STRUCTURAL STUDIES OF DNA CONTAINING BASE AND SUGAR MODIFICATIONS

JACK HARDWICK

CHRIST CHURCH

UNIVERSITY OF OXFORD

Supervisor: Prof Tom Brown



A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2019

To Mari, whose capacity for patience and understanding has,
over the last four years, been rigorously tested.

ABSTRACT

Chemical modifications to DNA are ubiquitous in nature, with roles ranging from bacterial warfare to epigenetics. Unnatural modications can also be introduced to DNA and constitute important research tools. In both cases, the effect of chemical modifications on DNA structure is key to their function. This work describes structural studies of DNA containing natural and unnatural DNA modifications.

Several modified bases—5-hydroxymethylcytosine (^{hm}C), 5-formylcytosine (^{f}C), and 5-carboxylcytosine (^{ca}C)—were recently discovered in human DNA and appear to have crucial biological roles. The molecular basis of their function remains unclear, but structural changes to the DNA double helix have been implicated. To address this, the effect of cytosine modifications on DNA structure was probed by an array of techniques including X-ray crystallography, CD and NMR spectroscopy, DNase I and gel-mobility assays. Contrary to previous findings, no evidence was found for a unique, ^{f}C-dependent DNA conformation that is distinct from the A-, B-, or Z-forms. However, 5-methylcytosine (^{m}C), ^{hm}C, and ^{f}C were found to induce other structural changes, which can alter DNA-protein interactions and may thus be important to their biology.

The second part of this thesis describes the structural characterisation of DNA containing nitroxide spin labels, introduced via a recently developed 2′-alkynyl labelling strategy. Using X-ray crystallography, MD simulations, EPR and NMR spectroscopy, the spin labels were localised and found to cause only modest and local perturbation to DNA conformation. These results establish 2′-alkynyl spin-labelling as an accurate and minimally perturbing method to probe DNA structure.

ACKNOWLEDGEMENTS

First of all, I am extremely grateful to my supervisor, Tom Brown, for all his guidance, encouragement, and support. This DPhil has taught me a great deal and taken me around the world. From the Bahamas to Stevenage, it has been a privilege. I can only apologise for the state of my desk.

I thank the EPSRC, the University of Oxford, and ATDBio for funding this work. I'm also very grateful to Afaf El-Sagheer for all her help in the lab. Working together has been an absolute pleasure. I'll miss the falafel, mahshi, and moussaka. I won't miss *The Archers* omnibus. Thanks to Ed Anderson for further guidance, to Andrew Lane for stimulating discussions and a memorable trip to Kentucky, and to Loren Williams for inviting me to Atlanta. Thanks to Dorcas, Dan and the others at ATDBio for their generosity with oligos and expertise. Thanks to my college advisor, Jason Davis, and to Martin Grossel, Dirk Aarts, and all the other Christ Church Chemists. I'm very grateful to Chris Cardin, James Hall, Mike McDonough, and Steve Carr for their help and advice on X-ray crystallography, and to Nick Proudfoot and Joan Monks for radiation safety training and access to their hotlab. Thanks to Arun Shivalingam for his scientific counsel, and to all the other Brownies for four years of good times. In particular, I'm indebted to Dr Dysko for showing me the ropes, proofreading this thesis, and even housing me while I finished it—and to my dear friend and fumehood mate, Pumpui, for all the laughter. Thanks to Louis, my college co-brewer, and Raam for a little too much light relief. I am extremely grateful to Jon Watts for all his support and mentoring over the years, and to Pam Lewis and David Read—the Science Foundation Year changed everything.

I am deeply grateful to my family for all their love and support, especially my mum, Sue, for her frequent bouts of wisdom, to my dad, Gordon, for his tireless assault on 476, and to my brother, Will, for some unforgettable times—including his engagement on Hornby Island (sorry, Pippa).

Most of all, thanks to my long-suffering girlfriend, Mari, who I do not deserve.

Grazie, amore mio.

PUBLICATIONS & CONTRIBUTIONS

Some of the work presented in this thesis has been published, as outlined below. The contributions of other authors to the experimental work are also given. Some figures from these publications have been reproduced in accordance with the publishers' license agreements.

Chapter 1

Some ideas and figures have appeared in the following publication:

> Jack S Hardwick, Andrew N Lane, Tom Brown, 'Epigenetic modifications of cytosine: biophysical properties, regulation, and function in mammalian DNA,' *BioEssays*, 40, 1700199 (2018).

Chapter 2

This chapter is based on the following publication:

> Jack S Hardwick, Denis Ptchelkine, Afaf H El-Sagheer, Ian Tear, Daniel Singleton, Simon E V Phillips, Andrew N Lane, Tom Brown, '5-Formylcytosine does not change the global structure of DNA,' *Nat. Struct. and Mol. Biol.*, 24, 544–552 (2017).

Contributions

AHES, IT, DS, and JSH prepared oligonucleotides. JSH performed crystallisations, collected diffraction data, and solved crystal structures of the formylated and unmodified duplexes. DP collected diffraction data and solved the crystal structure of the brominated duplex. JSH refined and analysed all structures, built DNA models, and collected CD and UV spectra. ANL acquired and analysed NMR spectra.

Chapter 3

This work is unpublished.

Contributions

AHES and JSH prepared oligonucleotides. All other work performed by JSH.

Chapter 4

This chapter is based on a collaboration led by Prof Ed Anderson (Oxford Chemistry) and has been described in the following publication:

> Jack S Hardwick, Marius M Haugland, Afaf H El-Sagheer, Denis Ptchelkine, Frank R Beierlein, Andrew N Lane, Tom Brown, Janet E Lovett, Edward A Anderson, '2′-Alkynyl spin-labelling is a minimally perturbing tool for DNA structural analysis,' *Nucleic Acids Res.*, advance online publication, 13 Feb 2020 (DOI 10.1093/nar/gkaa086).

Contributions

MMH prepared 2′-alkynylnucleosides. AHES, MMH, and JSH prepared oligonucleotides and performed spin-labelling. MMH, DP, and JSH performed crystallisations and collected diffraction data. JSH solved and analysed crystal structures. ANL acquired and analysed NMR spectra. FRB performed and analysed MD simulations. JEL acquired and analysed EPR spectra.

CONTENTS

1. Introduction **2**

1.1. The DNA Double Helix 3

1.1.1. B-DNA 4

1.1.2. A-DNA 6

1.1.3. Z-DNA 7

1.1.4. Sequence-dependent structural variation in B-DNA 8

1.2. Non-Canonical Bases in DNA 11

1.2.1. 5-Methylcytosine 11

1.2.2. The discovery of further non-canonical DNA bases in vertebrates 17

1.2.3. How do cytosine modifications alter DNA-protein interactions? 19

1.2.4. Can cytosine modifications be read indirectly via changes to DNA conformation? 25

1.3. Methods to Study DNA Structure 26

1.3.1. X-ray crystallography 27

1.3.2. NMR spectroscopy 27

1.3.3. Molecular rulers 29

1.4. Aims of this work 31

I. Structural studies of DNA containing epigenetic cytosine modifications **45**

2. F-DNA: a new form of the DNA double helix? **46**

2.1. Introduction 46

2.2. Results & Discussion 53

2.2.1. Crystal structure of $d(CTACGCGCGTAG)_2$ 53

2.2.2. Crystal structure of $d(CTA^fCG^fCG^fCGTAG)_2$ 59

2.2.3. The effect of ^{f}C on DNA groove geometry, coiling and helical trajectory 62

2.2.4. Comparing the structural consequences of formylation to those of DNA sequence . . . 65
2.2.5. CD spectroscopy of formylated DNA . . . 68
2.2.6. NMR analysis of formylated DNA . . . 72
2.3. Conclusions . . . 74

3. Do cytosine modifications alter DNA structure? 81
3.1. Introduction . . . 81
3.2. DNase I Cleavage Assay . . . 82
3.2.1. Background . . . 82
3.2.2. Results & discussion . . . 83
3.3. Electrophoretic Mobility Assay . . . 89
3.3.1. Background . . . 89
3.3.2. Results & Discussion . . . 92
3.4. Conclusions and Future Work . . . 108
3.5. Summary of Part I . . . 111

II. Structural studies of DNA containing spin labels 122

4. Calibrating a molecular ruler: structural studies of spin-labelled DNA 123
4.1. Introduction . . . 123
4.2. Results & Discussion . . . 126
4.2.1. DNA synthesis . . . 126
4.2.2. EPR spectroscopy . . . 126
4.2.3. X-ray crystallography . . . 127
4.2.4. Molecular dynamics simulations . . . 133
4.2.5. NMR spectroscopy . . . 137
4.2.6. Further discussion . . . 140
4.3. Conclusions & future work . . . 141

5. Experimental 147
5.1. Oligonucleotide Preparation, Purification and Analysis . . . 147
5.2. Polyacrylamide Gel Electrophoresis . . . 153
5.3. Circular Dichroism (CD) Spectroscopy . . . 153
5.4. UV Absorption Spectroscopy . . . 154
5.5. X-Ray Crystallography . . . 154

5.6. Gel Mobility 156
5.6.1. Polyacrylamide gel electrophoresis 156
5.6.2. Analysis of gel images 157
5.7. DNase I Cleavage 159
5.8. Maxam–Gilbert Sequencing 161
5.9. Hydroxyl Radical Cleavage 164

III. Appendix **168**

A. Supplementary information for Chapter 2 **169**
A.1. Supplementary figures 170
A.2. Supplementary tables 182

B. Supplementary information for Chapter 3 **205**
B.1. Supplementary figures 206
B.2. Supplementary tables 211

C. Supplementary information for Chapter 5 **215**
C.1. Supplementary figures 216
C.1.1. Mass spectra of spin-labelled oligonucleotides 220
C.2. Supplementary tables 225

LIST OF FIGURES

1.1. The chemical structure of the four canonical nucleotides of DNA 3
1.2. Base pairing and the double helix 4
1.3. A comparison of B- and A-DNA 5
1.4. The Z-DNA conformation 7
1.5. Inter- and intra-base pair structural parameters of DNA 9
1.6. Chemical information on sequence available from the major and minor grooves 10
1.7. Indirect sequence readout performed by the nucleosome 12
1.8. Maintenance methylation and passive demethylation 14
1.9. 5-Methylcytosine is favoured for structural reasons 16
1.10. Putative active DNA demethylation pathway 19
1.11. Crystal structures of MeCP2 and the SRA domain of UHRF1 bound to methylated DNA 21
1.12. Crystal structures of TET2 and TDG 23
1.13. The tyranny of the lattice 28
1.14. Molecular rulers for probing DNA conformation 30

2.1. Overlaid crystal structures of the Dickerson-Drew dodecamer containing either C, ^{m}C, ^{hm}C, ^{f}C, or ^{ca}C 48
2.2. Crystal structure of a DNA duplex containing 5-formylcytosine, which has unusual structural features 49
2.3. Comparison of F-DNA with A- and B-DNA 50
2.4. Schematic CD spectra of B-, A-, Z-DNA, and real spectra of the putative F-DNA conformation 52
2.5. Crystal structure of d(CTACGCGCGTAG)$_2$ 53
2.6. Overlaid crystal structures of the previously reported formylated duplex and its unmodified counterpart 56
2.7. Major-groove hydration patterns of 4qkk and 5mvk 58

2.8. Crystal structure of d(CTA^{f}CG^{f}CG^{f}CGTAG)$_2$ 60
2.9. Comparison of the low-salt ^{f}C structure (5mvu) with its unmodified counterpart (5mvk) 63
2.10. A comparison of the crystal contacts of the two formylated structures and their unmodified counterpart 64
2.11. Groove dimensions of the two ^{f}C structures and their unmodified counterpart, and modelled junctions with ideal B- and A-DNA 66
2.12. X-ray crystal structures of sequence variants of d(CTACGCGCGTAG)$_2$ 69
2.13. CD and UV absorption spectra of d(CTACGCGCGTAG)$_2$ and formylated variants 70
2.14. CD spectra of DNA containing ^{f}C and ^{ca}C in different sequence contexts 71
2.15. Chemical shift and NOE differences between duplexes of varying formylation levels 74

3.1. Cytosine modifications alter the relative cleavage rate of DNase I 85
3.2. The influence of cytosine modification on DNase I cleavage susceptibility of the opposite strand 86
3.3. DNA curvature and anomalous electrophoretic mobility 90
3.4. Ligation and analysis of the A-tract containing sequence, into which modified cytosines were incorporated 94
3.5. The effect of cytosine methylation on the electrophoretic mobility of A-tract DNA 96
3.6. The effect of increased methylation on electrophoretic mobility 98
3.7. The effect of cytosine methylation on the electrophoretic mobility of A-tract and G_3C_3 sequences 99
3.8. The effect of CpA methylation and altered sequence on electrophoretic mobility 101
3.9. The effect of CpG hydroxymethylation on electrophoretic mobility . . . 103
3.10. The effect of CpG formylation on the electrophoretic mobility of A-tract DNA 105
3.11. Two possible ways by which cytosine modifications could induce bending into the minor groove 107

4.1. Different for incorporating spin labels into DNA 125
4.2. EPR-derived distance distributions of 6-Me- and 5-Me-spin-labelled DNA 127
4.3. Crystal structure of the 6-Me-labelled duplex 128
4.4. Crystal structure of the 5-Me spin-labelled duplex 129

4.5. Crystal structures of spin-labelled duplexes overlaid with their unmodified counterparts 132
4.6. A comparison of the puckering of the 2′-spin-labelled uridines (PDBs 6qjs, 6qjr) with 2′-substituted ribo- and arabinonucleotides 133
4.7. Molecular dynamics simulations of spin-labelled duplexes 135
4.8. Sequence of the spin-labelled duplexes characterised by NMR 138
4.9. DQF-COSY and NOESY spectra of the 6-Me-labelled duplex 139

A.1. Crystal contacts between step 7 of 5mvk and a symmetry mate 170
A.2. Crystal structures containing either 5-formylcytosine or cytosine overlaid with ideal A-DNA 171
A.3. Structural parameters of formylated crystal structures 5mvu and 4qkk and their unmodified counterpart 5mvk 172
A.4. 60-bp models of junctions between ideal A- or B-DNA and crystal structures of the same sequence containing either 5-formylcytosine or cytosine 173
A.5. Overlay of unmodified crystal structures 5mvp, 5mvq, and 5mvt 174
A.6. Supplementary UV and CD spectra 175

B.1. Example gels showing Maxam-Gilbert sequencing and DNase I cleavage products 207
B.2. The reference peaks used to determine relative DNase I cleavage rates . . 208
B.3. 2D polyacrylamide gels to detect cyclised ligation products 208
B.4. The electrophoretic mobilities of the unmodified A-tract and G_3C_3 constructs, relative to an unmodified, random sequence 209
B.5. Interactions of the 5-hydroxymethyl group in the major groove 210

C.1. Electron density of the spin-labelled crystal structures 216
C.2. Interactions in the 6-Me-labelled crystal structure 6qjs 217
C.3. Crystal packing of the spin-labelled duplexes 218
C.4. Groove widths of the spin-labelled structures and their unmodified counterparts 219

LIST OF TABLES

1.1. Average helix parameters of the three major conformational classes of DNA 6

2.1. Data collection and refinement statistics of PDBs 5mvu, 5mvk, 5mvp, 5mvt, and 5mvq 54
2.2. All-atom RMSDs between PDBs 4qkk, 5mvu and 5mvk, and ideal A- and B-DNA of the same sequence 61
2.3. Comparison of averaged structural parameters of the formylated duplex crystallised under high- and low-salt conditions, its unmodified counterpart, and those of ideal A- and B-DNA. 61
2.4. Sequences, resolution, and PDB codes of the unmodified DNA crystal structures determined in this study. 68

4.1. Data collection and refinement statistics of PDBs 6qjs and 6qjr 130
4.2. Interspin distances derived from MD simulations, X-ray crystallography, and EPR 134

A.1. Data collection and refinement statistics of PDB 5mvl 183
A.2. Phasing statistics of 5mvl (MAD) 184
A.3. Twinning statistics of 5mvu 184
A.4. Crystallisation conditions of the six structures determined in Chapter 2 . 184
A.5. Sequences and masses of oligonucleotides used in this study 185
A.6. All-atom RMS deviation of the six crystal structures determined in this study 186
A.7. Sugar puckering and glycosyl torsion angles of 4qkk, 5mvu and 5mvk . . 187
A.8. α- and β-torsion angles of 4qkk, 5mvu and 5mvk 188
A.9. γ- and δ-torsion angles of 4qkk, 5mvu and 5mvk 189
A.10. ϵ- and ζ-torsion angles of 4qkk, 5mvu and 5mvk 190
A.11. Shift, slide, roll, and twist values of PDBs 4qkk, 5mvu, and 5mvk 191
A.12. Rise and tilt values of values of PDBs 4qkk, 5mvu, and 5mvk 192

A.13. Shear, strech, stagger, and buckle values of PDBs 4qkk, 5mvu, and 5mvk 193
A.14. Propeller twist and base-pair opening values of PDBs 4qkk, 5mvu, and 5mvk 194
A.15. X-displacement, Y-displacement, helical rise, and inclination values of 4qkk, 5mvu, and 5mvk 195
A.16. Tip and helical twist of 4qkk, 5mvu & 5mvk 196
A.17. Shift, slide, roll, and twist values of PDBs 5mvq, 5mvp, and 5mvt 197
A.18. Rise and tilt values values of PDBs 5mvq, 5mvp, and 5mvt 198
A.19. X-displacement, Y-displacement, helical rise, and inclination values of PDBs 5mvq, 5mvp, 5mvt 199
A.20. Tip and helical twist values of PDBs 5mvq, 5mvp, and 5mvt 200
A.21. Sugar puckering and glycosyl torsion angles of 201
A.22. Chemical shift assignments of the formylated duplexes 202
A.23. Chemical shift assignments of the unmodified duplex 203

B.1. Relative DNase I cleavage rates 5′ to the modified cytosine 211
B.2. DNase I cleavage rates on the strand opposing the modified cytosine . . 212
B.3. Sequences and masses of oligonucleotides used in the DNase I cleavage study 212
B.4. Sequences and masses of oligonucleotides used in the electrophoretic mobility study 213

C.1. Sequences of the spin-labelled oligonucleotides used in this study 225
C.2. Crystallisation conditions of the six structures determined in Chapter 2 . 225
C.3. Twinning statistics of PDB 6qjs 226
C.4. Sugar puckering of the 6-Me-labelled crystal structure and its unmodified counterpart 226
C.5. Sugar puckering of the 5-Me-labelled crystal structure and its unmodified counterpart 227
C.6. Glycosyl torsion angles of the spin-labelled crystal structures and their unmodified counterparts 228
C.7. Chemical shift assignments of the reduced 6-Me-labelled duplex 229
C.8. Chemical shift assignments of the reduced 5-Me-labelled duplex 230
C.9. Chemical shift assignments of the 6-Me-labelled duplex in the paramagnetic state 231
C.10. Chemical shift assignments of the reduced 5-Me-labelled duplex 232

ABBREVIATIONS

bp	Base pair
C	Cytosine
^{ca}C	5-Carboxylcytosine
^{f}C	5-Formylcytosine
^{hm}C	5-Hydroxymethylcytosine
^{m}C	5-Methylcytosine
CD	Circular dichroism
CE	Capillary electrophoresis
CpG	5′-CG-3′
DEER	Double electron-electron resonance
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EPR	Electron paramagnetic resonance
FRET	Förster/fluorescence resonance energy transfer
G	Guanine
HFIP	Hexafluoroisopropanol
HPLC	High-performance liquid chromatography
MPD	2-Methyl-2,4-pentanediol
NMR	Nuclear magnetic resonance
Nt	Nucleotide
PAGE	Polyacrylamide gel electrophoresis

PBS	Phosphate-buffered saline
PDB	Protein Data Bank
RP-HPLC	Reversed-phase high-performance liquid chromatography
T	Thymine
TDG	Thymine DNA glycosylase
TEA	Triethylamine
TEAB	Triethylammonium bicarbonate
TEMED	Tetramethylethylenediamine
TET	Ten-eleven translocation
U	Uracil
^{b}U	5-Bromouracil
UV	Ultraviolet
Vis	Visible

The capacity to blunder slightly is the real marvel of DNA. Without this special attribute, we would still be anaerobic bacteria and there would be no music.

—Lewis Thomas, *The Medusa and the Snail*, 1979

1

INTRODUCTION

Encoded within the structure of DNA lies the genetic information and evolutionary history of all known cellular life. But despite the exquisite complexity for which DNA is responsible, its composition is simple: a polymer chain comprising just four basic units. Each unit, or nucleotide, consists of an identical sugar-phosphate group and one of four canonical nucleobases—adenine (A), guanine (G), cytosine (C), or thymine (T). The sugar-phosphate groups of different nucleotides can form stable covalent linkages with one another, creating the backbone of the resulting polymer chain (Figure 1.1). Owing to the chemical stability of these linkages, DNA molecules can reach remarkable lengths. Chromosome 3B of bread wheat, for example, is a billion nucleotides long,[1] or about 0.3 metres.

Another important chemical property of nucleotides is the unique hydrogen-bonding pattern of each base (Figure 1.1). The acceptor-donor pattern of C is complementary to that of G, as are those of A and T. Consequently, these 'Watson-Crick' base pairs interact preferentially with each other, which, along with electrostatic and hydrophobic interactions, gives rise to a remarkable three-dimensional structure: the DNA double helix (Figure 1.2).

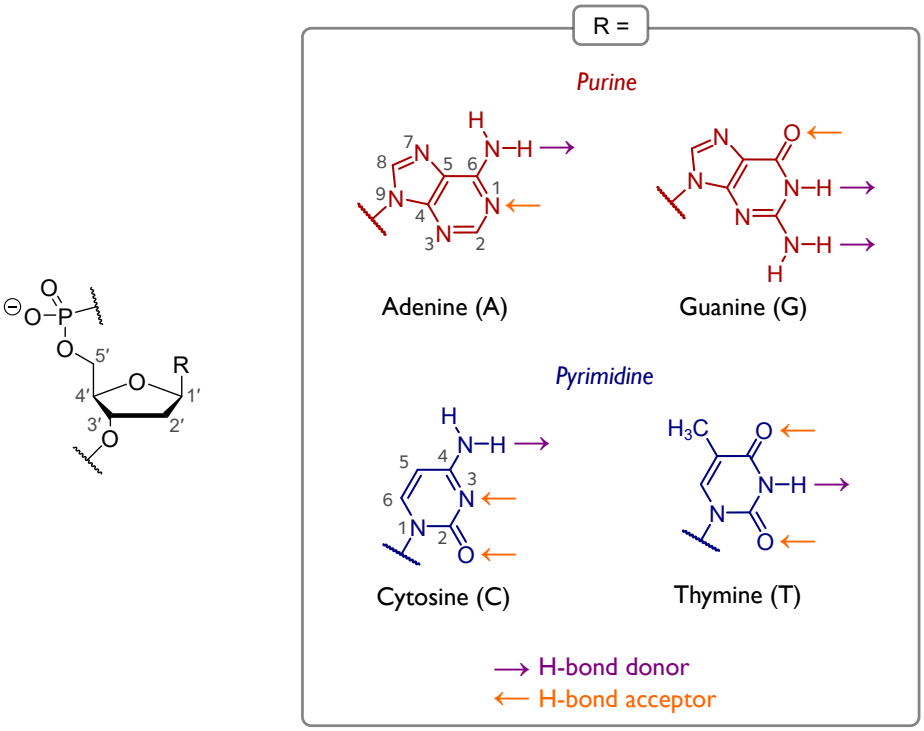


Figure 1.1. The chemical structure of the four canonical nucleotides of DNA. The hydrogen-bonding donor and acceptor properties of each base are indicated by purple and orange arrows.

1.1 The DNA Double Helix

In 1953, Watson and Crick published a seminal manuscript in *Nature*.[2] In a single page, they proposed the specific pairing of A-T and G-C bases, the double-helical structure of DNA, and how this could provide a copying mechanism for the genetic material. They even found space to denounce rival structural models. Watson and Crick's structure, based largely on experimental work by Franklin, Gosling, and Wilkins,[3,4] is known as B-DNA. Since then, two other major helical conformations of DNA have been characterised (A- and Z-DNA). In all three conformations the two DNA strands run in opposite directions (*antiparallel*), but other features are markedly different (Figures 1.3, 1.4, Table 1.1).

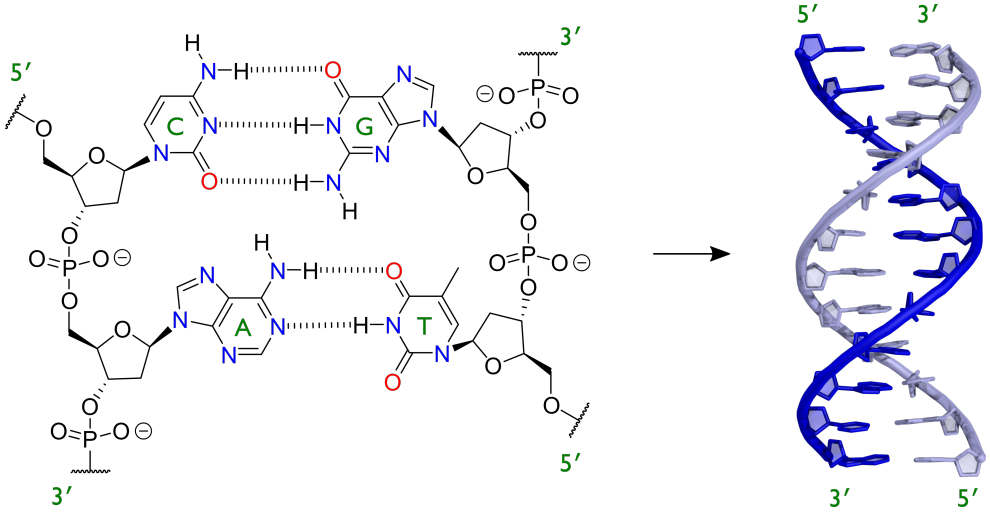


Figure 1.2. Base pairing of C with G, and A with T, which enables the formation of the double helix. The two strands align antiparallel, so that the 5′ end of one strand faces the 3′ end of the other.

1.1.1 B-DNA

DNA predominantly adopts the B-DNA conformation in solution (Figure 1.3). In the B form, each base pair sits roughly perpendicular to the helix axis, and its vertical separation from its nearest neighbour (*rise*) is 3.4 Å. The angle between each adjacent base pair (*twist*) is 36°about the helix axis, so that the helix repeats every 10 base pairs, giving a helical pitch of 34 Å. The direction of twist leads to a right-handed helix. The deoxyribose sugar of each nucleotide is twisted out-of-plane by the repulsive, non-bonding interactions of its substituents (*puckering*).[5,6] In B-DNA, nucleotide sugars generally adopt the C2′-*endo* pucker—an envelope conformation in which the C2′ atom protrudes from the plane formed by the other ring atoms, *syn* to the C5′ atom (Figure 1.3*c*). The torsion angle of the glycosidic bond (χ-angle) is typically around −102°.[6]

Another set of features characteristic to B-DNA are the dimensions of the grooves—the cavities between the spiralling phosphodiester backbones. The sugar-phosphate groups lie closer to one side of each base pair, producing two grooves of unequal size—termed

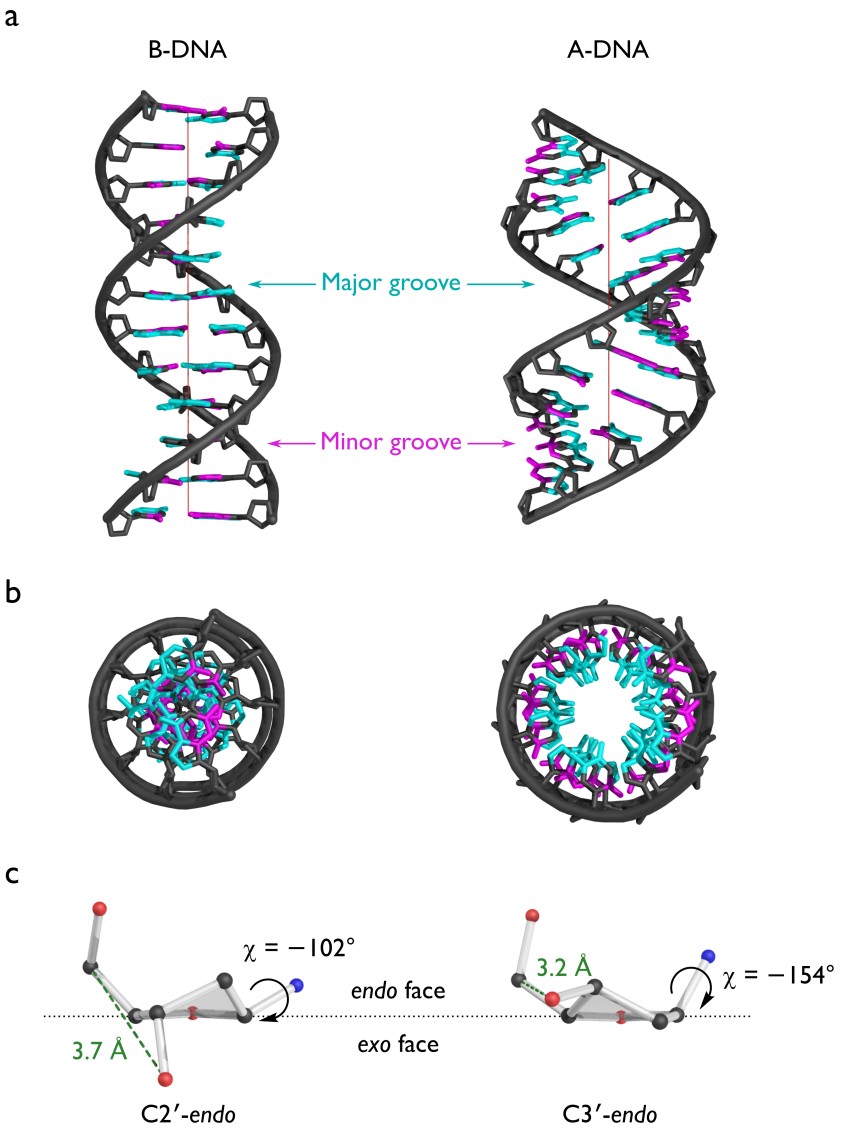


Figure 1.3. A comparison of B- and A-DNA. (*a*) Side view of ideal B- and A-DNA dodecamer duplexes. In B-DNA, the base pairs are roughly perpindicular to the helix axis (red line), which runs between the bases. In A-DNA, the base pairs are inclined and displaced from the helix axis, spiralling around it. (*b*) The base pairs' displacement from the helix axis towards the minor groove leads to an apparent hole when viewed from above. (*c*) The different sugar puckering of B- and A-DNA. In B-DNA, the deoxyribose sugar typically adopts an envelope conformation, in which the C2′ atom lies out-of-plane, on the *endo* side (C2′-*endo*), whereas in A-DNA, the C3′ atom does (C3′-*endo*). The relative positions of the ring substituents differ considerably in the two puckers. The torsion angle of the base-sugar bond (χ-angle) also distinguishes B- from A-DNA (defined as O4′-C1′–N1-C2 for pyrimidines and O4′-C1′–N9-C4 for purines).

Table 1.1. Average helix parameters of the three major conformational classes of DNA.[6]

					Groove width (Å)	
Form	*Helix sense*	*Bp / turn*	*Twist* (°)	*Rise* (Å)	*Minor*	*Major*
B	R	10	36.0	3.35	5.7	11.7
A	R	11	32.7	2.56	11.0	2.7
Z	L	12	−30.0	3.7	2.0	8.8

the major and minor grooves (Figure 1.3). The grooves of B-DNA have similar depths, but the major groove is wide and the minor groove narrow.

1.1.2 A-DNA

DNA can also adopt an alternative right-handed helical conformation known as A-DNA, first observed by Rosalind Franklin.[7] The A-form is commonly found in DNA crystal structures—likely due to a combination of crystal-packing effects and the dehydrating conditions required for crystallisation.[5]

In A-DNA, the base pairs are displaced and tilted by 20° with respect to the helix axis, resulting in a wider helix with a characteristic hole running through its centre (Figure 1.3*b*). The average twist and rise values are lower than in B-DNA (32.7° and 2.56 Å),[6] which reduces the helical pitch (28.2 Å) and increases the number of base pairs per helical turn (11 bp per turn). A-DNA sugars typically adopt the C3′-*endo* pucker (Figure 1.3*c*), with χ-angles of around −154°.[6] The altered geometry of A-form base-pairs also gives rise to different groove dimensions; the minor groove is much wider and shallower than that of B-DNA, while the major groove is narrower and deeper (Figure 1.3, Table 1.1).[6]

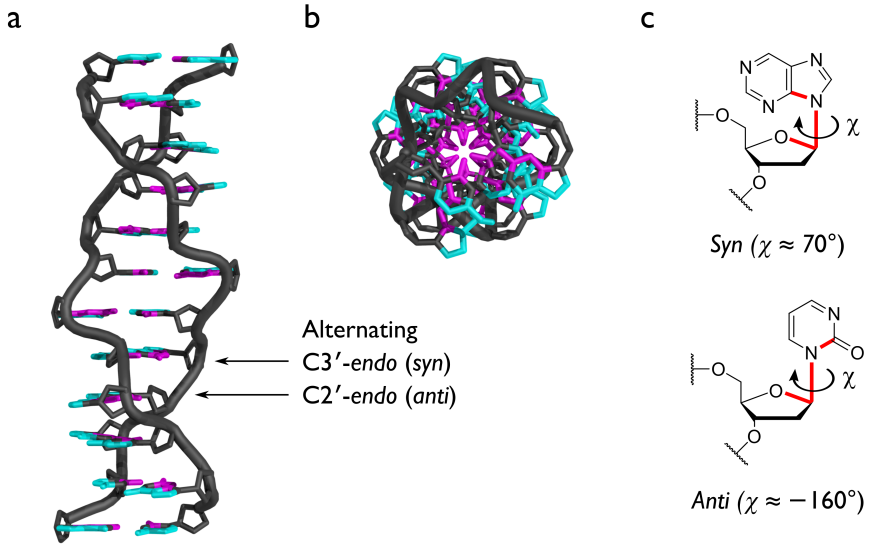


Figure 1.4. The Z-DNA conformation. (*a*) Unlike A- and B-DNA, the Z-DNA helix is left-handed. Its sugar puckers alternate between C3′*endo* (*syn*) and C2′-*endo* (*anti*). (*b*) Birds-eye view of the helix, showing the zig-zagging backbone. (*c*) The *syn* and *anti* ranges of conformers, which alternate in Z-DNA, and the approximate χ-angles adopted. Typically, purines favour *syn*, pyrimidines *anti*.[6]

1.1.3 Z-DNA

In the 1970s a third and radically different helical conformation of DNA was discovered. The first clues came from unusual CD spectra of DNA duplexes containing pyrimidine-purine repeats, under conditions of high salt.[8,9] A left-handed helical conformation was proposed to account for the unusual CD spectra,[8,9] whose atomic structure was later revealed by X-ray crystallography.[10] Notably, this was the first ever single-crystal X-ray structure of a DNA fragment.[11] The differences between this conformation and the A- and B-forms are striking: it has a left-handed helix, whose phosphodiester groups follow a zig-zag course—earning it the name Z-DNA.[10] Other distinctive structural features are given in Table 1.1.

