, B. R. (1993) DNA methylation of the fragile X locus in somatic and germ cells during fetal development: relevance to the fragile X syndrome and X inactivation, *Somat. Cell. Mol. Genet.* 19, 393–404.
- (52) Dion, V., and Wilson, J. H. (2009) Instability and chromatin structure of expanded trinucleotide repeats, *Trends Genet.* 25, 288–297.
- (53) Wang, T., Bray, S. M., and Warren, S. T. (2012) New perspectives on the biology of fragile X syndrome, *Curr. Opin. Genet. Dev.* 22, 256–263.
- (54) Garber, K., Smith, K. T., Reines, D., and Warren, S. T. (2006) Transcription, translation and fragile X syndrome, *Curr. Opin. Genet. Dev.* 16, 270–275.
- (55) Bassell, G. J., and Warren, S. T. (2008) Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function, *Neuron* 60, 201–214.
- (56) Wang, Y.-H. (2007) Chromatin structure of repeating CTG/CAG and CGG/CCG sequences in human disease, *Front. Biosci.* 12, 4731–4741.
- (57) Mirkin, S. M. (2006) DNA structures, repeat expansions and human hereditary disorders, *Curr. Opin. Struct. Biol.* 16, 351–358.
- (58) Gacy, A. M., Goellner, G., Juranić, N., Macura, S., and McMurray, C. T. (1995) Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*, *Cell* 81, 533–540.
- (59) Jarem, D. A., Wilson, N. R., and Delaney, S. (2009) Structure-dependent DNA

- damage and repair in a trinucleotide repeat sequence, *Biochemistry* 48, 6655–6663.
- (60) Jarem, D. A., Huckaby, L. V., and Delaney, S. (2010) AGG interruptions in (CGG)_n DNA repeat tracts modulate the structure and thermodynamics of non-B conformations in vitro., *Biochemistry* 49, 6826–6837.
 - (61) Jarem, D. A., and Delaney, S. (2011) Premutation huntingtin allele adopts a non-B conformation and contains a hot spot for DNA damage, *Biochem. Biophys. Res. Comm.* 416, 146–152.
 - (62) Amrane, S., and Mergny, J.-L. (2006) Length and pH-dependent energetics of (CCG)_n and (CGG)_n trinucleotide repeats, *Biochimie* 88, 1125–1134.
 - (63) Amrane, S., Saccà, B., Mills, M., Chauhan, M., Klump, H. H., and Mergny, J.-L. (2005) Length-dependent energetics of (CTG)_n and (CAG)_n trinucleotide repeats, *Nucleic Acids Res.* 33, 4065–4077.
 - (64) Zheng, M., Huang, X., Smith, G. K., Yang, X., and Gao, X. (1996) Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding. An NMR and UV spectroscopic study, *J. Mol. Biol.* 264, 323–336.
 - (65) Paiva, A. M., and Sheardy, R. D. (2004) Influence of sequence context and length on the structure and stability of triplet repeat DNA oligomers, *Biochemistry* 43, 14218–14227.
 - (66) Pearson, C. E., and Sinden, R. R. (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci, *Biochemistry* 35, 5041–5053.
 - (67) Avila Figueroa, A., and Delaney, S. (2010) Mechanistic studies of hairpin to

- duplex conversion for trinucleotide repeat sequences, *J. Biol. Chem.* 285, 14648–14657.
- (68) Avila Figueroa, A., Cattie, D., and Delaney, S. (2011) Structure of even/odd trinucleotide repeat sequences modulates persistence of non-B conformations and conversion to duplex, *Biochemistry* 50, 4441–4450.
- (69) Völker, J., Makube, N., Plum, G. E., Klump, H. H., and Breslauer, K. J. (2002) Conformational energetics of stable and metastable states formed by DNA triplet repeat oligonucleotides: implications for triplet expansion diseases, *Proc. Natl. Acad. Sci. USA* 99, 14700–14705.
- (70) Volle, C. B., Jarem, D. A., and Delaney, S. (2012) Trinucleotide repeat DNA alters structure to minimize the thermodynamic impact of 8-oxo-7,8-dihydroguanine., *Biochemistry* 51, 52–62.
- (71) Pearson, C. E., Edamura, K. N., and Cleary, J. D. (2005) Repeat instability: mechanisms of dynamic mutations, *Nat. Rev. Genet.* 6, 729–742.
- (72) Pearson, C. E., Wang, Y. H., Griffith, J. D., and Sinden, R. R. (1998) Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)_n•(CAG)_n repeats from the myotonic dystrophy locus, *Nucleic Acids Res.* 26, 816–823.
- (73) Völker, J., Klump, H. H., and Breslauer, K. J. (2007) DNA metastability and biological regulation: conformational dynamics of metastable omega-DNA bulge loops, *J. Am. Chem. Soc.* 129, 5272–5280.
- (74) Völker, J., Klump, H. H., and Breslauer, K. J. (2008) DNA energy landscapes via calorimetric detection of microstate ensembles of metastable macrostates and triplet repeat diseases, *Proc. Natl. Acad. Sci. USA* 105, 18326–18330.

- (75) Völker, J., Plum, G. E., Klump, H. H., and Breslauer, K. J. (2009) DNA repair and DNA triplet repeat expansion: the impact of abasic lesions on triplet repeat DNA energetics, *J. Am. Chem. Soc.* *131*, 9354–9360.
- (76) Völker, J., Plum, G. E., Klump, H. H., and Breslauer, K. J. (2010) Energetic coupling between clustered lesions modulated by intervening triplet repeat bulge loops: allosteric implications for DNA repair and triplet repeat expansion, *Biopolymers* *93*, 355–369.
- (77) Völker, J., Plum, G. E., Klump, H. H., and Breslauer, K. J. (2010) Energy crosstalk between DNA lesions: implications for allosteric coupling of DNA repair and triplet repeat expansion pathways, *J. Am. Chem. Soc.* *132*, 4095–4097.
- (78) Degtyareva, N. N., Barber, C. A., Sengupta, B., and Petty, J. T. (2010) Context dependence of trinucleotide repeat structures, *Biochemistry* *49*, 3024–3030.
- (79) Degtyareva, N. N., Barber, C. A., Reddish, M. J., and Petty, J. T. (2011) Sequence length dictates repeated CAG folding in three-way junctions, *Biochemistry* *50*, 458–465.
- (80) Liu, G., Chen, X., Bissler, J. J., Sinden, R. R., and Leffak, M. (2010) Replication-dependent instability at (CTG) x (CAG) repeat hairpins in human cells, *Nat. Chem. Biol.* *6*, 652–659.
- (81) Pearson, C. E. (2003) Slipping while sleeping? Trinucleotide repeat expansions in germ cells, *Trends Mol. Med.* *9*, 490–495.
- (82) Schweitzer, J. K., and Livingston, D. M. (1998) Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation,

Hum. Mol. Genet. 7, 69–74.

- (83) Freudenreich, C. H., Kantrow, S. M., and Zakian, V. A. (1998) Expansion and length-dependent fragility of CTG repeats in yeast, *Science* 279, 853–856.
- (84) White, P. J., Borts, R. H., and Hirst, M. C. (1999) Stability of the human fragile X (CGG)_n triplet repeat array in *Saccharomyces cerevisiae* deficient in aspects of DNA metabolism., *Mol. Cell. Biol.* 19, 5675–5684.
- (85) Ireland, M. J., Reinke, S. S., and Livingston, D. M. (2000) The impact of lagging strand replication mutations on the stability of CAG repeat tracts in yeast, *Genetics* 155, 1657–1665.
- (86) Schweitzer, J. K., and Livingston, D. M. (1999) The effect of DNA replication mutations on CAG tract stability in yeast, *Genetics* 152, 953–963.
- (87) Liu, Y., Zhang, H., Veeraraghavan, J., Bambara, R. A., and Freudenreich, C. H. (2004) *Saccharomyces cerevisiae* flap endonuclease 1 uses flap equilibration to maintain triplet repeat stability, *Mol. Cell. Biol.* 24, 4049–4064.
- (88) Yang, J., and Freudenreich, C. H. (2007) Haploinsufficiency of yeast FEN1 causes instability of expanded CAG/CTG tracts in a length-dependent manner, *Gene* 393, 110–115.
- (89) Spiro, C., Pelletier, R., Rolfsmeier, M. L., Dixon, M. J., Lahue, R. S., Gupta, G., Park, M. S., Chen, X., Mariappan, S. V., and McMurray, C. T. (1999) Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats, *Mol. Cell* 4, 1079–1085.
- (90) Lee, S., and Park, M. S. (2002) Human FEN-1 can process the 5'-flap DNA of CTG/CAG triplet repeat derived from human genetic diseases by length and

- sequence dependent manner, *Exp. Mol. Med.* 34, 313–317.
- (91) Veeraraghavan, J., Rossi, M. L., and Bambara, R. A. (2003) Analysis of DNA replication intermediates suggests mechanisms of repeat sequence expansion, *J. Biol. Chem.* 278, 42854–42866.
- (92) Henricksen, L. A., Tom, S., Liu, Y., and Bambara, R. A. (2000) Inhibition of flap endonuclease 1 by flap secondary structure and relevance to repeat sequence expansion, *J. Biol. Chem.* 275, 16420–16427.
- (93) Mirkin, E. V., and Mirkin, S. M. (2007) Replication fork stalling at natural impediments, *Microbiol. Mol. Biol. Rev.* 71, 13–35.
- (94) Voineagu, I., Surka, C. F., Shishkin, A. A., Krasilnikova, M. M., and Mirkin, S. M. (2009) Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility, *Nat. Struct. Mol. Biol.* 16, 226–228.
- (95) Samadashwily, G. M., Raca, G., and Mirkin, S. M. (1997) Trinucleotide repeats affect DNA replication *in vivo*, *Nat. Genet.* 17, 298–304.
- (96) Cleary, J. D., Nichol, K., Wang, Y.-H., and Pearson, C. E. (2002) Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells, *Nat. Genet.* 31, 37–46.
- (97) Nenguke, T., Aladjem, M. I., Gusella, J. F., Wexler, N. S., Arnheim, N., Venezuela HD Project. (2003) Candidate DNA replication initiation regions at human trinucleotide repeat disease loci, *Hum. Mol. Genet.* 12, 1021–1028.
- (98) Kang, S., Jaworski, A., Ohshima, K., and Wells, R. D. (1995) Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*, *Nat. Genet.* 10, 213–218.

- (99) Freudenreich, C. H., Stavenhagen, J. B., and Zakian, V. A. (1997) Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome, *Mol. Cell. Biol.* 17, 2090–2098.
- (100) Miret, J. J., Pessoa-Brandão, L., and Lahue, R. S. (1998) Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. USA* 95, 12438–12443.
- (101) Balakumaran, B. S., Freudenreich, C. H., and Zakian, V. A. (2000) CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in *Saccharomyces cerevisiae*, *Hum. Mol. Genet.* 9, 93–100.
- (102) Lindahl, T. (1993) Instability and decay of the primary structure of DNA, *Nature* 362, 709–715.
- (103) Turrens, J. F. (2003) Mitochondrial formation of reactive oxygen species, *J. Physiol.* 552, 335–344.
- (104) Tahara, E. B., Navarete, F. D. T., and Kowaltowski, A. J. (2009) Tissue-, substrate-, and site-specific characteristics of mitochondrial reactive oxygen species generation, *Free Radic. Biol. Med.* 46, 1283–1297.
- (105) Turrens, J. F. (1997) Superoxide production by the mitochondrial respiratory chain, *Biosci. Rep.* 17, 3–8.
- (106) Fleury, C., Mignotte, B., and Vayssière, J.-L. (2002) Mitochondrial reactive oxygen species in cell death signaling, *Biochimie* 84, 131–141.
- (107) Klaunig, J. E., and Kamendulis, L. M. (2004) The role of oxidative stress in carcinogenesis, *Annu. Rev. Pharmacol. Toxicol.* 44, 239–267.
- (108) Balaban, R. S., Nemoto, S., and Finkel, T. (2005) Mitochondria, oxidants, and

aging, *Cell* 120, 483–495.

- (109) Boveris, A., and Chance, B. (1973) The mitochondrial generation of hydrogen peroxide. General properties and effect of hyperbaric oxygen, *Biochem. J.* 134, 707–716.
- (110) Hernández-García, D., Wood, C. D., Castro-Obregón, S., and Covarrubias, L. (2010) Reactive oxygen species: A radical role in development?, *Free Radic. Biol. Med.* 49, 130–143.
- (111) Brand, M. D. (2005) The efficiency and plasticity of mitochondrial energy transduction, *Biochem. Soc. Trans.* 33, 897–904.
- (112) Fridovich, I. (1978) The biology of oxygen radicals, *Science* 201, 875–880.
- (113) Halliwell, B., and Gutteridge, J. M. (1990) Role of free radicals and catalytic metal ions in human disease: an overview, *Meth. Enzymol.* 186, 1–85.
- (114) Liochev, S. I., and Fridovich, I. (1999) Superoxide and iron: partners in crime, *IUBMB Life* 48, 157–161.
- (115) Starkov, A. A., Fiskum, G., Chinopoulos, C., Lorenzo, B. J., Browne, S. E., Patel, M. S., and Beal, M. F. (2004) Mitochondrial alpha-ketoglutarate dehydrogenase complex generates reactive oxygen species, *J. Neurosci.* 24, 7779–7788.
- (116) Esterbauer, H., Eckl, P., and Ortner, A. (1990) Possible mutagens derived from lipids and lipid precursors, *Mutat. Res.* 238, 223–233.
- (117) Powers, S. K., Duarte, J., Kavazis, A. N., and Talbert, E. E. (2010) Reactive oxygen species are signaling molecules for skeletal muscle adaptation, *Exp. Physiol.* 95, 1–9.

- (118) Rigoulet, M., Yoboue, E. D., and Devin, A. (2011) Mitochondrial ROS generation and its regulation: mechanisms involved in H₂O₂ signaling, *Antioxid. Redox. Signal.* *14*, 459–468.
- (119) Jackson, S. P., and Bartek, J. (2009) The DNA-damage response in human biology and disease, *Nature* *461*, 1071–1078.
- (120) Hussain, S. P., Hofseth, L. J., and Harris, C. C. (2003) Radical causes of cancer, *Nat. Rev. Cancer.* *3*, 276–285.
- (121) Huie, R. E., and Padmaja, S. (1993) The reaction of NO with superoxide, *Free Radic. Res. Commun.* *18*, 195–199.
- (122) Georgakilas, A. G., Mosley, W. G., Georgakila, S., Ziech, D., and Panayiotidis, M. I. (2010) Viral-induced human carcinogenesis: an oxidative stress perspective, *Mol. Biosyst.* *6*, 1162–1172.
- (123) Regulus, P., Spessotto, S., Gateau, M., Cadet, J., Favier, A., and Ravanat, J.-L. (2004) Detection of new radiation-induced DNA lesions by liquid chromatography coupled to tandem mass spectrometry, *Rapid Commun. Mass Spectrom.* *18*, 2223–2228.
- (124) Regulus, P., Duroux, B., Bayle, P.-A., Favier, A., Cadet, J., and Ravanat, J.-L. (2007) Oxidation of the sugar moiety of DNA by ionizing radiation or bleomycin could induce the formation of a cluster DNA lesion, *Proc. Natl. Acad. Sci. USA* *104*, 14032–14037.
- (125) Candeias, L. (1992) Candeias, L. P., and Steenken, S. (1992) Ionization of purine nucleosides and nucleotides and their components by 193-nm laser photolysis in aqueous solution: model studies for oxidative damage of DNA, *J.*

- Am. Chem. Soc.* **114**, 699–704.
- (126) Candeias, L. P., and Steenken, S. (1989) Structure and acid-base properties of one-electron-oxidized deoxyguanosine, guanosine, and 1-methylguanosine, *J. Am. Chem. Soc.* **111**, 1094–1099.
- (127) Pouget, J.-P., Frelon, S., Ravanat, J.-L., Testard, I., Odin, F., and Cadet, J. (2002) Formation of modified DNA bases in cells exposed either to gamma radiation or to high-LET particles, *Radiat. Res.* **157**, 589–595.
- (128) Cadet, J., Douki, T., Ravanat, J.-L., and Di Mascio, P. (2009) Sensitized formation of oxidatively generated damage to cellular DNA by UVA radiation, *Photochem. Photobiol. Sci.* **8**, 903–911.
- (129) Cadet, J., Sage, E., and Douki, T. (2005) Ultraviolet radiation-mediated damage to cellular DNA, *Mutat. Res.* **571**, 3–17.
- (130) Cadet, J., Berger, M., Buchko, G. W., Joshi, P. C., Raoul, S., and Ravanat, J.-L. (1994) 2, 2-Diamino-4-[(3, 5-di-O-acetyl-2-deoxy- β -D-erythro-pentofuranosyl) amino]-5-(2H)-oxazolone: a novel and predominant radical oxidation product of 3', 5'-Di-O-acetyl-2'-deoxyguanosine, *J. Am. Chem. Soc.* **116**, 7403–7404.
- (131) Kino, K., and Sugiyama, H. (2001) Possible cause of G-C $\rightarrow$ C-G transversion mutation by guanine oxidation product, imidazolone, *Chem. Biol.* **8**, 369–378.
- (132) Kino, K., Saito, I., and Sugiyama, H. (1998) Product analysis of GG-specific photooxidation of DNA via electron transfer: 2-aminoimidazolone as a major guanine oxidation product, *J. Am. Chem. Soc.* **120**, 7373–7374.
- (133) Steenken, S., and Jovanovic, S. (1997) How easily oxidizable is DNA? One-

electron reduction potentials of adenosine and guanosine radicals in aqueous solution, *J. Am. Chem. Soc.* **119**, 617–618.

- (134) Steenken, S. (1989) Purine bases, nucleosides, and nucleotides: aqueous solution redox chemistry and transformation reactions of their radical cations and e-and OH adducts, *Chem. Rev.* **89**, 503–520.
- (135) Gates, K. S. (2009) An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals, *Chem. Res. Toxicol.* **22**, 1747–1760.
- (136) Neeley, W. L., and Essigmann, J. M. (2006) Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products, *Chem. Res. Toxicol.* **19**, 491–505.
- (137) Cadet, J., Douki, T., and Ravanat, J.-L. (2010) Oxidatively generated base damage to cellular DNA, *Free Radic. Biol. Med.* **49**, 9–21.
- (138) Sato, M. H., Ura, K., Hohmura, K. I., Tokumasu, F., Yoshimura, S. H., Hanaoka, F., and Takeyasu, K. (1999) Atomic force microscopy sees nucleosome positioning and histone H1-induced compaction in reconstituted chromatin, *FEBS Lett.* **452**, 267–271.
- (139) Sugiyama, H., and Saito, I. (1996) Theoretical studies of GG-specific photocleavage of DNA via electron transfer: significant lowering of ionization potential and 5'-localization of HOMO of stacked GG bases in B-form DNA, *J. Am. Chem. Soc.* **118**, 7063–7068.
- (140) Henderson, P., Jones, D., Hampikian, G., Kan, Y., and Schuster, G. (1999) Long-distance charge transport in duplex DNA: the phonon-assisted polaron-

- like hopping mechanism, *Proc. Natl. Acad. Sci. USA* 96, 8353–8358.
- (141) Giese, B. (2002) Long-distance electron transfer through DNA, *Annu. Rev. Biochem.* 71, 51–70.
 - (142) Núñez, M. E., Rajski, S. R., and Barton, J. K. (2000) Damage to DNA by long-range charge transport, *Meth. Enzymol.* 319, 165–188.
 - (143) Steenken, S., Jovanovic, S., Bietti, M., and Bernhard, K. (2000) The trap depth (in DNA) of 8-Oxo-7,8-dihydro-2'-deoxyguanosine as derived from electron-transfer equilibria in aqueous solution, *J. Am. Chem. Soc.* 122, 2373–2374.
 - (144) Margolin, Y., Cloutier, J.-F., Shafirovich, V., Geacintov, N. E., and Dedon, P. C. (2006) Paradoxical hotspots for guanine oxidation by a chemical mediator of inflammation, *Nat. Chem. Biol.* 2, 365–366.
 - (145) Niles, J. C., Wishnok, J. S., and Tannenbaum, S. R. (2004) Mass spectrometric identification of 4-hydroxy-2,5-dioxo-imidazolidine-4-carboxylic acid during oxidation of 8-oxoguanosine by peroxynitrite and KHSO₅/CoCl₂, *Chem. Res. Toxicol.* 17, 1501–1509.
 - (146) Niles, J. C., Wishnok, J. S., and Tannenbaum, S. R. (2004) Spiroiminodihydantoin and guanidinohydantoin are the dominant products of 8-oxoguanosine oxidation at low fluxes of peroxynitrite: mechanistic studies with ¹⁸O, *Chem. Res. Toxicol.* 17, 1510–1519.
 - (147) Burrows, C. J., and Muller, J. G. (1998) Oxidative nucleobase modifications leading to strand scission, *Chem. Rev.* 98, 1109–1152.
 - (148) Candeias, L. P., and Steenken, S. (2000) Reaction of HO• with guanine derivatives in aqueous solution: formation of two different redox-active OH-

- adduct radicals and their unimolecular transformation reactions. Properties of G(-H), *Chemistry* 6, 475–484.
- (149) Cadet, J., Delatour, T., Douki, T., Gasparutto, D., Pouget, J.-P., Ravanat, J.-L., and Sauvaigo, S. (1999) Hydroxyl radicals and DNA base damage, *Mutat. Res.* 424, 9–21.
 - (150) Nishimura, S. (2002) Involvement of mammalian OGG1(MMH) in excision of the 8-hydroxyguanine residue in DNA, *Free Radic. Biol. Med.* 32, 813–821.
 - (151) Nakamoto, H., Kaneko, T., Tahara, S., Hayashi, E., Naito, H., Radak, Z., and Goto, S. (2007) Regular exercise reduces 8-oxodG in the nuclear and mitochondrial DNA and modulates the DNA repair activity in the liver of old rats, *Exp. Gerontol.* 42, 287–295.
 - (152) Roots, R., and Okada, S. (1975) Estimation of life times and diffusion distances of radicals involved in X-ray-induced DNA strand breaks of killing of mammalian cells, *Radiat. Res.* 64, 306–320.
 - (153) European Standards Committee on Oxidative DNA Damage ESCODD. (2003) Measurement of DNA oxidation in human cells by chromatographic and enzymic methods, *Free Radic. Biol. Med.* 34, 1089–1099.
 - (154) Gedik, C. M., Collins, A., and ESCODD European Standards Committee on Oxidative DNA Damage. (2005) Establishing the background level of base oxidation in human lymphocyte DNA: results of an interlaboratory validation study, *FASEB J.* 19, 82–84.
 - (155) Delaney, S., Jarem, D. A., Volle, C. B., and Yennie, C. J. (2012) Chemical and biological consequences of oxidatively damaged guanine in DNA, *Free Radic.*

Res. 46, 420–441.