In addition to conditions of high salt and alternating pyrimidine-purine repeats (in-

cluding the CpG dinucleotide), modifications to the 5-position of cytosine, such as the methyl group, have also been shown to stabilise Z-DNA.[12] However, despite decades of research, the biological significance of Z-DNA remains to be established.[5]

1.1.4 Sequence-dependent structural variation in B-DNA

In the early 1980s—almost 30 years after Watson and Crick proposed the structure of B-DNA—the first single-crystal X-ray structures of B-DNA were reported by Richard Dickerson's group at Caltech.[13,14] The self-complementary sequence of these structures, CGCGAATTCGCG, readily crystallises in the B-form and thus remains the most widely studied B-DNA fragment (118 crystal structures at the time of writing). It is now commonly referred to as the Dickerson-Drew dodecamer, named in part after Horace Drew—the PhD student who obtained the first structure. After presenting his new DNA structure to Francis Crick, who happened to be visiting the lab in 1980, Drew was apparently pleased when Crick smiled and said to him, 'So *that's* what it looks like!'[15]

As well as confirming Watson and Crick's model,[5] these structures provided the first atomic-level glimpse of sequence-dependent effects on B-DNA conformation, which had previously been obscured by the averaging inherent to X-ray fibre diffraction.[5,13,15]

Many structural features of DNA have been found to exhibit marked sequence-dependence. Although DNA contains around 700 atoms per helical turn, describing such structural features can be greatly simplified by treating either each base, or the entire base pair, as a rigid body. This allows the spatial relationship between two bases within a base pair, or two adjacent base pairs in a dinucleotide step, to be described by just six parameters: 3 rotations and 3 translations (Figure 1.5).[16] Variation in these structural features can encode information on DNA sequence, which may be read by DNA binding proteins.[17]

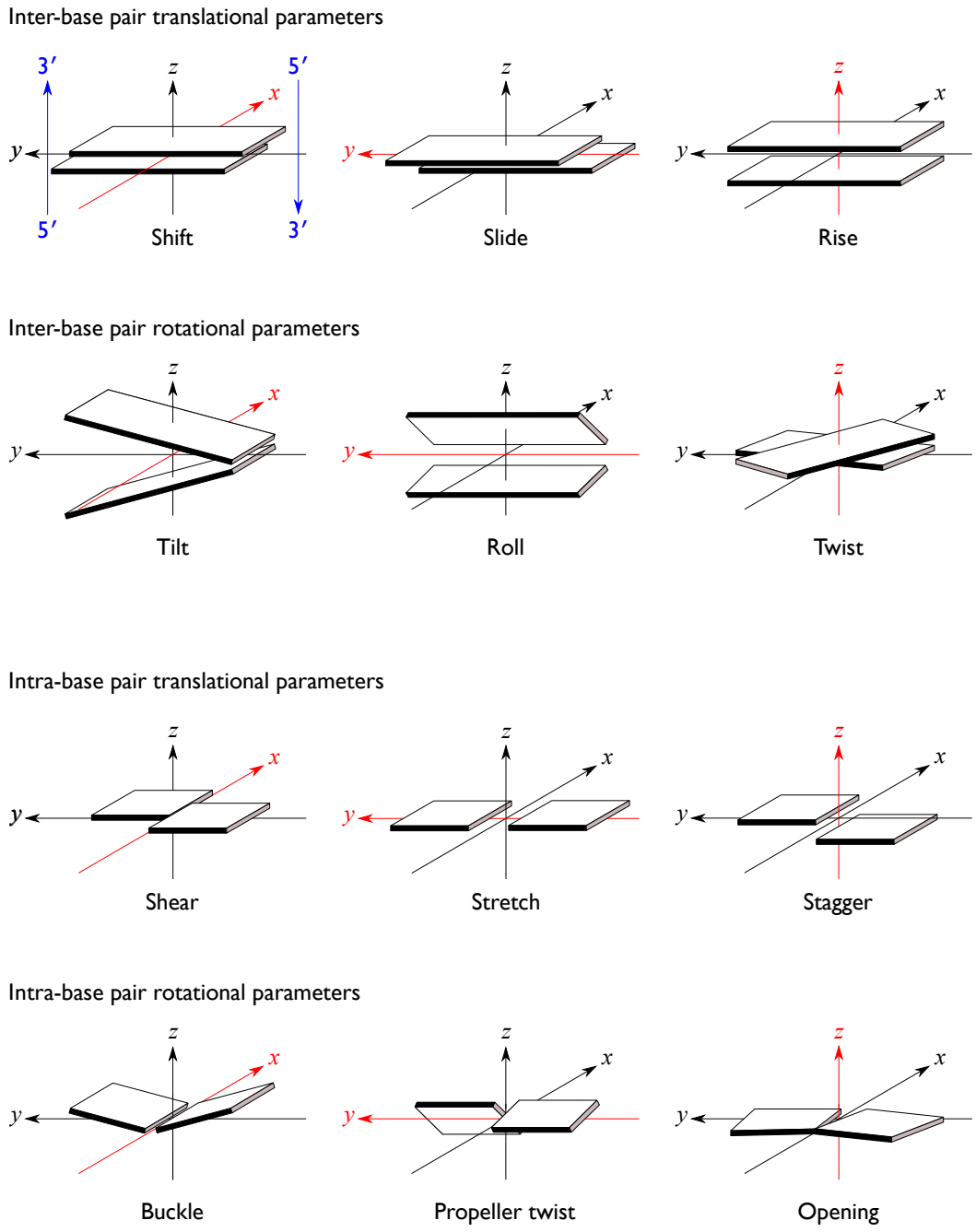


Figure 1.5. Definitions of parameters describing the geometry of dinucleotide steps and individual base pairs. The minor-groove side (black) points in the opposite direction to the x-axis. Blue arrows indicate the orientation of the two strands. The inter-base pair parameters treat each base pair as a rigid body, describing the six degrees of freedom between two base pairs in a dinucleotide step. The intra-base pair parameters describe the six degrees of freedom within a base pair.

Direct vs *indirect readout*

Proteins that bind to the major groove can detect DNA sequence without disrupting base-pairing. They do so by reading the unique chemical signature of each base pair, termed *direct readout* (Figure 1.6).[17] But less chemical information is available in the minor groove (Figure 1.6), so some minor-groove binding proteins take an indirect approach, elegantly exploiting the sequence-dependent variation in DNA shape (*indirect readout*).[17]

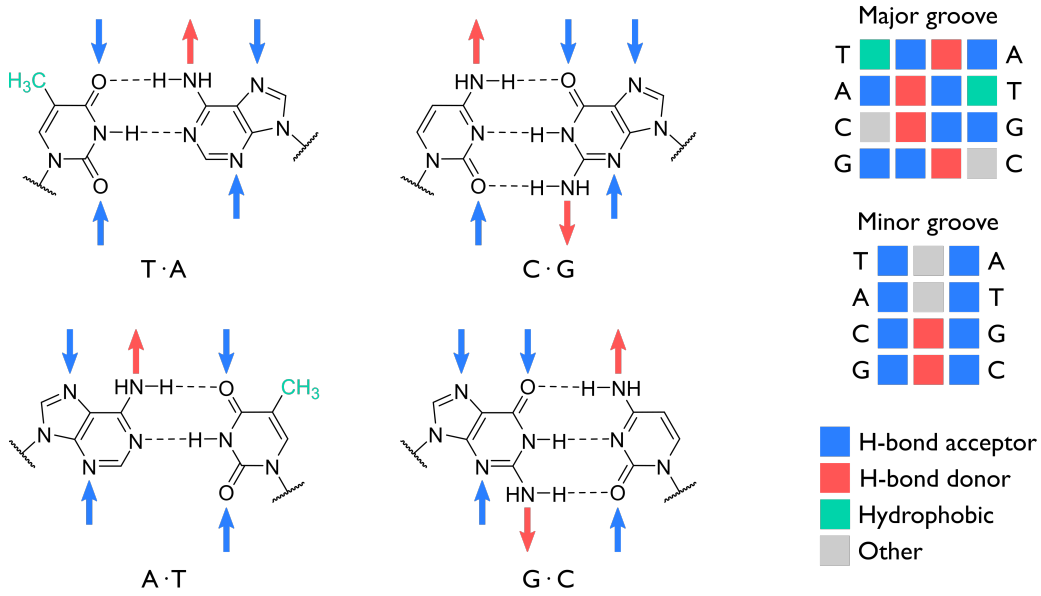


Figure 1.6. Chemical information on base sequence available from the major and minor grooves. Each base pair has a unique chemical signature, accessible from the major groove, that can be read by proteins without disrupting base pairing (direct readout). However, this information cannot be accessed via the minor groove. Proteins that bind to the minor groove can instead read sequence information via sequence-dependent variation in the structural features of DNA (indirect readout).[17] Blue and red arrows indicate hydrogen-bond donors and acceptors, respectively. Redrawn from Raiber *et al.*[18]

Indirect readout plays an important role in determining nucleosome stability

Many DNA-binding proteins perform indirect readout.[17] For example, the process is key to determining the positioning and stability of the nucleosome core particle (NCP)—the

basic repeating unit of chromatin.[17,19,20] NCPs occur roughly every 200 bp in the human genome, each comprising 147 bp of DNA, wrapped around the histone octamer (Figure 1.7*a*).[19] To form the nucleosome, the DNA must bend sharply around the protein core, circling it 1.7 times.[21] This equates to a bend of about 45° every helical turn. Consequently, nucleosome stability is closely linked to the propensity of the DNA to undergo this extreme bending, which is, in turn, strongly dependent on nucleotide sequence.[19,20,22]

The main way by which DNA bends is *roll*—a deformation between two base pairs that leads to a bend into either the major or minor groove (Figure 1.7*b*).[24] As a result, DNA bending in the nucleosome is not distributed evenly throughout the DNA. Instead, it is accommodated by key sites where either the major or minor groove faces the histone core, which alternate roughly every 5 bp (Figure 1.7*a*).[21] Because these sites are forced to accept a disproportionate share of the bending, and because some dinucleotides preferentially bend into either the major or minor groove, the DNA sequence at these sites has a disproportionately large effect on nucleosome stability (Figure 1.7*a*).[19,22]

1.2 Non-Canonical Bases in DNA

Modifications to the four canonical bases of DNA occur naturally in all genomes. Many of these are the unwanted and mutagenic products of DNA damage,[25,26] whereas others are introduced enzymatically, enabling a further layer of information to be stored alongside the genetic code.[18,27,28]

1.2.1 5-Methylcytosine

One of the most common modified bases in DNA is 5-methylcytosine (^{m}C). Nature has found many uses for ^{m}C, ranging from bacterial warfare to the control of gene expression.[18,28,29] The presence of ^{m}C in eukaryotic genomes was known even before the double-helical structure of DNA.[30] Accounting for 4% of all cytosines, ^{m}C is considered

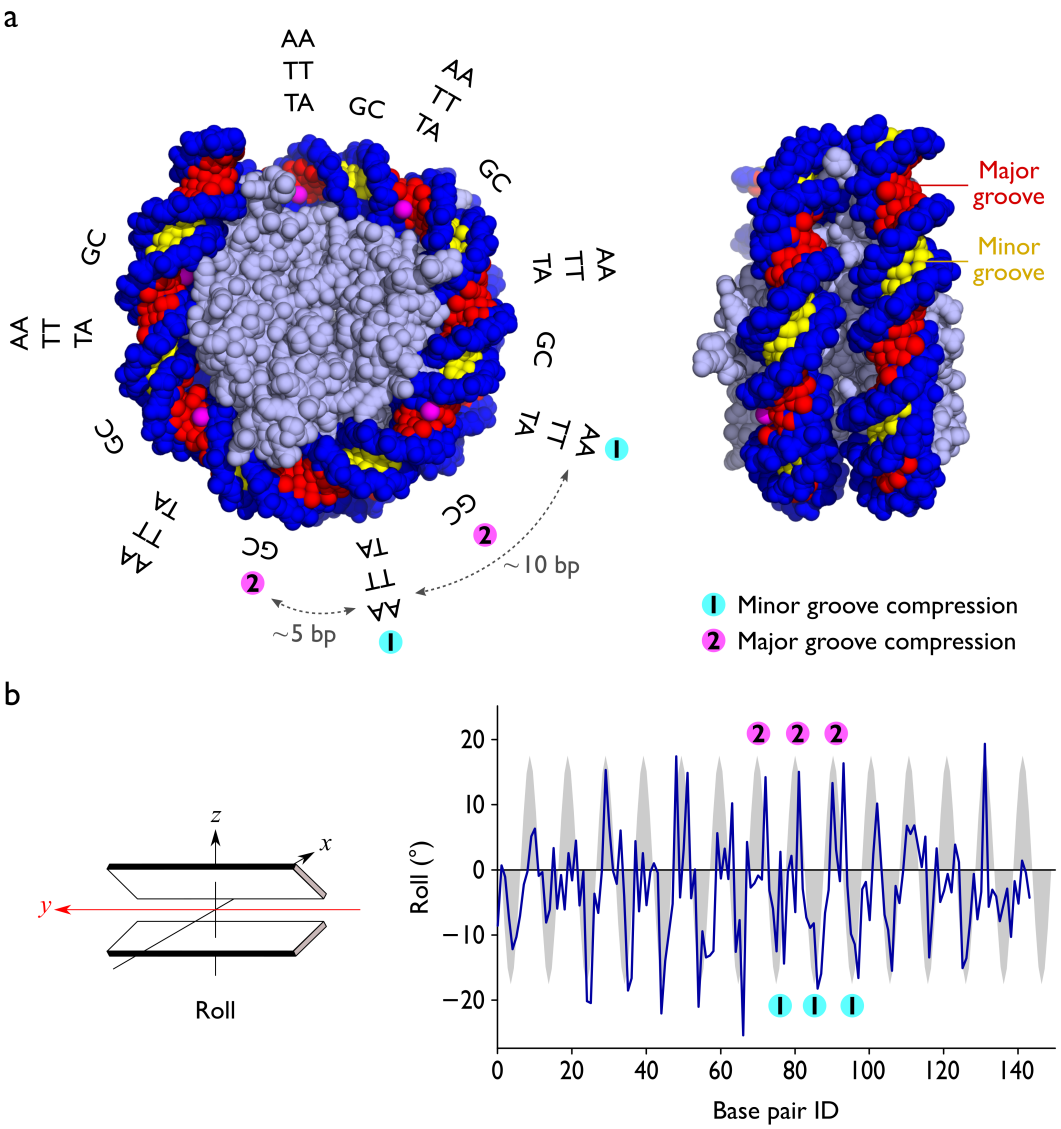


Figure 1.7. Indirect sequence readout performed by the nucleosome. (*a*) Front and side view of the nucleosome core particle (PDB 3mvd).[23] Nucleosome formation forces the DNA (blue, major and minor grooves red and yellow) to bend sharply around the histone octamer (grey). The propensity of DNA to bend is strongly sequence-dependent, resulting in certain dinucleotide steps being favoured at sites where the grooves face the histone core, which occur every ~5 bp. (*b*) The DNA bending is primarily achieved through roll—a deformation in which the helix axis of DNA bends into either the major or minor groove (corresponding to positive and negative values, respectively).[21,24] A plot of the roll angles in the nucleosome (PDB 3mvd)[23] shows that the bending is not evenly distributed throughout the dinucleotide steps. An even distribution of roll angles would follow a sinusoidal pattern in-phase with the helix repeat (shaded grey). Instead, it is often accommodated by one or two dinucleotide steps whose major or minor grooves lie in the centre of the bend—facing towards the histone octamer. Thus, dinucleotides at these positions would have a disproportionately large effect on nucleosome stability and organisation. Part *a* redrawn from Struhl and Segal.[22]

the 'fifth base' of the human genome.[31]

5-Methylcytosine is associated with transcriptional repression in humans

In humans, ^{m}C typically correlates with transcriptional repression.[32] The methyl marks are established in the genome by *de novo DNA methyltransferase* enzymes and occur primarily at CpG sites, of which 60–80% are symmetrically methylated in humans.[33] The symmetric methylation of CpG sites offers a simple means of transmitting methylation patterns from parent cell to daughter cell; upon replication, methyl marks are passed on to one of the two strands of each newly formed daughter duplex. The resulting hemimethylated sites are then recognised by *maintenance DNA methyltransferases* (primarily DNMT1), which restore symmetric CpG methylation (Figure 1.8).[34] This process is essential for cell differentiation in humans.[34,35] Cytosine methylation also plays key roles in genomic imprinting and X-chromosome inactivation,[34,36] but exactly how this occurs is not well understood.[29,37]

Cytosine methylation must be tightly regulated by the cell. Aberrant methylation is a common hallmark of cancer, reportedly being the most frequent molecular lesion in cancer cells.[38] Furthermore, cytosine methylation raises the risk of C→T point mutations. This appears to arise both from its higher rate of deamination compared to cytosine,[39] and from the reduced efficiency with which T·G mismatches can be repaired compared to the U·G mismatches that result from cytosine deamination.[40,41]

Removing 5-methylcytosine from the genome

Although methyl groups can be introduced to the genome with ease, removing them is less straightforward. Direct removal of the methyl group from cytosine is unfavourable, owing to the high stability of the covalent bond connecting them.[42] Moreover, even the entire base cannot be removed by any known mammalian enzyme.[43] To circumvent these obstacles, mammalian cells instead use indirect approaches, one of which is *passive*

demethylation. In this process, the mechanism that restores symmetrical CpG methylation after replication is inhibited, leading to the dilution and, ultimately, the loss of methylation (Figure 1.8).[44]

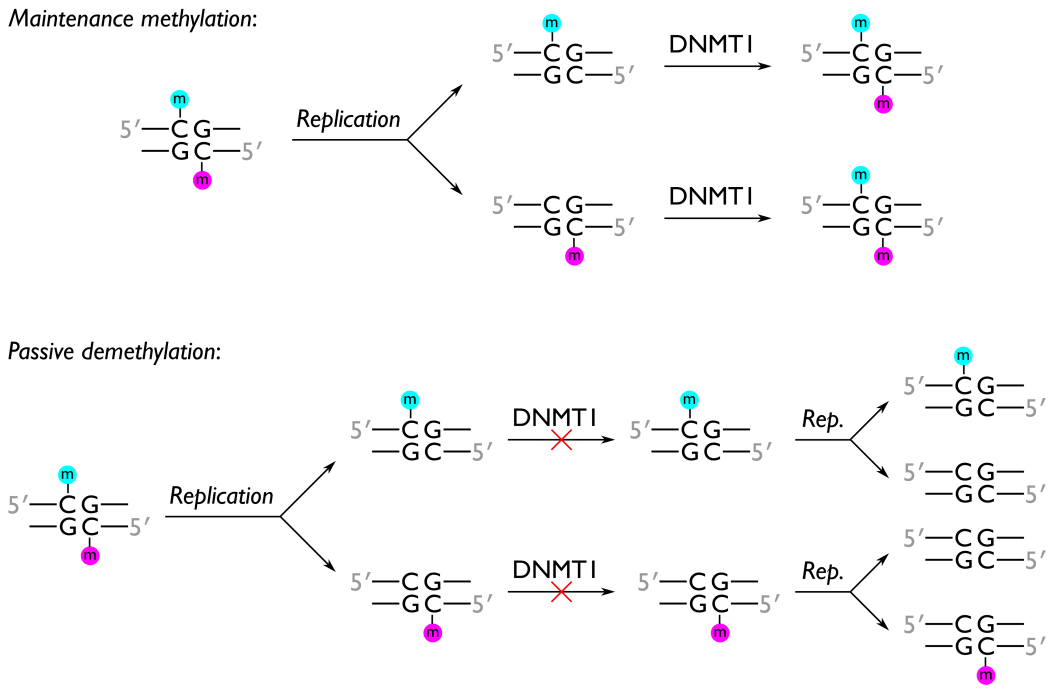


Figure 1.8. Maintenance methylation and passive demethylation. The symmetric methylation (cyan and magenta circles) of CpG steps provides a simple mechanism for the transmission of methylation patterns from parent cell to daughter cells. Following replication, the resulting hemimethylated sites are recognised by DNA methyltransferase 1 (DNMT1), which restores symmetric CpG methylation (maintenance methylation). Inhibiting the maintenance methylation process leads to the dilution and, ultimately, the loss of methylation upon replication (passive demethylation).[34]

Why 5-methylcytosine?

5-Methylcytosine's prevalence throughout nature suggests that it was not selected for arbitrarily. The methyl group itself is small, uncharged, and chemically stable. Furthermore, compared to other common aromatic-ring substituents, the methyl group only weakly perturbs electronic structure. Thus, it does not interfere with a base's normal

hydrogen-bonding pattern, which is essential for the two units of information encoded within the modified base to remain orthogonal. Base modifications that do interfere with normal Watson-Crick base pairing can lead to mutations during replication and must therefore be avoided (Figure 1.9*a,b*).[25,26]

As for the methylation site itself, the requirements imposed by chemical stability, Watson-Crick base-pairing, and the three-dimensional helical structure of DNA, all favour the 5-position of cytosine. Cytosine has 5 potential modification sites, but only C5 and C6 could accommodate a methyl group without disturbing base pairing (Figure 1.9*c*). Moreover, in the case of 6-methylcytosine, backbone atoms would clash with the methyl group, causing perturbations to normal B- and A-form helices (Figure 1.9*e*). Conversely, the methyl group of 5-methylcytosine sits more comfortably in the major groove (Figure 1.9*e*). Thus, the restrictions imposed by base pairing and DNA's three-dimensional structure strongly favour modification at the C5 position of cytosine.

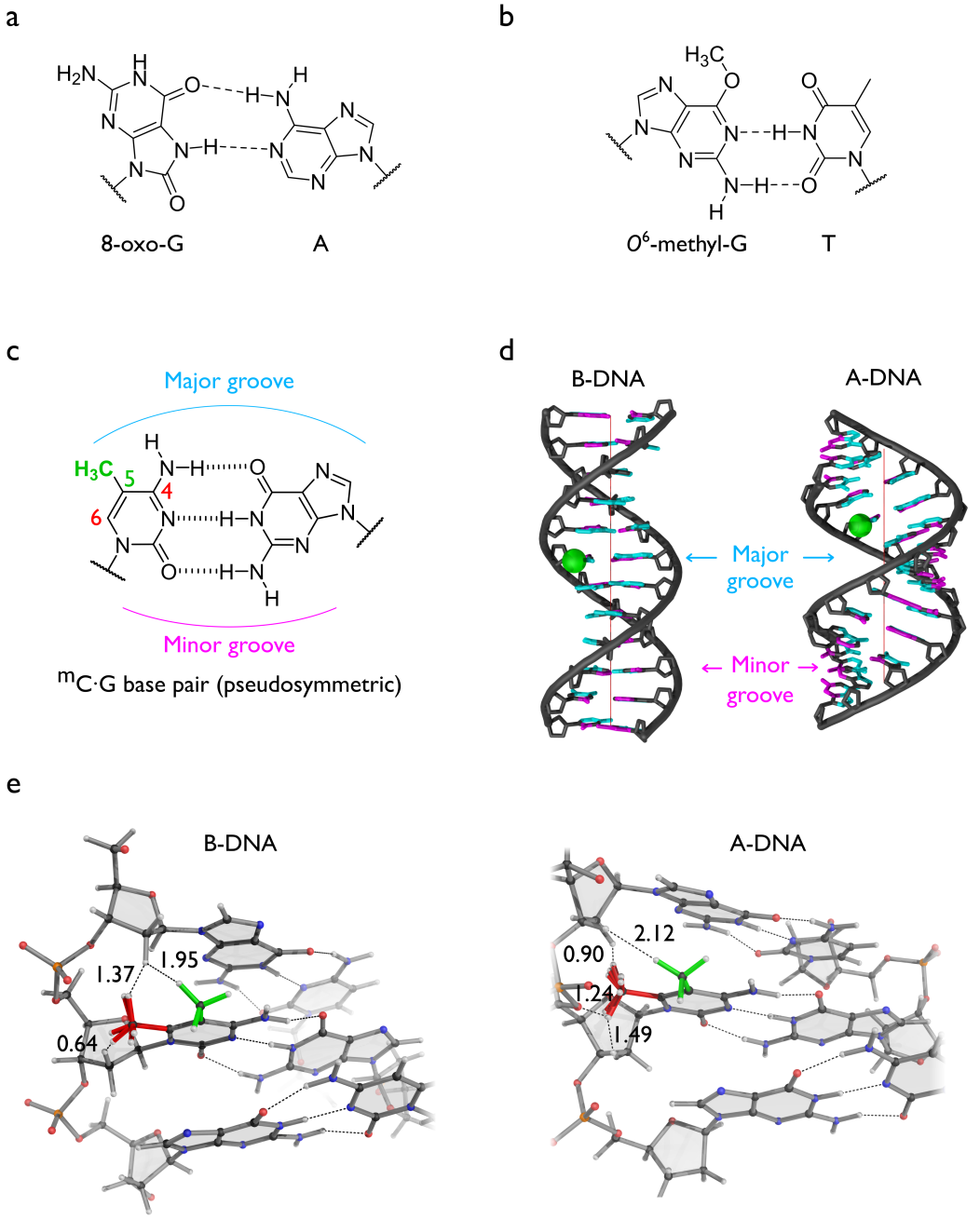


Figure 1.9. 5-Methylcytosine is favoured for structural reasons, which may partly explain its ubiquity in nature. (*a*) Substituting guanine's H8 for an oxygen atom alters its electronic structure, stabilising the mutagenic *keto* form. This leads to G→T mutations.[25] (*b*) Methylation of guanine's O6 alters its base-pairing properties, leading to G→A mutations.[26] (*c*) Methylation of cytosine's C5 or C6 positions does not perturb hydrogen-bonding. The methyl group of 5-methylcytosine (green sphere) resides in the major groove (*d*) and is better tolerated by B- and A-DNA helices than would be 6-methylcytosine (*e*). Models of 5- and 6-methylcytosine were built using UCSF Chimera.[45] *d* and *e* reproduced from Hardwick *et al.*[28]

1.2.2 The discovery of further non-canonical DNA bases in vertebrates

In the last decade, the discovery of other modified bases in the human genome has uncovered a potential pathway enabling replication-independent DNA demethylation (*active demethylation*). Remarkably, recent work suggests that these newly discovered bases might act as distinct signals in their own right, forming a rich layer of regulatory information alongside the genetic code.

5-Hydroxymethylcytosine (^{hm}C)

In 1972, ^{hm}C was reported to be present at high levels in mammalian DNA.[46] However, a later study did not reproduce these findings,[47] and thus genomic ^{hm}C had previously been thought to arise primarily from oxidative DNA damage of ^{m}C.[27,48]

In 2009, Kriaučionis and Heintz reported that the modified nucleobase 5-hydroxymethylcytosine (^{hm}C) is present at high levels in the brain—constituting up to 0.6% of all DNA bases in certain cell types.[49] Simultaneously, Tahiliani *et al* found that ^{hm}C can be formed via the oxidation of ^{m}C, mediated by the *ten-eleven translocase* (TET) family of enzymes. They also found that ^{hm}C is present in mouse embryonic stem cells.[50] Subsequent analysis has found that steady-state levels of ^{hm}C may reach 1.2% of all cytosine variants in the adult human brain.[51] By contrast, ^{hm}C levels are depleted in most types of cancer cells.[27,52]

5-Formylcytosine (^{f}C) and 5-carboxylcytosine (^{ca}C)

2011 saw the discovery of two further modified bases in the mammalian genome: 5-formylcytosine (^{f}C)[53] and 5-carboxylcytosine (^{ca}C).[54,55] These were also shown to be formed by the TET family of dioxygenase enzymes. As with ^{hm}C, the presence of ^{f}C and ^{ca}C in the genome was previously believed to occur only from oxidative DNA damage.[48]

The abundance of ^{f}C and ^{ca}C varies considerably among different cell types. The highest levels of ^{f}C have been observed in embryonic stem cells and the brain, where it accounts for around 0.002–0.02% of all cytosines.[56] Levels of ^{ca}C are typically about 10–100 times lower and are sometimes only observable in TDG-knockout cells.[27,55]

^{hm}C, ^{f}C, and ^{ca}C may be intermediates in active DNA demethylation

Both ^{f}C and ^{ca}C were found to be substrates of the DNA repair enzyme thymine DNA glycosylase (TDG), which selectively removes these bases from the genome by cleaving the base-sugar bond.[55,57] The resulting abasic site can ultimately be replaced with unmodified cytosine, via the base-excision repair (BER) pathway.[55,58] In the BER pathway, AP endonuclease 1 (APE1) converts the abasic site to a single-strand break. DNA polymerase β then inserts deoxycytidine monophosphate into the vacant site, after which XRCC1 and DNA ligase 3 ligate the nicks to regenerate double-stranded DNA.[59] Thus, the TET-mediated oxidation of ^{m}C, coupled with BER, appears to constitute a replication-independent mechanism for DNA demethylation. The pathway has been successfully reconstituted *in vitro*,[58] and its existence *in vivo* is supported by the accumulation of ^{m}C and depletion of ^{hm}C in TET-depleted cells, and by the enrichment of ^{f}C and ^{ca}C in TDG-depleted cells.[59] Furthermore, TDG depletion causes embryonic lethality in mice, suggesting essential roles for active DNA demethylation during development.[60]

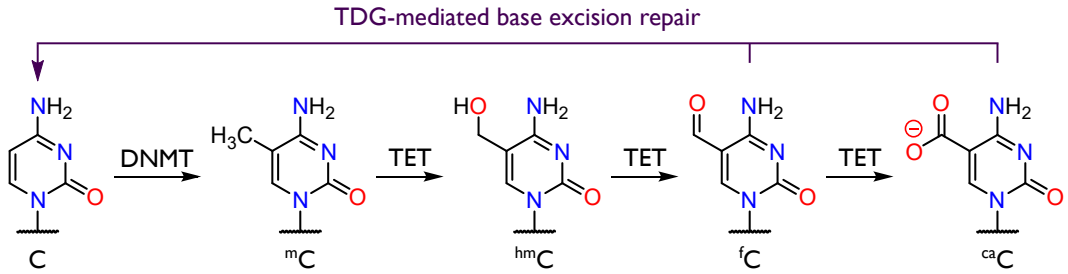


Figure 1.10. The putative active demethylation pathway of 5-methylcytosine (^{m}C), which is introduced to the genome by the DNA methyltransferase (DNMT) enzymes. First, ^{m}C is oxidised by the ten-eleven translocase (TET) enzymes, forming 5-hydroxymethylcytosine (^{hm}C), 5-formylcytosine (^{f}C), or 5-carboxylcytosine (^{ca}C). Of these ^{m}C derivatives, only ^{f}C and ^{ca}C can be recognised and removed by the enzyme thymine DNA glycosylase (TDG). The resulting abasic site is then ultimately replaced with unmodified cytosine (C) via base-excision repair. The basis of TDG's remarkable selectivity is unclear.

Oxidised ^{m}C derivatives: more than mere demethylation intermediates?

In addition to their putative role in active DNA demethylation, recent work indicates that oxidised ^{m}C derivatives may act as distinct epigenetic signals. Both ^{hm}C and ^{f}C are stable at certain genomic sites,[61–63] and many proteins selectively bind to ^{hm}C, ^{f}C, and ^{ca}C—including transcriptional regulators and chromatin remodellers.[64,65]

1.2.3 How do cytosine modifications alter DNA-protein interactions?

To elucidate the biological effects of 5-methylcytosine derivatives, a clear understanding of how they alter DNA-protein interactions is needed—for this must ultimately form the basis of any biological activity. Some key structural studies of proteins that recognise cytosine modifications are discussed below.

Recognition of 5-methylcytosine

Several protein families that recognise 5-methylcytosine have been structurally characterised, including the MBD (methyl-CpG binding domain), Kaiso-like, and SRA domain families. To highlight the different means by which ^{m}C can be recognised, two examples are considered here. For further information, the interested reader is referred to several recent reviews.[56,66,67]

The first example is MeCP2, whose importance to neuronal function is highlighted by Rett syndrome, a late-onset autism spectrum disorder primarily caused by mutations in the *MECP2* gene.[68] MeCP2 belongs to the MBD family and binds to the major groove of DNA, recognising symmetrically methylated CpG sites without disrupting base-pairing (Figure 1.11*a*). A 2.5-Å crystal structure shows that arginine fingers form short hydrogen-bonds with the two guanine residues of the methylated CpG site (Figure 1.11*b*).[69] The environment around the methyl groups is surprisingly hydrophilic: both are ≤ 3.5 Å from several tightly bound water molecules, leading the authors to speculate that the observed hydration pattern plays a key role in recognition.[69] Another notable interaction is a DNA-protein contact between a thymidine residue (dT31), two base pairs from the ^{m}CpG site, and T158 (Figure 1.11*c*). T158M is the most frequent missense mutation causing Rett syndrome. The team previously found that a run of four or more A/T base pairs adjacent to the mCpG site was essential for MeCP2 to bind with high affinity. Indeed, this thymidine residue lies within the AATT tract, whose unusually narrow minor groove (~2.9 Å) appears to favour this contact. The interaction may provide a structural basis for the importance of the A/T tract for binding affinity and highlights the importance of sequence-dependent variation in DNA-backbone geometry to protein binding.[69] More recent reports indicate that MeCP2 can also recognise ^{m}C, and even ^{hm}C, in non-CpG contexts, as well as certain unmethylated motifs.[71–73]

Another key example is the protein UHRF1,[70] which is essential for maintenance methylation in mammals. UHRF1 recognises hemimethylated CpG sites through its

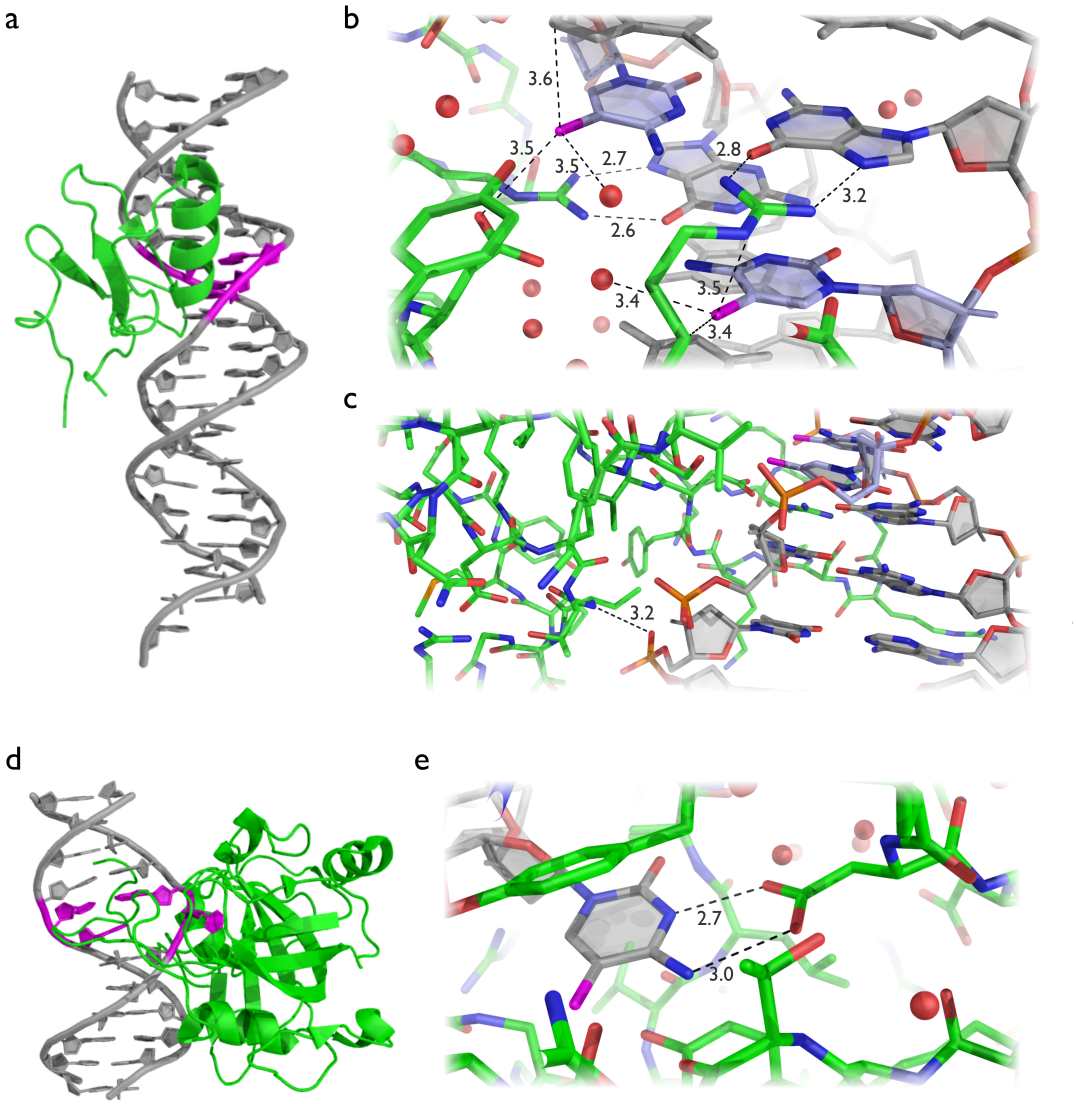


Figure 1.11. Crystal structures of MeCP2 and the SRA domain of UHRF1 bound to methylated DNA.[69] (*a*) MeCP2 binds to the major groove of DNA, interacting directly with the symmetrically methylated CpG site (magenta, PDB 3c2i).[69] (*b*) The MeCP2-DNA complex is stabilised by close interactions between argining fingers and the two guanine residues of the CpG site. The methyl groups (magenta) of the ^{m}C residues occupy a surprisingly polar environment. (*c*) The interaction between T158 of MeCP2 and dT31, whose formation may require a narrow minor groove around dT31. (*d*) The SRA domain of UHRF1 bound to hemimethylated DNA (PDB 2zkd).[70]. (*e*) The flipped-out ^{m}C residue is stabilised by interactions between its Watson-Crick face and an aspartate residue. Distances are in Å.

SRA domain (Figure 1.11*d*) and recruits DNMT1 at replication forks to restore symmetric methylation.[67] The SRA domain of UHRF1 penetrates both the minor and major grooves around the hemimethylated site and, unlike MeCP2, flips out the ^{m}C residue into a binding pocket. This flipped-out conformation is stabilised in part by interactions between an aspartate residue (D474) and the Watson–Crick face of ^{m}C (Figure 1.11*e*). Specificity for ^{m}C over C appears to arise from encapsulation of the methyl group, which fits snugly inside a hemispheric-like space.[67]

Recognition of 5-methylcytosine's oxidised derivatives

Proteomics studies have revealed many proteins that interact selectively with ^{hm}C, ^{f}C, and ^{ca}C,[64,65] but relatively few of these interactions have been well characterised. The most widely studied are complexes of TDG and the TET enzymes—both of which bind to the minor groove, sharply bend the DNA duplex, and flip out the modified base into the active site (Figure 1.12).[74–77] How TDG recognises ^{f}C and ^{ca}C with such selectivity is remarkable, given that other cytosine variants outnumber them by a factor of 1,000,000. Crystal structures of TDG in complex with formylated and carboxylated DNA indicate that hydrogen-bonds form between active site residues and the respective formyl and carboxylate groups, which probably help stabilise the flipped-out conformation (Figure 1.12*c,d*).[76,78] However, since TDG binds to abasic sites in DNA with such affinity that it is inhibited (product inhibition),[79] other factors must also strongly influence recognition. These may include variation in the physical properties of DNA, such as changes in minor-groove width, or an increased propensity to adopt the sharply bent, flipped-out conformation.