- (156) David, S. S., and Williams, S. D. (1998) Chemistry of Glycosylases and Endonucleases Involved in Base-Excision Repair, *Chem. Rev.* 98, 1221–1262.
- (157) David, S. S., O'Shea, V. L., and Kundu, S. (2007) Base-excision repair of oxidative DNA damage, *Nature* 447, 941–950.
- (158) Sathasivam, K., Hobbs, C., Mangiarini, L., Mahal, A., Turmaine, M., Doherty, P., Davies, S. W., and Bates, G. P. (1999) Transgenic models of Huntington's disease, *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 354, 963–969.
- (159) Kovtun, I. V., Liu, Y., Bjoras, M., Klungland, A., Wilson, S. H., and McMurray, C. T. (2007) OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells, *Nature* 447, 447–452.
- (160) Robertson, A. B., Klungland, A., Rognes, T., and Leiros, I. (2009) DNA repair in mammalian cells: Base excision repair: the long and short of it, *Cell Mol. Life. Sci.* 66, 981–993.
- (161) Guan, L., Bebenek, K., Kunkel, T. A., and Greenberg, M. M. (2010) Inhibition of short patch and long patch base excision repair by an oxidized abasic site, *Biochemistry* 49, 9904–9910.
- (162) Liu, Y., Prasad, R., Beard, W. A., Hou, E. W., Horton, J. K., McMurray, C. T., and Wilson, S. H. (2009) Coordination between polymerase beta and FEN1 can modulate CAG repeat expansion, *J. Biol. Chem.* 284, 28352–28366.
- (163) Liu, Y., Kao, H.-I., and Bambara, R. A. (2004) Flap endonuclease 1: a central component of DNA metabolism., *Annu. Rev. Biochem.* 73, 589–615.
- (164) Goula, A.-V., Berquist, B. R., Wilson, D. M., Wheeler, V. C., Trottier, Y., and

- Merienne, K. (2009) Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum In Huntington's disease transgenic mice, *PLoS Genet.* 5, e1000749.
- (165) Goula, A.-V., Pearson, C. E., Maria, Della, J., Trottier, Y., Tomkinson, A. E., Wilson, D. M., and Merienne, K. (2012) The Nucleotide Sequence, DNA Damage Location, and Protein Stoichiometry Influence the Base Excision Repair Outcome at CAG/CTG Repeats, *Biochemistry* 51, 3919–3932.
- (166) Luger, K., Dechassa, M. L., and Tremethick, D. J. (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?, *Nat. Rev. Mol. Cell. Biol.* 13, 436–447.
- (167) Kornberg, R. D., and Lorch, Y. (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome, *Cell* 98, 285–294.
- (168) Korolev, N., Fan, Y., Lyubartsev, A. P., and Nordenskiöld, L. (2012) Modeling chromatin structure and dynamics: status and prospects, *Curr. Opin. Struct. Biol.* 22, 151–159.
- (169) Widom, J. (1998) Structure, dynamics, and function of chromatin *in vitro*, *Annu. Rev. Biophys. Biomol. Struct.* 27, 285–327.
- (170) Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution, *Nature* 389, 251–260.
- (171) Akey, C. W., and Luger, K. (2003) Histone chaperones and nucleosome assembly, *Curr. Opin. Struct. Biol.* 13, 6–14.
- (172) Mello, J. A., and Almouzni, G. (2001) The ins and outs of nucleosome

- assembly, *Curr. Opin. Genet. Dev.* *11*, 136–141.
- (173) Verreault, A. (2000) *De novo* nucleosome assembly: new pieces in an old puzzle, *Genes Dev.* *14*, 1430–1438.
 - (174) Andrews, A. J., and Luger, K. (2011) Nucleosome Structure(s) and Stability: Variations on a Theme, *Annu. Rev. Biophys.* *40*, 99–117.
 - (175) Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A. S., Imamoto, N., Ishikawa, T., and Maeshima, K. (2012) Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure, *EMBO J.* *31*, 1644–1653.
 - (176) Adolph, K. W., Cheng, S. M., and Laemmli, U. K. (1977) Role of nonhistone proteins in metaphase chromosome structure, *Cell* *12*, 805–816.
 - (177) Kouzarides, T. (2007) Chromatin modifications and their function, *Cell* *128*, 693–705.
 - (178) Margueron, R., and Reinberg, D. (2010) Chromatin structure and the inheritance of epigenetic information, *Nat. Rev. Genet.* *11*, 285–296.
 - (179) Burgess, R. J., and Zhang, Z. (2013) Histone chaperones in nucleosome assembly and human disease, *Nat. Struct. Mol. Biol.* *20*, 14–22.
 - (180) Lowary, P. T., and Widom, J. (1997) Nucleosome packaging and nucleosome positioning of genomic DNA, *Proc. Natl. Acad. Sci. USA* *94*, 1183–1188.
 - (181) Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I. K., Wang, J.-P. Z., and Widom, J. (2006) A genomic code for nucleosome positioning, *Nature* *442*, 772–778.
 - (182) Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D.,

- Field, Y., LeProust, E. M., Hughes, T. R., Lieb, J. D., Widom, J., and Segal, E. (2009) The DNA-encoded nucleosome organization of a eukaryotic genome, *Nature* 458, 362–366.
- (183) Workman, J. L., and Kingston, R. E. (1998) Alteration of nucleosome structure as a mechanism of transcriptional regulation, *Annu. Rev. Biochem.* 67, 545–579.
- (184) Saveliev, A., Everett, C., Sharpe, T., Webster, Z., and Festenstein, R. (2003) DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing, *Nature* 422, 909–913.
- (185) Coffee, B., Zhang, F., Warren, S. T., and Reines, D. (1999) Acetylated histones are associated with FMR1 in normal but not fragile X syndrome cells, *Nat. Genet.* 22, 98–101.
- (186) Coffee, B., Zhang, F., Ceman, S., Warren, S. T., and Reines, D. (2002) Histone modifications depict an aberrantly heterochromatinized FMR1 gene in fragile X syndrome, *Am. J. Hum. Genet.* 71, 923–932.
- (187) Wang, Y. H., and Griffith, J. (1995) Expanded CTG triplet blocks from the myotonic dystrophy gene create the strongest known natural nucleosome positioning elements, *Genomics* 25, 570–573.
- (188) Wang, Y. H., Amirhaeri, S., Kang, S., Wells, R. D., and Griffith, J. D. (1994) Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene, *Science* 265, 669–671.
- (189) Godde, J. S., and Wolffe, A. P. (1996) Nucleosome assembly on CTG triplet repeats, *J. Biol. Chem.* 271, 15222–15229.

- (190) Mulvihill, D. J., Edamura, K. N., Hagerman, K. A., Pearson, C. E., and Wang, Y.-H. (2005) Effect of CAT or AGG interruptions and CpG methylation on nucleosome assembly upon trinucleotide repeats on spinocerebellar ataxia, type 1 and fragile X syndrome, *J. Biol. Chem.* 280, 4498–4503.
- (191) Wang, Y. H., Gellibolian, R., Shimizu, M., Wells, R. D., and Griffith, J. (1996) Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes, *J. Mol. Biol.* 263, 511–516.
- (192) Godde, J. S., Kass, S. U., Hirst, M. C., and Wolffe, A. P. (1996) Nucleosome assembly on methylated CGG triplet repeats in the fragile X mental retardation gene 1 promoter, *J. Biol. Chem.* 271, 24325–24328.
- (193) Hu, Y., Chopra, V., Chopra, R., Locascio, J. J., Liao, Z., Ding, H., Zheng, B., Matson, W. R., Ferrante, R. J., and Rosas, H. D. (2011) Transcriptional modulator H2A histone family, member Y (H2AFY) marks Huntington disease activity in man and mouse, *Proc. Natl. Acad. Sci. USA* 108, 17141–17146.
- (194) Butler, R., and Bates, G. P. (2006) Histone deacetylase inhibitors as therapeutics for polyglutamine disorders, *Nat. Rev. Neurosci.* 7, 784–796.
- (195) Kazantsev, A. G., and Thompson, L. M. (2008) Therapeutic application of histone deacetylase inhibitors for central nervous system disorders, *Nat. Rev. Drug. Discov.* 7, 854–868.
- (196) Shimizu, M., Fujita, R., Tomita, N., Shindo, H., and Wells, R. D. (2001) Chromatin structure of yeast minichromosomes containing triplet repeat sequences associated with human hereditary neurological diseases, *Nucleic Acids Res. Suppl.* 71–72.