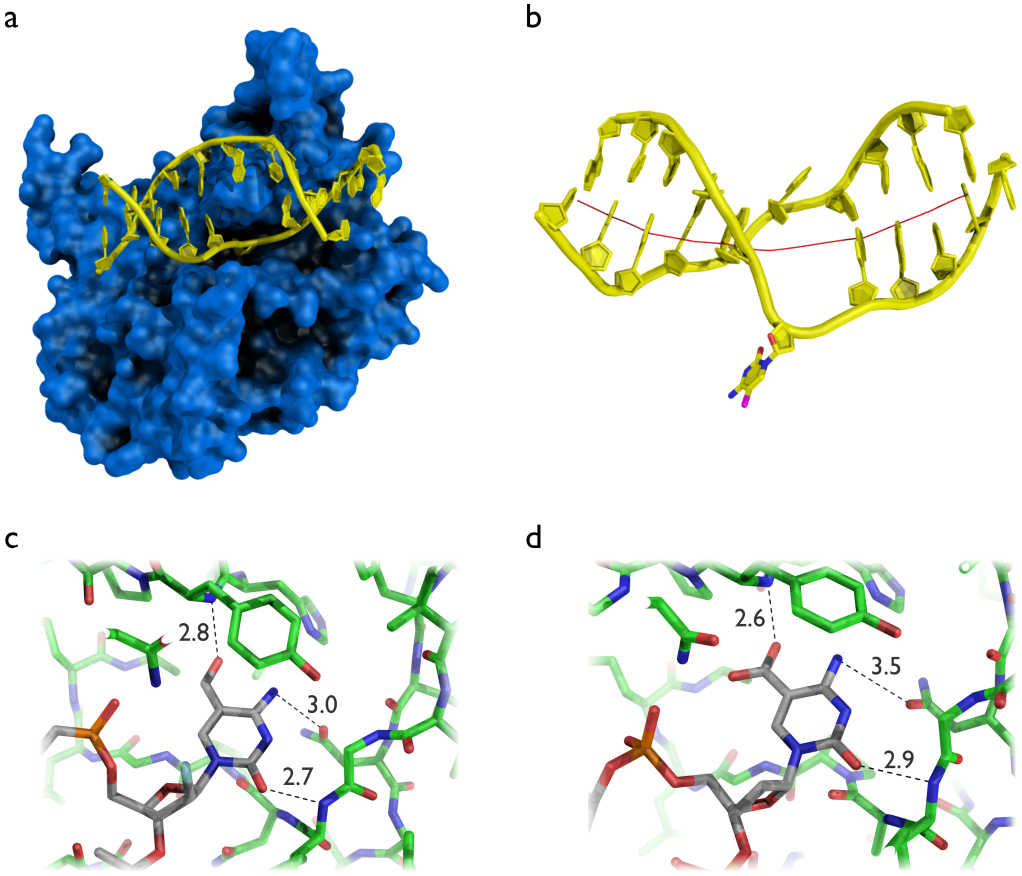


Figure 1.12. Crystal structures of TET2 and TDG. (*a*) Structure of methylated DNA in complex with TET2 (PDB 4nm6). (*b*) Close-up of the DNA, which is strongly bent. (*c*) Close-up of the interaction between TDG (green) and formylated DNA (PDB 5t2w).[76] The formyl group of ^{f}C is within hydrogen-bonding distance of the Y152 backbone N–H. (*d*) TDG in complex with ^{ca}C-DNA (PDB 3u07),[78] which forms similar interactions. Parts *a* and *b* reproduced from Hardwick *et al.*[28]

Does weakened base-pairing account for TDG's selectivity?

Whether TDG actively induces base flipping, or whether it does so passively by trapping the flipped out base following spontaneous base-pair opening, has been the subject of considerable debate.[28,80,81] For a passive base-flipping mechanism to predominate, a given cytosine modification would be expected to reduce base-pair stability. Several groups

have investigated how cytosine modifications influence duplex stability, as determined by their effect on T_m, the duplex melting temperature (reviewed by Hardwick *et al*).[28] Although some reports are contradictory, even when identical sequences were used,[82,83] they indicate that cytosine modifications have, at most, only a modest impact on duplex stability.[28] Contradictory reports may thus simply reflect differences in experimental conditions, such as the choice of buffer. In general, these studies indicate that ^{m}C exerts a weakly stabilising effect.[84–87] Conversely, hydroxylation of ^{m}C appears to reverse this modest stabilisation, with ^{hm}C-containing duplexes being of comparable stability to their unmodified analogues.[83–85,88–92] Dai *et al* recently reported that both ^{f}C and ^{ca}C significantly weaken base pairing relative to C, and they proposed that this enables their selective recognition by TDG.[83] However, the destabilising effect of ^{f}C was modest ($\Delta T_m \approx -0.5$ °C per modification),[83] which is well within the range of sequence-dependent effects on T_m for duplexes of the same length, under similar conditions.[93] In the case of ^{ca}C, the T_m of the carboxylated duplex was identical to that of its unmodified counterpart, even though 60% of its base pairs contained ^{ca}C. While Dai *et al* found that at pH 3–5, the T_m of the ^{ca}C duplex was reduced ($\Delta T_m \approx 0.7$ °C per modification at pH 4.5), the same experiments were not performed on the unmodified duplex. Thus, with no control, it is unclear whether the effect is ^{ca}C-specific. Given that the *N*3 of unmodified cytosine has a pK_a of 4.5,[83] it would likely be significantly protonated at pH 3–5, thereby disrupting base pairing. One might therefore expect similar destabilisation of the unmodified duplex at low pH.

Notably, other groups have found that ^{f}C and ^{ca}C have no effect on,[94,95] or even increase,[89,94,96] duplex stability. Szulik *et al* used NMR spectroscopy to probe the opening dynamics of individual base pairs in the Dickerson-Drew dodecamer, d(CGCGAATTX-GCG)$_2$, where X = C, ^{m}C, ^{hm}C, ^{f}C, or ^{ca}C.[89] The authors observed only modest differences in the opening rate of the modified base pair X·G, with ^{m}C or ^{ca}C causing a slight reduction, and ^{hm}C or ^{f}C causing a modest increase. Differences in both the rate

and equilibrium constant of base-pair opening arising from cytosine modification were smaller than those observed between canonical C·G and A·T base pairs.[89]

Overall, these reports indicate that differences in the openining dynamics and stability of base-pairs arising from cytosine modification are comparable to sequence-dependent effects. It therefore seems unlikely that base-pair stability plays a major role in determining TDG's selectivity.

1.2.4 Can cytosine modifications be read indirectly via changes to DNA conformation?

In the examples highlighted previously, cytosine modifications have been found to alter DNA-protein interactions via direct contacts with the modification. But, as described above, an alternative means of molecular recognition, *indirect readout*, is widely used by DNA-binding proteins.[17] For cytosine modifications to alter DNA-protein interactions through indirect readout, they must perturb DNA conformation. Structural studies of modified DNA—in the absence of protein—are therefore essential to understanding how cytosine modifications may do so. Prior to this work, crystallographic studies of isolated DNA containing ^{m}C, ^{hm}C, ^{f}C, or ^{ca}C had indicated that, at least in hemimodified CpG sites, these modifications do not significantly alter DNA structure.[28,84,88,89] By contrast, ^{f}C had been found to dramatically alter DNA conformation when present in consecutive, symmetrically modified CpG sites. The authors concluded that this conformation represented a new helical form of DNA, which they termed F-DNA.[82] However, experimental artefacts and effects of the underlying sequence could not be ruled out, as detailed in Chapter 2. Thus, more work is needed to determine the extent to which ^{f}C alters DNA structure.

Reports have also found differences in the physical properties DNA containing other derivatives of ^{m}C (reviewed by Hardwick *et al*).[28] For instance, ^{m}C has been found to

impact the susceptibility of DNA to cleavage by the minor-groove binding endonuclease DNase I.[97–99] This demonstrates that methylation can influence DNA-protein interactions solely through its effects on DNA conformation, since DNase I cannot interact directly with cytosine's C5 modifications. The influence of other cytosine modifications on DNase I cleavage susceptibility has not yet been investigated. Recently, Ngo *et al* found that ^{hm}C and ^{f}C increase the flexibility of duplex DNA, which increases its propensity to form stable nucleosomes.[100] However, the study did not account for any stable curvature (or anisotropic flexibility) associated with the modifications. Such effects may be significant; earlier studies using gel mobility assays have indicated that cytosine methylation induces bending into the minor groove.[101] The effects of ^{hm}C, ^{f}C, and ^{ca}C on DNA bending remain unclear. Given that key interactions between modified DNA and proteins (such as TET, TDG, the nucleosome core particle) involve sharp bending of the DNA, perturbations to DNA bending arising from these modifications could be biologically important and should be investigated.

1.3 Methods to Study DNA Structure

Structural studies of DNA and its complexes with protein are essential for a mechanistic understanding of biology. X-ray crystallography and NMR spectroscopy have long been the predominant methods of DNA structure determination, as demonstrated by a search of the PDB. So far, 7,388 atomic structures containing DNA have been deposited. The vast majority of these have been determined by X-ray crystallography (83%, 6,111 structures), followed by NMR spectroscopy (12%, 917 structures).* However, shortcomings in both techniques—particularly in reliably measuring long-range distances—have given rise to the development of alternative methods, which are discussed in this section.

*From a total of 160,233 biomolecular structures deposited at the time of writing.

1.3.1 X-ray crystallography

X-ray crystallographic studies underpin much of what is known about the structure of DNA, providing information of unrivalled detail and abundance. But, despite its strengths, X-ray crystallography has two key limitations—both of which were encountered during the course of this research. First, crystal structures cannot be obtained if well-diffracting crystals of the target molecule cannot be grown (see Section 4.2.3). In addition to hampering countless structural studies, this has greatly limited the number of sequences that have been studied. Of the 1,109 crystal structures of naked DNA currently deposited in the PDB, over 10% of these (118) are of a single sequence, the Dickerson-Drew dodecamer. Furthermore, the sequence diversity of B-DNA crystal structures must be even worse, given that a large number of DNA crystal structures are of the A- or Z-form.

Another major limitation of X-ray crystallography is that even if a crystal structure can be determined, it may not be representative of the molecule's preferred conformation in solution (explored in Chapters 2 and 3). X-ray crystal structures provide only conformational snapshots that may be strongly affected by crystal-packing forces, otherwise known as 'the tyranny of the lattice' (Figure 1.13).[102]

1.3.2 NMR spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy can also yield highly detailed information on nucleic acid structure and dynamics. And because the technique does not require crystals to be grown, NMR studies are not limited to certain crystallisable sequences and are free from crystal-packing artefacts. Furthermore, the technique probes the conformational ensemble, rather than an individual snapshot. As such, NMR provides valuable information that complements X-ray crystallographic studies (see Chapters 2 and 4).

Particularly important for characterising DNA structure is the 2D NMR experiment nuclear Overhauser effect spectroscopy (NOESY). NOESY experiments enable the

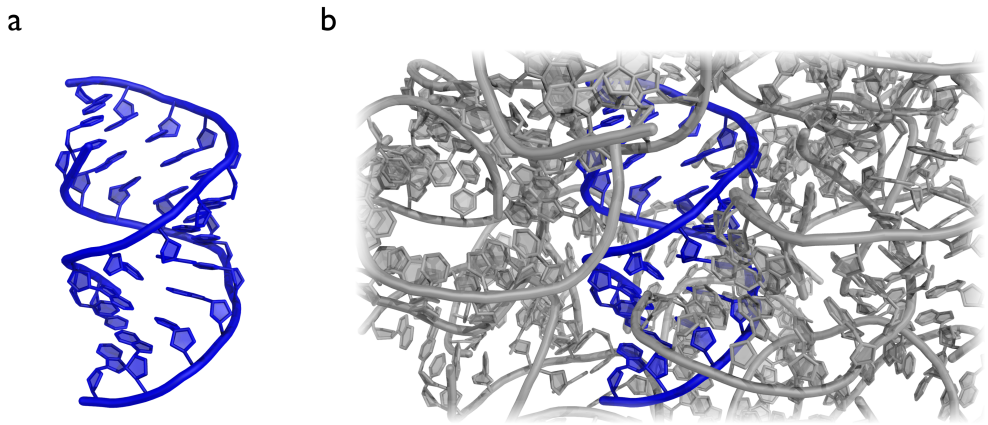


Figure 1.13. The tyranny of the lattice. (*a*) How a DNA crystal structure is usually potrayed. (*b*) The claustrophobic reality: DNA duplexes within a crystal are densely packed together, which can lead to structural features that are not representative of the molecule's conformation in solution (see Chapter 2). The crystal structure shown is PDB 5mvk.[103]

distance between two nuclei to be calculated, by measuring the through-space coupling interaction between them. Because the strength of this coupling interaction is proportional to r^{-6} (where r is the internuclear separation),[104] NOESY measurements are extremely sensitive to small changes in distance, allowing a nucleotide's glycosyl torsion angle to be readily determined. However, measurable distances are typically limited to within 5 Å.[104] Another important 2D experiment is double-quantum filtered correlation spectroscopy (DQF-COSY). The geminal J-coupling values derived from DQF-COSY experiments are sensitive to the torsion angle of a given bond, enabling the puckering of deoxyribose sugars to be determined. Additionally, a simple comparison of chemical shift differences can be used to gain a qualitative understanding of the difference in conformation between two DNA duplexes.

Overall, NMR can be used to determine the global conformation of a given DNA duplex, and to probe local structural differences between two DNA duplexes within a conformational class. However, using NMR to study DNA structure also has some caveats. The relatively high spatial separation of neighbouring protons in DNA, coupled

with the lack of long-range translational information afforded by the technique, poses significant challenges to obtaining accurate NMR structures of DNA.[105] Furthermore, NMR studies are generally limited to short DNA fragments (typically less than 25 bp).[6]

1.3.3 Molecular rulers

The propensity of DNA to bend in a given direction plays a key role in its interactions with protein, such as governing nucleosome stability and organisation (see Section 1.1.4). However, the limitations of X-ray crystallography and NMR outlined above make it extremely challenging to characterise DNA bending, and other structural features, in a conformational ensemble. Consequently, so-called molecular rulers have found widespread application in structural biology.[106–108] The molecular ruler approach typically involves incorporating two probes into the molecule of interest. The distance-dependent interaction between these probes enables their separation to be calculated, from which the distance between two specific points in the biomolecule can be inferred (Figure 1.14*a*). But the accuracy and precision of the result strongly depend on the probe's physicochemical properties, as well as detailed knowledge of its effect on biomolecular structure.

Perhaps the most widely used molecular ruler approach is Förster resonance energy transfer (FRET), which exploits the distance-dependent energy transfer between a pair of spectrally overlapping fluorophores. Despite its ubiquity,[109,110] FRET does not yield quantitatively accurate distance values due to uncertainties in the location, orientation, and dynamics of the fluorophores.[111] Moreover, when several distinct conformations exist in an ensemble, FRET cannot accurately determine the distribution of distances.[111,112]

Recently, an alternative molecular ruler approach involving electron paramagnetic resonance (EPR) spectroscopy has shown promise in addressing these limitations. Using the EPR technique *double electron-electron resonance* (DEER), inter-probe distances can be measured with sub-nanometre precision across a similar distance range to FRET (~2–10 nm). In place of fluorophores, DEER uses a pair of stable free-radical probes

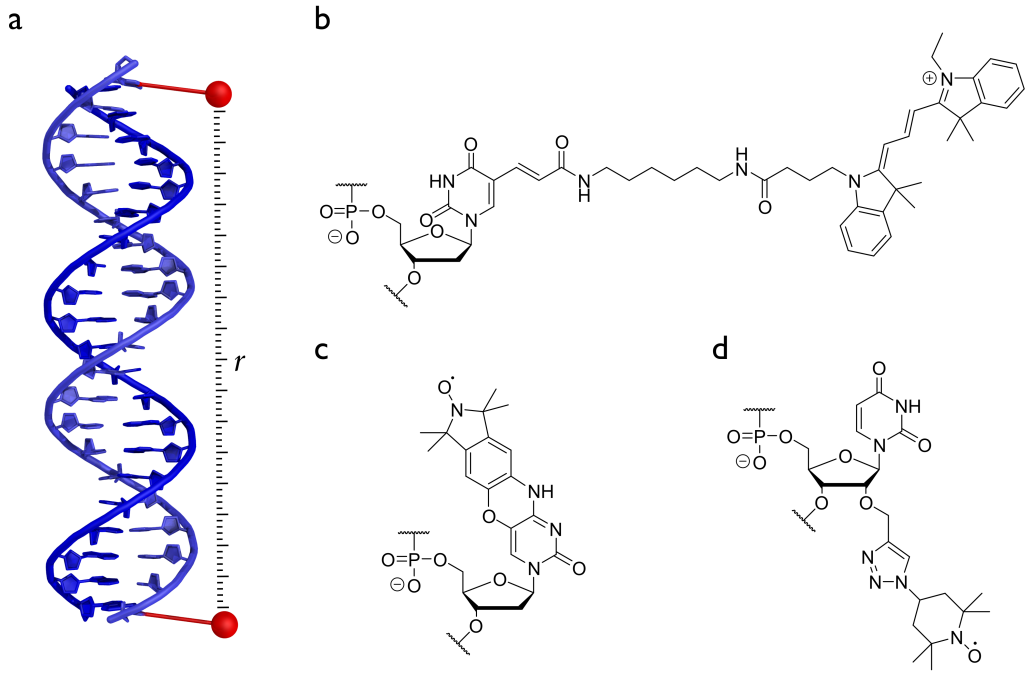


Figure 1.14. Molecular rulers for probing DNA conformation. (*a*) Schematic of the molecular ruler approach. The distance-dependent interaction between two probes bound to a DNA duplex enables their separation, *r*, to be calculated. From this, the distance between two points in the helix can be inferred. But the accuracy and precision of the result strongly depend on the properties of the probe—and detailed knowledge of its effect on DNA structure. (*b*) An example of one such probe used in FRET: the 'eCy3' fluorophore, tethered to the base via a flexible linker.[113] (*c,d*) Examples of spin labels, whose separation can be determined by EPR spectroscopy. Typically, spin labels are smaller and more rigidly bound than fluorophores, affording measurements of higher precision.[107]

(*spin labels*), which are incorporated at defined positions—known as site-directed spin labelling (SDSL). Because spin labels are typically smaller and more rigidly bound than are fluorophores (Figure 1.14*b*–*e*), their use enables higher measurement precision and may reduce structural perturbation.[107,114] Furthermore, rather than giving a single, averaged distance measurement, SDSL in conjunction with DEER enables distance distributions to be determined, shedding light on the structures and populations of different states in a conformational ensemble.[115]

To accurately interpret DEER-derived distance measurements, the location and orientation of the spin labels, and their effect on DNA structure, must be known.[111,116] However, spin-labelled DNA has not yet been fully characterised, limiting the method's reliability. A comprehensive structural study of spin-labelled DNA would thus advance the field.

1.4 Aims of this work

This thesis is concerned with how chemical modifications to DNA alter its structure. Two classes of modification were investigated: naturally occurring cytosine modifications, and unnatural modifications to the deoxyribose sugar. As such, the thesis is divided into two parts, each with a distinct aim.

Part I of this thesis describes biophysical investigations into the effects of naturally occurring cytosine modifications on DNA structure. Although the function of these recently discovered ^{m}C derivatives is currently unknown, they have been implicated in development,[59] learning and memory,[117] and their misregulation is associated with cancer and other diseases.[59] Understanding the role of ^{m}C derivatives, and the mechanisms by which they are introduced and removed from the genome, could therefore expand our knowledge of fundamental biology and may help to address important clinical needs. To elucidate the biological effects of these modified bases, a precise understanding of how they alter DNA-protein interactions is needed. While the direct readout of

cytosine modifications by proteins is well established, how cytosine modifications may alter DNA-protein interactions indirectly—through perturbations to DNA structure—remains unclear. Recently, ^{f}C was reported to give rise to a unique helical conformation of DNA, which was proposed to enable its specific recognition by proteins.[82] However, key conclusions were based on a single crystal structure, and effects of the underlying sequence could not be ruled out. Chapter 2 explores the extent to which ^{f}C alters DNA structure, with results from X-ray crystallography, NMR, CD and UV-absorption spectroscopy. Reports also indicate that other derivatives of ^{m}C can alter the physical properties of duplex DNA, though it remains unclear whether these modifications can be sensed indirectly through changes to DNA bending, minor-groove geometry, or both. This was investigated using DNase I and gel mobility assays and is described in Chapter 3.

Part II describes the structural characterisation of DNA containing recently developed 2′-alkynyl-tethered nitroxide spin labels. Owing to the limitations of established biophysical techniques, obtaining accurate, long-range distance information in conformational ensembles remains an important goal in structural biology. Site-directed spin labelling coupled with EPR spectroscopy is a promising method that can yield such information. However, its use is limited in part by a lack of knowledge of the spin labels' location and effects on DNA structure. The work described in Chapter 4 sought to use complementary biophysical techniques to characterise the structure of DNA containing 2′-alkynyl-tethered nitroxide spin labels, to establish their use as an accurate tool to probe DNA structure.

BIBLIOGRAPHY

1. Paux, E. *et al.* A Physical Map of the 1-Gigabase Bread Wheat Chromosome 3B. *Science* **322,** 101–104 (2008).

2. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171,** 737–738 (1953).

3. Franklin, R. E. & Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **171,** 740–741 (1953).

4. Wilkins, M. H. F., Stokes, A. R. & Wilson, H. R. Molecular Structure of Deoxypentose Nucleic Acids. *Nature* **171,** 738–740 (1953).

5. Neidle, S. *Principles of Nucleic Acid Structure* 1st Ed. (Elsevier, London, 2008).

6. Blackburn, M. G., Gait, M. J., Loakes, D. & Williams, D. M. *Nucleic Acids in Chemistry and Biology* (Royal Society of Chemistry, Cambridge, 2007).

7. Franklin, R. E. & Gosling, R. G. The structure of sodium thymonucleate fibres. I. The influence of water content. *Acta Crystallogr.* **6,** 673–677 (1953).

8. Mitsui, Y. *et al.* Physical and Enzymatic Studies on Poly d(I–C).Poly d(I–C), an Unusual Double-helical DNA. *Nature* **228,** 1166–1169 (1970).

9. Pohl, F. M. & Jovin, T. M. Salt-induced co-operative conformational change of a synthetic DNA: Equilibrium and kinetic studies with poly(dG-dC). *J. Mol. Biol.* **67,** 375–396 (1972).

10. Wang, A. H. *et al.* Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282,** 680–686 (1979).

11. Rich, A. & Zhang, S. Z-DNA: the long road to biological function. *Nat. Rev. Genet.* **4,** 566–572 (2003).

12. Fujii, S., Wang, A. H.-J., van der Marel, G., van Boom, J. H. & Rich, A. Molecular structure of (m5 dC-dG)3 : the role of the methyl group on 5-methyl cytosine in stabilizing Z-DNA. *Nucleic Acids Res.* **10,** 7879–7892 (1982).

13. Wing, R. *et al.* Crystal structure analysis of a complete turn of B-DNA. *Nature* **287,** 755–758 (1980).

14. Drew, H. R. *et al.* Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 2179–2183 (1981).

15. Neidle, S. *Oxford Handbook of Nucleic Acid Structure* (Oxford University Press, Oxford, 1999).

16. Calladine, C. R., Drew, H., Luisi, B. & Travers, A. *Understanding DNA: The Molecule and How it Works* (Elsevier, 2004).

17. Rohs, R. *et al.* The role of DNA shape in protein-DNA recognition. *Nature* **461,** 1248–1253 (2009).

18. Raiber, E.-A., Hardisty, R., van Delft, P. & Balasubramanian, S. Mapping and elucidating the function of modified bases in DNA. *Nat. Rev. Chem.* **1,** 0069 (2017).

19. Widom, J. Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* **34,** 269–324 (2001).

20. Chua, E. Y. D., Vasudevan, D., Davey, G. E., Wu, B. & Davey, C. A. The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Res.* **40,** 6338–6352 (2012).

21. Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett.* **589,** 2914–2922 (2015).

22. Struhl, K. & Segal, E. Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **20,** 267–273 (2013).

23. Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467,** 562–566 (2010).

24. Travers, A. A. The structural basis of DNA flexibility. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **362,** 1423–1438 (2004).

25. Grollman, A. P. & Moriya, M. Mutagenesis by 8-oxoguanine: an enemy within. *Trends Genet.* **9,** 246–249 (1993).

26. Warren, J. J., Forsberg, L. J. & Beese, L. S. The structural basis for the mutagenicity of O6-methyl-guanine lesions. *Proc. Natl. Acad. Sci.* **103,** 19701–19706 (2006).

27. Carell, T., Kurz, M. Q., Müller, M., Rossa, M. & Spada, F. Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA. *Angew. Chemie Int. Ed.* **57,** 4296–4312 (2018).

28. Hardwick, J. S., Lane, A. N. & Brown, T. Epigenetic Modifications of Cytosine: Biophysical Properties, Regulation, and Function in Mammalian DNA. *BioEssays,* 1700199 (2018).

29. Deaton, A. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25,** 1010–1022 (2011).

30. Wyatt, G. R. Occurrence of 5-Methyl-Cytosine in Nucleic Acids. *Nature* **166,** 237–238 (1950).

31. Breiling, A. & Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* **8,** 24 (2015).

32. Illingworth, R. S. & Bird, A. P. CpG islands – 'A rough guide'. *FEBS Lett.* **583,** 1713–1720 (2009).

33. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14,** 204–220 (2013).

34. Alberts, B. *et al. Molecular Biology of the Cell* 6th (Garland Science, New York, 2015).

35. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13,** 484–492 (2012).

36. Schübeler, D. Function and information content of DNA methylation. *Nature* **517,** 321–326 (2015).

37. Moarii, M., Boeva, V., Vert, J.-P. & Reyal, F. Changes in correlation between promoter methylation and gene expression in cancer. *BMC Genomics* **16,** 873 (2015).

38. Esteller, M. Aberrant DNA methylation as a cancer-inducing mechanism. *Annu. Rev. Pharmacol. Toxicol.* **45,** 629–656 (2005).

39. Shen, J.-C., Rideout, W. M. & Jones, P. A. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* **22,** 972–976 (1994).

40. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).

41. Schmutte, C., Yang, A. S., Beart, R. W. & Jones, P. A. Base excision repair of U:G mismatches at a mutational hotspot in the p53 gene is more efficient than

base excision repair of T:G mismatches in extracts of human colon tumors. *Cancer Res.* **55,** 3742–6 (1995).

42. Kriukienė, E., Liutkevičiūtė, Z. & Klimašauskas, S. 5-Hydroxymethylcytosine – the elusive epigenetic mark in mammalian DNA. *Chem. Soc. Rev.* **41,** 6916–6930 (2012).

43. Bhutani, N., Burns, D. M. M. & Blau, H. M. M. DNA Demethylation Dynamics. *Cell* **146,** 866–872 (2011).

44. Chen, Z.-x. & Riggs, A. D. DNA Methylation and Demethylation in Mammals. *J. Biol. Chem.* **286,** 18347–18353 (2011).

45. Pettersen, E. F. *et al.* UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

46. Penn, N. W., Suwalski, R., O'Riley, C., Bojanowski, K. & Yura, R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.* **126,** 781–790 (1972).

47. Kothari, R. M. & Shankar, V. 5-methylcytosine content in the vertebrate deoxyribonucleic acids: Species specificity. *J. Mol. Evol.* **7,** 325–329 (1976).

48. Privat, E. & Sowers, L. C. Photochemical Deamination and Demethylation of 5-Methylcytosine. *Chem. Res. Toxicol.* **9,** 745–750 (1996).

49. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324,** 929–930 (2009).

50. Tahiliani, M. *et al.* Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324,** 930–935 (2009).

51. Wagner, M. *et al.* Age-Dependent Levels of 5-Methyl-, 5-Hydroxymethyl-, and 5-Formylcytosine in Human and Mouse Brain Tissues. *Angew. Chemie Int. Ed.* **54,** 12511–12514 (2015).

52. Pfeifer, G. P. & Szabó, P. E. Gene body profiles of 5-hydroxymethylcytosine: potential origin, function and use as a cancer biomarker. *Epigenomics* **10,** 1029–1032 (2018).

53. Pfaffeneder, T. *et al.* The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA. *Angew. Chemie Int. Ed.* **50,** 7008–7012 (2011).

54. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–1303 (2011).

55. He, Y.-F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **333,** 1303–1307 (2011).

56. Rausch, C., Hastert, F. D. & Cardoso, M. C. DNA modification readers and writers and their interplay. *J. Mol. Biol.* (2019).

57. Maiti, A. & Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine. *J. Biol. Chem.* **286,** 35334–35338 (2011).

58. Weber, A. R. *et al.* Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat. Commun.* **7,** 10806 (2016).

59. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18,** 517–534 (2017).

60. Cortellino, S. *et al.* Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. *Cell* **146,** 67–79 (2011).

61. Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* **6,** 1049–1055 (2014).

62. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11,** 3–6 (2015).

63. Su, M. *et al.* 5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites. *Angew. Chemie Int. Ed.* **55,** 11797–11800 (2016).

64. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14,** R119 (2013).

65. Spruijt, C. G. *et al.* Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* **152,** 1146–1159 (2013).

66. Ludwig, A. K., Zhang, P. & Cardoso, M. C. Modifiers and readers of DNA modifications and their impact on genome structure, expression, and stability in disease. *Front. Genet.* **7,** 1–24 (2016).

67. Patel, D. J. A Structural Perspective on Readout of Epigenetic Histone and DNA Methylation Marks. *Cold Spring Harb. Perspect. Biol.* **8,** a018754 (2016).

68. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23,** 185–188 (1999).

69. Ho, K. L. *et al.* MeCP2 Binding to DNA Depends upon Hydration at Methyl-CpG. *Mol. Cell* **29,** 525–531 (2008).

70. Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. & Shirakawa, M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* **455,** 818–821 (2008).

71. Gabel, H. W. *et al.* Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522,** 89–93 (2015).

72. Lagger, S. *et al.* MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain. *PLoS Genet.* **13,** 1–26 (2017).

73. Lei, M., Tempel, W., Chen, S., Liu, K. & Min, J. Plasticity at the DNA recognition site of the MeCP2 mCG-binding domain. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1862,** 194409 (2019).

74. Hu, L. *et al.* Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **155,** 1545–1555 (2013).

75. Hu, L. *et al.* Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527,** 118–122 (2015).

76. Pidugu, L. S. *et al.* Structural Basis for Excision of 5-Formylcytosine by Thymine DNA Glycosylase. *Biochemistry* **55,** 6205–6208 (2016).

77. Hashimoto, H., Hong, S., Bhagwat, A. S., Zhang, X. & Cheng, X. Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* **40,** 10203–10214 (2012).

78. Zhang, L. *et al.* Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* **8,** 1–27 (2012).

79. Morgan, M. T., Maiti, A., Fitzgerald, M. E. & Drohat, A. C. Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA. *Nucleic Acids Res.* **39,** 2319–2329 (2011).

80. Dodd, T., Yan, C., Kossmann, B. R., Martin, K. & Ivanov, I. Uncovering universal rules governing the selectivity of the archetypal DNA glycosylase TDG. *Proc. Natl. Acad. Sci.* **115,** 5974–5979 (2018).

81. Da, L.-T. & Yu, J. Base-flipping dynamics from an intrahelical to an extrahelical state exerted by thymine DNA glycosylase during DNA repair process. *Nucleic Acids Res.* **46,** 5410–5425 (2018).

82. Raiber, E.-A. *et al.* 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22,** 44–49 (2014).

83. Dai, Q. *et al.* Weakened N3 Hydrogen Bonding by 5-Formylcytosine and 5-Carboxylcytosine Reduces Their Base-Pairing Stability. *ACS Chem. Biol.* **11,** 470–477 (2016).

84. Renciuk, D., Blacque, O., Vorlickova, M. & Spingler, B. Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Res.* **41,** 9891–9900 (2013).

85. Thalhammer, A., Hansen, A. S., El-Sagheer, A. H., Brown, T. & Schofield, C. J. Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation. *Chem. Commun.* **47,** 5325 (2011).

86. Nardo, L. *et al.* Effects of non-CpG site methylation on DNA thermal stability: a fluorescence study. *Nucleic Acids Res.* **43,** 10722–10733 (2015).

87. Wojdacz, T. K., Dobrovic, A. & Hansen, L. L. Methylation-sensitive high-resolution melting. *Nat. Protoc.* **3,** 1903–1908 (2008).

88. Lercher, L. *et al.* Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem. Commun.* **50,** 1794–1796 (2014).

89. Szulik, M. W. *et al.* Differential Stabilities and Sequence-Dependent Base Pair Opening Dynamics of Watson–Crick Base Pairs with 5-Hydroxymethylcytosine, 5-Formylcytosine, or 5-Carboxylcytosine. *Biochemistry* **54,** 1294–1305 (2015).

90. Münzel, M., Globisch, D., Trindler, C. & Carell, T. Efficient Synthesis of 5-Hydroxymethylcytosine Containing DNA. *Org. Lett.* **12,** 5671–5673 (2010).

91. Carson, S., Wilson, J., Aksimentiev, A., Weigele, P. R. & Wanunu, M. Hydroxymethyluracil modifications enhance the flexibility and hydrophilicity of double-stranded DNA. *Nucleic Acids Res.* **44,** 2085–2092 (2016).

92. Wanunu, M. *et al.* Discrimination of Methylcytosine from Hydroxymethylcytosine in DNA Molecules. *J. Am. Chem. Soc.* **133,** 486–492 (2011).

93. Owczarzy, R. *et al.* Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers* **44,** 217–239 (1997).

94. Münzel, M. *et al.* Improved Synthesis and Mutagenicity of Oligonucleotides Containing 5-Hydroxymethylcytosine, 5-Formylcytosine and 5-Carboxylcytosine. *Chem. - A Eur. J.* **17,** 13782–13788 (2011).

95. Karino, N., Ueno, Y. & Matsuda, A. Synthesis and properties of oligonucleotides containing 5-formyl-2'-deoxycytidine: in vitro DNA polymerase reactions on DNA templates containing 5-formyl-2'-deoxycytidine. *Nucleic Acids Res.* **29,** 2456–2463 (2001).

96. Sumino, M., Ohkubo, A., Taguchi, H., Seio, K. & Sekine, M. Synthesis and properties of oligodeoxynucleotides containing 5-carboxy-2'-deoxycytidines. *Bioorganic Med. Chem. Lett.* **18,** 274–277 (2008).

97. Lazarovici, A. *et al.* Probing DNA shape and methylation state on a genomic scale with DNase I. *Proc. Natl. Acad. Sci.* **110,** 6376–6381 (2013).

98. Kochanek, S., Renz, D. & Doerfler, W. Differences in the accessibility of methylated and unmethylated DNA to DNase I. *Nucleic Acids Res.* **21,** 5843–5845 (1993).

99. Fox, K. R. The effect of Hha I methylation on DNA local structure. *Biochem. J.* **234,** 213–216 (1986).

100. Ngo, T. T. M. *et al.* Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **7,** 10813 (2016).

101. Hodges-Garcia, Y. & Hagerman, P. J. Cytosine methylation can induce local distortions in the structure of duplex DNA. *Biochemistry* **31,** 7595–7599 (1992).

102. Dickerson, R. E., Goodsell, D. S. & Neidle, S. "...the tyranny of the lattice...". *Proc. Natl. Acad. Sci.* **91,** 3579 LP –3583 (1994).

103. Hardwick, J. S. *et al.* 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.* **24,** 544–552 (2017).

104. Jacobsen, N. E. *NMR spectroscopy explained: simplified theory, applications and examples for organic chemistry and structural biology* (John Wiley & Sons, 2007).

105. Kuszewski, J., Schwieters, C. & Clore, G. M. Improving the Accuracy of NMR Structures of DNA by Means of a Database Potential of Mean Force Describing Base-Base Positional Interactions. *J. Am. Chem. Soc.* **123,** 3903–3918 (2001).

106. Craggs, T. D. & Kapanidis, A. N. Six steps closer to FRET-driven structural biology. *Nat. Methods* **9,** 1157–1158 (2012).

107. Haugland, M. M., Lovett, J. E. & Anderson, E. A. Advances in the synthesis of nitroxide radicals for use in biomolecule spin labelling. *Chem. Soc. Rev.* **47,** 668–680 (2018).

108. Shi, X., Bonilla, S., Herschlag, D. & Harbury, P. in *Methods Enzymol.* 1st ed. 1, 75–97 (Elsevier Inc., 2015).

109. Selvin, P. R. The renaissance of fluorescence resonance energy transfer. *Nat. Struct. Biol.* **7,** 730–734 (2000).

110. Lerner, E. *et al.* Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. *Science* **359,** eaan1133 (2018).

111. Mathew-Fenn, R. S., Das, R., Silverman, J. A., Walker, P. A. & Harbury, P. A. B. A Molecular Ruler for Measuring Quantitative Distance Distributions. *PLoS One* **3,** e3229 (2008).

112. Shi, X., Herschlag, D. & Harbury, P. A. B. Structural ensemble and microscopic elasticity of freely diffusing DNA by direct measurement of fluctuations. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E1444–51 (2013).

113. Lee, W., von Hippel, P. H. & Marcus, A. H. Internally labeled Cy3/Cy5 DNA constructs show greatly enhanced photo-stability in single-molecule FRET experiments. *Nucleic Acids Res.* **42,** 5967–5977 (2014).

114. Azarkh, M. *et al.* Site-directed spin-labeling of nucleotides and the use of in-cell EPR to determine long-range distances in a biologically relevant environment. *Nat. Protoc.* **8,** 131–147 (2013).

115. Jeschke, G. The contribution of modern EPR to structural biology. *Emerg. Top. Life Sci.* **2,** 9–18 (2018).

116. Edwards, T. E. *et al.* Crystal structure of a DNA containing the planar, phenoxazine-derived bi-functional spectroscopic probe. *Nucleic Acids Res.* **39,** 4419–4426 (2011).

117. Li, X. *et al.* Neocortical Tet3-mediated accumulation of 5-hydroxymethylcytosine promotes rapid behavioral adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 7120–7125 (2014).

Part I.

Structural studies of DNA containing epigenetic cytosine modifications

2

F-DNA: A NEW FORM OF THE DNA DOUBLE HELIX?

2.1 Introduction

5-Methylcytosine (^{m}C) cannot be removed directly from the human genome by any known enzyme. Instead, ^{m}C is oxidised sequentially to 5-hydroxymethylcytosine (^{hm}C),[1,2] 5-formylcytosine (^{f}C),[3] and 5-carboxylcytosine (^{ca}C).[4,5] Of these cytosine derivatives, only ^{f}C and ^{ca}C can be recognised and removed from the genome—by the enzyme thymine DNA glycosylase (TDG).[6,7] Following their removal, ^{f}C and ^{ca}C may then be replaced by unmodified cytosine via the base excision repair pathway, completing the process of active demethylation.[8]

^{f}C is a key intermediate in active DNA demethylation, being the first cytosine derivative in the pathway that can be excised from the genome. Furthermore, recent work suggests that the biological function of ^{f}C might go beyond its role in demethylation. ^{f}C is stable at certain genomic sites,[9,10] and its genomic profile is distinct from other cytosine derivatives.[9] It has also been found to bind specifically to a variety of proteins involved in regulating transcription and chromatin remodelling, interacting with more

proteins than either ^{m}C or ^{hm}C.[11,12] But how it can be recognised and removed with such remarkable selectivity by TDG and other proteins remains poorly understood.

^{f}C reportedly induces a unique helical form of DNA

In 2015, Raiber *et al* reported that ^{f}C, when present in consecutive, symmetrically modified CpG steps, gives rise to a unique conformation of the DNA double helix—distinct from A- or B-DNA.[13] The authors named this conformation F-DNA and proposed that it could enable specific recognition of genomic ^{f}C and regulate transcription through its effect on chromatin structure. This was the first report that ^{m}C derivatives can cause large structural changes to DNA. Previous crystallographic studies of non-consecutive, hemimodified CpG steps containing either ^{m}C, ^{hm}C, ^{f}C, or ^{ca}C had found that the modifications did not have a significant effect on DNA structure (Figure 2.1).[14–17] To support their claim, Raiber *et al* presented a high-resolution crystal structure of a DNA duplex containing 3 consecutive, diformylated CpG steps. They also provided evidence that F-DNA is adopted in solution—CD spectra that are consistent with a unique helical conformation (Figure 2.4).

Crystal structure of DNA containing ^{f}C has unusual features

Raiber *et al* determined the crystal structure of the duplex $d(CTA^{f}CG^{f}CG^{f}CGTAG)_2$ at 1.4-Å resolution (Figure 2.2*a*). Like B- and A- DNA, the helix is right-handed. But unlike the B and A forms, the putative F-DNA conformation is significantly underwound, featuring 13 base pairs per turn. B- and A-DNA, by contrast, have 10 and 11 base pairs per turn, respectively.

The authors also observed unique geometries of the major and minor grooves in the F-DNA structure, proposing that they could act as potential protein-recognition sites. These differences in global conformation observed in the F-DNA structure were attributed to altered base-stacking geometries of the ^{f}CpG steps, caused by bridging water molecules

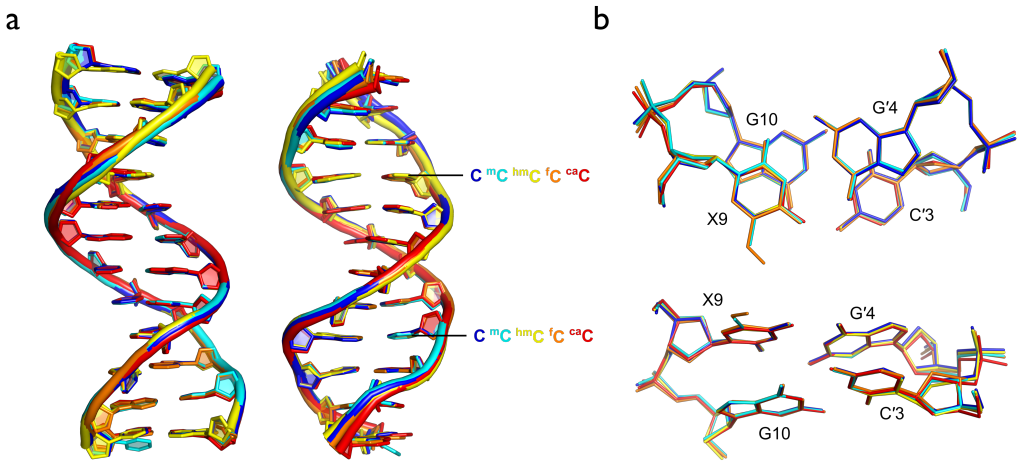


Figure 2.1. Overlaid crystal structures of the Dickerson–Drew dodecamer containing either C, ^{m}C, ^{hm}C, ^{f}C, or ^{ca}C (PDBs 4c64, 4c63, 4c5x, 4qc7, or 4pwm).[16,17] Cytosine modification has a negligible effect on global structure (*a*), and on the base-stacking geometries of CpG steps (*b*). The modified sites are denoted by 'X'. Reproduced from Hardwick *et al.*[14]

between the formyl groups and the O6 atoms of adjacent guanines (Figure 2.2*b*,*c*). These structural features would thus occur only in a CpG-specific context. Finally, the authors modelled junctions between B-DNA and the F-DNA crystal structure, finding that ^{f}C alters helical coiling and trajectory (Figure 2.3*e*). Such effects could lead to biological activity through altering chromatin structure.

CD spectra of formylated DNA are consistent with a unique helical structure

To investigate whether the F-DNA structure is adopted in solution, the authors used circular dichroism (CD) spectroscopy—a widely used technique to characterise DNA conformation, owing to the distinct CD spectra of A-, B-, and Z-DNA (Figure 2.4*a*). Formylation of CpG sites reportedly gave rise to dramatic spectral differences, consistent with a unique and previously unknown helical conformation (Figure 2.4*b*). These unusual spectral features of formylated DNA were also observed for the duplex in its crystallisation buffer, leading the authors to conclude that the crystal structure is representative of the

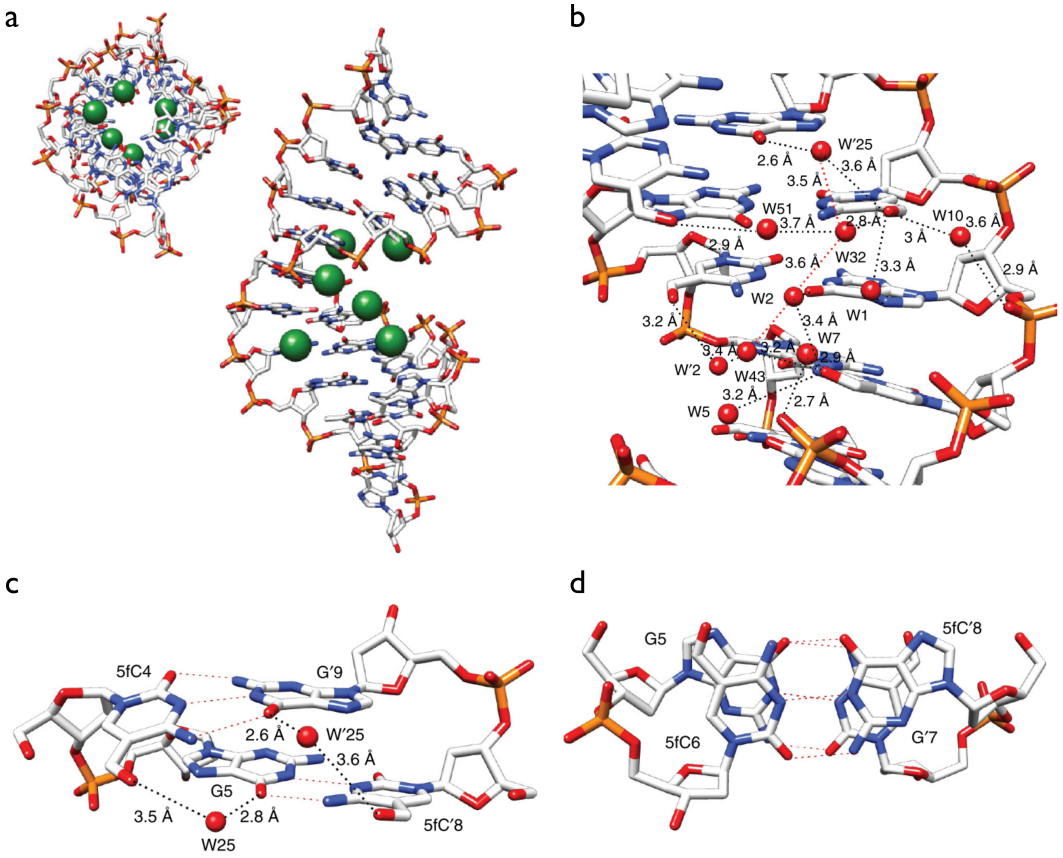


Figure 2.2. Crystal structure of a duplex containing 5-formylcytosine, which has unusual structural features. (*a*) Top and side view of the dodecamer, which contains three consecutive, diformylated CpG steps in the centre. The formyl groups are shown as green spheres. (*b*) Close-up of the formylated CpG repeats. A network of water molecules (red spheres) link the formyl group of each ^{f}C to the O6 of an adjacent guanine. (*c*) Side view of an ^{f}CpG step, showing the bridging water molecules that were proposed to induce unusual base-stacking geometries. (*d*) Top view of an ^{f}CpG step, showing the overlap between π-systems of ^{f}C and G. The altered base-stacking geometries of ^{f}CpG steps reportedly result in an unusual twist of the helix. Reproduced by permission from Springer Nature: *Nature Structural & Molecular Biology* (5-Formylcytosine alters the structure of the DNA double helix, Raiber *et al*), © 2014.[13]

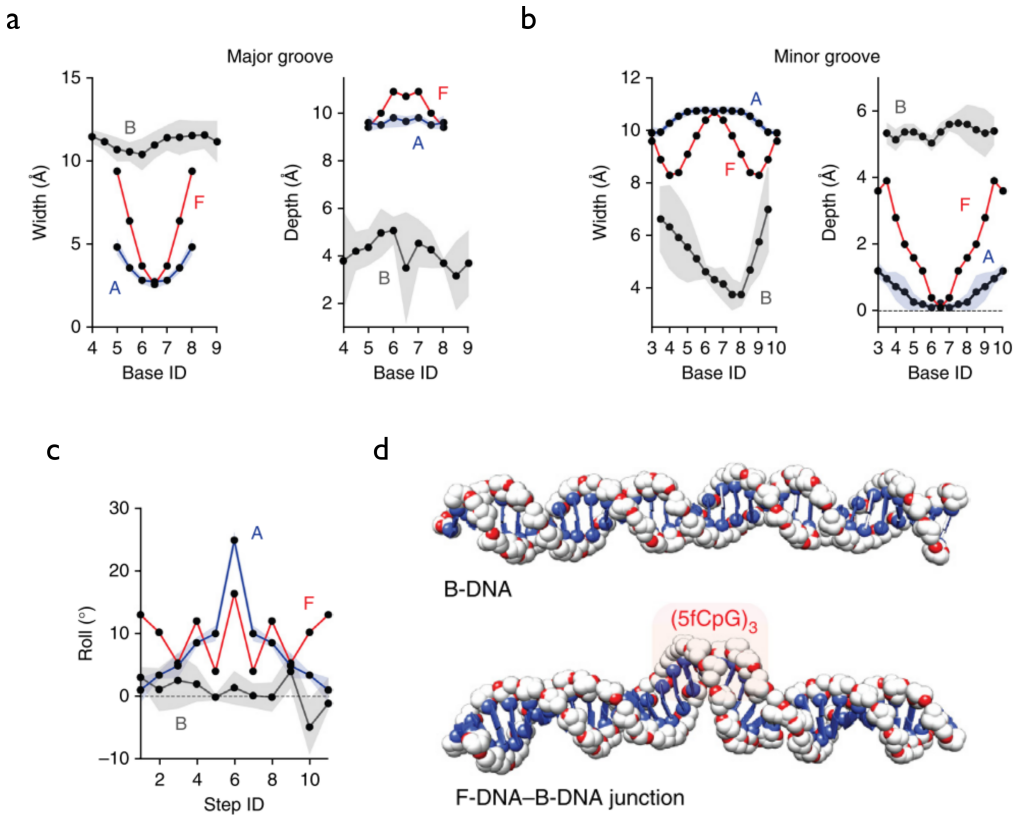


Figure 2.3. Comparison of F-DNA to A- and B-DNA. Major- and minor-groove dimensions of the ^{f}C structure (*a*,*b*) were found to be distinct from A- or B-DNA. (*c*) 'F-DNA' was found to exhibit periodic roll angles, in contrast to A- or B-DNA. (*d*) Modelling of a 36-bp B-DNA duplex alongside that of the junction formed between the formylated structure and B-DNA. The authors found that formylation alters the groove dimensions and helical trajectory of DNA, which could modulate DNA-protein interactions. Reproduced by permission from Springer Nature: *Nature Structural & Molecular Biology* (5-Formylcytosine alters the structure of the DNA double helix, Raiber *et al*), © 2014.[13]

conformation adopted in solution. Furthermore, following reduction of the formyl groups using sodium borohydride—or the addition of spermine (a DNA condensation agent)—the CD spectrum resembled that of normal B-DNA. This was interpreted as a structural transition from F- to B-DNA (Figure 2.4*c,d*).[13]

Could the unusual features of the crystal structure relate to sequence rather than formylation?

The key conclusions of the study were that ^{f}C alters base-stacking geometries specifically within CpG repeats, giving rise to a DNA conformation distinct from the A, B, or Z forms. This unique conformation, F-DNA, was proposed to enable specific recognition of ^{f}C by DNA-binding proteins and alter chromatin structure. However, a crystal structure of the formylated duplex's unmodified counterpart had not been determined. Furthermore, no crystal structures had, at that time, been reported of any unmodified, right-handed DNA duplexes containing three consecutive CpG steps. It was thus unclear if some of the unusual features observed in the F-DNA crystal structure could relate to the underlying sequence, rather than formylation. Additionally, the crystallisation buffer contained a high level of salt, including 1.8-M lithium sulphate. High cation concentrations can cause dramatic conformational changes to DNA, particularly for sequences containing several consecutive CpG steps.[19] Thus, the effect of the crystallisation buffer also could not be ruled out.

To understand better the consequences of formylation on DNA structure, we sought to untangle its effects from those of the underlying sequence and the crystallisation buffer. To do so, crystal structures were determined of the unmodified sequence, the same formylated duplex crystallised under low-salt conditions, and three unmodified structures in which the sequence of the central $(CpG)_3$ tract was altered. Formylated duplexes of different sequence and levels of formylation were also characterised in solution by NMR, CD, and UV spectroscopy.

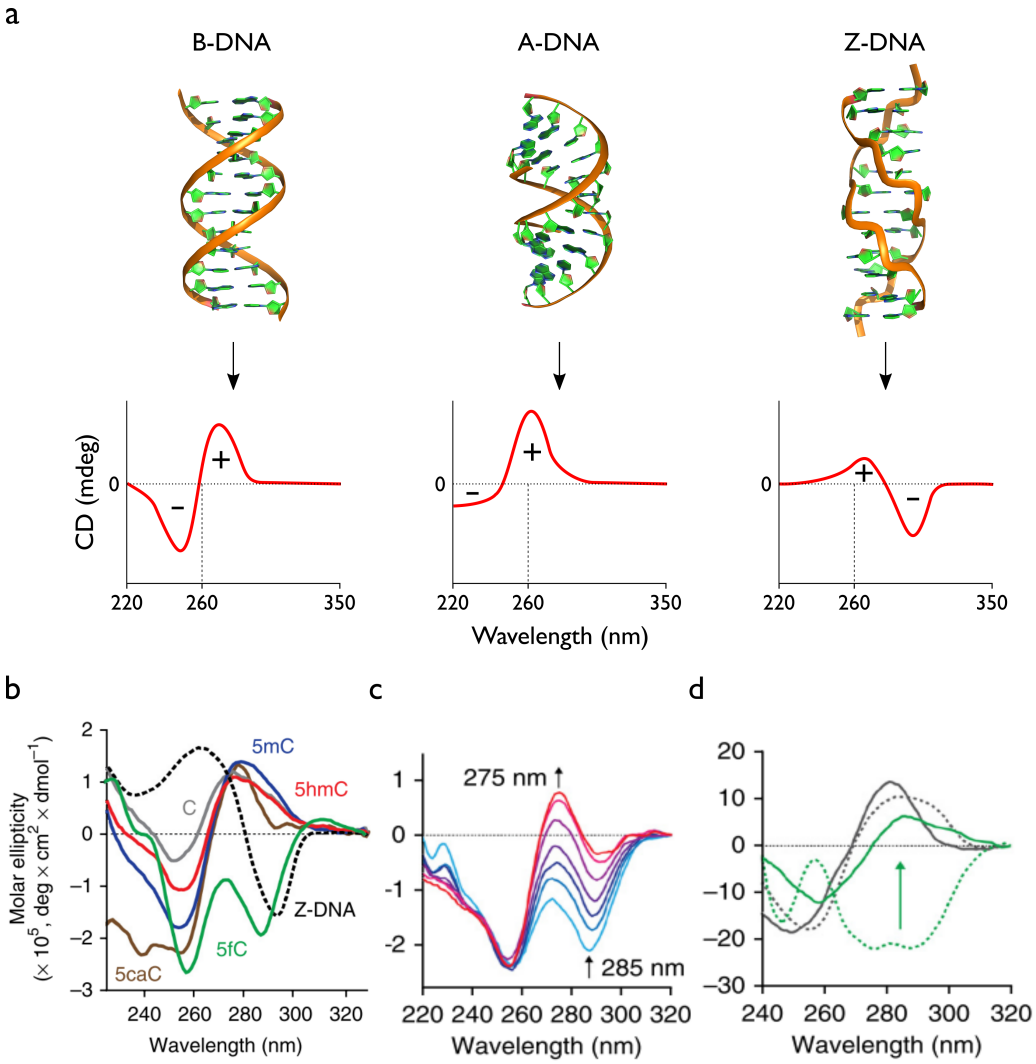


Figure 2.4. Schematic CD spectra of B-, A-, and Z-DNA (based on real spectra from Kypr *et al*),[18] and real spectra of the putative F-DNA conformation, reproduced by permission from Springer Nature: *Nature Structural & Molecular Biology* (5-Formylcytosine alters the structure of the DNA double helix, Raiber *et al*), © 2014.[13] (*a*) The three major conformational classes of the DNA double helix are readily distinguishable in solution using CD spectroscopy. (*b*) The CD spectrum of DNA containing formylated CpG repeats (green) shows marked differences to the CD spectra of known DNA duplex conformations. (*c*) Upon reduction of the formyl groups to form ^{hm}C, the CD spectrum shifts from that of 'F-DNA' (light blue) to B-DNA (red)—interpreted as a transition between the two helical forms. (*d*) The addition of 200-mM spermine, a known DNA condensation agent, was also reported to induce a structural transition from F-DNA (green dashed line) to B-DNA (green solid line).

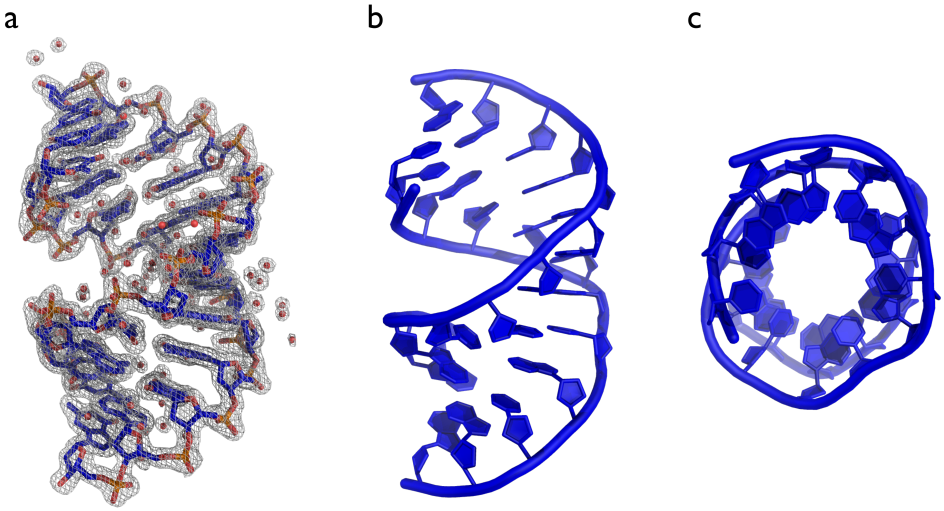


Figure 2.5. Crystal structure of d(CTACGCGCGTAG)$_2$, determined at 1.5-Å resolution (PDB 5mvk). (*a*) The structural model and $2F_o$–F_c electron density map, contoured at 1.0 σ. (*b*) Side view of the duplex. The base pairs are displaced from the helix axis and spiral around it—a hallmark of A-DNA. (*c*) Bird's-eye view, looking down the helix axis. In A-DNA, the base-pairs' displacement from the helix axis leads to a characteristic 'hole' in the duplex centre.

2.2 Results & Discussion

2.2.1 Crystal structure of d(CTACGCGCGTAG)$_2$

To separate the structural features associated with CpG formylation from those of the underlying sequence, a crystal structure of the unmodified sequence was determined (Figure 2.5, PDB 5mvk). The sequence crystallised in the $P3_221$ space group—unlike that of the formylated duplex ($P4_32_12$)—with a single duplex in the asymmetric unit. The structure was solved by molecular replacement (Section 5.5) and refined to 1.5-Å resolution. Data collection and refinement statistics are given in Table 2.1.

Table 2.1. Data collection and refinement statistics of PDBs 5mvu, 5mvk, 5mvp, 5mvt, and 5mvq. Each dataset was collected from a single crystal. Values in parentheses are for the highest-resolution shell.

	5mvu	**5mvk**	**5mvp**	**5mvt**	**5mvq**
Data collection					
Space group	$P3_221$	$P3_221$	$P3_221$	$P3_221$	$P3_221$
Cell dimensions:					
a, *b*, *c* (Å)	44.2, 44.2, 61.7	43.7, 43.7, 58.9	43.2, 43.2, 61.5	43.4, 43.3, 57.8	43.4, 43.4, 61.7
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	38.2–2.3 (2.42–2.3)	37.8–1.5 (1.56–1.53)	37.4–1.6 (1.63–1.61)	38.4–1.9 (1.94–1.9)	36.7–1.6 (1.63–1.6)
R_{meas}	7.6 (107)	5.2 (113)	3.4 (60.7)	6.0 (102)	3.7 (60.9)
$I/\sigma(I)$	15.4 (2.1)	19.7 (2.0)	23.0 (2.3)	21.0 (2.3)	18.3 (2.4)
$CC_{1/2}$	0.996 (0.566)	0.998 (0.764)	0.999 (0.837)	1.000 (0.730)	0.999 (0.744)
Completeness (%)	100 (100)	99.8 (99.8)	99.7 (94.7)	99.9 (99.1)	91.3 (89.3)
Redundancy	8.9 (9.2)	9.2 (6.9)	6.3 (5.8)	9.2 (8.5)	4.3 (3.8)
Refinement					
Resolution (Å)	32.5–2.3	37.8–1.5	37.5–1.6	38.4–1.9	36.7–1.6
Reflections	3,170	10,206	9,075	5,518	7,901
R_{work}/R_{free} (%)	14.6 / 15.8	18.1 / 18.8	15.3 / 20.9	20.2 / 22.5	17.9 / 19.4
Atoms:					
DNA	498	486	486	486	486
K^+	0	0	1	0	0
Co_{3+}	0	0	0	1	0
Mg^{2+}	0	0	0	0	2
Water	3	67	75	31	75
B factors:					
DNA	64.9	32.2	33.9	36.0	28.3
Water	57.4	40.6	44.3	39.8	37.0
RMS deviations:					
Bond lengths (Å)	0.008	0.014	0.014	0.011	0.011
Bond angles (°)	1.51	1.38	1.19	1.18	1.21

The formylated and unmodified structures are similar

The unmodified structure adopts typical A-form geometry (Table 2.3). As noted in Chapter 1, B-DNA predominates in solution, but the A-form is often observed in crystal structures.

Superposition of the unmodified duplex and its formylated analogue shows that the central regions of the two structures are conformationally similar, but the duplex ends are not (Figure 2.6*a*). This was unexpected: the composition of the duplexes differs in the centre, where the modification sites are located, but it is identical at the duplex ends.

To shed light on these findings, quantitative structural analysis was performed to compare the two structures at the single-nucleotide and base-pair level (using the packages 3DNA and Curves+).[20,21] Parameters such as sugar puckering, glycosyl torsion angle, slide, twist, and roll are used to distinguish between different helical forms of DNA (Chapter 1). Throughout the unmodified structure, these parameters are typical of A-DNA. Similarly, in the formylated region of the ^{f}C structure, these parameters are also characteristic of A-DNA (Table 2.3, Figure 2.6). As shown in the overlays, it is the duplex ends that are anomalous. For example, the twist at the terminal dinucleotide steps is 10.1° (Figure 2.6*d*). This value is anomalously low: a duplex comprising only these steps would have 36 base pairs per helical turn. It is only by including these distorted, unmodified steps that the unusual value reported for F-DNA (13 base-pairs per turn) can be obtained. If these terminal steps are excluded, the average helical twist becomes 33.9°, which corresponds to 10.6 base pairs per turn. Thus, contrary to the previous report, this region is slightly overwound compared to the unmodified structure (31.9°, 11.3 bp per turn).

Similarly, nucleotides in the formylated region all adopt the C3′-*endo* pucker and their glycosyl torsion angles range from 154–177°. These values are typical of A-DNA[22] and are similar to those of the unmodified structure (Tables 2.3 and A.7). Roll and slide values of this region are also characteristic of A-type geometry (Figure 2.6*e*,*f*). Raiber

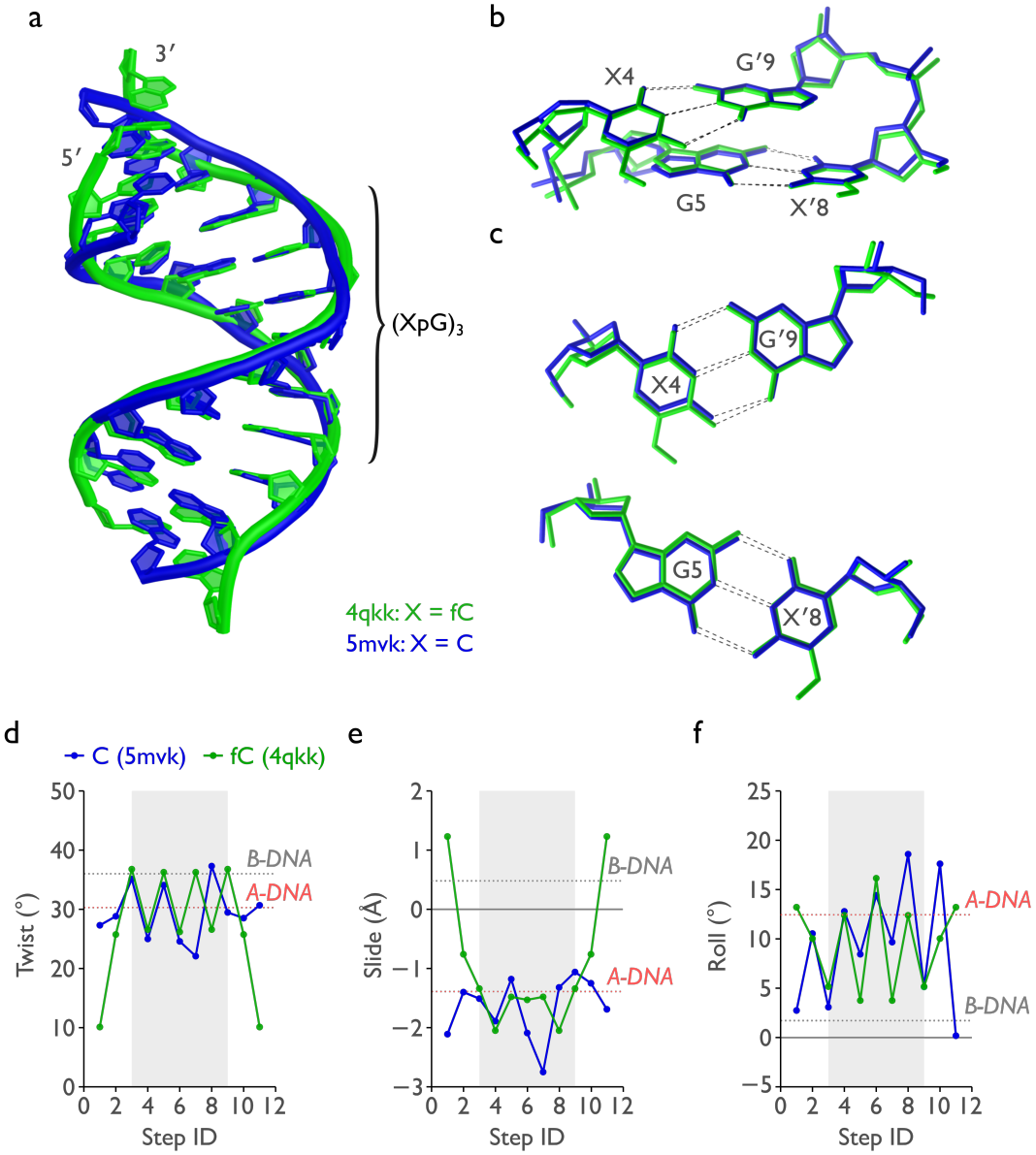


Figure 2.6. Overlaid crystal structures of d(CTA^{f}CG^{f}CG^{f}CGTAG)$_2$ (green) and its unmodified analogue (blue). (*a*) The formylated region of 4qkk is conformationally similar to that of its unmodified counterpart, but the duplex ends are not. (*b*) Base stacking of diformylated CpG step ^{f}C4G5/^{f}C′8G9 and overlays of its constituent base pairs (*c*). (*d*,*e*) Twist and slide values of the two structures. The terminal dinucleotide steps of 4qkk are anomalous, but the central, formylated region is characteristic of A-DNA. (*f*) Roll values of the two structures. Both exhibit a periodic pattern, indicating that it is a consequence of the underlying sequence rather than formylation. Shaded regions correspond to the octamer core containing the (XpG)$_3$ tract.

et al noted that the roll angles of the formylated structure exhibit a periodic pattern, attributing this to the high propeller twist of ^{f}C·G base pairs. However, this appears to be an effect of sequence rather than formylation. The same periodic pattern is observed in the unmodified duplex (Figure 2.6*f*)—which also has similar propeller twist angles (Table A.14). Periodic roll patterns have previously been observed in sequences with alternating pyrimidines and purines.[23]

Despite the overall similarity between the formylated and unmodified CpG tracts, there are some modest conformational differences between them (Figure 2.6*d*–*f*). These are most pronounced at step 7 and may be explained by differences in crystal packing. The sequence of both structures is self-complementary, with each duplex comprising two identical strands. Each duplex therefore could, in principle, crystallise so that its two strands are symmetrically equivalent. This is how the formylated sequence of 4qkk crystallised, producing a symmetric distribution of twist, slide, and roll values (Figure 2.6*d*–*f*). However, the unmodified sequence did not. The resulting asymmetric distribution of twist, slide, and roll values therefore relates directly to crystal-packing effects,[24] which may explain the anomalous values at step 7. Inspection of the crystal structure reveals that groups on the minor-groove side of this step do indeed form close contacts with the terminal residues of a neighbouring duplex (Figure A.1). Such end-groove packing is often seen in A-DNA crystal structures.[25]

The hydration patterns in the two structures differ

Raiber *et al* proposed that the unusual features of F-DNA arise from a distinctive hydration pattern, where bridging water molecules connect the formyl oxygens of ^{f}C to the O6 atoms of adjacent guanines, distorting the base-stacking geometries of ^{f}CpG steps. The hydration patterns of the two structures are indeed different (Figure 2.7). Despite this, the stacking geometries of CpG steps in both structures are similar, suggesting that the hydration pattern does not play a dominant role in determining the observed base-

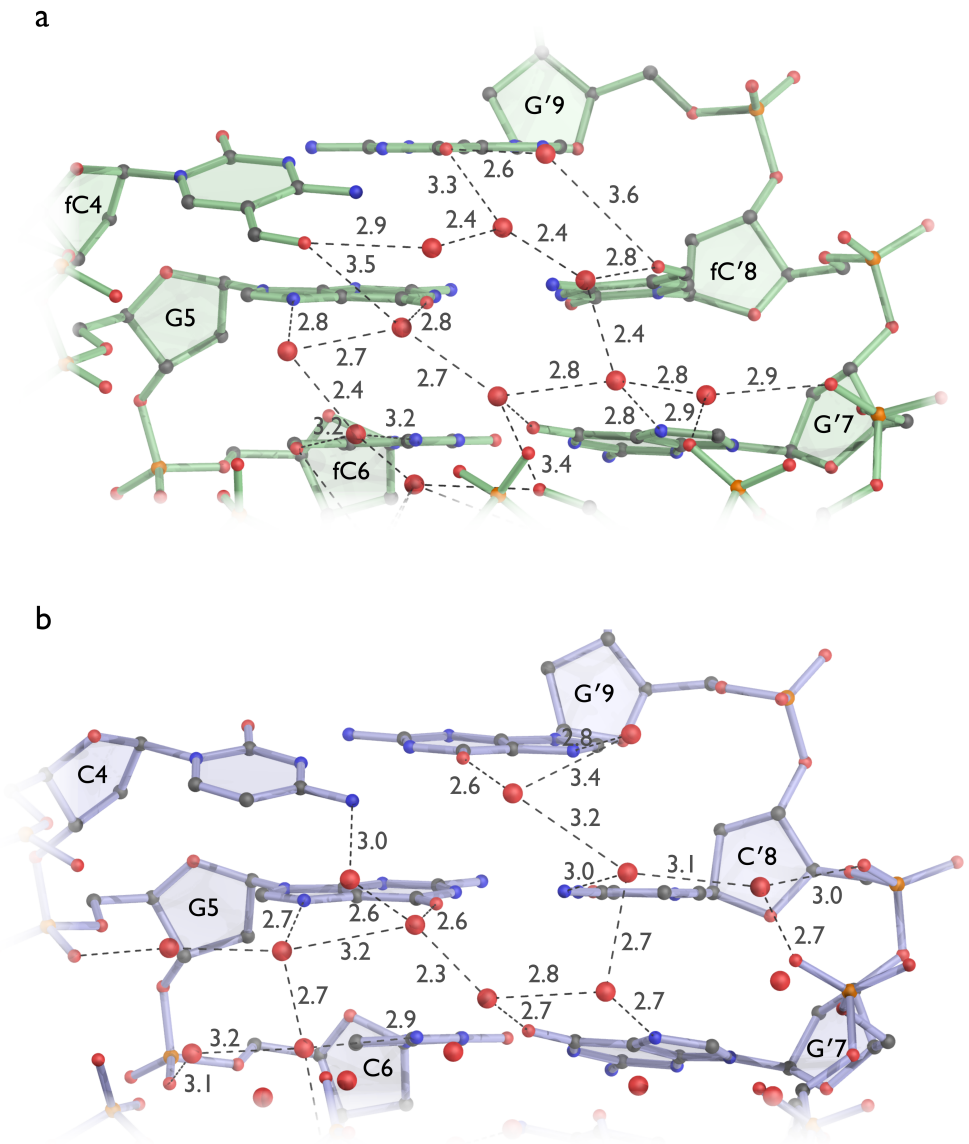


Figure 2.7. Major-groove hydration patterns of crystal structures 4qkk and 5mvk. (*a*) A section of the formylated CpG tract of the previously reported ^{f}C structure 4qkk. The hydration pattern reportedly distorts the base-stacking geometry of formylated CpG steps, leading to a unique global conformation termed F-DNA. (*b*) The same region of the unmodified analogue 5mvk. The hydration pattern is somewhat different to that of 4qkk. Despite this, base stacking geometries are similar, indicating that the specific hydration pattern observed in 4qkk is not the dominant factor governing conformation.

stacking geometries. Furthermore, hydration patterns observed in A-form structures are unlikely to be relevant in solution, where B-DNA predominates.

Overall, a comparison of the 'F-DNA' structure and its unmodified counterpart revealed that the formylated region is conformationally similar to the unmodified CpG tract, both of which adopt A-form geometry. There are, however, modest differences between these regions of the structures, which could be explained either by the presence of ^{f}C, or by differences in crystal packing.

2.2.2 Crystal structure of d(CTA^{f}CG^{f}CG^{f}CGTAG)$_2$

To differentiate between the effects of formylation and crystal packing, a crystal structure was determined of the same formylated sequence used in the previous study. Importantly, crystals were obtained under low-salt conditions (<150 mM). The structure was solved by molecular replacement, using PDB 5mvk as a search model, and refined to 2.3-Å resolution (Figure 2.8). Under these conditions, the ^{f}C duplex crystallised in the same space group as the unmodified duplex and formed similar crystal contacts. Structural differences resulting from formylation could therefore be separated from the effects of crystallisation to a greater extent.