- (197) Tomita, N., Fujita, R., Kurihara, D., Shindo, H., Wells, R. D., and Shimizu, M. (2002) Effects of triplet repeat sequences on nucleosome positioning and gene expression in yeast minichromosomes, *Nucleic Acids Res. Suppl.* 231–232.
- (198) Liu, Y., and Wilson, S. H. (2012) DNA base excision repair: a mechanism of trinucleotide repeat expansion, *Trends Biochem. Sci.* 37, 162–172.
- (199) Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., and Hudson, T. (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member, *Cell* 69, 385.
- (200) Wang, Y.-H. (2006) Chromatin structure of human chromosomal fragile sites, *Cancer Lett.* 232, 70–78.
- (201) Kornberg, R. D. (1974) Chromatin structure: a repeating unit of histones and DNA, *Science* 184, 868–871.
- (202) Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., and Klug, A. (1984) Structure of the nucleosome core particle at 7 Å resolution, *Nature* 311, 532–537.
- (203) Hayes, J. J., Tullius, T. D., and Wolffe, A. P. (1990) The structure of DNA in a nucleosome, *Proc. Natl. Acad. Sci. USA* 87, 7405–7409.
- (204) Hayes, J. J., Clark, D. J., and Wolffe, A. P. (1991) Histone contributions to the structure of DNA in the nucleosome, *Proc. Natl. Acad. Sci. USA* 88, 6829–6833.
- (205) Wang, Y. H., and Griffith, J. (1996) Methylation of expanded CCG triplet

- repeat DNA from fragile X syndrome patients enhances nucleosome exclusion, *J. Biol. Chem.* 271, 22937–22940.
- (206) Hayes, J. J., and Lee, K. M. (1997) *In vitro* reconstitution and analysis of mononucleosomes containing defined DNAs and proteins, *Methods* 12, 2–9.
 - (207) Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M., and Richmond, T. J. (1997) Characterization of nucleosome core particles containing histone proteins made in bacteria, *J. Mol. Biol.* 272, 301–311.
 - (208) Shrader, T. E., and Crothers, D. M. (1989) Artificial nucleosome positioning sequences, *Proc. Natl. Acad. Sci. USA* 86, 7418–7422.
 - (209) Ober, M., and Lippard, S. J. (2007) Cisplatin damage overrides the predefined rotational setting of positioned nucleosomes, *J. Am. Chem. Soc.* 129, 6278–6286.
 - (210) Studitsky, V. M., Clark, D. J., and Felsenfeld, G. (1994) A histone octamer can step around a transcribing polymerase without leaving the template, *Cell* 76, 371–382.
 - (211) Metzenberg, S. (1996) On the formation of nucleosomes within the FMR1 trinucleotide repeat, *Am. J. Hum. Genet.* 59, 252–253.
 - (212) Satchwell, S. C., and Travers, A. A. (1989) Asymmetry and polarity of nucleosomes in chicken erythrocyte chromatin, *EMBO J.* 8, 229–238.
 - (213) Sivolob, A. V., and Khrapunov, S. N. (1995) Translational positioning of nucleosomes on DNA: the role of sequence-dependent isotropic DNA bending stiffness, *J. Mol. Biol.* 247, 918–931.
 - (214) Chastain, P. D., and Sinden, R. R. (1998) CTG repeats associated with human

genetic disease are inherently flexible, *J. Mol. Biol.* 275, 405–411.

- (215) Lowary, P. T., and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning, *J. Mol. Biol.* 276, 19–42.
- (216) Hsu, Y. Y., and Wang, Y.-H. (2002) Human fragile site FRA16B DNA excludes nucleosomes in the presence of distamycin, *J. Biol. Chem.* 277, 17315–17319.
- (217) Eberhart, D. E., and Warren, S. T. (1996) Nuclease sensitivity of permeabilized cells confirms altered chromatin formation at the fragile X locus, *Somat. Cell. Mol. Genet.* 22, 435–441.
- (218) Wang, Y. H., and Griffith, J. (1996) Methylation of expanded CCG triplet repeat DNA from fragile X syndrome patients enhances nucleosome exclusion, *J. Biol. Chem.* 271, 22937–22940.
- (219) Lutter, L. C. (1979) Precise location of DNase I cutting sites in the nucleosome core determined by high resolution gel electrophoresis, *Nucleic Acids Res.* 6, 41–56.
- (220) Fox, K. R. (1998) DNase I Footprinting, *Meth. Mol. Biol.* 90, 1–22.
- (221) Vitolo, J. M., Thiriet, C., and Hayes, J. J. (2001) DNase I and hydroxyl radical characterization of chromatin complexes, *Curr. Protoc. Mol. Biol. Chapter 21*, Unit 21.4.
- (222) Lutter, L. C. (1989) Digestion of nucleosomes with deoxyribonucleases I and II, *Meth. Enzymol.* 170, 264–269.
- (223) Drew, H. R., and Travers, A. A. (1984) DNA structural variations in the E. coli

tyrT promoter, *Cell* 37, 491–502.

- (224) Drew, H. R. (1984) Structural specificities of five commonly used DNA nucleases, *J. Mol. Biol.* 176, 535–557.
- (225) Hogan, M. E., Roberson, M. W., and Austin, R. H. (1989) DNA flexibility variation may dominate DNase I cleavage, *Proc. Natl. Acad. Sci. USA* 86, 9273–9277.
- (226) Brukner, I., Jurukovski, V., and Savic, A. (1990) Sequence-dependent structural variations of DNA revealed by DNase I, *Nucleic Acids Res.* 18, 891–894.
- (227) Davis, W. B., Bjorklund, C. C., and Deline, M. (2012) Probing the effects of DNA-protein interactions on DNA hole transport: the N-terminal histone tails modulate the distribution of oxidative damage and chemical lesions in the nucleosome core particle, *Biochemistry* 51, 3129–3142.
- (228) Riley, D., and Weintraub, H. (1978) Nucleosomal DNA is digested to repeats of 10 bases by exonuclease III, *Cell* 13, 281–293.
- (229) Prunell, A. (1983) Periodicity of exonuclease III digestion of chromatin and the pitch of deoxyribonucleic acid on the nucleosome, *Biochemistry* 22, 4887–4894.
- (230) Archer, T. K., and Ricci, A. R. (1999) Exonuclease III as a probe of chromatin structure in vivo, *Meth. Enzymol.* 304, 584–599.
- (231) Linxweiler, W., and Hörz, W. (1982) Sequence specificity of exonuclease III from *E. coli*, *Nucleic Acids Res.* 10, 4845–4859.
- (232) Jain, S. S., and Tullius, T. D. (2008) Footprinting protein-DNA complexes

- using the hydroxyl radical, *Nat. Protoc.* 3, 1092–1100.
- (233) Tullius, T. D., and Dombroski, B. A. (1986) Hydroxyl radical “footprinting”: high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein, *Proc. Natl. Acad. Sci. USA* 83, 5469–5473.
- (234) Otten, A. D., and Tapscott, S. J. (1995) Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure, *Proc. Natl. Acad. Sci. USA* 92, 5465–5469.
- (235) Klesert, T. R., Otten, A. D., Bird, T. D., and Tapscott, S. J. (1997) Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP, *Nat. Genet.* 16, 402–406.
- (236) Thornton, C. A., Wymer, J. P., Simmons, Z., McClain, C., and Moxley, R. T. (1997) Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene, *Nat. Genet.* 16, 407–409.
- (237) Sobczak, K., and Krzyzosiak, W. J. (2005) CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts, *J. Biol. Chem.* 280, 3898–3910.
- (238) Allemand, J. F., Bensimon, D., Lavery, R., and Croquette, V. (1998) Stretched and overwound DNA forms a Pauling-like structure with exposed bases, *Proc. Natl. Acad. Sci. USA* 95, 14152–14157.
- (239) Randall, G. L., Zechiedrich, L., and Pettitt, B. M. (2009) In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form, *Nucleic Acids Res.* 37, 5568–5577.
- (240) Lionberger, T. A., Demurtas, D., Witz, G., Dorier, J., Lillian, T., Meyhöfer, E.,

- and Stasiak, A. (2011) Cooperative kinking at distant sites in mechanically stressed DNA, *Nucleic Acids Res.* 39, 9820–9832.
- (241) Napierala, M., Michalowski, D., de Mezer, M., and Krzyzosiak, W. J. (2005) Facile FMR1 mRNA structure regulation by interruptions in CGG repeats, *Nucleic Acids Res.* 33, 451–463.
- (242) Jarem, D. A., Wilson, N. R., Schermerhorn, K. M., and Delaney, S. (2011) Incidence and persistence of 8-oxo-7,8-dihydroguanine within a hairpin intermediate exacerbates a toxic oxidation cycle associated with trinucleotide repeat expansion, *DNA Repair (Amst)* 10, 887–896.
- (243) Völker, J., Gindikin, V., Klump, H. H., Plum, G. E., and Breslauer, K. J. (2012) Energy landscapes of dynamic ensembles of rolling triplet repeat bulge loops: implications for DNA expansion associated with disease states, *J. Am. Chem. Soc.* 134, 6033–6044.
- (244) Slean, M. M., Reddy, K., Wu, B., Nichol Edamura, K., Kekis, M., Nelissen, F. H. T., Aspers, R. L. E. G., Tessari, M., Schärer, O. D., Wijmenga, S. S., and Pearson, C. E. (2013) Interconverting conformations of slipped-DNA junctions formed by trinucleotide repeats affect repair outcome, *Biochemistry* 52, 773–785.
- (245) Herr, W. (1985) Diethyl pyrocarbonate: a chemical probe for secondary structure in negatively supercoiled DNA, *Proc. Natl. Acad. Sci. USA* 82, 8009–8013.
- (246) Degtyareva, N. N., Reddish, M. J., Sengupta, B., and Petty, J. T. (2009) Structural studies of a trinucleotide repeat sequence using 2-aminopurine,

Biochemistry 48, 2340–2346.

- (247) Cullis, P., Malone, M., and Merson-Davies, L. (1996) Guanine radical cations are precursors of 7,8-Dihydro-8-oxo-2'-deoxyguanosine but are not precursors of immediate strand breaks in DNA, *J. Am. Chem. Soc.* 118, 2775–2781.
- (248) Bodepudi, V., Shibutani, S., and Johnson, F. (1992) Synthesis of 2'-deoxy-7,8-dihydro-8-oxoguanosine and 2'-deoxy-7,8-dihydro-8-oxoadenosine and their incorporation into oligomeric DNA, *Chem. Res. Toxicol.* 5, 608–617.
- (249) Delaney, S., Delaney, J. C., and Essigmann, J. M. (2007) Chemical-biological fingerprinting: probing the properties of DNA lesions formed by peroxynitrite, *Chem. Res. Toxicol.* 20, 1718–1729.
- (250) Plum, G. E., Grollman, A. P., Johnson, F., and Breslauer, K. J. (1995) Influence of the oxidatively damaged adduct 8-oxodeoxyguanosine on the conformation, energetics, and thermodynamic stability of a DNA duplex, *Biochemistry* 34, 16148–16160.
- (251) Chinyenetere, F., and Jamieson, E. R. (2008) Impact of the oxidized guanine lesion spiroiminodihydantoin on the conformation and thermodynamic stability of a 15-mer DNA duplex, *Biochemistry* 47, 2584–2591.
- (252) Singh, S. K., Szulik, M. W., Ganguly, M., Khutsishvili, I., Stone, M. P., Marky, L. A., and Gold, B. (2011) Characterization of DNA with an 8-oxoguanine modification, *Nucleic Acids Res.* 39, 6789–6801.
- (253) Maxam, A. M., and Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages, *Meth. Enzymol.* 65, 499–560.
- (254) Paiva, A. M., and Sheardy, R. D. (2005) The influence of sequence context and