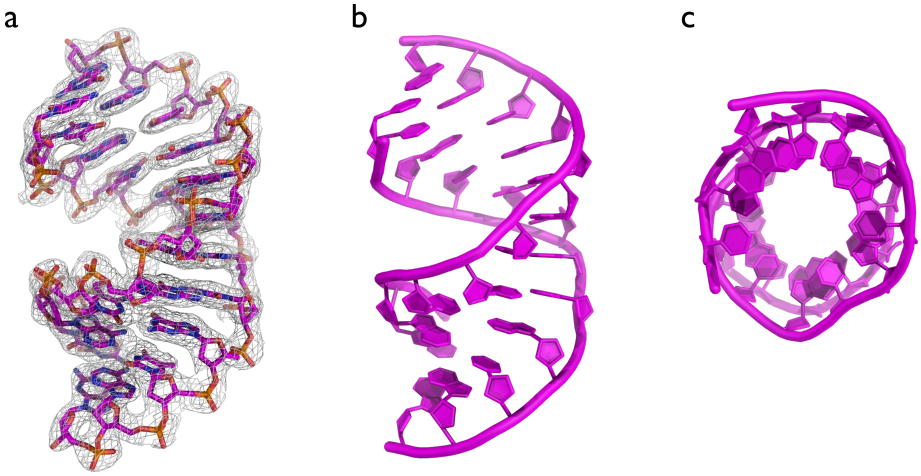


Figure 2.8. Crystal structure of d(CTA^{f}CG^{f}CG^{f}CGTAG)$_2$, determined at 2.3-Å resolution (PDB 5mvu). (*a*) The structural model and $2F_o$–F_c electron density map, contoured at 1.0 σ. (*b*) Side view and (*c*) top views of the structure. The base pairs are displaced from the helix axis, giving the appearance of a hole running through the duplex—characteristic of A-DNA.

The low-salt ^{f}C structure is similar to its unmodified counterpart

The low-salt ^{f}C structure adopts typical A-form geometry (Table 2.3). The structure closely resembles its unmodified counterpart 5mvk (Figure 2.9), with an all-atom RMSD of 0.65 Å (Table 2.2). Notably, it is much less similar to the structure of the same formylated sequence crystallised under high-salt conditions (PDB 4qkk, all-atom RMSD of 2.30 Å). Conformational differences of similar magnitude imparted by the crystalline environment have previously been observed for DNA duplex dodecamers—even within the same crystal lattice (RMSD of 1.93 Å).[26] The perturbations to the duplex ends observed in the previously published ^{f}C structure are absent, indicating that they arise not from formylation, but from differences in the crystallisation conditions. The duplex of 4qkk does indeed form different crystal contacts, which appear to cause the structural perturbation to its terminal residues (Figure 2.10). Presumably, these alternative interactions are stabilised by the high level of salt present in the crystallisation buffer (1.8 M lithium sulphate). Such distortion of the terminal base pairs is common in DNA crystal

Table 2.2. RMS deviations between PDBs 4qkk, 5mvu and 5mvk, and ideal A- and B-DNA of the same sequence.

	$d(CTAXGXGXGTAG)_2$ RMSD (Å, 486 atoms)		
	4qkk (X = ^{f}C)	5mvu (X = ^{f}C)	5mvk (X = C)
Ideal A-DNA	2.21	1.27	1.52
Ideal B-DNA	5.53	6.54	6.87
5mvk	2.65	0.65	–
5mvu	2.30	–	0.65
	$d(AXGXGXGT)_2$ RMSD (Å, 322 atoms)		
	4qkk (X = ^{f}C)	5mvu (X = ^{f}C)	5mvk (X = C)
Ideal A-DNA	1.08	1.24	1.44
Ideal B-DNA	4.19	4.39	4.56
5mvk	1.45	0.51	–
5mvu	1.21	–	0.51

Table 2.3. Comparison of averaged structural parameters of the formylated duplex (^{f}C) crystallised under high- and low-salt conditions, its unmodified counterpart (C), and those of ideal A- and B-DNA; the two outermost dinucleotide steps at each duplex end were omitted from the analysis to minimise the effect of crystal-packing artefacts.

Structure	Pucker	Helical twist (°)	Bp / turn	Roll (°)	Slide (Å)
4qkk (^{f}C-high salt)	C3′-*endo*	33.9	10.6	8.4	−1.6
5mvu (^{f}C-low salt)	C3′-*endo*	32.6	11.0	9.0	−1.8
5mvk (C)	C3′-*endo*	31.9	11.3	10.3	−1.7
Ideal A-DNA	C3′-*endo*	32.7	11.0	12.4	−1.4
Ideal B-DNA	C2′-*endo*	36.0	10.0	1.7	0.5

structures.[27–37]

Local features of the low-salt ^{f}C structure and the control are also similar (Figure 2.9). Anomalous features are again observed at step 7 (Figure 2.9*c,d*), consistent with their arising from crystal-packing effects rather than formylation. But potential ^{f}C-specific differences can also be seen. Most notably, the roll angles of the formylated structures are similar at steps 5 and 7, both of which are GpC steps. These values are considerably lower than those of the unmodified structure and typical A-DNA. Formylation might therefore alter the roll angles of GpC steps, though this is unclear owing to the small sample size (n = 2).

2.2.3 The effect of ^{f}C on DNA groove geometry, coiling and helical trajectory

The groove dimensions of F-DNA were reportedly distinct from those of A- and B-DNA, with CpG formylation narrowing the major groove and opening the minor groove (Figure 2.3*a,b*).[13] However, the groove-widths of the unmodified duplex are similar to those of the previously published, high-salt ^{f}C structure, and are very similar to those of the low-salt ^{f}C structure (Figure 2.11*a,b*). Compared to the unmodified duplex, the minor groove of the low-salt ^{f}C structure is slightly narrower and its major groove slightly wider. This could be due either to the presence of the formyl groups, or to differences in the crystallisation conditions. Overall, the crystallisation conditions exerted a larger influence on groove geometry than did CpG formylation, which has, at most, a small effect on the crystal structures.

^{f}C was found to alter helical coiling and trajectory, based on the modelling of junctions formed between F- and B-DNA (Figure 2.3).[13] To investigate this further, similar models were built from the two ^{f}C structures and compared with ideal B-DNA, and junctions formed between ideal A- and B-DNA (Figure 2.11*c,d*). To reduce the effects of crystal

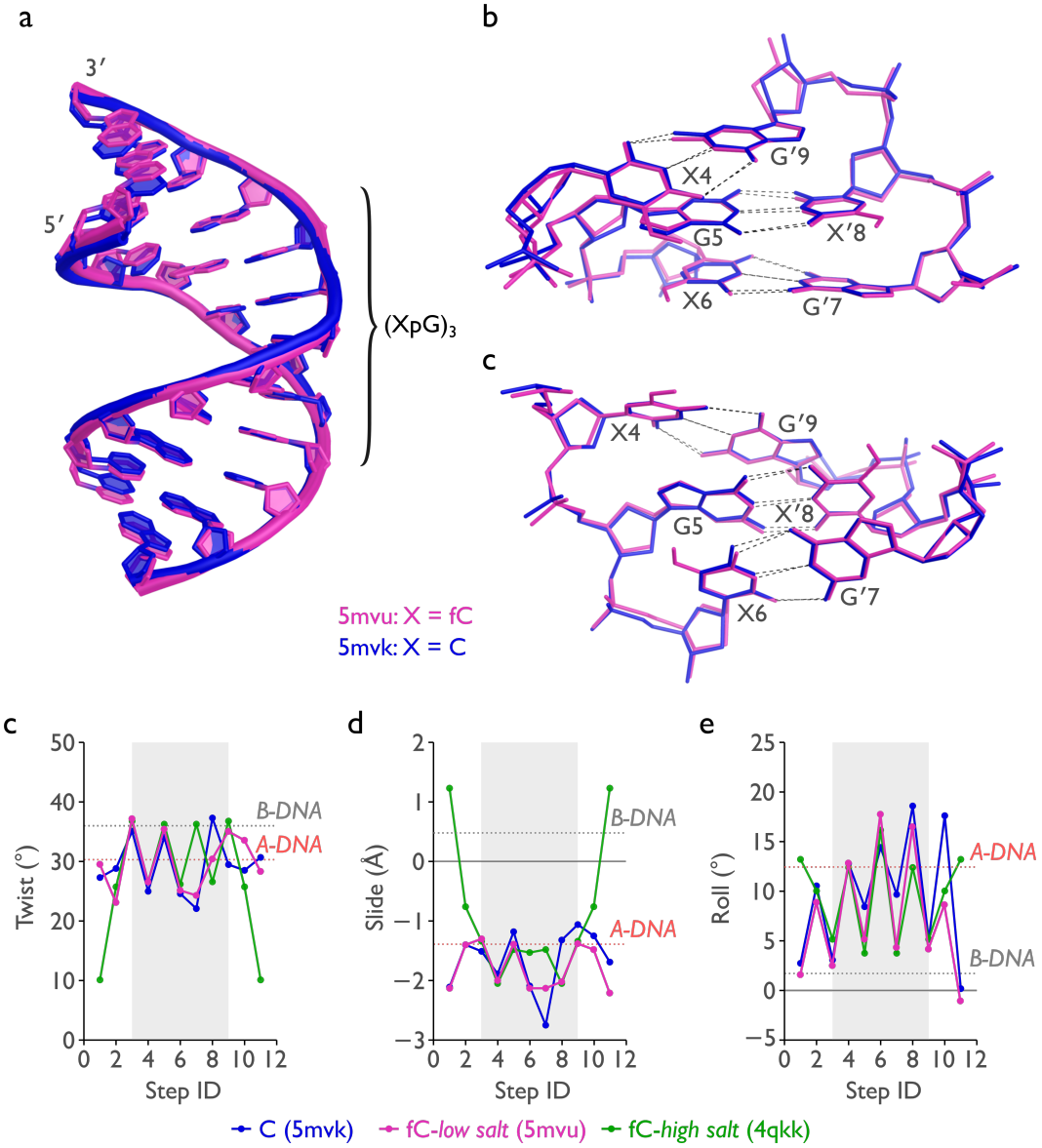


Figure 2.9. Comparison of the low-salt ^{f}C structure (magenta, PDB 5mvu) with its unmodified counterpart (blue, PDB 5mvk). (*a*) Overlaying the structures shows that they are similar in global conformation (RMSD of 0.65 Å, 486 atoms). (*b*) A comparison of base stacking of base pairs X4·G′9, G5·X′8, and X6·G′7, where X = ^{f}C (5mvu) or C (5mvk). CpG formylation does not exert a large influence on the observed base-stacking geometries. (*c*) Twist values of the two structures, along with those of the high-salt ^{f}C structure (green). Anomalous twist values of the terminal steps in the formylated duplex are not observed when crystallised under low-salt conditions. (*d*) Slide values of the low-salt ^{f}C structure are typical of A-DNA and are not anomalous at the duplex termini. (*e*) Roll angles of the low-salt ^{f}C structure are characteristic of A-DNA and, like the other two structures, exhibit a periodic pattern. In both formylated structures, the values of steps 5 and 7 (both Gp^{f}C steps) are similar and differ from those of the unmodified structure. Shaded regions correspond to the octamer core containing the $(XpG)_3$ tract.

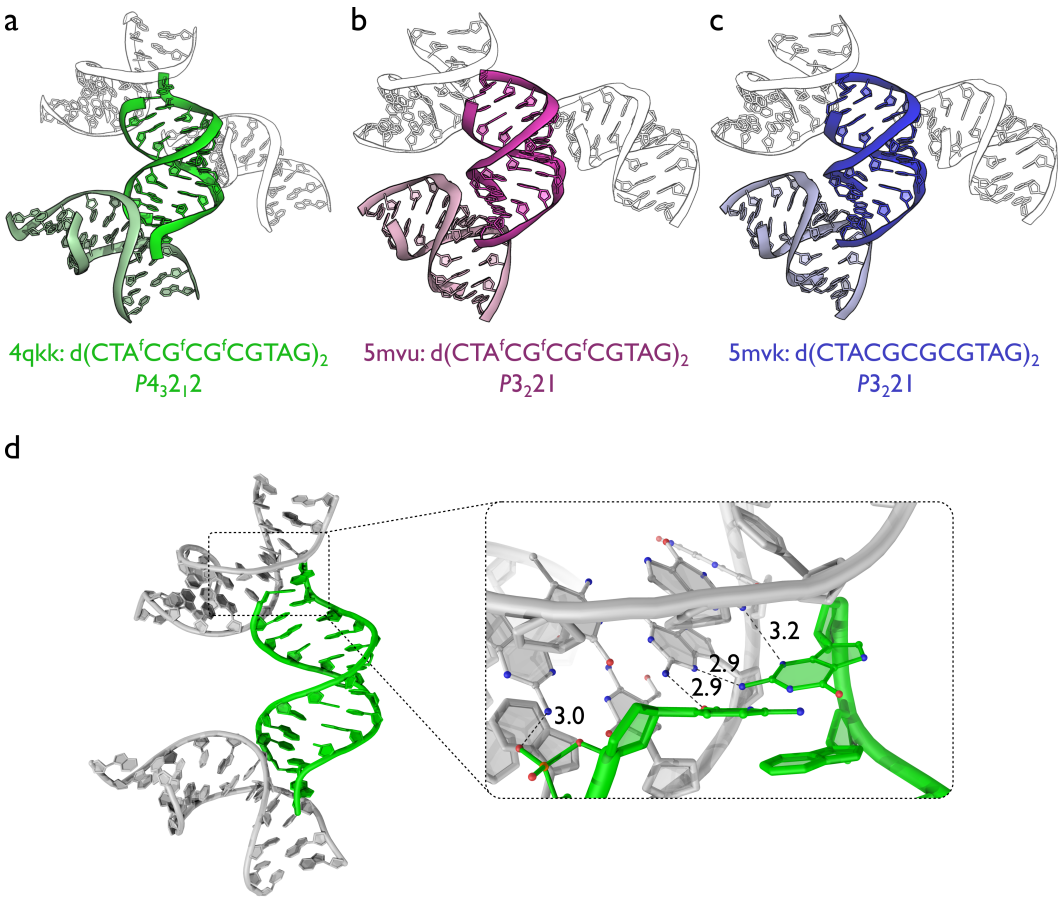


Figure 2.10. A comparison of the crystal contacts of the high- and low-salt structures of the formylated sequence (green and magenta, respectively) and the unmodified analogue (blue). (*a*) The formylated duplex forms different crystal contacts when crystallised under high-salt conditions, which appear to induce distortions at the terminal dinucleotide steps. Under low-salt conditions (*b*), the formylated sequence forms similar contacts to the unmodified structure (*c*). (*d*) Close-up of the interactions between symmetry mates in the 'F-DNA' structure (4qkk). *a–c* reproduced from Hardwick *et al*,[38] *d* from Hardwick *et al*.[14]

packing, the terminal dinucleotide steps were removed from each duplex (none of which contain ^{f}C). Models containing them are shown in Figure A.4. Indeed, such differences in helical trajectory were observed in the models derived from both formylated crystal structures compared to ideal B-DNA (Figure 2.11*c*). However, similar features were also observed in the model of the A-DNA–B-DNA junctions. This demonstrates that the observed effects are characteristic of junctions between A- and B-DNA and are not ^{f}C-dependent. Analogous models were then built in which the formylated crystal structures were flanked by ideal A-DNA. The models derived from both formylated structures resemble ideal A-DNA (Figure 2.11*d*), consistent with the structures' adopting typical A-form geometry.

These findings mirror the all-atom RMSD of the F-DNA structure and a modelled A-DNA duplex of the same sequence (Table 2.2). In both cases it was found that after removing the terminal dinucleotide steps from the crystal structures, the F-DNA structure closely resembles ideal A-DNA, even more so than does the unmodified crystal structure (RMSDs of 1.08 and 1.44 Å, respectively). Overlays of the crystal structures with ideal A-DNA models are shown in Figure A.2.

2.2.4 Comparing the structural consequences of formylation to those of DNA sequence

F-DNA, the helical conformation of DNA proposed by Raiber *et al*, was based on its having structural features distinct from A-, B-, or Z-DNA. In their analysis, the authors compared the F-DNA structure with six X-ray crystal structures—three in the B-form and three in the A-form. The three A-form structures that were chosen are all of the same space group ($P6_122$), which is different to that of the F-DNA structure ($P4_32_12$). Consequently, these three representative A-form structures are much more similar to each other than structures from different space groups typically would be. Thus, the

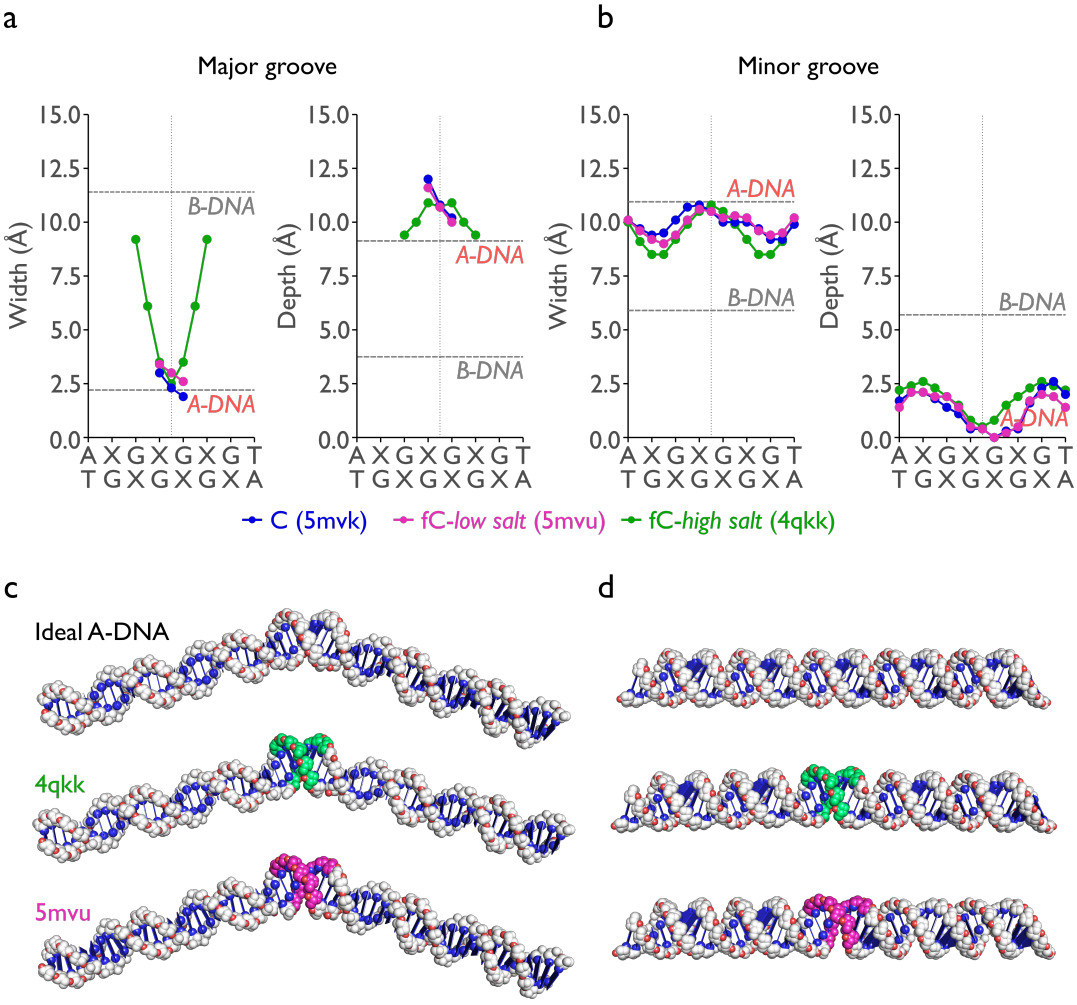


Figure 2.11. Groove dimensions of the high- and low-salt ^{f}C structures (PDBs 4qkk and 5mvk) and their unmodified counterpart (PDB 5mvk), and modelled junctions with ideal B- and A-DNA. Widths and depths of the major (*a*) and minor (*b*) grooves of the three structures. Groove dimensions of all structures are similar, especially those of the unmodified (5mvk) and low-salt ^{f}C structure (5mvu), which are of the same space group. (*c*) 60-bp models in which a central 8-bp region of ideal A-DNA or the formylated section of the ^{f}C structures is flanked by either B-DNA (*c*) or A-DNA (*d*); *c* and *d* reproduced from Hardwick *et al.*[38]

structural parameters of F- and A-DNA appear to be distinct (Figure 2.3). Conversely, the three B-DNA structures are all of different space groups, and therefore exhibit a greater range of values.

A more meaningful comparison would be between formylated and unmodified, A-DNA structures of the same space group. This would shed light on how the conformational effects of formylation compare in magnitude to sequence-dependent variation in DNA structure. It might also highlight any distinct conformational features arising from the underlying sequence. This is of interest because despite the prevalence of CpG repeats in gene promoters,[39] the crystal structure 5mvk obtained in this study remains the only reported example of an unmodified, right-handed DNA duplex containing three consecutive CpG steps.

To address this, the crystal structures of three analogues of $d(CTACGCGCGTAG)_2$ were determined, in which the central $(CpG)_3$ motif was altered (see Table 2.4 for sequences). All duplexes crystallised in the $P3_221$ space group and adopted the A-form (Figure 2.12).

As expected, local structural parameters vary according to sequence. For example, the roll angles of 5mvt exhibit a periodic pattern like that of 5mvk and its formylated analogues, which reflects its alternating pyrimidine-purine sequence. But overall, the duplexes were similar in structure, as demonstrated by overlays (Figure A.5) and RMSDs (Table 2.4, see Table A.6 for full comparison). Notably, the largest RMSD value observed for any of the structures obtained in this study was between the unmodified structures 5mvq and 5mvt (1.26 Å, 448 atoms)—twice that between the low-salt ^{f}C structure and 5mvk (0.65 Å, 486 atoms). Thus, global conformational differences between the formylated and unmodified structures are smaller than those between two unmodified A-DNA structures.

Local structural parameters and groove dimensions of the low-salt formylated structure also fall within, or close to, the range of values exhibited by the four unmodified structures (Figure A.3). Any perturbation arising from CpG formylation is therefore comparable in

magnitude to modest, sequence-dependent differences in conformation.

Table 2.4. Sequences, resolution, and PDB codes of the unmodified crystal structures determined in this study, along with the RMSD of each structure from $d(CTACGCGCGTAG)_2$.

Sequence	Resolution (Å)	PDB ID	RMSD (Å)
1. $d(CTACGCGCGTAG)_2$	1.5	5mvk	–
2. $d(CTACGGCCGTAG)_2$	1.6	5mvq	0.66
3. $d(CTAGGGCCCTAG)_2$	1.9	5mvp	0.64
4. $d(CTACGTACGTAG)_2$	1.6	5mvt	1.13

Overall, no crystallographic evidence was found for a unique, ^{f}C-dependent helical conformation of DNA. The unusual features in the previously reported structure (PDB 4qkk) are not associated with the formylated region and result from crystal-packing artefacts. Some local features in the formylated structures could be ^{f}C-dependent, but most appear to be dominated by the underlying sequence.

2.2.5 CD spectroscopy of formylated DNA

Although no crystallographic evidence was found for a unique, ^{f}C-dependent helical conformation of DNA, this did not explain the unusual CD spectra of formylated DNA reported by Raiber *et al*. The crystal structures obtained in this study were all of the A-form, which suggested they may not faithfully reflect the duplexes' conformation in solution. Perhaps, then, the CD spectra point to a unique helical form of DNA that was not observed by crystallography. To explore this, formylated and unmodified duplexes were characterised by CD and UV-absorption spectroscopy.

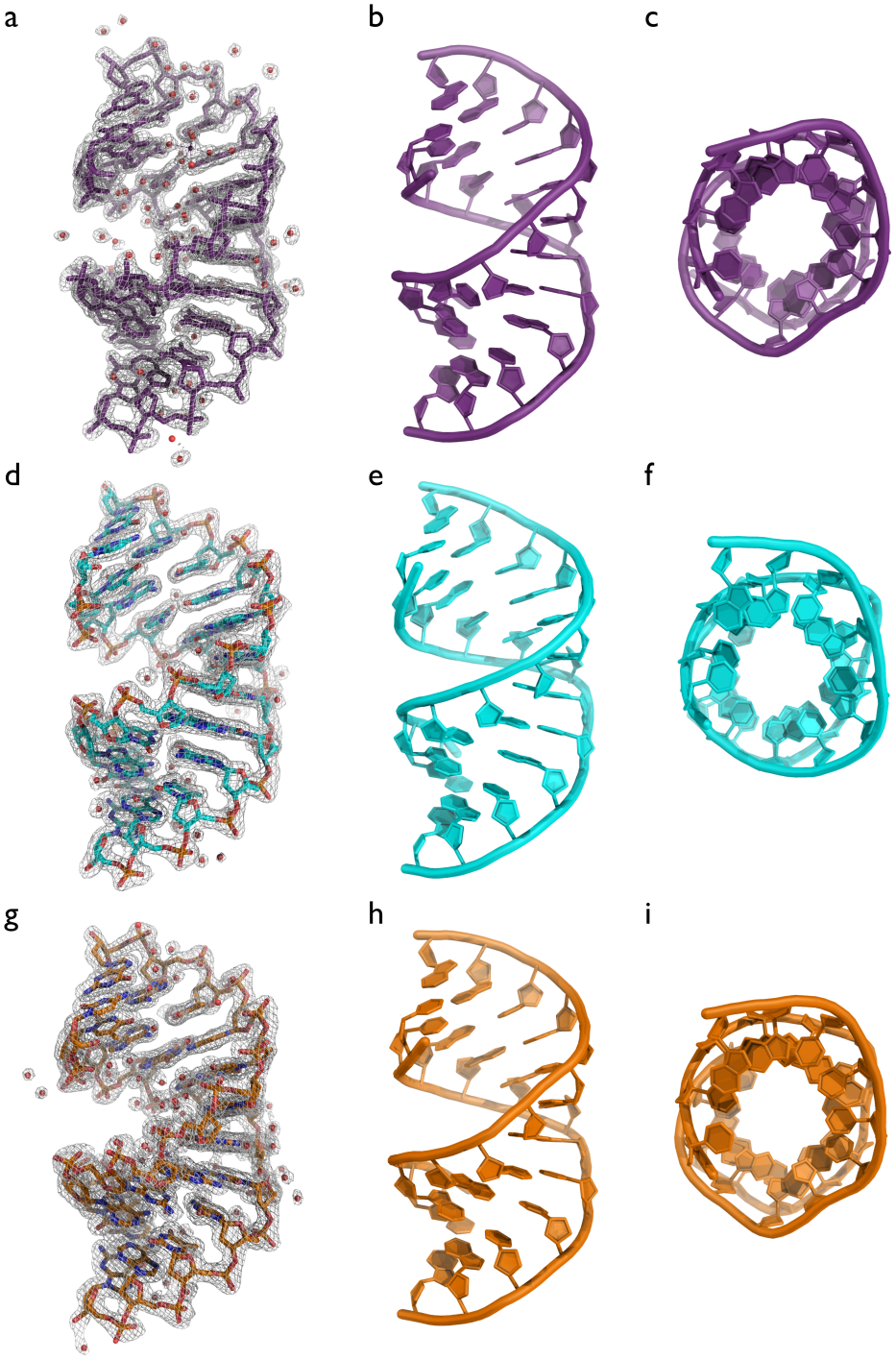


Figure 2.12. X-ray crystal structures of sequence variants of d(CTACGCGCGTAG)$_2$ in which the central (CpG)$_3$ was changed to CGGCCG (purple, PDB 5mvq), CGTACG (cyan, PDB 5mvt), or GGGCCC (orange, PDB 5mvp). (*a,b,c*), (*d,e,f*), and (*g,h,i*) show the $2F_o$–F_c electron density map (contoured at 1.0 σ), cartoon side-view and cartoon top-view of each structure.

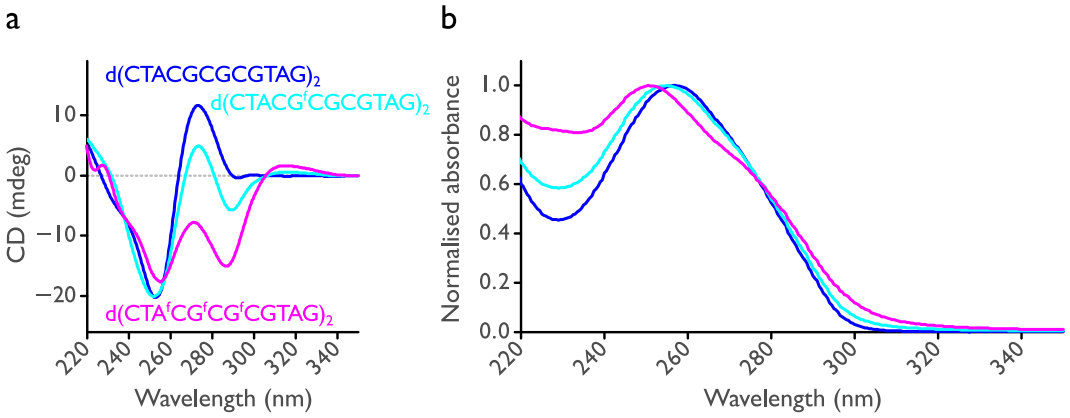


Figure 2.13. CD and UV absorption spectra of d(CTACGCGCGTAG)$_2$ and formylated variants. Formylation of CpG sites induces large changes in both CD (*a*) and UV absorption (*b*) spectra.

CD and UV absorption spectra of d(CTACGCGCGTAG)$_2$ and formylated variants

First, CD spectra were acquired of the duplex d(CTA^{f}CG^{f}CG^{f}CGTAG)$_2$, an analogue containing a single diformylated CpG step—d(CTACG^{f}CGCGTAG)$_2$—and the unmodified sequence. As expected, the CD spectrum of the unmodified duplex is consistent with B-DNA (Figure 2.13), even in the crystallisation buffer that yielded the A-DNA crystal structure (Figure A.6*d*). Thus, in this case at least, the crystal structure does not accurately reflect the duplex's conformation in solution.

The CD spectrum of d(CTA^{f}CG^{f}CG^{f}CGTAG)$_2$ (Figure 2.13) is similar to that obtained by Raiber *et al* (Figure 2.4) and is quite unlike spectra of A-, B-, or Z-DNA. However, contrary to the findings of Raiber *et al*, the addition of ≥200 mM spermine did not induce a spectral transition to that resembling B-DNA (Figure A.6*b*).

Curiously, the spectrum of the duplex containing a single, diformylated CpG step is also unusual, being somewhat intermediate between those of the unmodified and fully formylated duplexes. The magnitude of these spectral differences is certainly consistent with significant deviation from B-form structure. But an alternative theory was also considered: might ^{f}C's effects on CD spectra arise from altered photophysical properties,

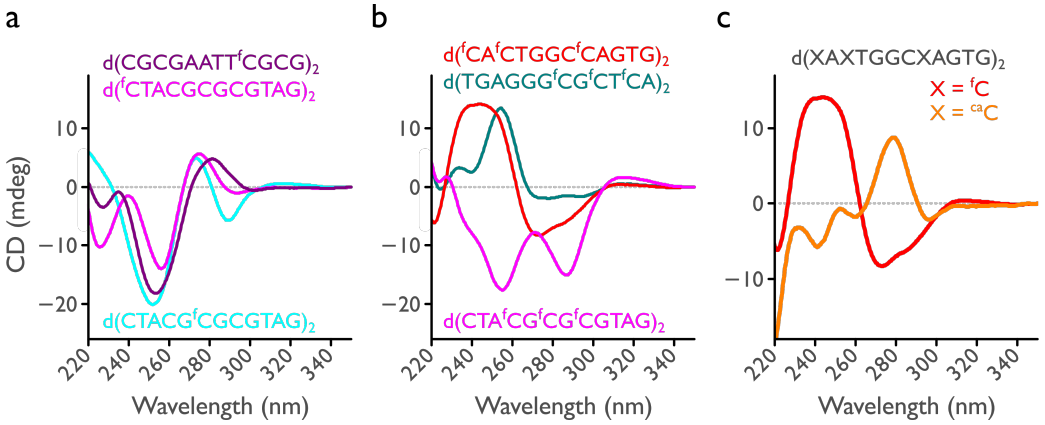


Figure 2.14. CD spectra of DNA containing ^{f}C and ^{ca}C in different sequence contexts. (*a*) Duplexes containing two ^{f}C residues. (*b*) Duplexes containing six ^{f}C residues. Perturbations to spectra occur even when ^{f}C is not in a CpG context. (*c*) CD spectra of the same sequence containing either six residues of ^{f}C or ^{ca}C. Both modifications lead to strongly perturbed spectra. Note the change in scale of the Y-axis to enlarge the spectra.

rather than conformation? This seemed plausible, since the π-electrons of the formyl group would be conjugated with those of the aromatic ring. To investigate this, UV absorption spectra of the duplexes were acquired (Figure 2.13). Increased perturbation of both UV and CD spectra were observed with increased formylation, suggesting that the effects could indeed be electronic.

Next, the effect of formylation on the CD spectra in different sequence contexts was investigated. This was important, as the proposed basis of F-DNA involves a specific hydration network that connects the formyl group of ^{f}C with a 3′ guanine. The F-DNA conformation would therefore form only when ^{f}C is present in CpG steps; in other sequence contexts, it would not be expected to perturb CD spectra. Conversely, if the unusual spectral features arise from an electronic effect, ^{f}C should perturb CD spectra regardless of sequence context.

For each sequence tested, the number of A, C, G, and T nucleotides was conserved—only their order and level of formylation was varied. In each case, significant perturbation of both the CD and UV spectra was observed, and the effect increased with increasing

levels of formylation (Figures 2.14 and A.6*a*). Notably, this was true even for duplexes containing no CpG sites. Moreover, not only are all spectra of the formylated duplexes perturbed, they are all different to each other—and those of any known conformational class of DNA. Of the two possibilities considered to explain these surprising results—(i) that each spectrum represents a unique DNA conformation, or (ii) that the effects are primarily electronic, not structural—the latter certainly seemed more likely, especially considering the perturbations to UV spectra.

Should an electronic effect be responsible, arising from the presence of a carbonyl group conjugated to the cytosine ring, it was reasoned that 5-carboxylcytosine might also perturb CD spectra. A carboxylated analogue of one of the formylated duplexes was therefore prepared, d(^{ca}CA^{ca}CTGGC^{ca}CAGTG)$_2$. Its CD spectrum is dramatically different both to its formylated analogue, and to that of any known helical form of DNA (Figure 2.14*c*). And, as with ^{f}C, its UV absorption spectrum was also perturbed (Figure A.6*a*).

Together, the data strongly suggest that differences observed in the CD spectra of formylated and carboxylated DNA arise from the modifications' influence on cytosine's electronic structure, rather than on helical conformation. But this interpretation cannot be confirmed by CD and UV spectroscopy alone, and a combination of electronic and structural effects could not be ruled out. To address this ambiguity, we turned to NMR.

2.2.6 NMR analysis of formylated DNA

Like CD, NMR spectroscopy can be used to probe DNA conformation in solution. But unlike CD, NMR can yield detailed information on molecular structure. For example, the glycosyl torsion angles and sugar puckers can be readily determined and compared with those derived by X-ray crystallography (see Section 1.3).[40] Thus, NMR can distinguish between different conformational classes of DNA in solution, and determine small, local differences within a conformational class.[40] Furthermore, unlike X-ray crystallography,

NMR experiments can be carried out under similar conditions to CD.

NMR data were collected and analysed for the following duplexes: d(CTACGCGCGTAG)$_2$, d(CTACG^{f}CGCGTAG)$_2$, and d(CTA^{f}CG^{f}CG^{f}CGTAG)$_2$. All protons in each duplex were uniquely assigned, except H5′ and H5″, using a combination of DQF-COSY and NOESY experiments (see Tables A.22 and A.23). CD spectra of the three duplexes were also acquired in the same buffer used for NMR experiments and were not appreciably different to those acquired in PBS (Figure A.6*e*).

The protons of ^{f}C exhibited the largest change in chemical shift, with H6 resonances moving ~0.9 ppm downfield (Figure 2.15*a*). Conversely, resonances of the sugar protons of ^{f}C residues and their nearest-neighbour bases were only slightly perturbed (Figure 2.15*a,b*). These observations are consistent with changes to the electronic structure of ^{f}C, but suggest only minor changes to conformation, local to the modification sites. All residues formed Watson-Crick base pairs, and all non-terminal base pairs were stable (indicated by inter-imino and TN3H-AC2H NOEs and hydrogen-deuterium exchange). Coupling constants obtained from 1D, DQF-COSY, and NOESY spectra for all non-terminal residues are consistent with the C2′-*endo* pucker, which is characteristic of B-DNA. Glycosyl torsion angles ranged from −90 to −110° for all three duplexes (determined from the NOE peak volumes of H8/H6(i)-H2′(i)), typical of B-DNA. Those of A-DNA are around −150°.[22] Variation in RMSDs of intra- and internucleotide NOEs at the nucleotide level are consistent with modest differences in sugar puckering and glycosyl torsion angles arising from formylation (Figure 2.15*c,d*).

Overall, the NMR analysis shows conclusively that both formylated duplexes adopt the B-form in solution. Conformational differences between the modified and unmodified duplexes are small and local to the vicinity of the modification sites.

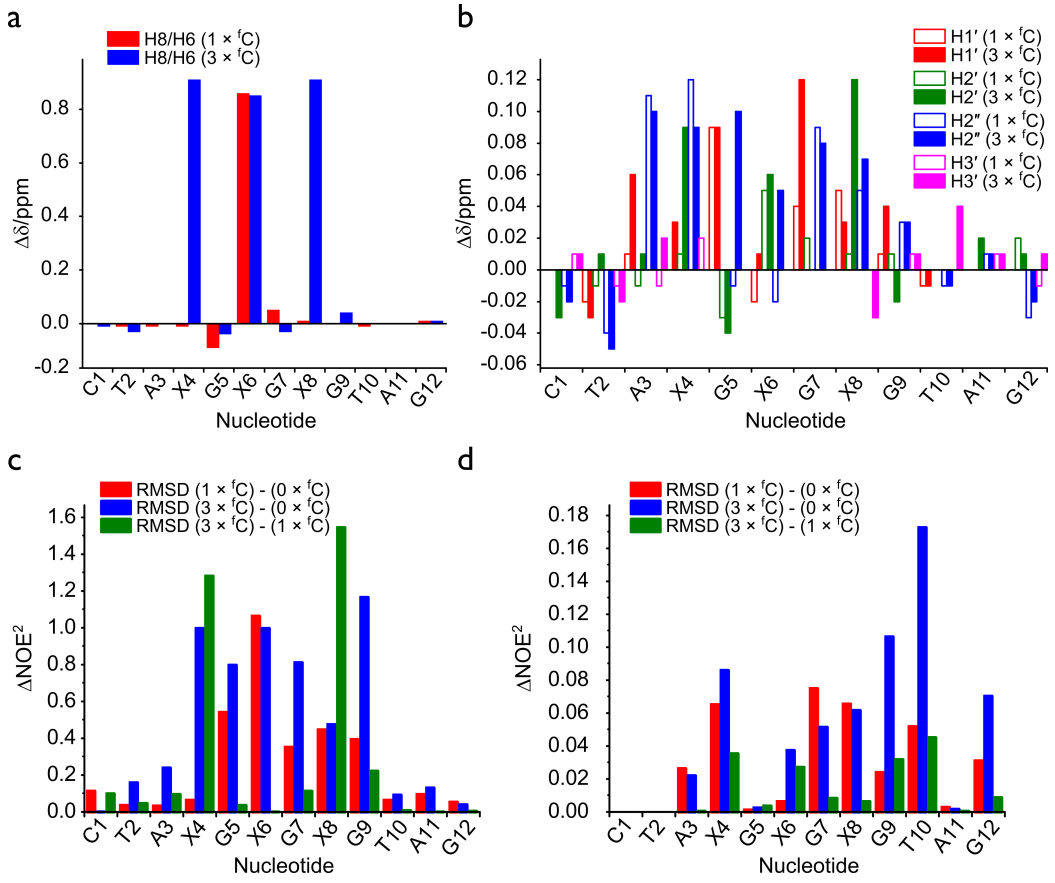


Figure 2.15. Chemical shift and NOE differences between duplexes of varying formylation levels. Differences in chemical shift between each formylated sequence and its unmodified counterpart are shown for (*a*) base protons H8 or H6, and (*b*) sugar protons H1′, H2′, H2″ and H3′. RMSDs at the nucleotide level are shown for (*c*) intranucleotide NOEs H8/H6 to H1′, H2′, H2″ and H3′, and (*d*) internucleotide NOEs H8/H6 to preceding H2′ and H2″. Reproduced from Hardwick *et al.*[38]

2.3 Conclusions

The existence of F-DNA—an ^{f}C-dependent, unique helical conformation—was proposed based on unusual features of CD spectra and an X-ray crystal structure of DNA containing 5-formylcytosine. However, reanalysis of this crystal structure, along with newly obtained structures of the same formylated duplex crystallised under different conditions and its

unmodified counterpart, revealed that there is no crystallographic evidence for F-DNA. In the original structure, the entire formylated region adopts the A-form. And while there are distortions at the duplex-ends, they appear to arise not from the presence of ^{f}C, but from crystal-packing artefacts resulting from the high concentration of salt in the crystallisation buffer. When an alternative, low-salt buffer was used, the same duplex adopted the A-form throughout and closely resembled the unmodified structure.

CD spectra of the duplex from the 'F-DNA' crystal structure, and other formylated sequences, did contain unusual features that could be consistent with a distinct helical form of DNA. But UV-absorption spectra suggest that these differences reflect electronic, rather than conformational, changes. This was confirmed by NMR analysis of two formylated duplexes, which showed that they adopt the B-form in solution.

Overall, it was concluded that there is no evidence for a unique helical conformation of DNA contingent on CpG formylation.

BIBLIOGRAPHY

1. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324,** 929–930 (2009).

2. Tahiliani, M. *et al.* Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324,** 930–935 (2009).

3. Pfaffeneder, T. *et al.* The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA. *Angew. Chemie Int. Ed.* **50,** 7008–7012 (2011).

4. He, Y.-F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **333,** 1303–1307 (2011).

5. Carell, T., Kurz, M. Q., Müller, M., Rossa, M. & Spada, F. Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA. *Angew. Chemie Int. Ed.* **57,** 4296–4312 (2018).

6. Maiti, A. & Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine. *J. Biol. Chem.* **286,** 35334–35338 (2011).

7. Raiber, E.-A. *et al.* Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.* **13,** R69 (2012).

8. Weber, A. R. *et al.* Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat. Commun.* **7,** 10806 (2016).

9. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11,** 3–6 (2015).

10. Su, M. *et al.* 5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites. *Angew. Chemie Int. Ed.* **55,** 11797–11800 (2016).

11. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14,** R119 (2013).

12. Spruijt, C. G. *et al.* Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* **152,** 1146–1159 (2013).

13. Raiber, E.-A. *et al.* 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22,** 44–49 (2014).

14. Hardwick, J. S., Lane, A. N. & Brown, T. Epigenetic Modifications of Cytosine: Biophysical Properties, Regulation, and Function in Mammalian DNA. *BioEssays,* 1700199 (2018).

15. Renciuk, D., Blacque, O., Vorlickova, M. & Spingler, B. Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Res.* **41,** 9891–9900 (2013).

16. Lercher, L. *et al.* Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem. Commun.* **50,** 1794–1796 (2014).

17. Szulik, M. W. *et al.* Differential Stabilities and Sequence-Dependent Base Pair Opening Dynamics of Watson–Crick Base Pairs with 5-Hydroxymethylcytosine, 5-Formylcytosine, or 5-Carboxylcytosine. *Biochemistry* **54,** 1294–1305 (2015).

18. Kypr, J., Kejnovská, I., Renčiuk, D. & Vorlíčková, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **37,** 1713–1725 (2009).

19. Neidle, S. *Principles of Nucleic Acid Structure* 1st Ed. (Elsevier, London, 2008).

20. Lu, X.-J. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31,** 5108–5121 (2003).

21. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **37,** 5917–5929 (2009).

22. Blackburn, M. G., Gait, M. J., Loakes, D. & Williams, D. M. *Nucleic Acids in Chemistry and Biology* (Royal Society of Chemistry, Cambridge, 2007).

23. Šponer, J. & Kypr, J. Different intrastrand and interstrand contributions to stacking account for roll variations at the alternating purine-pyrimidine sequences in A-DNA and A-RNA. *J. Mol. Biol.* **221,** 761–764 (1991).

24. Dickerson, R. E., Goodsell, D. S., Kopka, M. L. & Pjura, P. E. The effect of crystal packing on oligonucleotide double helix structure. *J. Biomol. Struct. Dyn.* **5,** 557–579 (1987).

25. Wahl, M. C. & Sundaralingam, M. Crystal structures of A-DNA duplexes. *Biopolymers* **44,** 45–63 (1997).

26. DiGabriele, A. D., Sanderson, M. R. & Steitz, T. A. Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract. eng. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 1816–1820 (1989).

27. Spink, N., Nunn, C. M., Vojtechovsky, J., Berman, H. M. & Neidle, S. Crystal structure of a DNA decamer showing a novel pseudo four-way helix-helix junction. *Proc. Natl. Acad. Sci.* **92,** 10767–10771 (1995).

28. Liu, J., Malinina, L., Huynh-Dinh, T. & Subirana, J. A. The structure of the most studied DNA fragment changes under the influence of ions: a new packing of d(CGCGAATTCGCG). *FEBS Lett.* **438,** 211–214 (1998).

29. Liu, J. & Subirana, J. A. Structure of d(CGCGAATTCGCG) in the presence of Ca(2+) ions. *J. Biol. Chem.* **274,** 24749–52 (1999).

30. Seth, P. P. *et al.* Synthesis and Antisense Properties of Fluoro Cyclohexenyl Nucleic Acid (F-CeNA), a Nuclease Stable Mimic of 2'-Fluoro RNA. *J. Org. Chem.* **77,** 5074–5085 (2012).

31. Robeyns, K., Herdewijn, P. & Van Meervelt, L. Direct observation of two cyclohexenyl (CeNA) ring conformations in duplex DNA. *Artif. DNA PNA XNA* **1,** 2–8 (2010).

32. Valls, N., Steiner, R. A., Wright, G., Murshudov, G. N. & Subirana, J. A. Variable role of ions in two drug intercalation complexes of DNA. *J. Biol. Inorg. Chem.* **10,** 476–482 (2005).

33. Valls, N., Wright, G., Steiner, R. A., Murshudov, G. N. & Subirana, J. A. DNA variability in five crystal structures of d(CGCAATTGCG). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 680–685 (2004).

34. Yang, X.-l., Robinson, H., Gao, Y.-G. & Wang, A. H.-J. Binding of a Macrocyclic Bisacridine and Ametantrone to CGTACG Involves Similar Unusual Intercalation Platforms. *Biochemistry* **39,** 10950–10957 (2000).

35. Adams, A., Guss, J. M., Collyer, C. A., Denny, W. A. & Wakelin, L. P. A novel form of intercalation involving four DNA duplexes in an acridine-4-carboxamide complex of d(CGTACG)(2). *Nucleic Acids Res.* **28,** 4244–53 (2000).

36. Thorpe, J. H. *et al.* Guanine-Specific Binding at a DNA Junction Formed by d[CG(5-BrU)ACG]2 with a Topoisomerase Poison in the Presence of Co2+ Ions. *Biochemistry* **39,** 15055–15061 (2000).

37. Canals, A. *et al.* Intercalative DNA binding of the marine anticancer drug variolin B. *Sci. Rep.* **7,** 39680 (2017).

38. Hardwick, J. S. *et al.* 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.* **24,** 544–552 (2017).

39. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9,** 465 (2008).

40. Gyi, J. I., Lane, A. N., Conn, G. L. & Brown, T. Solution Structures of DNA-RNA Hybrids with Purine-Rich and Pyrimidine-Rich Strands: Comparison with the Homologous DNA and RNA Duplexes. *Biochemistry* **37,** 73–80 (1998).

3

DO CYTOSINE MODIFICATIONS ALTER DNA STRUCTURE?

3.1 Introduction

Any biological effect of ^{m}C and its oxidised derivatives must ultimately arise through altering DNA's interactions with protein. Many important examples of proteins whose interactions with DNA are perturbed by cytosine modifications have recently been identified, including RNA polymerase II, transcription factors and chromatin remodellers.[1–4] But how cytosine modifications alter DNA-protein interactions remains unclear. Understanding this process is therefore essential for decoding the molecular basis for any biological function.

In Chapter 2, it was established that ^{f}C does not cause dramatic changes to DNA structure—and there is no evidence that ^{hm}C or ^{ca}C do so either. However, major changes to DNA conformation are not required to modulate DNA-protein interactions. More subtle, sequence-dependent variations in DNA structure are widely exploited by DNA-binding proteins (*indirect readout*, see Section 1.1.4). Thus, if ^{m}C derivatives perturb DNA structure to a comparable extent, they too could be sensed via indirect readout.

Such effects, however, remain unknown. With this in mind, the work described in this chapter was undertaken to address the following: do ^{m}C derivatives perturb DNA structure, and, if so, could such perturbations influence DNA–protein interactions?

Owing to the difficulty in using X-ray crystallography and NMR spectroscopy to probe sequence-dependent variation in DNA structure (Section 1.3), alternative techniques have been widely used—notably DNase I and electrophoretic mobility assays.[5–8] Both methods have also been used to characterise DNA containing ^{m}C,[9–12] indicating that methylation perturbs DNA structure, likely involving bending towards the minor groove (discussed below).[12] Given these findings, and that neither technique has yet been used to characterise DNA containing ^{hm}C, ^{f}C, or ^{ca}C, DNase I and electrophoretic mobility assays were used to probe the effects of ^{m}C derivatives on DNA structure.

3.2 DNase I Cleavage Assay

3.2.1 Background

Bovine pancreatic deoxyribonuclease I (DNase I) is an endonuclease that is specific to double-stranded DNA, with a footprint of ~6 bp.[13] It introduces single-strand nicks to the phosphodiester backbone by cleaving the P–O3′ bond.[14] Although DNase I is not sequence-specific, its cleavage rate is strongly sequence dependent, owing to sequence-dependent variation in the structure and flexibility of DNA.[15,16] Because of its sensitivity to these features, DNase I has been widely used as a probe of DNA conformation.[7,9,15–19]

In a typical cleavage experiment, a DNA duplex is subjected to controlled digestion by DNase I. This produces a range of different fragments, whose relative abundance directly reflects each site's susceptibility to DNase I cleavage. These products are then separated by electrophoresis and quantified, allowing the relative cleavage rate at each position to be determined (Figure 3.1).

Because DNase I binds to the minor groove of DNA, it cannot directly sense sub-

stituents in the major groove—such as the C5 modifications of cytosine (Figure 3.1*a*). Therefore, differences in cleavage rate at modified sites must arise from conformational changes to the DNA imparted by the modification.[11] Furthermore, DNase I is similar in size to a typical human transcription factor.[9] It is thus an ideal probe to investigate whether changes to the intrinsic properties of DNA caused by cytosine modifications could be sensed by DNA-binding proteins.

Previously, cytosine methylation has been shown to influence DNase I cleavage rates, as discussed below.[9–11] But oxidised derivatives of 5-methylcytosine have not been investigated. Here, the DNase I susceptibility of duplexes containing C, ^{m}C, ^{hm}C, ^{f}C, and ^{ca}C was compared, and the structural and potential biological implications of the results are discussed.

3.2.2 Results & discussion

Preparation of constructs

A random 32-nucleotide sequence was generated, into which two consecutive CpG steps were inserted at the centre (Figure 3.2*b*). Variants of this sequence were prepared containing either C, ^{m}C, ^{hm}C, ^{f}C, or ^{ca}C at position X, and each was fluorescently labelled at the 5′ end with 6-carboxyfluorescein (FAM). Oligonucleotides were purified by both HPLC and PAGE to ensure that no detectable amounts of degraded fragments were present in the uncleaved samples, since this could produce artefacts in the analysis (for an example gel, see Figure B.1*b*).

DNase I cleavage assay

First, reaction conditions were adapted from published procedures and optimised.[17,20,21] Particularly important was achieving a high signal-to-noise ratio of weaker bands, while ensuring at least 70% of the substrate remained uncleaved. The latter requirement is

so that each duplex was, on average, cut only once; multiple cuts on the same strand would otherwise bias the distribution towards shorter cleavage products.[22] Following optimisation, a variant of the Maxam–Gilbert 'G+A' sequencing reaction was performed to determine the sequences of the cleavage fragments, enabling the relative cleavage rates of dinucleotide steps around the modification site to be determined (Figure B.1). This method has been widely used for sequencing DNase I cleavage products.[14,23,24] Finally, cleavage reactions were carried out on duplexes containing C, ^{m}C, ^{hm}C, ^{f}C, or ^{ca}C in the fluorescently labelled strand. The unlabelled complement contained the same modification, ^{m}C, or C (Figure 3.1*b,d*). Experiments were performed in duplicate.

Results

A significant increase in susceptibility to DNase I cleavage was observed upon the incorporation of ^{m}C, ^{hm}C, or ^{f}C in the fluorescently labelled strand. The effect was localised primarily at the phosphodiester directly 5′ of the modified base, whose relative cleavage rate increased by a factor of ~2.5 (Figure 3.1*d*, Table B.1). The modification state of the opposite cytosine in the unlabelled strand did not appear to influence significantly the cleavage rate of this site. For ^{ca}C, however, the enhancement in cleavage rate was negligible in all cases (Figure 3.1*d*).

Next, the effect of cytosine modification on the opposite strand was investigated using two sets of constructs: one in which the cytosine modification was incorporated only in the unlabelled strand, and another in which both strands contained the same modification. Small decreases in the relative cleavage rate were observed for ^{m}C and ^{hm}C, 4–5 nucleotides downstream of site 'X' in the labelled strand, though these may not be significant (Figure 3.2, Table B.2). These differences were enhanced for ^{f}C and abrogated for ^{ca}C.

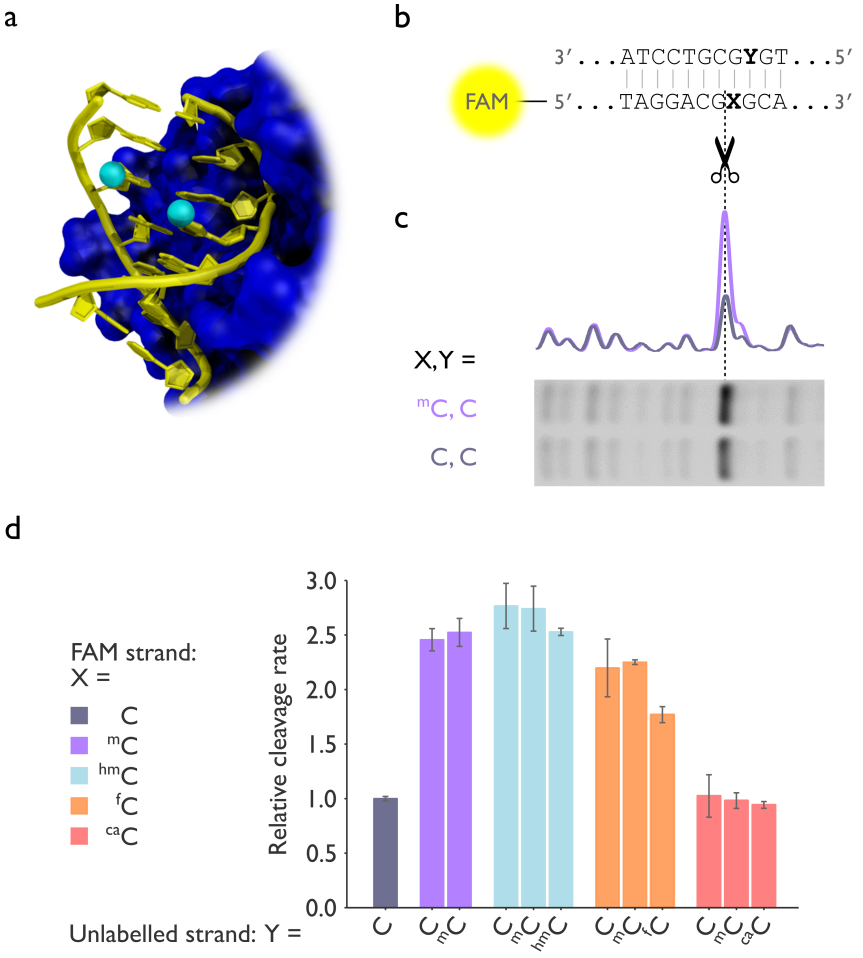


Figure 3.1. Cytosine modifications alter the relative cleavage rate of DNase I. (*a*) A crystal structure of an unmodified DNA duplex in complex with DNase I (PDB: 1dnk). The enzyme binds to the minor groove and cannot directly sense major-groove substituents, *eg* the C_5 modifications of cytosine, whose approximate positions are depicted by cyan spheres. (*b*) Section of the 5′-FAM-labelled 32-nt sequence and its unlabelled complement used in this study. The modification sites of the central CpG step are denoted by 'X'. The modification site in the opposite strand is denoted by 'Y'. (*c*) Section of densitograms of the cleavage products of two constructs: one unmodified, the other containing a hemimethylated CpG site. The altered cleavage rate is local to the modification site. (*d*) Mean relative cleavage rates, at the 5′ position of the modified residue, for duplexes containing different cytosine modifications. Reactions of the modified duplexes were performed in duplicate. Error bars show standard deviation.

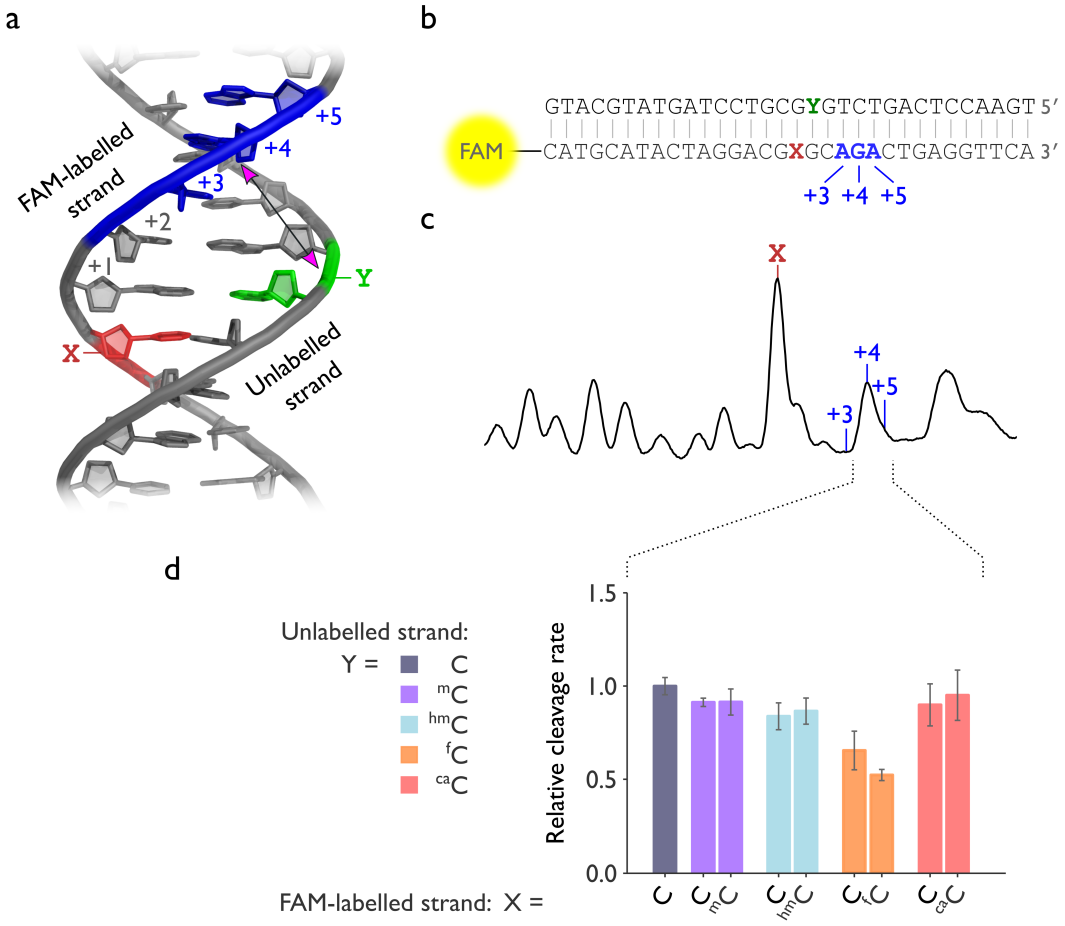


Figure 3.2. The influence of cytosine modification on DNase I cleavage susceptibility of the opposite strand. (*a*) If the structural effects from the modification spread across the minor groove (pink arrow), differences in cleavage rate would be expected on the opposite strand, ~3 nucleotides in the 3′ direction. Thus, if the unlabelled strand were modified at position Y, differences in cleavage rate would be expected on the labelled strand, ~4 nucleotides downstream of position X (the previous site of interest). (*b*) Sequence of the construct, with nucleotides 3, 4, and 5 downstream of position X highlighted in blue. (*c*) Example densitogram of the unmodified duplex, showing the corresponding cleavage pattern. Because the peaks of nucleotides 4 and 5 could not be resolved, the combined peak was integrated. No appreciable cleavage occurred at nucleotide +3. (*d*) Relative intensites of the combined peak of nucleotides +4 and +5 upon the introduction of ^{m}C, ^{hm}C, ^{f}C, or ^{ca}C into the unlabelled strand. The labelled strand was either unmodified, or contained the same modification. For the former, the control was the doubly unmodified duplex; for the latter, the control duplex contained the modification in the labelled strand and unmodified cytosine in the unlabelled strand. Error bars show standard deviation.

Discussion

In this study, CpG methylation was found to increase its susceptibility to DNase I cleavage. These observations are consistent with previous reports,[9–11] all of which show that differences in cleavage rate are localised primarily to the phosphodiester directly 5′ of the methylated cytosine. Furthermore, the magnitude of rate enhancement is in close agreement with that determined for the same tetramer motif (5′-G^{m}CGC) by Fox,[11] who observed a 2–3-fold increase in relative cleavage rate.

In all previous studies, differences in DNase I sensitivity were attributed to structural changes in the DNA. As discussed above, this is because DNase I binds to the minor groove of DNA and so cannot directly sense the methyl group. But the precise nature of these structural changes remains unclear. Lazarovici *et al* concluded, based on simulations, that the enhanced cleavage rate is caused by methylation-induced narrowing of the minor groove.[9] By contrast, Fox reasoned that changes in minor-groove width are unlikely to account for the effects; if minor-groove width were affected by methylation, an enhanced cleavage rate would also be expected on the opposite strand, about 2–3 nucleotides in the 3′ direction (Figure 3.2). Fox did not observe such changes, concluding that the differences in cleavage rate probably arise from changes in the orientation of the Gp^{m}C phosphodiester group—with methylation shifting it to a position that favours nucleophilic attack by the enzyme.[11] In this study, enhanced cleavage rates in the opposite strand 2–3 nucleotides downstream of the methylation site were also not observed, where, if anything, a slight reduction in cleavage rate was found (Figure 3.2*d*).

Hydroxymethylation of the CpG site induced a similar enhancement of the cleavage rate (2.5 to 2.8-fold) directly 5′ to the modified base (Figure 3.1*d*). Again, the modification state (C, ^{m}C or ^{hm}C) of the unlabelled strand appeared to have no significant influence on the cleavage rate at the Gp^{hm}C site. And only a small reduction in relative cleavage rate was observed across the minor groove (Figure 3.2*d*). Overall, ^{hm}C behaved similarly to ^{m}C, suggesting that their effects on DNase I susceptibility could be partly influenced

by the steric bulk of the C5 substituent.

Formylation also increased susceptibility of the Gp^{f}C site to cleavage, though seemingly to a lesser extent (Figure 3.1*d*). The mean cleavage rate was lowest for the diformylated CpG duplex (relative enhancement of 1.78 ± 0.07). Cleavage rates of the nucleotides on the opposite strand, across the minor groove, were significantly reduced—suggesting that formylation could alter minor-groove width.

Curiously, carboxylation appeared to have a negligible effect on the Gp^{ca}C site's susceptibility to DNase I cleavage (relative enhancement of 0.94–1.02). This was surprising, given that the modification is similar in size and shape to the formyl group, and places a negative charge in the major groove. This was expected to induce considerable structural distortion through electrostatic repulsion with backbone phosphodiester groups. And even greater perturbation was anticipated with the dicarboxylated CpG step, whose carbonyl carbon atoms would be separated by only ~7.5 Å in regular B-DNA.[25] It may be that the carboxylate charges in ^{ca}CpG steps are efficiently shielded by counterions, counteracting the expected perturbation. Observations that ^{ca}C increases duplex stability are consitent with this interpretation.[26,27] Alternatively, it is possible that ^{ca}C is structurally perturbing in a different way to ^{m}C and ^{hm}C, but that DNase I happens to be insensitive to these particular changes.

The findings presented here indicate that ^{m}C, ^{hm}C, and ^{f}C alter DNA conformation in the minor groove, but what is the nature of these effects? Many studies have tried to pinpoint the conformational cause of sequence- and methylation-dependent differences in DNase I cleavage rate.[9,13,16–18,28,29] Several point to minor-groove width being the predominant factor,[9,17] while others have concluded that DNA flexibility,[16,18] or local structural changes, are the main determinant.[11,28,29] That no single factor appears to explain all experimental observations suggests that the behaviour of DNase I may not be ascribed to an individual structural or mechanical feature of DNA. This interpretation seems reasonable, considering the many degrees of freedom involved in the binding

of a protein to DNA. Understanding the precise structural nature of these effects is thus beyond the scope of this work. But these results do indicate that ^{hm}C and ^{f}C cause conformational changes to DNA—changes that have not previously been identified. Moreover, these data show that such changes can influence DNA–protein interactions and may thus be biologically significant.

3.3 Electrophoretic Mobility Assay

The DNase I experiments indicate that ^{m}C, ^{hm}C and ^{f}C alter the structure of DNA, although the nature of these changes could not be ascertained. Given the importance of DNA bending to the stability of nucleosomes and other DNA–protein complexes, the effect of cytosine modifications on DNA bending was investigated by analysis of their influence on electrophoretic mobility.

3.3.1 Background

The discovery of sequence-directed DNA bending came from observations that bent DNA migrates unusually slowly through polyacrylamide gels.[5,6,30,31] The effect occurs because the electrophoretic mobility of a DNA fragment depends on its shape, as well as its contour length.[5,32] Curved DNA, with its smaller end-to-end distance and greater cross-sectional area, encounters more drag when passing through a polyacrylamide gel than does straight DNA.[33] Thus, for a given contour length, a curved DNA duplex migrates more slowly, behaving as though it were a larger fragment (Figure 3.3*a*).[33]

Since its role in the discovery of intrinsically bent DNA molecules, PAGE-based analysis of electrophoretic mobility has been widely used to characterise DNA bending—both in free DNA[7,8,12,34–36] and in complex with protein.[32,37–40] The method has revealed multiple sequence motifs that impart stable curvature into the double helix,[5,7,8,31] the most widely studied example being A-tracts: runs of $A_n{\cdot}T_n$, where n equals 4–6.[30] Each

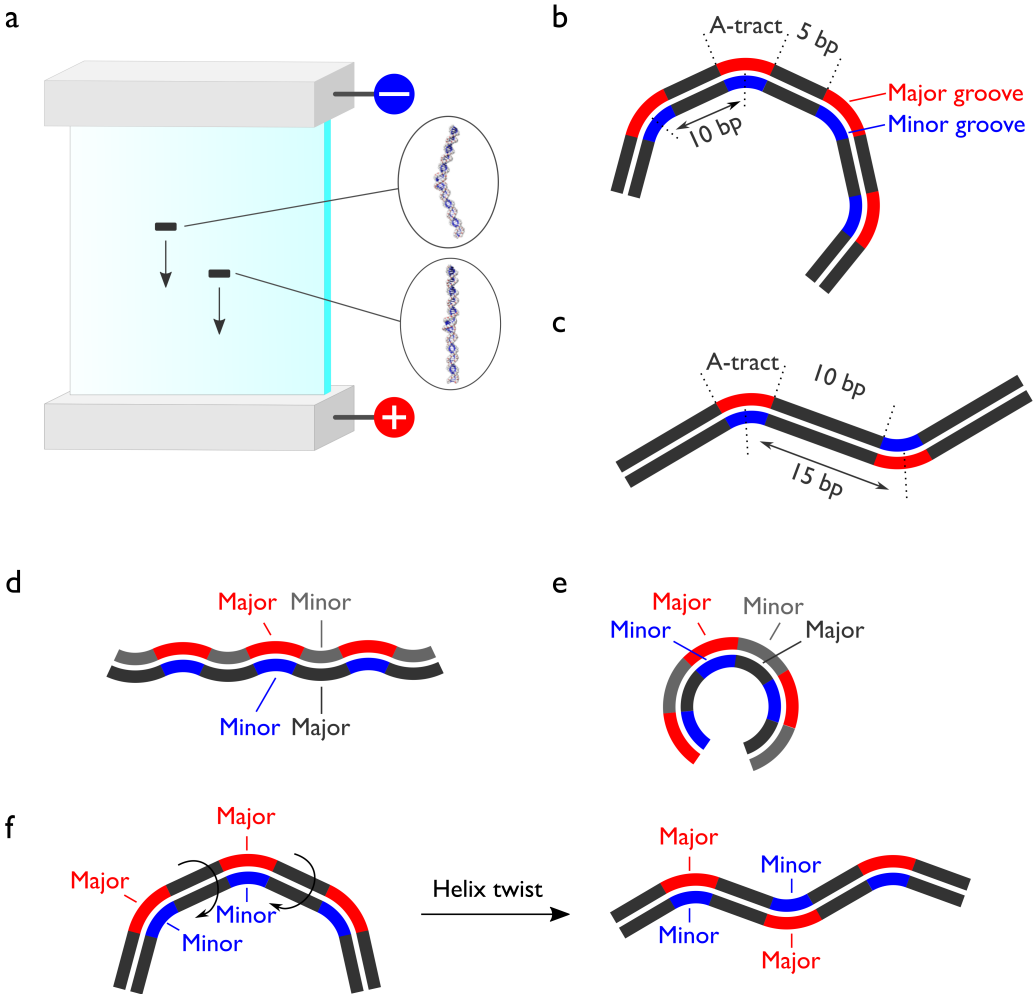


Figure 3.3. DNA curvature and anomalous electrophoretic mobility. (*a*) Non-denaturing polyacrylamide gel electrophoresis of bent *vs* straight dsDNA fragments comprising the same number of base pairs. The bent DNA encounters more drag as it passes through the gel, causing it to migrate more slowly compared to the straight fragment. (*b*) Schematic of duplex DNA containing A-tracts in-phase with the helix repeat (every ~10 bp). Each A-tract ($A_5{\cdot}T_5$) induces a strong bend into the minor groove. When their centres are separated by 10 bp, the bends of the A-tracts reside on the same side of the double helix and add constructively, resulting in net curvature of the fragment. (*c*) When the centres of the A-tracts are out-of-phase with the helix repeat (every ~15 bp), the bend of any given A-tract opposes that of its nearest neighbour and no net curvature is observed. (*d*) If the 5-bp region in between A-tracts also bends into the minor groove, the fragment's net curvature would be reduced. Conversely, if the non-A-tract region bends into the major groove (*e*), the fragment's net curvature would be enhanced. (*f*) Altering the twist between A-tracts could also alter the fragment's curvature without inducing any local bending. Shown here is the extreme case of a twist perturbation of 180° spread across 5 bp, causing the bending of each A-tract to oppose that of its nearest neighbour.

A-tract is associated with a bend into the minor groove, the angle of which has been estimated to be 11–28°, depending on the experimental technique and conditions.[30] When repeated every helical turn (~10.5 bp), each A-tract resides on the same side of the helix. The bending of multiple A-tracts therefore adds constructively, resulting in substantial curvature and thus reduced electrophoretic mobility (Figure 3.3*b*).[41] Conversely, when A-tracts are repeated every 1.5 helical turns, they are out-of-phase with the helical repeat; each bend opposes its neighbour, resulting in a zig-zag structure with no net curvature. Thus, despite the presence of local bends, the lack of net curvature in the molecule results in normal electrophoretic mobility (Figure 3.3*c*).[41]

Electrophoretic mobility analysis of A-tract sequences can also be used to investigate the effects of base modification on DNA structure.[12,42] Previous work has shown that cytosine methylation alters the electrophoretic mobility, and therefore the net curvature, of DNA containing A-tracts.[12] The authors tested several different sequences comprising the repeating motif A_5N_5 (where N_5 is a non-A-tract sequence containing ^{m}C at various different positions).[12] The most likely cause of these differences in mobility was concluded to be the methylation-induced bending of the DNA (Figure 3.3*d*,*e*). However, the authors could not rule out an alternative explanation—that ^{m}C exerted its effect through altered helix twist (Figure 3.3*f*). Additionally, they found that methylation of non-A-tract DNA did not substantially alter electrophoretic mobility. Thus, it was not clear if A-tracts simply amplify the methylation-induced effects to a level detectable by the gel mobility assay, or if the changes in net curvature arise from specific interactions between ^{m}C and the A-tract region.[12] Given that the largest difference in mobility occurred when ^{m}C was furthest (3 bp) from the A-tract, the latter explanation seems unlikely—though it cannot be ruled out from the previous study.

The aim of these experiments was to investigate qualitatively how cytosine modifications alter DNA curvature through changes in electrophoretic mobility. The ambiguities highlighted in earlier work on ^{m}C were first addressed. The effects of hydroxymethyla-

tion and formylation were then explored.

3.3.2 Results & Discussion

Construct design, ligation, and electrophoresis

The design of constructs was adapted from Ref [12]. Briefly, a set of 10-mer oligonucleotides were prepared by solid-phase synthesis and HPLC-purified, each of which contained an A-tract, the modified cytosine of interest, and a 5′-terminal phosphate (Figure 3.4*a*). Each strand and its complement contained an 8-bp complementary region, leaving a 2-nt overhang at each 5′ end. The overhanging region of one duplex fragment is complementary to another, enabling their enzymatic ligation by T4 DNA ligase (Figure 3.4*a*).

The sequence itself also came from Ref [12] and was reported to exhibit the largest change in mobility upon methylation (Figure 3.4*a*). Importantly, the modification site was in a CpG context—the most common sequence context of cytosine methylation in vertebrates.[43] Additionally, modifications were introduced to the same heptamer motif that was used in the DNase I studies (ACGXGCA, where X denotes the modified cytosine). Thus, the local conformation around the modification site in these constructs is likely to have been similar to that encountered by DNase I, which has a footprint of 6 base pairs.[13]

Ligation reaction conditions (based on Refs [7, 12]) were optimised to yield ladders containing suitable proportions of shorter and longer bands (Figure 3.4*b*). Because the ends of each fragment were complementary and could potentially enable its cyclisation, the presence of cyclised products was investigated and ruled out using the procedure of Ulanovsky *et al* (Figure B.3).[44]

Non-denaturing polyacrylamide gels (10%) were cast and run at room temperature. For reasons discussed below, gels were run in two different buffers—one without divalent

cations (TBE: 90 mM Tris-borate, 2 mM EDTA) and one with magnesium (TBMg: 90 mM Tris-borate, 10 mM magnesium acetate).