- length on the kinetics of DNA duplex formation from complementary hairpins possessing (CNG) repeats, *J. Am. Chem. Soc.* **127**, 5581–5585.
- (255) Pearson, C. E., Tam, M., Wang, Y.-H., Montgomery, S. E., Dar, A. C., Cleary, J. D., and Nichol, K. (2002) Slipped-strand DNAs formed by long (CAG)*(CTG) repeats: slipped-out repeats and slip-out junctions, *Nucleic Acids Res.* **30**, 4534–4547.
- (256) Mitas, M. (1997) Trinucleotide repeats associated with human disease, *Nucleic Acids Res.* **25**, 2245–2254.
- (257) Huertas, D., Bellolell, L., Casasnovas, J. M., Coll, M., and Azorín, F. (1993) Alternating d(GA)_n DNA sequences form antiparallel stranded homoduplexes stabilized by the formation of G.A base pairs, *EMBO J.* **12**, 4029–4038.
- (258) Bhattacharya, P. K., Cha, J., and Barton, J. K. (2002) ¹H NMR determination of base-pair lifetimes in oligonucleotides containing single base mismatches, *Nucleic Acids Res.* **30**, 4740–4750.
- (259) Marky, L. A., and Breslauer, K. J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves, *Biopolymers* **26**, 1601–1620.
- (260) McAuley-Hecht, K. E., Leonard, G. A., Gibson, N. J., Thomson, J. B., Watson, W. P., Hunter, W. N., and Brown, T. (1994) Crystal structure of a DNA duplex containing 8-hydroxydeoxyguanine-adenine base pairs, *Biochemistry* **33**, 10266–10270.
- (261) Oda, Y., Uesugi, S., Ikehara, M., Nishimura, S., Kawase, Y., Ishikawa, H., Inoue, H., and Ohtsuka, E. (1991) NMR studies of a DNA containing 8-

hydroxydeoxyguanosine, *Nucleic Acids Res.* 19, 1407–1412.

- (262) Lee, B. J., Barch, M., Castner, E. W., Völker, J., and Breslauer, K. J. (2007) Structure and dynamics in DNA looped domains: CAG triplet repeat sequence dynamics probed by 2-aminopurine fluorescence, *Biochemistry* 46, 10756–10766.
- (263) SantaLucia, J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc. Natl. Acad. Sci. USA* 95, 1460–1465.
- (264) Petruska, J., Hartenstine, M. J., and Goodman, M. F. (1998) Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease, *J. Biol. Chem.* 273, 5204–5210.
- (265) Heidenfelder, B. L., Makhov, A. M., and Topal, M. D. (2003) Hairpin formation in Friedreich's ataxia triplet repeat expansion, *J. Biol. Chem.* 278, 2425–2431.
- (266) Ohshima, K., and Wells, R. D. (1997) Hairpin formation during DNA synthesis primer realignment in vitro in triplet repeat sequences from human hereditary disease genes, *J. Biol. Chem.* 272, 16798–16806.
- (267) Møllersen, L., Rowe, A. D., Larsen, E., Rognes, T., and Klungland, A. (2010) Continuous and periodic expansion of CAG repeats in Huntington's disease R6/1 mice, *PLoS Genet.* 6, e1001242.
- (268) Entezam, A., Lokanga, A. R., Le, W., Hoffman, G., and Usdin, K. (2010) Potassium bromate, a potent DNA oxidizing agent, exacerbates germline repeat expansion in a fragile X premutation mouse model, *Hum. Mutat.* 31, 611–616.

- (269) Mergell, B., Everaers, R., and Schiessel, H. (2004) Nucleosome interactions in chromatin: Fiber stiffening and hairpin formation, *Phys. Rev. E* **70**, 011915.
- (270) Nilsen, H., Lindahl, T., and Verreault, A. (2002) DNA base excision repair of uracil residues in reconstituted nucleosome core particles, *EMBO J.* **21**, 5943–5952.
- (271) Hinz, J. M., Rodriguez, Y., and Smerdon, M. J. (2010) Rotational dynamics of DNA on the nucleosome surface markedly impact accessibility to a DNA repair enzyme, *Proc. Natl. Acad. Sci. USA* **107**, 4646–4651.
- (272) Menoni, H., Gasparutto, D., Hamiche, A., Cadet, J., Dimitrov, S., Bouvet, P., and Angelov, D. (2007) ATP-dependent chromatin remodeling is required for base excision repair in conventional but not in variant H2A.Bbd nucleosomes, *Mol. Cell. Biol.* **27**, 5949–5956.
- (273) Odell, I. D., Barbour, J.-E., Murphy, D. L., Della-Maria, J. A., Sweasy, J. B., Tomkinson, A. E., Wallace, S. S., and Pederson, D. S. (2011) Nucleosome disruption by DNA ligase III-XRCC1 promotes efficient base excision repair, *Mol. Cell. Biol.* **31**, 4623–4632.
- (274) Rodriguez, Y., and Smerdon, M. J. (2013) The structural location of DNA lesions in nucleosome core particles differentially influence accessibility of base excision repair enzymes, *J. Biol. Chem.* DOI: 10.1074/jbc.M112.441444
- (275) Volle, C. B., and Delaney, S. (2012) CAG/CTG repeats alter the affinity for the histone core and the positioning of DNA in the nucleosome, *Biochemistry* **51**, 9814–9825.
- (276) Meersseman, G., Pennings, S., and Bradbury, E. M. (1992) Mobile

- nucleosomes- a general behavior, *EMBO J.* 11, 2951–2959.
- (277) Wong, H. M., and Huppert, J. L. (2009) Stable G-quadruplexes are found outside nucleosome-bound regions, *Mol. Biosyst.* 5, 1713.
 - (278) Halder, K., Halder, R., and Chowdhury, S. (2009) Genome-wide analysis predicts DNA structural motifs as nucleosome exclusion signals, *Mol. Biosyst.* 5, 1703–1712.
 - (279) Huppert, J. L. (2008) Hunting G-quadruplexes, *Biochimie* 90, 1140–1148.
 - (280) Aoyagi, S., and Hayes, J. J. (2002) hSWI/SNF-catalyzed nucleosome sliding does not occur solely via a twist-diffusion mechanism, *Mol. Cell. Biol.* 22, 7484–7490.
 - (281) Aoyagi, S., Wade, P. A., and Hayes, J. J. (2003) Nucleosome sliding induced by the xMi-2 complex does not occur exclusively via a simple twist-diffusion mechanism, *J. Biol. Chem.* 278, 30562–30568.
 - (282) Menoni, H., Shukla, M. S., Gerson, V., Dimitrov, S., and Angelov, D. (2012) Base excision repair of 8-oxoG in dinucleosomes, *Nucleic Acids Res.* 40, 692–700.
 - (283) Routh, A., Sandin, S., and Rhodes, D. (2008) Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure, *Proc. Natl. Acad. Sci. USA* 105, 8872–8877.
 - (284) Huynh, V. A. T., Robinson, P. J. J., and Rhodes, D. (2005) A method for the in vitro reconstitution of a defined “30 nm” chromatin fibre containing stoichiometric amounts of the linker histone, *J. Mol. Biol.* 345, 957–968.
 - (285) Robinson, P. J. J., Fairall, L., Huynh, V. A. T., and Rhodes, D. (2006) EM

measurements define the dimensions of the “30-nm” chromatin fiber: evidence for a compact, interdigitated structure, *Proc. Natl. Acad. Sci. USA* 103, 6506–6511.

Appendix A: Protocols

Isolation of Nucleosome Core Particles

Nucleosome core particles (NCPs) are composed of 146 base pairs of DNA wrapped around the core histone octamer. The histone core is described as two dimers of dimers: histones H3 and H4 form a dimer and then dimerize with a second H3/H4 dimer; histones H2A and H2B form a dimer and the H3/4 dimer recruits two H2A/H2B dimers to form the histone core. The NCP is the most basic unit of chromatin packaging in cells. In the cell NCPs are further packaged using an additional histone protein, the linker histones H1/H5 (H5 is only found in avian erythrocytes).

We can isolate chromatin from natural sources, replace the native DNA with our DNA of interest, and use these constructs to ask questions about our DNA of interest in a more cellular context. In the Delaney lab, we isolate NCPs from chicken red blood cells (which are nucleated, unlike their mammalian counterparts), but these methods would need only minimal modification for other biological sources. This protocol is written such that the entire yield of isolated nuclei is used for NCP isolation, but the isolated nuclei can be aliquoted, stored at -20°C, and used as needed. Just make sure to adjust the volumes of each added solution accordingly.

Materials:

500 mL Chicken Blood (Pel Freeze Biologicals, comes stored in sodium citrate)

Buffer A (pH 7.5):

15 mM Tris HCl

60 mM KCl

15 mM NaCl

0.15 mM spermine

0.5 mM spermidine

Solution X (pH 7.5):

Buffer A

340 mM sucrose

*0.2 mM phenylmethylsulfonyl fluoride (PMSF)

*1 mM benzamidine

2 mM EDTA

0.5 mM EGTA

*15 mM β -mercaptoethanol

(* added fresh before use)

Solution L (pH 7.5):

Solution X

0.1% (v:v) Nonidet P40

Solution N (pH 7.5):

Buffer A

340 mM sucrose

*0.2 mM phenylmethylsulfonyl fluoride (PMSF)

*15 mM β -mercaptoethanol

(* added fresh before use)

Solution R (pH 8):

10 mM Tris

0.2 mM EDTA

*0.2 mM phenylmethanesulfonyl fluoride (PMSF)

(* added fresh before use)

Solution S (pH 7.5):

630 mM NaCl

20 mM Tris

0.2 mM EDTA

*0.2 mM phenylmethanesulfonyl fluoride (PMSF)

(* added fresh before use)

Solution D (pH 7.5):

10 mM Tris

0.2 mM EDTA

*0.2 mM phenylmethanesulfonyl fluoride (PMSF)

(* added fresh before use)

4 M NaCl

Micrococcal nuclease

1 M CaCl_2

100 mM EDTA

2 heavy-duty centrifuge bottles

Large capacity centrifuge (Cane lab)

Bench top centrifuges

Concentrators

Tall glass chromatography column

Sepharose 4B

20% ethanol

4-20% tris-glycine SDS-PAG (Invitrogen)

1X tris-glycine running buffer (pH 8.3)

10X from Invitrogen

25 mM Tris

250 mM Glycine

0.1% (w/v) SDS

2X SDS loading buffer

Coomassie Brilliant Blue

0.25 g Coomassie brilliant blue

50 mL methanol

40 mL H₂O

10 mL glacial acetic acid

Coomassie Destain

50 mL methanol

40 mL H₂O

10 mL glacial acetic acid

PART A: Isolation of Nuclei

1. Pour chicken blood into two heavy-duty centrifuge bottles.
2. Centrifuge for 12 minutes; 6000 rpm, 4°C.

3. Discard supernatant (plasma) and resuspend each pellet in 250 mL cold solution **X**
4. Repeat steps 2 and 3 until the supernatant is clean (2-3 rounds).
5. Resuspend each pellet in 250 mL of cold solution **L** to lyse the cell membranes (but not the nuclear membranes).
6. Centrifuge for 5 minutes, 6000 rpm, 4°C.
7. Discard supernatant and resuspend each pellet in 250 mL cold solution **X**.
8. Repeat steps 6 and 7 three times, or until all the hemoglobin is washed away and the pellet (washed nuclei) is white.
9. Resuspend each pellet in 250 mL of cold solution **N**.
10. Centrifuge for 5 minutes, 6000 rpm, 4°C.
11. Discard the supernatant and either resuspend each pellet in 50 mL solution **N** (100 mL total; for immediate use) or 50 mL 50% solution **N**, 50% glycerol (100 mL total; for storage at -20°C).