Analysis of gels

The relative electrophoretic mobility of a modified fragment to its unmodified counterpart, μ_R, was determined for each resolvable fragment and is defined as follows:

$$\mu_R = \frac{d_B}{d_A} \tag{3.1}$$

where d_B and d_A are the distances migrated by the modified and unmodified fragments, respectively (see Section 5.6 for derivation). According to theory (Section 5.6), the value μ_R of a given fragment was expected to be independent of the gel, provided that all gels are run under identical conditions. Indeed, the same relative trends in electrophoretic mobility were consistently observed. However, owing to small variations in μ_R across gels, quantitative comparisons of μ_R were made only between values obtained from the same gel, unless otherwise stated. For each comparison shown below (*eg* Figure 3.5), the data derive from two separate gels, each run in one of two different buffers: TBE or TBMg. Despite the differing buffers, the relative trends in electrophoretic mobility are similar, indicating that the observed effects represent intrinsic differences in the physical properties of DNA arising from the modifications.

Because of the large number of individual measurements required in each gel lane, and the difficulty in manually determining the peak maximum of each band, a semi-automated analysis procedure was developed, enabling rapid and unbiased analysis of gel images (see Section 5.6).

Cytosine methylation: axial bending vs *helix twist*

First, efforts were made to determine how ^{m}C alters net curvature—to differentiate between its potential effects on axial bending and helix twist. In the same sequence, the

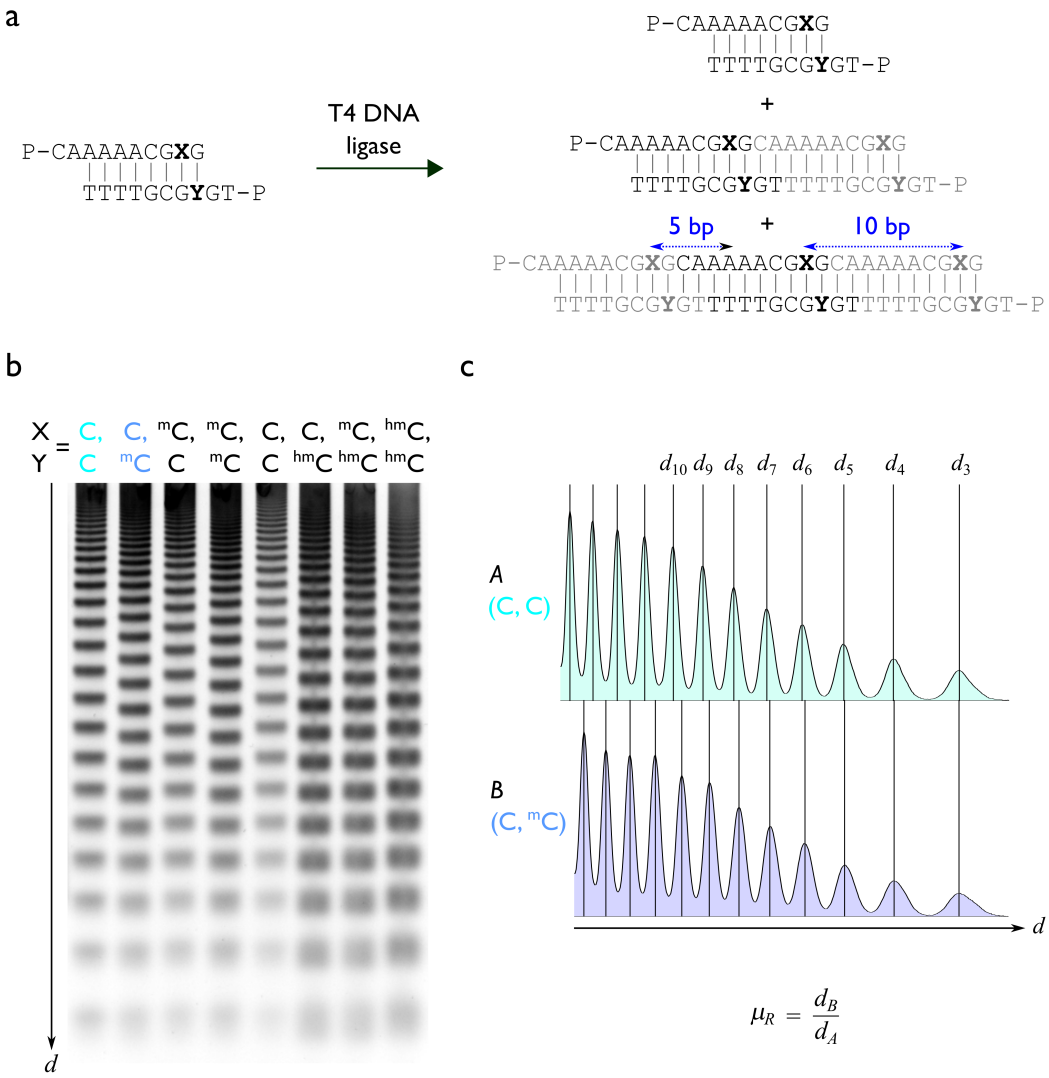


Figure 3.4. Ligation and analysis of the A-tract containing sequence, into which modified cytosines were incorporated. (*a*) T4 DNA ligase was used to ligate the duplex monomer, which contains 5′ phosphates at each end (denoted as P) and an 8-bp complementary region containing the modified cytosines (denoted X and Y). The modified residues repeat every 10 bp and are ~5 bp from the A-tract centre. Each ligation reaction produced a distribution of fragments, differing in length by 10-bp intervals. (*b*) Example image of a 10%, non-denaturing polyacrylamide gel used to determine the electrophoretic mobility of the ligation products. (*c*) For each fragment, the distance, *d*, from the origin to the peak maximum was determined using a custom *R* script. Mobilities of each modified fragment, relative to its unmodified counterpart were then calculated by taking the ratio of the two distances.

methylated site was moved from position 'X' (Figure 3.5) to the adjacent base pair on the complementary strand (position 'Y', Figure 3.5). If the effect of ^{m}C on net curvature arise from altered helix twist, then the two constructs would be expected to have similar electrophoretic mobilities. This is because the relative orientation between A-tract bends would be altered by the same angle, regardless of the position of ^{m}C. Conversely, if the effect arises from axial bending, the precise position of ^{m}C would be key; moving the methylation site to the adjacent base pair in the opposite strand (from position 'X' to 'Y', Figure 3.5) would change its orientation with respect to the A-tract bend by ~35°. A change in mobility would thus be expected—the effect being largest when the modification site is exactly half a helical turn away from the centre of the A-tract bend. Indeed, moving the ^{m}C site from position 'X' to 'Y' significantly affected mobility, which was further increased compared to the unmethylated variant (Figure 3.5).

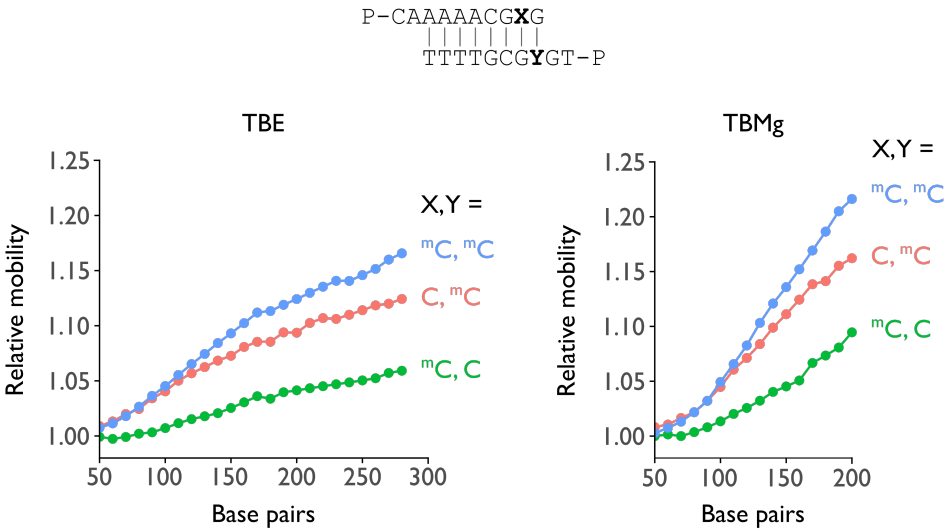


Figure 3.5. The effect of cytosine methylation on the electrophoretic mobility of A-tract DNA. Incorporating ^{m}C ~5 bp (half a helical turn) from the centre of the A-tract reduces the anomalous mobility of the fragments, indicating that ^{m}C reduces the net curvature of the fragments. Moving ^{m}C to the adjacent base pair on the other strand has a large effect on mobility. Symmetrical methylation of the CpG site caused the largest structural perturbation. The presence of 10 mM magnesium in the running buffer amplifies the effects but does not alter the trend. The data from each plot are derived from an individual gel.

Next, a further experiment to distinguish between helix twist and bending was devised, in which every C was replaced with ^{m}C (Figure 3.6). If each addition of ^{m}C were to increase helix twist by, say, 10°, then 5 additions between two A-tracts would alter their relative orientation by ~50°. In other words, perturbations to helix twist would add constructively, so that multiple additions would have a greater effect on mobility than would a single addition. Perturbations to the helix axis, however, can add either constructively (if in-phase with the helix repeat) or can cancel each other out (if out-of-phase with the helix repeat).

Strikingly, incorporating a single ^{m}C residue had a greater effect on mobility than did incorporating five (Figure 3.6*a*). Thus, in the constructs containing 5 consecutive additions of ^{m}C, the outermost methyl groups would sit on opposite sides of the helix

(Figure 3.6*b*). Therefore, any bending associated with the two residues would at least partially cancel. Conversely, the bending of a single methyl group would not be opposed in such a way (Figure 3.6*c*). These observations are therefore consistent with methylation-induced changes to axial bending, rather than perturbations to helix twist.

Are the effects on DNA curvature specific to A-tract sequences?

In previous work, it was found that incorporating ^{m}C into electrophoretically normal, non-A-tract sequences did not significantly alter mobility. It thus remained unclear if the effects of ^{m}C on net curvature are specific to A-tract sequences, or if the A-tracts just amplify the effects to an observable level. To resolve this, ^{m}C was incorporated into a sequence that contained an alternative bending motif—the GGGCCC (G_3C_3) tract—which has been shown to exhibit stable curvature in the presence of divalent cations.[7,8] Thus, if ^{m}C induces DNA bending independently of A-tracts, then ^{m}C should also alter the electrophoretic mobility of constructs containing the G_3C_3 motif. Furthermore, the G_3C_3 motif bends into the major groove—the opposite direction to the A-tract.[7,8] The effect of methylation on the electrophoretic mobility of G_3C_3 constructs should therefore be opposite to that of A-tracts. *Eg*, if ^{m}C causes bending into the minor groove, it should increase the mobility of A-tract constructs and further reduce those of G_3C_3 (Figure 3.3*d,e*).

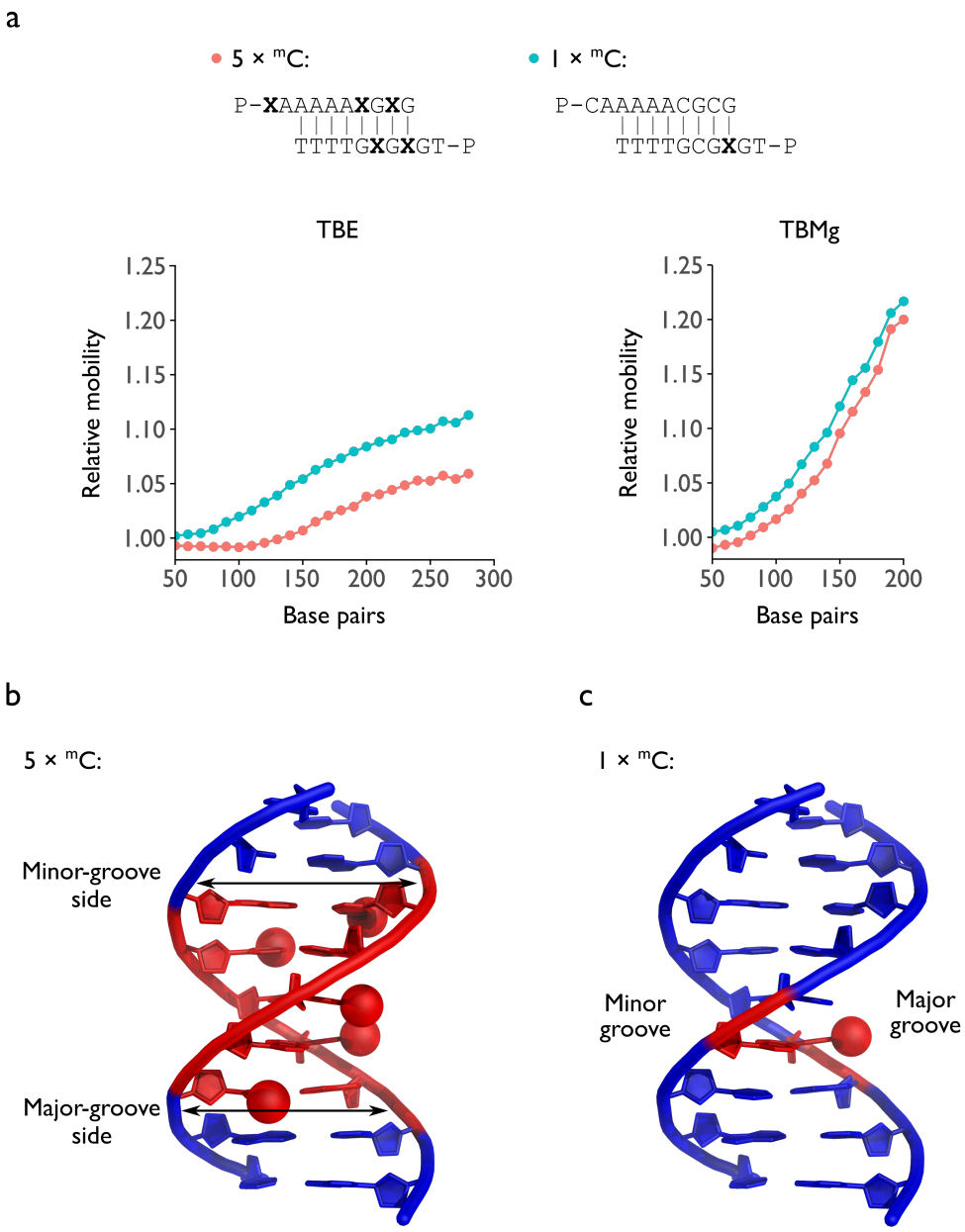


Figure 3.6. The effect of increased methylation on electrophoretic mobility of the A-tract fragments. (*a*) Electrophoretic mobilities of the fully and singly methylated constructs. A single addition of ^{m}C had a larger effect than did incoporating five ^{m}C residues consecutively, indicating that the perturbations arising from methylation partly cancelled each other out. This is consistent with methylation-induced bending rather than altered helix twist. The data from each plot are derived from an individual gel. (*b*) A model of the fully methylated CGCGC tract, containing five consecutive ^{m}C-containing base pairs (red), shows that the bending of the outermost ^{m}C base pairs would be in opposition and would thus partially cancel. (*c*) Model of the singly methylated CGCGC tract, in which none of the bending would be opposed.

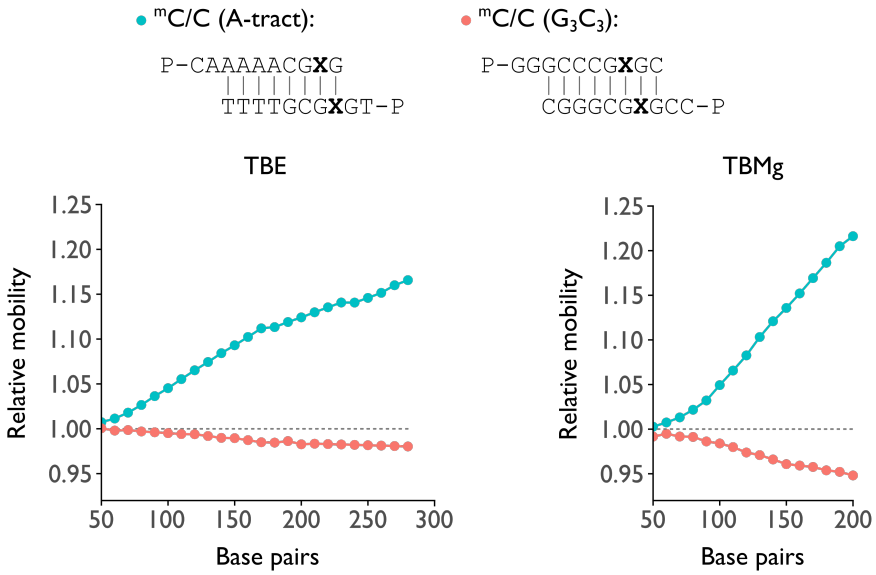


Figure 3.7. The effect of cytosine methylation on the electrophoretic mobility of A-tract and G_3C_3 sequences. Symmetric CpG methylation ~5 bp from the G_3C_3 centre decreases mobility, whereas it increases mobility of the A-tract fragments. There was a measurable effect on mobility of the G_3C_3 fragments even in the absence of divalent cations (TBE buffer), under which conditions the G_3C_3 tract is not intrinsically bent (Figure B.4).[8] In the presence of 10 mM magnesium (TBMg), the G_3C_3 tract exhibits modest bending into the major groove (Figure B.4),[8] further enhancing the fragments' net curvature. Thus, the effect of methylation on gel mobility is not A-tract-specific and is consistent with methylation-induced bending of DNA into the minor groove. Note that the A-tract and G_3C_3 fragments were run on different gels, so the two sets should only be compared qualitatively.

Incorporating ^{m}C into constructs containing the G_3C_3 motif significantly altered electrophoretic mobility, showing that the effect is not specific to A-tract sequences. Methylation of a CpG site ~5 bp away from the centre of the G_3C_3 motif reduced the electrophoretic mobility of the constructs, corresponding to an enhancement of net curvature (Figure 3.7).

The effect of methylation on the electrophoretic mobility of the G_3C_3 constructs was opposite to that of the A-tract sequence. As discussed above, the A_5 and G_3C_3 motifs are associated with bending in the opposite direction (into the minor and major grooves,

respectively). Because the ^{m}C sites were situated ~5 bp away from the centre of the bending motif, these results are consistent with CpG methylation inducing a bend into the minor groove (Figure 3.3*d,e*).

The reduction in mobility resulting from methylation was observed even in the absence of divalent cations (Figure 3.7), which are reportedly required to induce the bending of the G_3C_3-tract into the major groove. Thus, the effects of cytosine methylation are detectable even without intrinsically bent constructs. This was confirmed by comparing the mobility of the unmodified G_3C_3 fragments, in both the presence and absence of 10 mM magnesium, with that of two different unmodified sequences whose electrophoretic mobilities are normal (Figure B.4). The analysis showed that in the absence of magnesium, the mobility of the G_3C_3-containing fragments was not retarded, and thus that they were not intrinsically curved (Figure B.4). In the presence of 10 mM magnesium acetate, a small difference in mobility of the G_3C_3 fragments was observed—considerably less than that of the A-tract seqence. Thus, the unmodified G_3C_3 fragments possess much less intrinsic curvature than did those containing A-tracts. This may explain why methylation had a stronger effect on the mobility of the A-tract fragments.

The effect of CpA methylation on DNA bending

Cytosine methylation in vertebrates occurs primarily at CpG sites, of which around 70–80% are methylated.[43] But in the mammalian brain and embryonic stem cells, a considerable amount of non-CpG methylation occurs—reaching levels of up to 25%.[45,46] Being the most common site of non CpG methylation in pluripotent cells,[47] the effect of CpA methylation on DNA curvature was explored.

The same A-tract sequence was used, except that GX was replaced with A (Figure 3.8). Electrophoretic analysis indicates that CpA methylation has only a small effect on helical curvature. The biological significance of this observation is unclear, but it is interesting that CpG sites appear to be more perturbed by methylation.

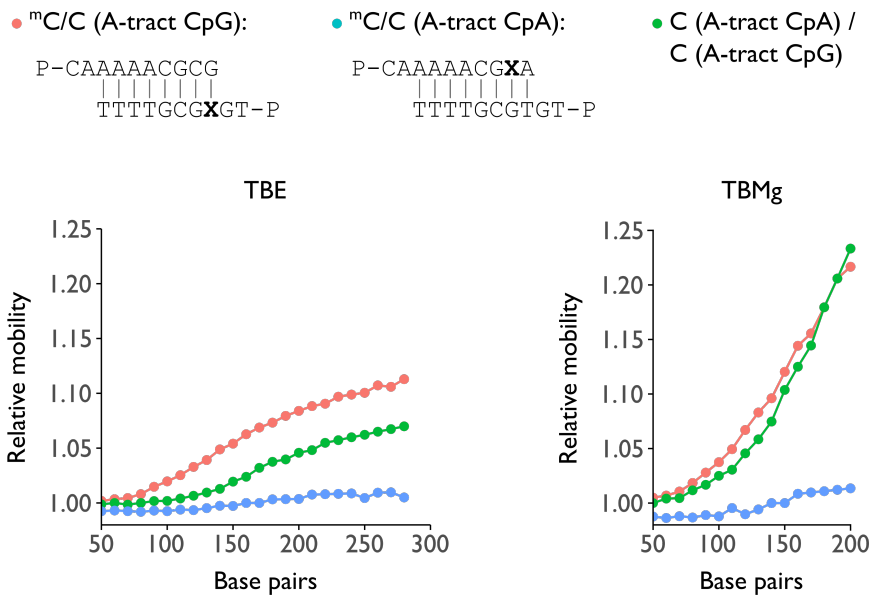


Figure 3.8. The effect of CpA methylation and altered sequence on electrophoretic mobility in TBE and TBMg running buffers. CpA methylation had only a weak effect on mobility (blue), indicating that the structural influence of cytosine methylation depends on the sequence around the modification site. By contrast, incorporating a single methyl group into a CpG site (red) typically had a greater effect on mobility than did changing the entire base pair. This suggests that the influence of CpG methylation on DNA bending is comparable in magnitude to sequence-dependent effects. The data from each plot are derived from an individual gel.

The unmodified CpA construct also enabled a comparison between the effects of CpG methylation and sequence on DNA bending. CpG methylation typically caused greater perturbation to mobility than did substituting G with A (Figure 3.8). In other words, introducing a single methyl group into the duplex appears to have a larger effect on curvature than does changing an entire base pair. Thus, the structural consequences of CpG methylation appear to be comparable with those of sequence-dependent effects.

Hydroxymethylation

Having established that DNA bending is the most likely cause of altered mobility, and that the effects are not specific to A-tract sequences, the influence of hydroxymethylation

on gel mobility was investigated. CpG hydroxymethylation in the A-tract constructs was probed in three different contexts: opposite unmodified cytosine, ^{m}C, or ^{hm}C.

As with ^{m}C, hydroxymethylation increased electrophoretic mobility of the A-tract fragments in all cases (Figure 3.9*a*). This indicates that ^{hm}C also induces bending into the minor groove, though the effect of ^{hm}C was weaker than that of ^{m}C. An unexpected trend was also observed; the presence of a single ^{hm}C residue in a CpG site was more perturbing than was the presence of two. Thus, in symmetrically hydroxymethylated CpG sites, there may be a cooperative interaction between the ^{hm}C groups that partially resists bending into the minor groove. To investigate potential structural causes, a B-DNA model was built containing a symmetrically hydroxymethylated CpG step (Figure B.5). The model indicates that the two hydroxymethyl (hm) groups are too far apart to form a direct hydrogen-bond across the major groove. Instead, the two hm groups could interact through a network of water molecules. Alternatively, the effect may arise from enhanced interactions with the 3′ guanine residues of the CpG site (Figure B.5).

The interpretation of the above results was supported by additional experiments in which ^{hm}C residues were incorporated into constructs containing the G_3C_3 motif (Figure 3.9*b*). Placing ^{hm}C opposite ^{m}C in CpG sites perturbed mobility, but to a lesser extent than dimethylated CpG sites, while symmetric hydroxymethylation of CpG sites had the weakest effect. This behaviour qualitatively mirrors that of the A-tract sequences (Figure 3.9*a*). Thus, the results support the notion that hydroxymethylation of CpG sites induces bending into the minor groove—though to a lesser extent than does ^{m}C.

Formylation

Finally, the effect of CpG formylation on the electrophoretic mobility of the A-tract constructs was investigated. ^{f}C typically had a weaker effect on mobility than did ^{m}C or ^{hm}C (Figure 3.10). Additionally, for the symmetrically formylated constructs, the relationship between construct length and mobility was more complex; in the range

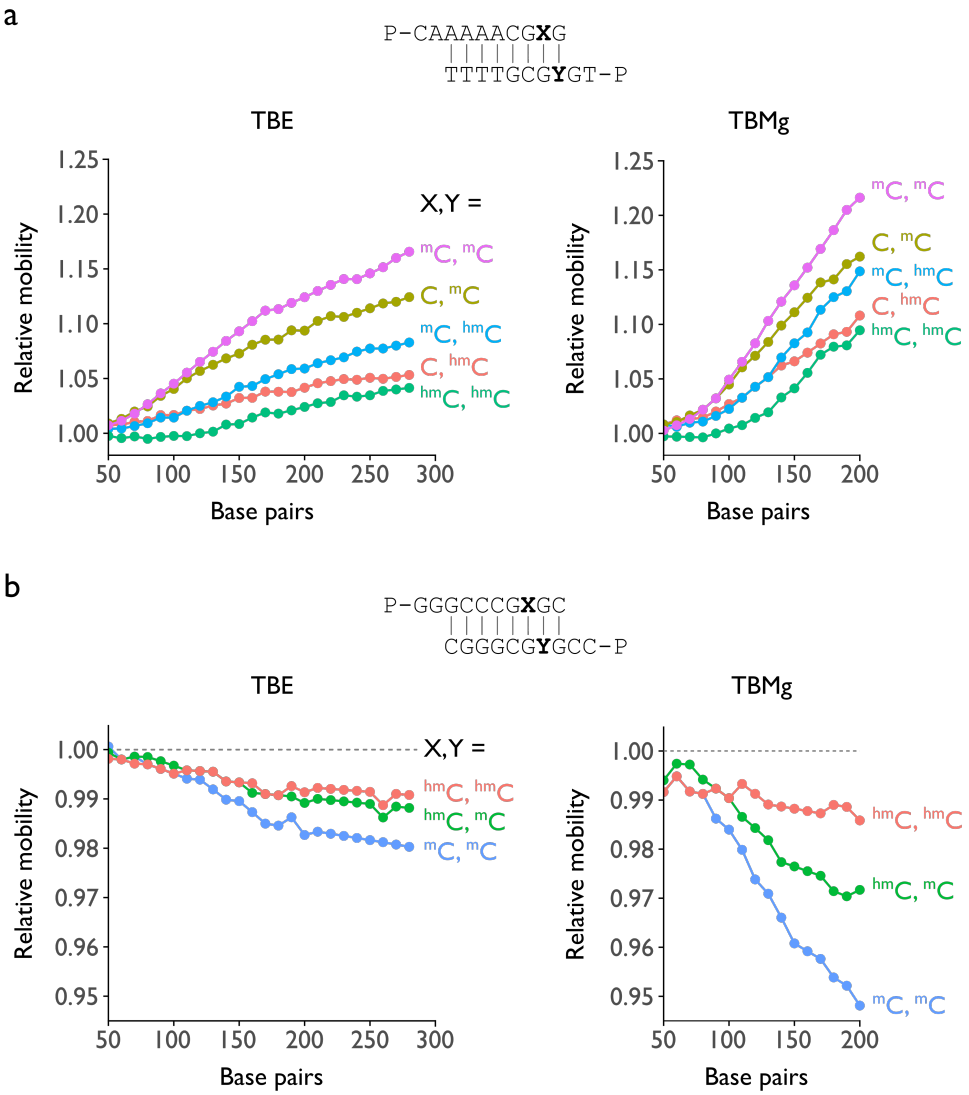


Figure 3.9. The effect of CpG hydroxymethylation on electrophoretic mobility. (*a*) Modified variants of the A-tract sequence containing ^{hm}C, ^{m}C, or both. ^{hm}C had a weaker effect on mobility than did ^{m}C. Surprisingly, incorporating two ^{hm}C residues into the CpG site was less perturbing than incorporating one. For all five variants, the same trend was observed in both TBE and TBMg buffers, though typically perturbations to mobility were greater in TBMg buffer. (*b*) The inverse trend was observed for the G_3C_3-containing constructs—consistent with ^{hm}C inducing bending into the minor groove. As with ^{m}C, ^{hm}C also altered the electrophoretic mobility of the G_3C_3 fragments in the absence of divalent cations, *ie* in electrophoretically normal DNA; the effects were amplified in the presence of 10 mM magnesium acetate. The data from each plot are derived from an individual gel.

of about 50–150 bp, the relative mobility of the formylated fragments decreased with increasing length. Beyond ~150 bp, the trend reversed, and relative mobility increased with increasing length. This effect was repeatedly observed under these conditions, but its origin is unclear. One explanation could be that the effects of DNA bending and cation coordination compete; if CpG formylation were to promote cation coordination, this would reduce the constructs' net negative charge and thus their electrophoretic mobility. At greater fragment lengths (above ~150 bp) the effect on bending (which typically increases sharply above ~100 bp) may begin to dominate, reversing the trend. Owing to time constraints, it was not possible to investigate the effect of ^{f}C on the G_3C_3 constructs.

Further discussion

Previous work indicated that cytosine methylation induces distortions to the helix axis but could not rule out the possibility of altered helix twist. Furthermore, it was unclear if the effects were specific to A-tract sequences. In this work, strong evidence was found in favour of methylation-induced bending into the minor groove, and it was shown that the effects are not A-tract specific. The sequence context of the modification site, however, is important; methylation of CpA steps—which are one of the most abundant sites of non-CpG methylation—appears to have a much smaller influence on DNA bending.

Notably, these results also indicate that symmetrical methylation of CpG sites induces stronger bending into the minor groove than does hemimethylation, which has not previously been identified. It would be interesting to investigate whether such effects could contribute towards the molecular recognition of different CpG methylation states.

Unlike the DNase I cleavage studies, which report directly on DNA-protein interactions, the link between differences in electrophoretic mobility and DNA-protein interactions is less clear. However, the electrophoretic mobility experiments show that CpG modification produces structural differences of comparable magnitude to those

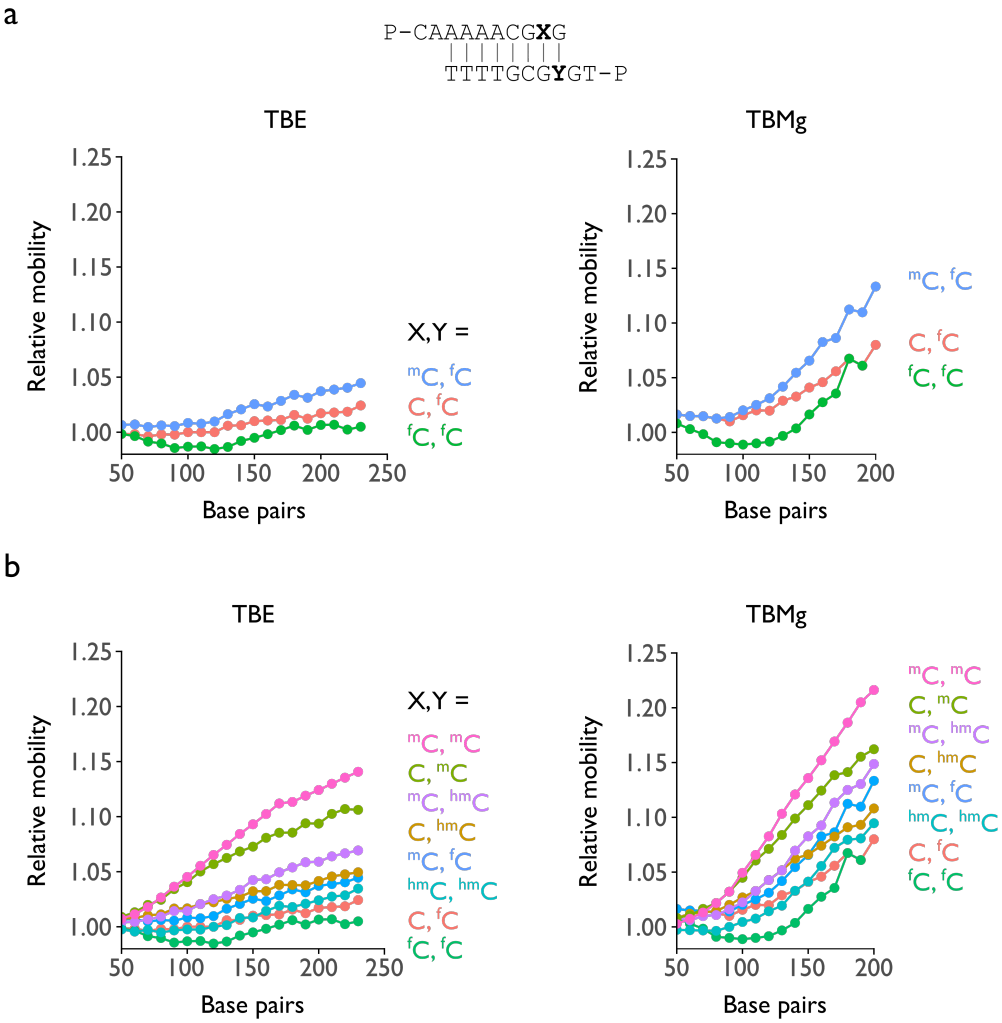


Figure 3.10. The effect of CpG formylation on electrophoretic mobility of A-tract DNA. (*a*) The mobility of the formylated variants followed the same trend as their hydroxymethylated counterparts, though ^{f}C was less perturbing. Again, incorporating two ^{f}C residues into the CpG site was less perturbing than incorporating one. The same trends were observed in both buffers, though TBMg amplified the effects, as was the case with ^{m}C and ^{hm}C. Under both buffer conditions, diformylation of the CpG site decreased mobility at lengths of ~50–150 bp. (*b*) A comparison of all modified variants of the A-tract sequence. For any given combination of modifications to the CpG site, the order of perturbation from greatest to least was ^{m}C > ^{hm}C > ^{f}C. In every case, the same trend was observed in both buffers—with TBMg consistently magnifying the effects. The data from each plot are derived from an individual gel.

arising from changes in sequence and thus could be biologically significant.

CpG hydroxymethylation also appears to induce bending into the minor groove, though to a lesser extent than does ^{m}C. The modification state of the other cytosine in the CpG site has a significant influence on electrophoretic mobility, indicating that each of the three variants has distinct conformational properties. CpG sites containing ^{hm}C opposite ^{m}C appear to exhibit greater bending into the minor groove than do those containing ^{hm}C opposite C. Curiously, symmetrically hydroxymethylated CpG sites were less perturbed than were hemihydroxymethylated sites. Thus, having two hydroxymethyl groups appears to counteract some of the bending—perhaps through interactions accross the major groove that resist its opening. All three variants of hydroxymethylated CpG steps (^{hm}C/C, ^{hm}C/^{m}C, ^{hm}C/^{hm}C) have been reported to occur *in vivo*.[48–50] The presence of ^{hm}C and ^{m}C within a CpG step was only recently demonstrated but has been estimated to account for ~60% of all ^{hm}C.[50] Despite their apparent prevalence, CpG sites containing both ^{hm}C and ^{m}C remain understudied, with most investigations into their biophysical properties focusing on either hemi- or fully hydroxymethylated states.[26,51–56] Thus, CpG sites containing both ^{hm}C and ^{m}C should be the focus of future structural and biochemical work, especially considering that they appear to be the most structurally perturbed of the ^{hm}C variants.

CpG formylation had a weaker and more complex effect on electrophoretic mobility than did methylation or hydroxymethylation. More work is needed to determine the nature of the observed trends, especially for the diformylated variants. Analysis of constructs containing the G_3C_3 motif would be particularly useful; if the observed trends arise predominantly from axial bending, inverse effects would be observed in the A-tract and G_3C_3 constructs—as was observed for ^{m}C and ^{hm}C. Conversely, if the trend is a combination of effects derived from bending into the minor groove and cation coordination (see above), both would reduce the mobility of G_3C_3 constructs.

Overall, it was concluded that both mC and hmC induce bending into the minor

groove. But why are these conformational changes not apparent in their crystal structures? Two possible explanations could account for this. It may be that lowest-energy conformer is altered by the modification (Figure 3.11*a*) but that structural changes imposed by the crystalline environment prevent this conformer from being adopted. Indeed, both crystallisation conditions and crystal-packing effects have been found to significantly alter DNA bending.[57,58] Alternatively, cytosine modifications could induce bending without altering the lowest-energy conformer. This could be achieved through steric repulsion of the C_5 modifications, which may resist DNA's bending into the major groove. Thus, the averaged structure would be biased towards one that bends into the minor groove (Figure 3.11*b*).