PART B: Isolation of Nucleosome Core Particles

1. If the nuclei are coming out of the freezer, add 400 mL solution **N** and re-pellet (centrifuge for 5 minutes, 6000 rpm, 4°C) the nuclei to remove the glycerol; resuspend both pellets in 100 mL (total) solution **N**. If the nuclei have not been frozen skip this step.
2. Place resuspended nuclei in an Erlenmeyer flask.
3. Add 1 M CaCl_2 to the nuclei to a final concentration of 1 mM (~100 μL).
4. Incubate at 37°C for 10 minutes.
5. Add micrococcal nuclease to 200 U/mL; digest for 5 minutes at 37°C.

6. Quench digestion with 100 mM EDTA to a final concentration of 2 mM (~2 mL)
7. Pour digested nuclei into centrifuge bottles.
8. Centrifuge for 12 minutes; 6000 rpm, 4°C.
9. Discard supernatant and resuspend the nuclei in 50 mL solution **R** (the low salt concentration of solution **R** lyses the nuclear membrane).
10. Let the suspension shake overnight in the cold room to bring the digested chromatin into solution. You should be able to see the nuclei lyse and the chromatin come out from the nucleus; the nuclei are very sticky and will clump together (it will look a bit like boogers) and the chromatin looks like fluff surrounding the sticky, nuclear mess.
11. Make sure the Sepharose column is equilibrated with solution **S**. The column is either stored dry, in which case it needs to be filled with Sepharose or contains washed Sepharose. In either case, the 20% ethanol that the Sepharose is stored in needs to be removed. Flush the column through with at least 300 mL of solution **S**. The column can then be stored overnight at 4°C.

PART C: Purification of Nucleosome Core Particles

1. Pour the lysed nuclei into centrifuge bottles.
2. Centrifuge for 12 minutes; 6000 rpm, 4°C.
3. SAVE THE SUPERNATANT, which contains the digested chromatin, and discard the pellet, which contains the remaining nuclear components.
4. Using the 25 mL Amicon filters, concentrate the chromatin to less than 5 mL.
5. Measure the volume of the chromatin solution and store the flask at 4°C.

6. Add 4 M NaCl dropwise to a final concentration of exactly 0.65 M. The solution should turn clear to cloudy to clear. This dissociates the linker histones from the NCP.
7. Apply the solution to a Sepharose 4B column pre-equilibrated solution S, using the flow adaptor. Collect and save the flow through. DO NOT LET THE COLUMN RUN DRY!!!!!! Once all the chromatin has been applied to the column, run the column with solution S, collecting a 20 mL fraction, then 20 5 mL fractions, then at least 2 more 20 mL fractions. Store the fractions at 4°C.
8. Clean the column by flushing it with at least 300 mL of 20% ethanol or, if the column will not be needed for a while, run the column dry and remove the Sepharose. Save the Sepharose since it can be used again; store it at 4°C in 20% ethanol.

PART D: Analysis of Column Fractions

1. Remove 500 µL from as many fractions as you can fit on an Invitrogen pre-poured 4-20% SDS-PAGE gel. Make sure to get good coverage across the fractions. Use the small Amicon concentrators to reduce the volume to ~20 µL (centrifuge at 14000 rpm, 25 minutes).
2. Mix sample with an equal volume SDS loading buffer.
3. Place gel in chamber and make sure it is properly sealed. Fill the entire gel rig with 1X tris-glycine buffer.
4. Load 20 µL into each well. Remember to leave a well for a size marker.
5. Run the gel at 125 V for 1.5 to 2 hours.

6. Remove the gel from the plastic casing and place in coomassie stain. Place in shaking incubator at 37 °C for ~5 hours to overnight.
7. Place gel in destain. Let shake overnight.
8. Image gel. Store in water.
9. Analyze column fractions for presence of histones H2A, H2B, H3, H4, and H5.

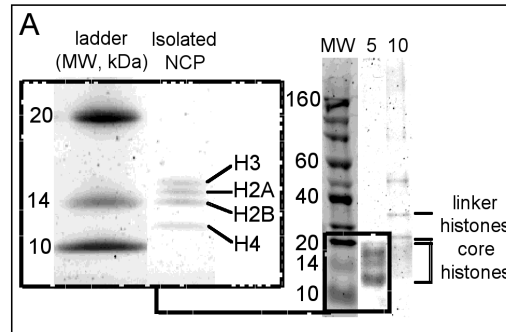


Figure 1: SDS-PAGE analysis of column fractions. Core histones run around 14 kD, while linker histones run at 21.5. Remember that nucleosomes come off the column early because the column is under native conditions. The chromatin is very large compared to the dissociated linker histones and other cellular proteins.

PART E: Second Chromatin Digest

1. Combine the fractions that contain only the core histones and concentrate them in the Amicon concentrators.
2. Repeat the micrococcal nuclease digestion (Part B, steps 3-6).
3. Reapply the digested chromatin to the pre-equilibrated Sepharose column.
4. Repeat the fraction collection and analysis process.

PART F: Storage of Nucleosomes

1. Combine all column fractions that contain only the core histones (no linker histones).
2. Dialyze against solution **D**.
3. Concentrate the sample using an Amicon concentrator.
4. Perform a Bradford assay to determine the mg/mL concentration of NCPs. The concentration should be 1 mg/mL or higher.
5. Store at 4°C.

Agarose and SDS-Agarose Gel Electrophoresis

Another form of electrophoresis used to separate DNA uses agarose in the gel matrix. This electrophoretic method is primarily used to separate larger pieces of DNA, usually PCR products and digested fragments isolated from plasmids. It can also be used to separate the DNA component from the protein component of nucleosome core particles, when the denaturant sodium dodecyl sulfate (SDS) is added to the loading dye. DNA will separate based on its size and then a chemical intercalater can be used to visualize each band.

Materials:

Agarose

1X TBE (National Diagnostics)

Agarose gel apparatus (horizontal electrophoresis unit)

Loading dye

SDS loading dye (Invitrogen)

Non-denaturing dye (See native PAGE protocol)

Appropriate size marker (10 bp, 100 bp, or 1 kb)

10% SDS

Visualization intercalater (Both used in lab are light sensitive; make sure they are covered in aluminum foil)

Sybr gold (10 μ L stock in 100 mL 1X TBE)

10 mg/mL Ethidium Bromide stock

Caution!!! Use proper personal protective equipment (gloves and goggles) when working with the DNA intercalaters. They are potential carcinogens (they intercalate into DNA; you don't want them to intercalate into yours).

PART A: Agarose matrix preparation

1. Measure out agarose for the correct percentage gel (we use a 1% gel: 1 g per 100 mL). Pour into an Erlenmeyer.
2. Add 100 mL 1X TBE to agarose.
3. Heat in microwave until matrix just begins to boil.
4. If you plan to stain with ethidium bromide add 5 μ L now, and swirl to mix. DO NOT add Sybr Gold to the gel matrix.
5. Pour the matrix into the gel tray, sealed in the electrophoresis rig.

PART B: Running the gel

1. Once the gel has hardened, turn the gel such that the wells are at the negative electrode (the DNA backbone is negatively charged and will migrate toward the positive electrode).
2. Mix equal volumes of sample and dye.
 - a. If the sample is resuspended in SDS dye (nucleosome samples), it needs to be heated at 90°C for ~10 minutes before loading. 10% SDS will need to be added to an aliquot of the ladders as well (1% SDS in the final sample) and heated as well.
3. Load sample into well.

4. Electrophorese samples at 100 V. The time will depend on the individual sample (~45 minutes for nucleosome samples). Once the samples are properly separated, the gel is ready for imaging.

PART C: Visualization

1. If the sample has been run with ethidium bromide, it can be placed directly in the UV light box and imaged.
2. If the sample needs to be stained with Sybr Gold, pour the dilution into an empty P1000 tip box. Cover with aluminum foil.
 - a. Remove the gel from the tray and place it directly into the Sybr Gold.
 - b. Place it on a shaking table for ~1 hour.
 - c. Remove the gel from the Sybr Gold and place it in the UV light box.
3. Turn on the UV light and camera.
 - a. Make sure the door is closed properly, or the UV light won't come on.
4. Adjust the camera settings (on the light box, not the camera) so that the gel is properly exposed and capture and save an image.

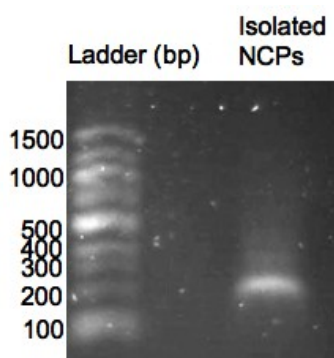
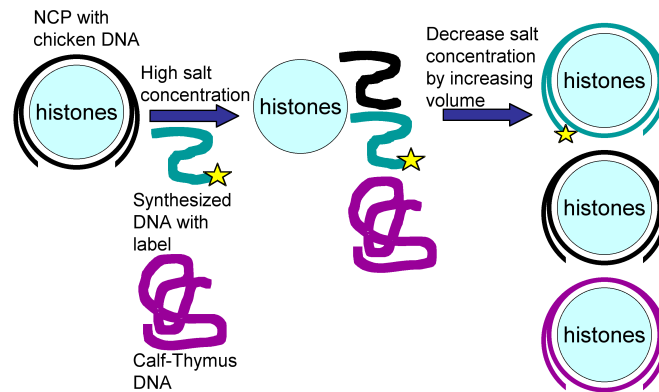


Figure 1: Semi-denaturing agarose gel of an isolated NCP fraction. The DNA runs at ~200 base pair, which is about the length associated with one NCP.

Competitive Exchange of DNA into Nucleosome Core Particles

To determine the relative efficiency of DNA incorporation into the nucleosome, we perform a competitive nucleosome exchange. This is just like the regular exchange procedure, but now there is calf thymus DNA added as a competitor.



Materials:

Purified NCPs

Radiolabeled 146-mer duplex

1 mg/mL CT-DNA

100 μ L 10X Exchange Buffer (EB, pH 8.0), 50 mM Tris can also be used

200 mM Tris

10 mM EDTA

10 mM β -mercaptoethanol

4 M NaCl

Amicon micro-concentrators

5% native gel

PART A: NCP exchange

1. Prepare 4 reactions per DNA sample: 1 duplex control where no NCPs are added to the reaction and 3 replicates with NCPs added.
2. Mix a 50 μ L exchange reaction. Incubate 1 hour at room temperature.
 - a. 12.5 μ L 4 M NaCl
 - b. 5 μ L radiolabeled duplex (~2 pmol or 50 ng)
 - c. 2.5 μ L CT-DNA (1 mg/mL)
 - d. 5 μ L 10X EB
 - e. 25 μ L digested NCPs or buffer for duplex control.
3. Add 16.5 μ L 1X EB. Incubate 1 hour RT.
4. Add 16.5 μ L 1X EB. Incubate 1 hour RT.
5. Add 17 μ L 1X EB. Incubate 1 hour RT.
6. Add 100 μ L 1X EB. Incubate 15 min RT.
7. Add 100 μ L 1X EB. Incubate 15 min RT.
8. Add 100 μ L 1X EB. Incubate 15 min RT.
9. Add 100 μ L 1X EB. Incubate 1 hour 37°C.
10. Concentrate samples, 30 min, 14000 rpm.
11. Add 1/6th the total volume of 6X non-denaturing dye.
12. Load ~ 12 μ L of each sample into a small 5% native gel; run 4 hrs at 100V.
13. Disassemble gel and place it in a cassette for ~30 minutes.
14. Image.

- a. Since these will be used for quantification, it's important that it not be over-exposed.

PART B: Quantification

1. Use the box quantitation method to draw boxes around the duplex and NCP.
Make sure to include a background box. Copy and paste the boxes so that they are all the same size.
 - a. Repeat for all samples.
2. Export the counts in each box as a text file.
3. Open the file in Excel.
4. Calculate the ratio of DNA in the NCP band to the DNA remaining as free duplex for each sample. Then calculate the standard deviation.
 - a. To get the most accurate results, it's best to repeat this experiment on a second and third day. Then calculate the standard error for all of the experiments.

Hydroxyl Radical Footprinting

Hydroxyl radical footprinting is a technique that allows reaction along a DNA backbone, except where a protein is bound, which protects the DNA backbone from cleavage. This allows us to map exactly where a protein contacts the DNA backbone. The hydroxyl radical is generated by the Fenton reaction, in which $[\text{Fe(II)(EDTA)}]^{2-}$ reacts with H_2O_2 producing the $(\bullet\text{OH})$, $(-\text{OH})$, and $[\text{Fe(II)(EDTA)}]^{1-}$. Sodium ascorbate is added to the reaction to reduce the $[\text{Fe(II)(EDTA)}]^{1-}$.

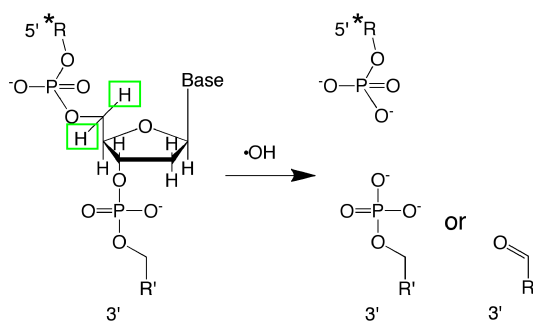


Figure 1: Hydroxyl radical footprinting creates a strand break through hydrogen abstraction from the C5'.

Hydrogen atom abstraction can occur at either of the hydrogens at C5'. In a minority of reactions, it can also occur at C4'. During the reaction, 90% of the strand breaks will form the strand break indicated in the figure above. It's important to note that the radiolabeled fragment visualized during phosphorimager is 1 nucleotide shorter than the position where the reaction occurred (reaction at nucleotide 50 will result in a 49 nucleotide fragment).

Unfortunately, the process of optimizing the footprinting reaction for a new substrate can be quite difficult. The protocol that follows is a good place to start, but may need some tweaking.

Materials:

Radiolabeled DNA/protein complex in Sodium Phosphate or 10 mM or less Tris buffer

(reaction will not occur efficiently in buffers more than 10 mM Tris)

Radiolabeled DNA without protein bound

$(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$

EDTA

Labeled aluminum foil caps (these are labeled with reagent names; do not discard them)

Sodium ascorbate

H_2O_2

Thiourea

1 mg/mL CT-DNA

10% SDS

Phenol:chloroform

3 M sodium acetate

Ethanol

Denaturing polyacrylamide gel and accompanying supplies

PART A: Reagent preparation**1. Iron(II)EDTA**

- a. Prepare a 20 mM Fe(II) stock by dissolving 40 mg $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ in 5 mL H_2O .
- b. Mix equal volumes of the 20 mM iron(II) stock and 40 mM EDTA.

- c. Dispense aliquots into Eppendorf tubes covered in aluminum foil (Fe(II)EDTA is light sensitive) and store them at -20°C.
 - d. Use a fresh tube each day you perform reactions.
 - e. Make a 1:10 dilution for use in the reactions.
2. Sodium ascorbate
- a. Prepare a 10 mM solution by dissolving 20 mg in 10 mL of water.
 - b. Dispense aliquots into Eppendorf tubes covered in aluminum foil (sodium ascorbate is light sensitive) and store them at -20°C.
 - c. Use a fresh tube each day you perform reactions.
3. H₂O₂
- a. Prepare a 0.6% (w:v) by mixing 98 µL water with 2 µL 30% H₂O₂. This can be made fresh before starting the reactions.
4. Stop solution
- a. Prepare a 100 mM thiourea solution by dissolving 7.6 mg in 1 mL of water.
 - b. Store at 4°C.

PART B: Hydroxyl radical footprinting of protein binding

1. Prepare samples for footprinting reaction. Make sure to perform the experiments on both free duplex and protein bound samples.
 - a. Control reaction, with no footprinting reagents added to the sample, and 3 experimental reactions incubated for 1, 2, or 5 minutes.
 - b. 9.5 µL DNA in each Eppendorf tube; add 0.5 µL 1 mg/mL CT-DNA.

2. To initiate the reaction, spot 1 μL of the Fe(II)EDTA , sodium ascorbate, and 0.6% H_2O_2 individually on the wall of the Eppendorf tube. Use a pipette tip to mix all the drops together and immediately vortex the sample to add the reagents to the DNA solution.
3. Incubate the reaction at room temperature for the appropriate amount of time.
4. Quench the reaction with 1 μL of thiourea.
5. Add 1.5 μL 10% SDS and heat at 90°C for 5 minutes.
6. Add 50 μL phenol:chloroform and vortex thoroughly (the mixture should turn white).
7. Centrifuge 1 minute at max speed.
8. Remove the aqueous phase (should be on top) being careful not to collect any of the insoluble layer or the organic phase.
 - a. Check both fractions for radioactivity; the radiolabeled DNA should be in the aqueous phase.
 - b. For most samples one round of phenol chloroform extraction is sufficient, but if you are concerned about carry over from the insoluble layer or organic phase, the process can be repeated.
9. Add 1.5 μL 3 M sodium acetate to the DNA. Vortex.
10. Add 500 μL ethanol and place in dry ice for at least 20 minutes.
11. Centrifuge 14,000 rpm, 5 minutes, 4°C .
12. Remove the supernatant, being careful not to disturb the pellet.
13. Dry the DNA. It is now ready to be counted, resuspended in denaturing dye and run out in a denaturing gel.

An excellent protocol was written by Jain and Tullius in Nature Protocols.¹ The footprinting reactions on NCPs can be found in the Ober and Lippard papers.^{2,3}

(1) Jain, S. S.; Tullius, T. D.; *Nat Protoc* **2008**, *3*, 1092.

(2) Ober, M.; Lippard, S. J.; *J Am Chem Soc* **2007**, *129*, 6278.

(3) Ober, M.; Lippard, S. J.; *J Am Chem Soc* **2008**, *130*, 2851.

Hydroxyl Radical Footprinting of DNA Within the Nucleosome

Hydroxyl radical footprinting is a technique that allows reaction along a DNA backbone, except where a protein is bound, which protects the DNA backbone from cleavage. This allows us to map exactly where a protein contacts the DNA backbone. The hydroxyl radical is generated by the Fenton reaction, in which $[\text{Fe(II)(EDTA)}]^{2-}$ reacts with H_2O_2 producing the $(\bullet\text{OH})$, $(-\text{OH})$, and $[\text{Fe(II)(EDTA)}]^{1-}$. Sodium ascorbate is added to the reaction to reduce the $[\text{Fe(II)(EDTA)}]^{1-}$.

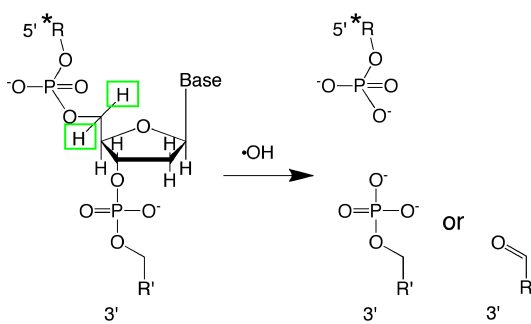


Figure 1: Hydroxyl radical footprinting creates a strand break through hydrogen abstraction from the C5'.

Hydrogen atom abstraction can occur at either of the hydrogens at C5'. In a minority of reactions, it can also occur at C4'. During the reaction, 90% of the strand breaks will form the strand break indicated in the figure above. It's important to note that the radiolabeled fragment visualized during phosphorimager is 1 nucleotide shorter than the position where the reaction occurred (reaction at nucleotide 50 will result in a 49 nucleotide fragment).

Unfortunately, the process of optimizing the footprinting reaction for a new substrate can be quite difficult. The protocol that follows is a good place to start, but may need some tweaking.

Materials:

Radiolabeled Duplex

NCPs

4 M NaCl

50 mM Tris

5 mM Tris

0.5 mL and 15 mL Amicon concentrators

Footprinting reagents

See previous HRF protocol for preparation

50% glycerol

100 mM EDTA

1 mg/mL CT-DNA

5% native gel and accompanying supplies

Phenol:Chloroform

100% Ethanol

7.5 M ammonium acetate

8% Denaturing polyacrylamide gel and accompanying supplies

PART A: Nucleosome preparation

1. Prepare exchange reaction
 - a. 12.5 μ L 4M NaCl
 - b. 25 μ L 1 mg/mL NCPs
 - c. 5 μ L 50 mM Tris

- d. 7.5 μ L radiolabeled duplex (10-100 pmol)
2. Incubate 1 h at room temperature.
3. Add 16.5 μ L 5 mM Tris. Incubate 1 h at room temperature.
 - a. Repeat step 3 twice more, adding a total of 50 μ L 5 mM Tris, bringing the reaction volume to 100 μ L
4. Add 100 μ L 5 mM Tris. Incubate 10 min at room temperature.
 - a. Repeat step 4 twice more, bringing the reaction volume to 400 μ L
5. Add 100 μ L 5 mM Tris. Incubate 1 h at 37 °C.
 - a. This brings the final reaction volume to 500 μ L
6. Transfer each reaction to a 0.5 mL concentrator
 - a. Centrifuge at maximum speed, 30 min, 4 °C

PART B: HRF reactions on nucleosomes

1. Transfer 52.5 μ L of the exchanged nucleosomes to a clean eppendorf tube
2. Add 7.5 μ L of each reagent to the side of the eppendorf tube; do not mix
3. Vortex the sample; this will mix both the reagents and the sample, starting the reaction
4. Incubate at RT for 5 minutes
5. Quench with 8 μ L 100 mM EDTA and 8 μ L 50% glycerol
6. Immediately run your entire sample in an 8 mm, 5% native gel
 - a. 150 V, 3 h
7. Expose the gel for ~20 minutes; use the image to identify your nucleosome band
8. Excise the nucleosomes and crush and soak them overnight at 37 °C

9. Centrifuge the crush and soak, collecting the supernatant
10. Concentrate the supernatant using a 15 mL amicon concentrator
 - a. $3220 \times g$, 50 min, leaves ~0.5 mL of recovered nucleosomes
11. Add SDS to 0.01%; heat at 90 °C for ~10 min
12. Extract samples twice with 400 μ L of phenol:chloroform
13. Ethanol precipitate the DNA with 50 μ L 7.5 M ammonium acetate and 700 μ L ethanol
14. Samples are now ready for evaluation in an 8% denaturing gel

PART C: HRF reactions on free duplex

1. Bring 5 pmol of free duplex to 52.5 μ L
2. Add and mix the HRF reagent as described for the nucleosome samples
3. Incubate at RT for 2 minutes
4. Quench with 8 μ L 50% glycerol and 8 μ L 1 mg/mL CT-DNA
5. Ethanol precipitate the DNA with 50 μ L 7.5 M ammonium acetate and 500 μ L ethanol
6. Samples are now ready for evaluation in an 8% denaturing gel

PART D: Evaluating the reactions

1. Pour a long, 8% denaturing gel (pour it the day you are going to use it)
 - a. The long gels leak. A lot. Make sure to have at least 100 mL of matrix even though the gel only takes about 40 mL.
2. Pre-run the gel at 80 W

3. Count your samples
 - a. Prepare each sample at 5000 counts/ μ L in 100% formamide
4. Load 15,000 counts per well (3 μ L)
 - a. You'll only be able to run ~10 samples per gel. The first two lanes on either side won't migrate properly
5. Run at 65 W for 2.5 hours (146-mers)
 - a. This will put the parent band just below the filter paper used in the gel dryer
6. Dry the gel
 - a. Using saran wrap, pull the gel off the glass plates
 - b. Cut a piece of filter paper in half along the long axis
 - c. Align the filter paper with the bottom edge of the gel and place flat against the gel
 - d. Cut off the top of the gel (the area not covered by the filter paper)
 - i. Check for radioactivity and discard
 - e. Trim the edges of the filter paper and place in the gel dryer
7. Use the gel to expose a phosphor plate for 4 days

PART E: Processing the Data

1. Import the gel file into the ImageLab software.
2. Define each lane, then each band within the lane.
 - a. To make sure you get the area under the curve you'll have to adjust the band manually using the lane profile.

3. Export the intensity of each band as an excel file.
4. Using the marker lane, determine which band corresponds to which base number.
5. Apply a moving average with a window of five to the data. This is the data that will be imported into Origin to determine the periodicity, as well as the translational and rotational positions.

PART F: Determining the Periodicity

1. Import the data into Origin and generate a graph. This should appear as an oscillating line.
2. Using the background subtraction feature, straighten the oscillating line by aligning the troughs.
3. Using the non-linear curve fitting function, fit a sine wave to the data start with a best guess of ~ 10 and allow the software to refine the fit until the chi-squared stops changing. You may also have to refine some of the other parameters of the fit to get a reasonable result.

PART G: Determining the Position of the Dyad Axis

1. Using the peak finder function in Origin, determine the base with maximal or minimal reactivity for each period; these may have to be adjusted based on your own analysis of the data.
2. Determine the number of nucleotides between two consecutive maxima ($y = \max_n - \max_{n-1}$) and plot that value against the average value of the maxima [$x = (\max_n + \max_{n-1})/2$]. Apply a moving average of 2 or 3.

3. Use these points to construct the curve, the highest peak of which will be the location of the dyad axis.

PART H: Determining the Rotational Position

1. Export the background subtracted data back into Excel.
2. Overlay the data from both strands of the duplex, with one strand displayed in the 5' to 3' direction and the other in the 3' to 5' direction.
3. Using the maxima from the 5' to 3' strand, determine the closest maxima from the 3' to 5' strand. This is the location of the minor groove.

Enzymatic Reactions of DNA Within the Nucleosome

While chemical footprinting of nucleosomes provides the most detailed analysis of DNA behavior, it is often beneficial to perform a more general analysis using enzymes such as DNase I and Exo III. DNase I is an endonuclease that creates single-strand nicks. While DNase I does not possess strict sequence dependency, it does cleave preferentially at some sequences, mainly depending on minor groove width and DNA flexibility. DNase I will only be able to cleave areas of the DNA facing away from the histone core, leading to a pattern similar to that of the hydroxyl radical, but complicated by the sequence preference of the enzyme. Exo III is a 3' to 5' exonuclease, which digests duplex DNA. When the enzyme encounters DNA that is interacting with the histone core it pauses, which results in intense strand cleavage. These locations of intense strand cleavage map the translational position of the DNA. Like with the hydroxyl radical footprinting, separate experiments should be carried out with free duplex and nucleosomal substrates and with each complementary strand of the duplex radiolabeled.

While reaction with S1 nuclease does not produce a footprint, the methodology is the same as the footprinting enzymes. S1 nuclease recognizes and cleaves single stranded DNA.

Materials:

Radiolabeled Duplex

NCPs

4 M NaCl

50 mM Tris

5 mM Tris

0.5 mL and 15 mL Amicon concentrators

Enzyme and enzyme buffer

DNase I buffer (100 mM Tris, 25 mM MgCl₂, 5 mM CaCl₂, pH 7.6), NEB1 buffer
(100 mM Bis Tris propane-HCl, 100 mM MgCl₂, 10 mM DTT, pH 7.0), or S1
nuclease buffer (500 mM sodium acetate, 2.8 M NaCl, 45 mM ZnSO₄, pH 4.5)

50% glycerol

100 mM EDTA

1 mg/mL CT-DNA

5% native gel and accompanying supplies

Phenol:Chloroform

100% Ethanol

7.5 M ammonium acetate

8% Denaturing polyacrylamide gel and accompanying supplies

PART A: Nucleosome preparation

1. Prepare exchange reaction
 - a. 12.5 μ L 4M NaCl
 - b. 25 μ L 1 mg/mL NCPs
 - c. 5 μ L 50 mM Tris
 - d. 7.5 μ L radiolabeled duplex (10-100 pmol)
2. Incubate 1 h at room temperature.
3. Add 16.5 μ L 5 mM Tris. Incubate 1 h at room temperature.

- a. Repeat step 3 twice more, adding a total of 50 μL 5 mM Tris, bringing the reaction volume to 100 μL
4. Add 100 μL 5 mM Tris. Incubate 10 min at room temperature.
 - a. Repeat step 4 twice more, bringing the reaction volume to 400 μL
5. Add 100 μL 5 mM Tris. Incubate 1 h at 37 $^{\circ}\text{C}$.
 - a. This brings the final reaction volume to 500 μL
6. Transfer each reaction to a 0.5 mL concentrator
 - a. Centrifuge at maximum speed, 30 min, 4 $^{\circ}\text{C}$

PART B: Enzymatic reactions on nucleosomes

1. Transfer 45-x μL of the exchanged nucleosomes to a clean eppendorf tube (x is the amount of enzyme needed)
2. Add 5 μL of the appropriate buffer (total reaction volume is 50 μL)
3. Add the optimized amount of enzyme (0.4 units DNase I (New England Biolabs), 50 units Exo III (New England Biolabs), or 42.5 units S1 nuclease (Promega)
 - a. To get 0.4 units of DNase I you'll need a 1:10 dilution of the stock from NEB
4. Incubate at 37 $^{\circ}\text{C}$ for 5 minutes
5. Quench with 5 μL 100 mM EDTA and 5 μL 50% glycerol
6. Immediately run your entire sample in an 8 mm, 5% native gel
 - a. 150 V, 3 h
7. Expose the gel for ~20 minutes; use the image to identify your nucleosome band
8. Excise the nucleosomes and crush and soak them overnight at 37 $^{\circ}\text{C}$

9. Centrifuge the crush and soak, collecting the supernatant
10. Concentrate the supernatant using a 15 mL amicon concentrator
 - a. $3220 \times g$, 50 min, leaves ~0.5 mL of recovered nucleosomes
11. Add SDS to 0.1%; heat at 90 °C for ~10 min
12. Extract samples twice with 400 μ L of phenol:chloroform
13. Ethanol precipitate the DNA with 50 μ L 7.5 M ammonium acetate and 700 μ L ethanol
14. Samples are now ready for evaluation in an 8% denaturing gel

PART C: Enzymatic reactions on free duplex

1. Bring 5 pmol of free duplex to 45-x μ L (x is the amount of enzyme needed)
2. Add 5 μ L of the appropriate buffer
3. Add the optimized amount of enzyme (0.04 units DNase I (New England Biolabs), 50 units Exo III (New England Biolabs), or 42.5 units S1 nuclease (Promega)
 - a. To get 0.04 units of DNase I you'll need a 1:100 dilution of the stock from NEB
4. Incubate at 37 °C for 2 minutes
5. Quench with 5 μ L 100 mM EDTA and 5 μ L 1 mg/mL CT-DNA
6. Ethanol precipitate the DNA with 50 μ L 7.5 M ammonium acetate and 500 μ L ethanol
7. Samples are now ready for evaluation in an 8% denaturing gel

PART D: Evaluating the reactions

1. Pour a long, 8% denaturing gel (pour it the day you are going to use it)
 - a. The long gels leak. A lot. Make sure to have at least 100 mL of matrix even though the gel only takes about 40 mL.
2. Pre-run the gel at 80 W
3. Count your samples
 - a. Prepare each sample at 5000 counts/ μ L in 100% formamide (except for S1 reactions, you'll only need about 1000 counts/ μ L)
4. Load 15,000 counts per well (3 μ L)
 - a. You'll only be able to run ~10 samples per gel. The last two lanes on either side won't migrate properly
5. Run at 65 W for 2.5 hours
 - a. This will put the parent band just below the filter paper used in the gel dryer
6. Dry the gel
 - a. Using saran wrap, pull the gel off the glass plates
 - b. Cut a piece of filter paper in half along the long axis
 - c. Align the filter paper with the bottom edge of the gel and place flat against the gel
 - d. Cut off the top of the gel (the area not covered by the filter paper)
 - i. Check for radioactivity and discard
 - e. Trim the edges of the filter paper and place in the gel dryer
7. Use the gel to expose a phosphor plate for 4 days