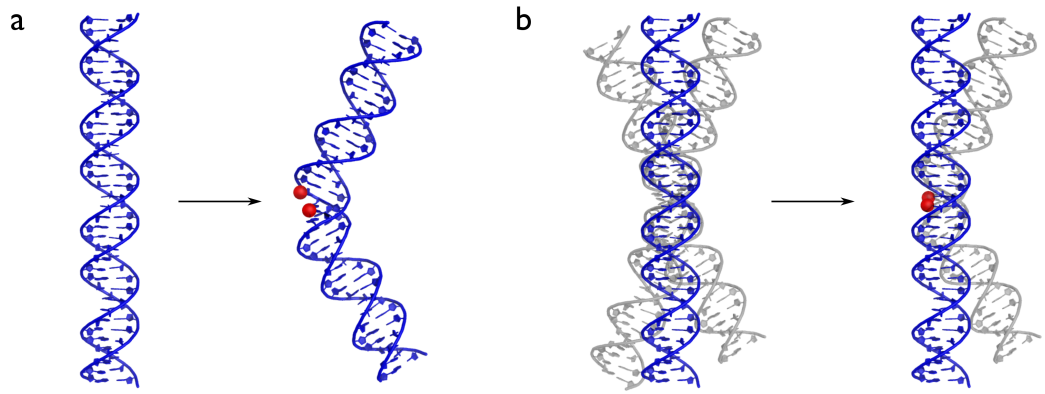


Figure 3.11. Cartoons depicting two ways by which cytosine modifications could induce bending into the minor groove. (*a*) Cytosine modifications (red spheres) could alter the lowest-energy conformer (blue duplex)—perhaps through steric repulsion of the C_5 groups—inducing a bend into the minor groove. (*b*) By preventing the DNA from adopting higher energy conformers (grey duplexes) that bend into the major groove, cytosine modifications could bias the conformational ensemble such that the averaged structure bends into the minor groove. Thus, bending could potentially occur without altering the lowest-energy conformer (blue duplex). Models built using w3DNA.[59]

Many proteins that bind DNA induce strong bending, including the histone octamer and transcription factors.[60,61] It is therefore possible that the bending of DNA induced by cytosine modification could alter such interactions and might thus be biologically

significant. As noted above, certain dinucleotide steps in the nucleosome must accept a disproportionately large share of the bending (into the major or minor groove). Therefore, the effect of cytosine modification at these sites would presumbaly be greater than at other sites. Moreover, the modification could have either a stabilising or destabilising effect, depending on whether the modification-induced bend follows or opposes the curvature of the DNA induced by the histones (Figure 1.7).[62]

Indeed, several reports indicate that the effect of the relative positions of modified CpG sites on nucleosome stability is important. But the results of major studies are controversial. For example, Choy *et al* found that CpG methylation leads to a more compact and rigid nucleosome structure, while Ngo *et al* later found the opposite.[51,63] Ngo *et al* attributed the differences in the previous study to sample contamination, but they did not point out that the two groups used different DNA sequences. These two sequences contain different numbers of CpG sites at different positions relative to the histone octamer. This is consistent with the idea that CpG methylation could either stabilise or destabilise the nucleosome, depending on their relative positions. Notably, genomics studies have also identified significant methylation at sites where the minor groove faces the histone octamer, and that this correlates with increased nucleosome occupancy.[64,65] These effects remain to be studied systematically.

3.4 Conclusions and Future Work

The DNase I cleavage and electrophoretic mobility assays both indicate that ^{m}C, ^{hm}C, and ^{f}C perturb the structure of the DNA double helix. The magnitude of these perturbations appears to be comparable with those of sequence-dependent effects. While the nature of the structural perturbations could not be determined from the DNase I experiments, gel mobility assays indicate that ^{m}C, ^{hm}C induce bending into the minor groove at CpG sites—with ^{m}C having the greatest effect. ^{f}C had a weaker and more complex influence on electrophoretic mobility, and further work is required to ensure that the observed

effects arise from axial bending.

As for the biological significance of these structural perturbations, the DNase I cleavage experiments demonstrate that ^{m}C, ^{hm}C, and ^{f}C can significantly alter DNA-protein interactions via the minor groove. But further investigation is needed to determine whether these structural changes are exploited by DNA-binding proteins *in vivo*. Of particular interest is the nucleosome, in which the intrinsic structural properties of DNA—especially bending preferences—are known to exert strong influence on its structure and stability.

In addition to the work suggested above, this investigation would benefit from the following experiments.

DNase I

Different sequence contexts around the modified cytosine should be explored. In particular, the region directly surrounding the modified CpG site, which has been shown to influence DNase I cleavage rates.[9] The effects of cytosine modification on the opposite strand should also be studied where ^{m}C is incorporated into the labelled strand. Additionally, the effect of CpA modification should be investigated.

Gel Mobility

Analysis of G_3C_3 constructs containing ^{f}C would aid the interpretation of the complex trends observed for the formylated A-tract sequences. Experiments with ^{ca}C should also be undertaken, though it should be noted that its negative charge could significantly alter electrophoretic mobility and thus may prevent reliable interpretation of the results. Exploring the effects of the sequence surrounding the modified CpG site is also recommended.

Site-directed spin labelling

Because of the indirect nature of the electrophoretic mobility analysis, quantitative measurements of the extent of bending induced by cytosine modifications could not be obtained. Site-directed spin labelling with 2′-nitroxides coupled with EPR spectroscopy could yield valuable information on the effects of cytosine modifications on DNA bending for conformational ensembles. Furthermore, the method is well-suited to studying interactions between and modified DNA and proteins—such as TDG and the histone octamer. See Chapter 4 for further details of the technique.

Atomic Force Microscopy

It is also recommended that the effect of these modifications on DNA bending be investigated by atomic force microscopy (AFM), which can directly visualise individual fragments of DNA. By altering the spacing of the modifications, such that they occur either in-phase or out-of-phase with the helical repeat, effects on curvature could be separated from isotropic flexibility. To explore this, a collaboration was recently started with the Lipfert group (Ludwig-Maximilians-Universität München).

Hydroxyl radical cleavage

The hydroxyl radical cleavage assay would complement the data acquired in this chapter and is conceptually similar to that of DNase I: a probe is used to degrade DNA, and the relative abundance of the resulting cleavage products provides information on the three-dimensional structure of the DNA duplex. But in contrast to the DNase I assay, which appears to be sensitive to a variety of structural features and flexibility, the hydroxyl radical cleavage assay depends primarily on DNA shape alone—specifically minor-groove width.[66]

Hydroxyl radical cleavage experiments were started as part of this study, using the same fluorescently labelled, 32-bp constructs used in the DNase I assay. Quantification of the

cleavage products was first attempted using PAGE. However, hydroxyl radical severely degraded the FAM fluorophore, preventing the cleavage products from being quantifiable. Attempts to optimise the conditions to find an acceptable compromise between sufficient DNA cleavage and fluorescence intensity were unsuccessful. Radiolabelling—an alternative means of quantifying DNA fragments—was then attempted by incorporating a ^{32}P-labelled phosphate to the 5′ terminus of the non-fluorescently labelled strand. Successful cleavage was achieved, but the gel images were too noisy for reliable densitometric analysis. Finally, successful hydroxyl radical cleavage of the fluorescently labelled constructs was achieved using capillary electrophoresis (CE). These experiments were undertaken during a research visit to the laboratory of Prof Loren Williams at the Georgia Institute of Technology, Atlanta, USA. As the work was started shortly before submission of this thesis, the experiments and their analysis are ongoing and could not be included here.

3.5 Summary of Part I

The central questions addressed in Part I of this thesis are: do derivatives of 5-methylcytosine alter DNA structure, and might such changes be biologically significant?

Prior to this work, crystal structures had been obtained of isolated DNA containing ^{hm}C, ^{f}C, and ^{ca}C, each in a hemimodified CpG site.[25,26,53,54] None of these modified structures differed significantly from their unmodified counterpart, suggesting that they do not significantly alter DNA structure. By contrast, a 2015 study found that ^{f}C dramatically alters DNA structure when present in consecutive, symmetrically modified CpG sites. This newly reported helical conformation, termed F-DNA, was found to be distinct from the previously known A-, B-, and Z-forms. Being unique to formylated DNA, the F-DNA conformation was proposed to form the basis for molecular recognition of ^{f}C clusters. A unique CD spectrum of formylated DNA was also found, suggesting that F-DNA also forms readily in solution. But because a crystal structure of the unmodified

sequence had not been obtained (or any other unmodified, right-handed structure containing three consecutive CpG sites) the effect of the underlying sequence was unclear. Furthermore, F-DNA had not been characterised in solution by any technique other than CD spectroscopy.

In Chapter 2, biophysical studies were performed to explore the effects of the underlying sequence, to rule out the presence of experimental artefacts, and to further characterise the F-DNA structure in solution. Crystal structures were obtained of the unmodified sequence, three unmodified sequence variants, and the identical formylated duplex crystallised under milder conditions. Additionally, the original F-DNA structure was reanalysed. These data revealed that—contrary to previous findings—there is no crystallographic evidence of a unique, ^{f}C-dependent DNA conformation. All structures, including that which reportedly adopted the F-DNA conformation, were found to adopt the A-form. Differences between the formylated sequence and its unmodified counterpart were smaller than those between different unmodified sequences, and between identical sequences crystallised under different conditions. While dramatic differences were observed in the CD spectra of formylated DNA, they were observed in any sequence context, with each different sequence producing a unique spectrum. This is incompatible with the hydrogen-bonding pattern proposed to form the structural basis of F-DNA, and it was deemed unlikely that each spectrum represents a unique helical conformation of DNA. These conclusions are consistent with UV-absorption studies, which indicate that the ^{f}C-dependent effects on CD spectra arise from perturbations to DNA's photophysical properties, rather than conformation. NMR analysis of formylated duplexes confirmed that they adopt a B-type conformation in solution and indicated that ^{f}C causes only modest conformational changes, in the vicinity of the modification sites. Overall, no evidence was found for F-DNA.

As established in Chapter 2, ^{m}C derivatives are unlikely to induce large changes to the global structure of DNA. But dramatic structural changes are not required for molecular

recognition. Subtle, sequence-dependent structural variation within the B-DNA family is known to play a key role in modulating DNA-protein interactions, such as changes in minor-groove geometry and DNA bending (see Section 1.1.4). Prior to this work, it was unclear if ^{hm}C, ^{f}C, or ^{ca}C could perturb DNA structure in a similar way. This was the subject of Chapter 3. Owing to technical limitations, such effects may not be apparent in the atomic structures obtained by X-ray crystallography and NMR (see Section 1.3). Thus, alternative, well-established techniques were used—DNase I and electrophoretic mobility assays. These had previously been used to show that ^{m}C causes structural changes to DNA that can influence DNA-protein binding,[9–12] but had not been used to study DNA containing oxidised ^{m}C derivatives. The effect ^{m}C on DNase I cleavage susceptibility was found to be in agreement with previous reports.[9–11] ^{hm}C and ^{f}C also enhanced the cleavage rate significantly, with ^{hm}C having the strongest overall effect. Curiously, ^{ca}C had no apparent influence. These findings demonstrate that both ^{hm}C and ^{f}C alter DNA conformation to an extent that can modulate DNA-protein interactions. Notably, ^{m}C and ^{hm}C also had the largest effect on DNA bending, as determined by electrophoretic mobility assays. Phasing analysis with A-tracts and the GGGCCC motif indicate that CpG sites containing either modification induce bending towards the minor groove. The case of ^{f}C was unclear and requires further investigation. The electrophoretic mobility assays also revealed conformational differences between symmetrically and asymmetrically modified CpG sites. For example, modifying the CpG site with ^{hm}C/^{m}C had a greater effect on bending than did ^{hm}C/^{hm}C. The symmetry of such modified CpG sites has received relatively little attention, but a recent report indicates that a significant proportion of ^{hm}C/^{m}C-modified CpG sites exist *in vivo*.[50]

Overall, Part I of this thesis demonstrates that ^{f}C does not cause dramatic changes to DNA structure, but it does—along with ^{hm}C—induce more modest structural perturbations. These effects appear to be of comparable magnitude to those associated with sequence-dependent variation in DNA conformation. Given that the latter plays an

important role in modulating DNA-protein interactions, the observed structural perturbations arising from hydroxymethylation and formylation could also be biologically significant.

To further characterise the effect of ^{m}C derivatives on DNA structure, several additional experiments are recommended. Hydroxyl radical cleavage studies would help elucidate the relationship between cytosine modifications and minor-groove width. This has been a useful tool in mapping the relationship between DNA sequence and minor-groove width,[66] but it has not yet been applied to ^{m}C derivatives. Further investigation into the modifications' effects on DNA bending is also recommended, since the extent of the observed bending was not quantified in this study. AFM would enable this, albeit in two dimensions; phasing analysis would enable the effects of isotropic and anisotropic flexibility, or stable curvature, to be separated. Finally, site-directed spin labelling of the DNA with 2′-nitroxides in conjunction with EPR may be an ideal technique for probing DNA bending, and would also be useful for studying DNA-protein complexes—such as the nucleosome core particle and TDG—where crystal-packing is likely to exert a large influence on DNA conformation. The validation of this method is described in Part II.

BIBLIOGRAPHY

1. Kellinger, M. W. *et al.* 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19,** 831–833 (2012).

2. Spruijt, C. G. *et al.* Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* **152,** 1146–1159 (2013).

3. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14,** R119 (2013).

4. Wang, L. *et al.* Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature* **523,** 621–625 (2015).

5. Wu, H. M. & Crothers, D. M. The locus of sequence-directed and protein-induced DNA bending. *Nature* **308,** 509–13 (1984).

6. Hagerman, P. J. Sequence-Directed Curvature of DNA. *Annu. Rev. Biochem.* **59,** 755–781 (1990).

7. Brukner, I. *et al.* Evidence for opposite groove-directed curvature of GGGCCC and AAAAA sequence elements. *Nucleic Acids Res.* **21,** 1332–1332 (1993).

8. Brukner, I., Susic, S., Dlakic, M., Savic, A. & Pongor, S. Physiological Concentration of Magnesium Ions Induces a Strong Macroscopic Curvature in GGGCCC-containing DNA. *J. Mol. Biol.* **236,** 26–32 (1994).

9. Lazarovici, A. *et al.* Probing DNA shape and methylation state on a genomic scale with DNase I. *Proc. Natl. Acad. Sci.* **110,** 6376–6381 (2013).

10. Kochanek, S., Renz, D. & Doerfler, W. Differences in the accessibility of methylated and unmethylated DNA to DNase I. *Nucleic Acids Res.* **21,** 5843–5845 (1993).

11. Fox, K. R. The effect of Hha I methylation on DNA local structure. *Biochem. J.* **234,** 213–216 (1986).

12. Hodges-Garcia, Y. & Hagerman, P. J. Cytosine methylation can induce local distortions in the structure of duplex DNA. *Biochemistry* **31,** 7595–7599 (1992).

13. Suck, D. DNA Recognition by DNase I. *J. Mol. Recognit.* **7,** 65–70 (1994).

14. Hampshire, A. J., Rusling, D. A., Broughton-Head, V. J. & Fox, K. R. Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. *Methods* **42,** 128–40 (2007).

15. Brukner, I., Jurukovski, V. & Savic, A. Sequence-dependent structural variations of DNA revealed by DNase I. *Nucleic Acids Res.* **18,** 891–894 (1990).

16. Heddi, B., Abi-Ghanem, J., Lavigne, M. & Hartmann, B. Sequence-Dependent DNA Flexibility Mediates DNase I Cleavage. *J. Mol. Biol.* **395,** 123–133 (2010).

17. Drew, H. R. & Travers, A. A. DNA structural variations in the E. coli tyrT promoter. *Cell* **37,** 491–502 (1984).

18. Hogan, M. E., Robersont, M. W. & Austint, R. H. DNA flexibility variation may dominate DNase I cleavage. *Proc. Natl. Acad. Sci.* **86,** 9273–9277 (1989).

19. Brukner, I., Sánchez, R., Suck, D. & Pongor, S. Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. *EMBO J.* **14,** 1812–1818 (1995).

20. Leblanc, B. & Moss, T. in *Methods Mol. Biol.* 31–38 (Humana Press, Totowa, NJ, 2001).

21. Fox, K. R. & Waring, M. J. DNA structural variations produced by actinomycin and distamycin as revealed by DNAase I footprinting. *Nucleic Acids Res.* **12,** 9271–9285 (1984).

22. Jain, S. S. & Tullius, T. D. Footprinting protein–DNA complexes using the hydroxyl radical. *Nat. Protoc.* **3,** 1092–1100 (2008).

23. Spana, C. & Corces, V. G. DNA bending is a determinant of binding specificity for a Drosophila zinc finger protein. *Genes Dev.* **4,** 1505–1515 (1990).

24. Leblanc, B. & Moss, T. in *Methods Mol. Biol.* 37–47 (Humana Press, Totowa, NJ, 2009).

25. Hardwick, J. S., Lane, A. N. & Brown, T. Epigenetic Modifications of Cytosine: Biophysical Properties, Regulation, and Function in Mammalian DNA. *BioEssays,* 1700199 (2018).

26. Szulik, M. W. *et al.* Differential Stabilities and Sequence-Dependent Base Pair Opening Dynamics of Watson–Crick Base Pairs with 5-Hydroxymethylcytosine, 5-Formylcytosine, or 5-Carboxylcytosine. *Biochemistry* **54,** 1294–1305 (2015).

27. Sumino, M., Ohkubo, A., Taguchi, H., Seio, K. & Sekine, M. Synthesis and properties of oligodeoxynucleotides containing 5-carboxy-2'-deoxycytidines. *Bioorganic Med. Chem. Lett.* **18,** 274–277 (2008).

28. Lomonossoff, G., Butler, P. & Klug, A. Sequence-dependent variation in the conformation of DNA. *J. Mol. Biol.* **149,** 745–760 (1981).

29. Herrera, J. E. & Chaires, J. B. Characterization of Preferred Deoxyribonuclease I Cleavage Sites. *J. Mol. Biol.* **236,** 405–411 (1994).

30. Haran, T. E. & Mohanty, U. The unique structure of A-tracts and intrinsic DNA bending. *Q. Rev. Biophys.* **42,** 41–81 (2009).

31. Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. Bent helical structure in kinetoplast DNA. *Proc. Natl. Acad. Sci.* **79,** 7664–7668 (1982).

32. Crothers, D. M., Gartenberg, M. R. & Shrader, T. E. DNA bending in protein-DNA complexes. *Methods Enzymol.* **208,** 118–146 (1991).

33. Stellwagen, N. C. & Stellwagen, E. Effect of the matrix on DNA electrophoretic mobility. *J. Chromatogr. A* **1216,** 1917–1929 (2009).

34. Merling, A., Sagaydakova, N. & Haran, T. E. A-tract polarity dominate the curvature in flanking sequences. *Biochemistry* **42,** 4978–4984 (2003).

35. Hagerman, P. J. Sequence-directed curvature of DNA. *Nature* **321,** 449–450 (1986).

36. Hagerman, P. J. Pyrimidine 5-Methyl Groups Influence the Magnitude of DNA Curvature. *Biochemistry* **29,** 1980–1983 (1990).

37. Zinkel, S. S. & Crothers, D. M. DNA bend direction by phase sensitive detection. *Nature* **328,** 178–181 (1987).

38. Thompson, J. F. & Landy, A. Empirical estimation of protein-induced DNA bending angles: applications to λ site-specific recombination complexes. *Nucleic Acids Res.* **16,** 9687–9705 (1988).

39. Kerppola, T. K. & Curran, T. Fos-Jun heterodimers and jun homodimers bend DNA in opposite orientations: Implications for transcription factor cooperativity. *Cell* **66,** 317–326 (1991).

40. Kim, J., Klooster, S. & Shapiro, D. J. Intrinsically Bent DNA in a Eukaryotic Transcription Factor Recognition Sequence Potentiates Transcription Activation. *J. Biol. Chem.* **270,** 1282–1288 (1995).

41. Crothers, D. M. & Drak, J. Global features of DNA structure by comparative gel electrophoresis. *Methods Enzymol.* **212,** 46–71 (1992).

42. Hagerman, P. J. Pyrimidine 5-methyl groups influence the magnitude of DNA curvature. *Biochemistry* **29,** 1980–1983 (1990).

43. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6–21 (2002).

44. Ulanovsky, L., Bodner, M., Trifonov, E. N. & Choder, M. Curved DNA: design, synthesis, and circularization. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 862–6 (1986).

45. Guo, J. U. *et al.* Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* **17,** 215–222 (2014).

46. Jang, H. S., Shin, W. J., Lee, J. E. & Do, J. T. CpG and non-CpG methylation in epigenetic gene regulation and brain function. *Genes (Basel).* **8,** 2–20 (2017).

47. Ziller, M. J. *et al.* Genomic distribution and Inter-Sample variation of Non-CpG methylation across human cell types. *PLoS Genet.* **7** (2011).

48. Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* **14,** 341–356 (2013).

49. Hahn, M. A., Szabó, P. E. & Pfeifer, G. P. 5-Hydroxymethylcytosine: A stable or transient DNA modification? *Genomics* **104,** 314–323 (2014).

50. Song, C.-X., Diao, J., Brunger, A. T. & Quake, S. R. Simultaneous single-molecule epigenetic imaging of DNA methylation and hydroxymethylation. *Proc. Natl. Acad. Sci.* **113,** 4338–4343 (2016).

51. Ngo, T. T. M. *et al.* Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **7,** 10813 (2016).

52. Wanunu, M. *et al.* Discrimination of Methylcytosine from Hydroxymethylcytosine in DNA Molecules. *J. Am. Chem. Soc.* **133,** 486–492 (2011).

53. Lercher, L. *et al.* Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem. Commun.* **50,** 1794–1796 (2014).

54. Renciuk, D., Blacque, O., Vorlickova, M. & Spingler, B. Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Res.* **41,** 9891–9900 (2013).

55. Raiber, E.-A. *et al.* 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22,** 44–49 (2014).

56. Dai, Q. *et al.* Weakened N3 Hydrogen Bonding by 5-Formylcytosine and 5-Carboxylcytosine Reduces Their Base-Pairing Stability. *ACS Chem. Biol.* **11,** 470–477 (2016).

57. Dlakic, M., Park, K., Griffith, J. D., Harvey, S. C. & Harrington, R. E. The Organic Crystallizing Agent 2-Methyl-2,4-pentanediol Reduces DNA Curvature by Means of Structural Changes in A-tracts. *J. Biol. Chem.* **271,** 17911–17919 (1996).

58. DiGabriele, A. D., Sanderson, M. R. & Steitz, T. A. Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract. eng. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 1816–1820 (1989).

59. Zheng, G., jun Lu, X. & Olson, W. K. Web 3DNA - A web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.* **37,** 240–246 (2009).

60. Rohs, R. *et al.* The role of DNA shape in protein-DNA recognition. *Nature* **461,** 1248–1253 (2009).

61. Van der Vliet, P. C. & Verrijzer, C. P. Bending of DNA by transcription factors. *BioEssays* **15,** 25–32 (1993).

62. Jimenez-Useche, I. *et al.* DNA Methylation Regulated Nucleosome Dynamics. *Sci. Rep.* **3,** 2121 (2013).

63. Choy, J. S. *et al.* DNA methylation increases nucleosome compaction and rigidity. *J. Am. Chem. Soc.* **132,** 1782–1783 (2010).

64. Chodavarapu, R. K. *et al.* Relationship between nucleosome positioning and DNA methylation. *Nature* **466,** 388–392 (2010).

65. Collings, C. K., Waddell, P. J. & Anderson, J. N. Effects of DNA methylation on nucleosome stability. *Nucleic Acids Res.* **41,** 2918–2931 (2013).

66. Bishop, E. P. *et al.* Map of minor groove shape and electrostatic potential from hydroxyl radical cleavage patterns of DNA. *ACS Chem. Biol.* **6,** 1314–1320 (2011).

Part II.

Structural studies of DNA containing spin labels

4

CALIBRATING A MOLECULAR RULER: STRUCTURAL STUDIES OF SPIN-LABELLED DNA

4.1 Introduction

Understanding the structure of DNA is crucial to elucidating its function at the molecular level. Part I highlights the central role of structural techniques in decoding the biological activity of ^{m}C and its oxidised derivatives, but it also illustrates the challenges of using techniques such as X-ray crystallography and NMR spectroscopy to reliably determine certain structural features of DNA, such as bending.

As outlined in Section 1.3, a molecular ruler approach involving site-directed spin labelling (SDSL) in conjunction with electron paramagnetic resonance (EPR) spectroscopy has shown promise in addressing these limitations, enabling intrahelical distance distributions to be measured in the range of ~2–10 nm.

SDSL requires a judicious choice of both the spin label, and the chemical strategy for its incorporation. The radical itself must be chemically resistant to common biological buffer components, particularly reducing agents.[1] Heterocyclic nitroxides have been

developed to fulfil these requirements and are thus the most widely used free radicals in biological EPR experiments.[1,2] As for the labelling strategy, the ideal spin label would have several key properties. First, it would not perturb the structure of the biomolecule to which it is bound.[2] Second, it would be rigidly bound to the biomolecule—to maximise the precision of the measurement.[2] Third, it would also be fully sequence-flexible, *ie* any site within any sequence could be labelled.[3] Finally, its incorporation would be compatible with standard solid-phase oligonucleotide synthesis.

In efforts to meet these stringent demands, a range of different labelling strategies have been developed, involving modifications to the phosphodiester,[4,5] the sugar,[6] and the nucleobase.[7–10] For example, by incorporating the spin label into a cytosine-like nucleobase, the nitroxide group can be rigidly bound, affording distance measurements of high precision (Figure 4.1*a*).[10] However, this places strict limitations on the sites that can be spin labelled. And while sequence-flexible approaches have been reported, such as tethering the nitroxide to the sugar's 2′ position via a propargyl ether-derived triazole linker (Figure 4.1*b*), the high flexibility of the linkers in such systems typically increases the spin label's mobility, reducing the precision of distance measurements.[3,4,6]

In 2016, Haugland *et al* reported a novel strategy for the site-specific spin-labelling of DNA, in which 2′-alkynylnucleotides are incorporated into DNA using standard solid-phase oligonucleotide synthesis and subsequently spin-labelled in high yield using click chemistry (Figure 4.1*c*). The method is sequence-flexible, yet enables measurements of sub-nanometre precision.[3] Furthermore, since the oligonucleotides are labelled post-synthetically, different spin labels—or any other azide-functionalised probes—can be incorporated according to the needs of the experiment. The authors demonstrated this by testing 6 different nitroxide spin labels: the well-established 6-Me and 5-Me labels (Figure 4.1) and other variants.[3]

In DEER experiments—and any other molecular ruler approach—once an interprobe distance, or distance distribution, has been obtained, it is then interpreted with reference

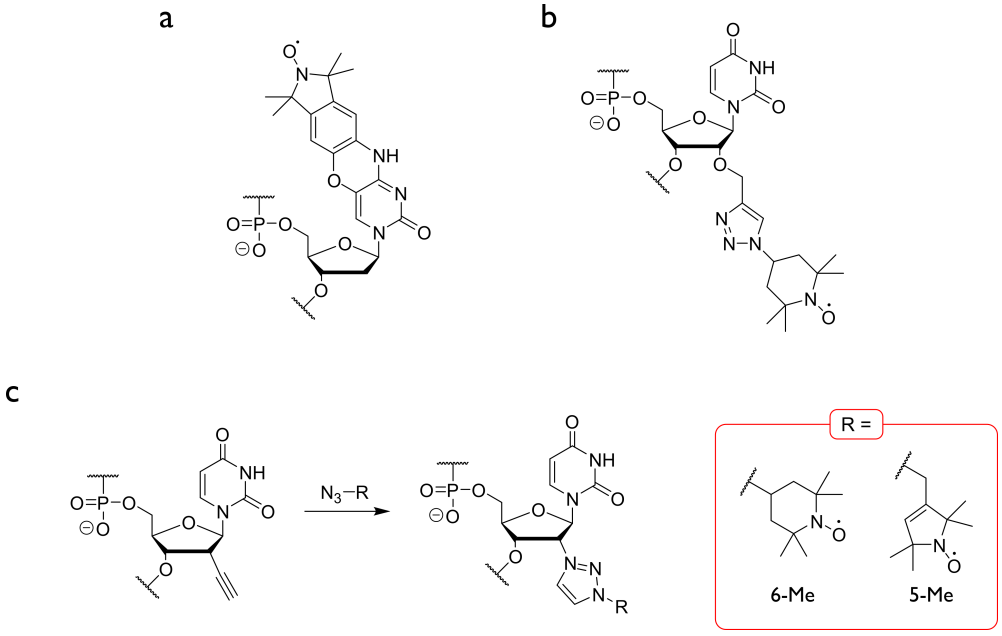


Figure 4.1. Different strategies for incorporating spin labels into DNA. (*a*) Incorporating the spin label into a cytosine-like base leads to precise measurements but is not sequence-flexible.[10] (*b*) Linking the spin label to the sugar via a propargyl linker is compatible with any base, but the linker flexibility in such systems significantly reduces measurement precision.[3,4,6] (*c*) Sequence-flexible spin-labelling strategy of Haugland *et al*, enabling distance measurements with sub-nanometre precision.[3] The DNA characterised in this study was spin-labelled with 6-Me or 5-Me using this approach.

to a structural model. For accurate, quantitative interpretation of the data, this model must account for each spin label's preferred location and orientation,[11] as well as any perturbation to DNA structure.[12] But no B-form crystal structures of spin-labelled DNA have been obtained thus far, and few crystal structures exist of any biomolecules covalently bound to spectroscopic probes. Thus, researchers have typically relied on indirect methods to characterise their effects on DNA conformation.[12]

In light of the above, we sought to pinpoint the positions of 6-Me and 5-Me spin labels introduced via the strategy of Haugland *et al*, and to characterise their effects on DNA structure and dynamics. We obtained detailed information on the spin labels' location and their effects on B-DNA structure using X-ray crystallography, then probed the dynamics

of the spin labels and their effect on the conformational ensemble using molecular dynamics (MD) simulations and NMR spectroscopy. Together, these results enable the accurate interpretation of EPR data determined using these spin labels, establishing their use as a minimally perturbing tool to study DNA conformation.

4.2 Results & Discussion

4.2.1 DNA synthesis

For crystallisation trials, EPR, and MD simulations, a variant of the Dickerson–Drew dodecamer sequence was chosen (Figure 4.2*a*). The unmodified, self-complementary sequence had previously been shown to crystallise in the B-form, resulting in a high-resolution structure (PDB 1s2r, 1.5 Å).[13] Position 9 was selected for the site of spin-labelling to allow adequate separation of the two spin labels, while ensuring that they were sufficiently far from the duplex termini to minimise end effects. For reasons discussed in Section 4.2.5, a different sequence was used for NMR analysis (Figure 4.8), whose unmodified form has been well characterised in previous work[14] (see Table C.1 for sequences and mass spectrometry data of all spin-labelled oligonucleotides).

4.2.2 EPR spectroscopy

Using DEER experiments, interspin distance distributions were determined for the 6-Me- and 5-Me-labelled dodecamer duplexes (Figure 4.2). The distributions were narrow in both cases, each with a full width at half maximum (FWHM) of < 0.5 nm. But that of 6-Me (FWHM 0.37 nm) was narrower than that of 5-Me (FWHM 0.48 nm). The distribution of the 6-Me labels was near symmetric, indicating the label's strong preference for a single conformation with a most probable interspin distance, *r*, of 2.99 nm. That of the 5-Me spin labels was less symmetric (r = 3.12 nm), suggesting that the label adopts a wider range of conformations than does 6-Me.

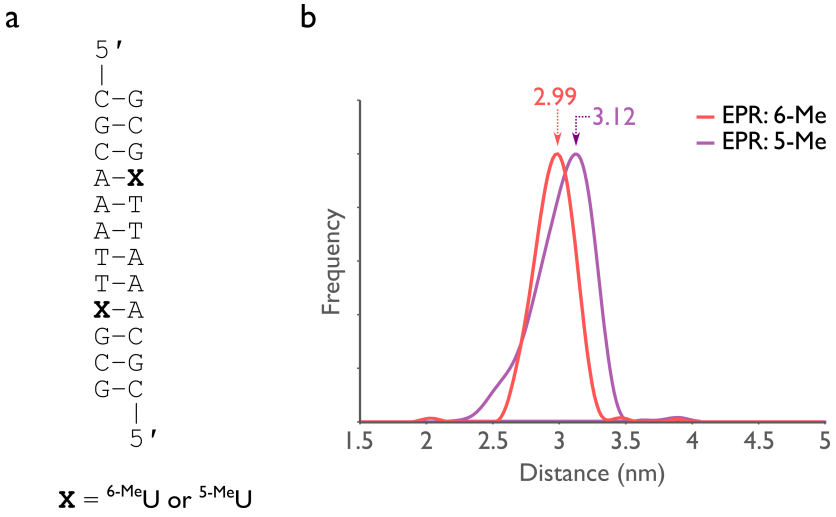


Figure 4.2. EPR-derived distance distributions of 6-Me- and 5-Me-spin-labelled DNA. (*a*) The self-complementary sequence. (*b*) DEER distance distributions of the 6-Me- (red) and 5-Me- (lilac) labelled duplexes; the most probable interspin distances were 2.99 and 3.12 nm, respectively.

4.2.3 X-ray crystallography

Following determination of the distance distributions by EPR, X-ray crystallography was used to investigate the effect of spin labelling on DNA structure and elucidate the spin labels' conformational preferences.

Crystal structure of a DNA duplex spin-labelled with 6-Me

A structure of the same 6-Me-labelled duplex studied by EPR was determined at 1.8 Å resolution. Notably, the duplex crystallised in the B-form. The asymmetric unit contains a single duplex (space group $P3_221$, see Table 4.1 for data collection, processing, and refinement statistics, Table C.3 for twinning statistics, and Table C.2 for crystallisation conditions).

The electron density of both 6-Me spin labels is well defined (Figure 4.3*c*). This was somewhat surprising, since the spin labels could, in principle, rotate freely about two bonds. Contacts formed between the triazole group of each label and a cobalt(III)

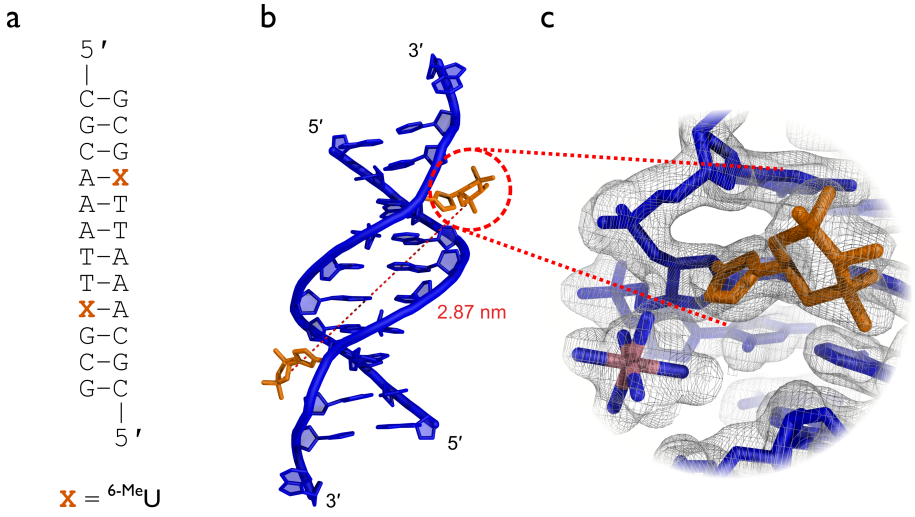


Figure 4.3. B-DNA crystal structure of the 6-Me-labelled duplex (PDB 6qjs). (*a*) The self-complementary sequence. (*b*) Side-view of the structure, showing the two 6-Me spin labels in the minor groove. The interspin distance is 2.87 nm. (*c*) Close-up of one of the spin labels, showing the $2F_o–F_c$ electron density map, contoured at 1.0 σ. The density of both spin labels is well defined. Adapted from Hardwick *et al.*[15]

hexammine ion likely contribute to their lack of mobility (Figure C.2*a*). Both labels reside in the minor groove, and their radical centres are separated by 2.87 nm. This agrees well with the most probable interspin separation measured by EPR (2.99 nm), suggesting that the crystal structure and the major conformer observed by EPR are similar. All other residues have well-defined electron density (Figure C.1*a*), except for the 2′-deoxycytidines at the 5′ termini, whose base pairs are broken in favour of forming crystal contacts with cobalt ions and neighbouring duplexes (Figure C.2*b*). These interactions arise from crystal-packing artefacts rather than from the presence of the spin labels, as evidenced by the occurrence of similar contacts in several other DNA crystal structures featuring cobalt ions and terminal CpG dinucleotides,[16–23] and two intact base pairs that separate the spin labels from these unpaired residues (Figure 4.3).

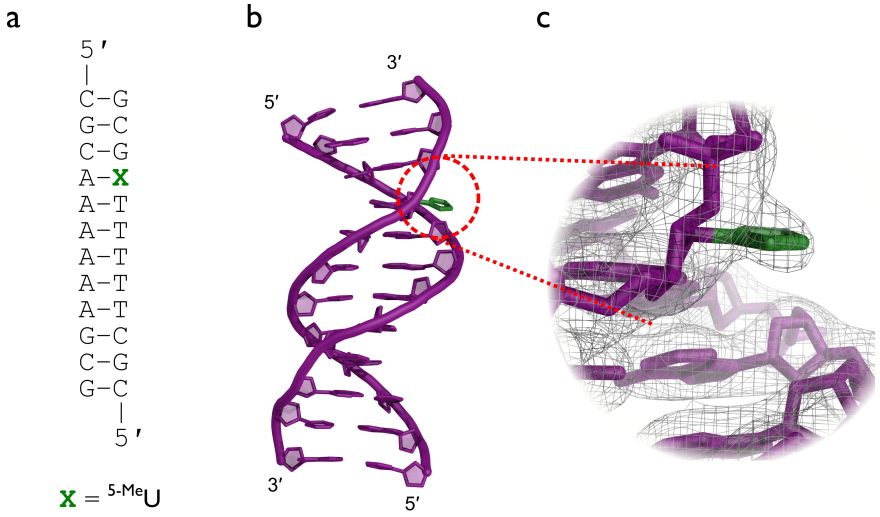


Figure 4.4. B-DNA crystal structure of the 5-Me spin-labelled duplex (PDB 6qjr). (*a*) The sequence, which contains a single spin label. (*b*) Side-view of the structure. (*c*) Close-up of the triazole group, showing the $2F_o$–F_c electron density map, contoured at 1.0 σ. Only the trazole group of the spin label is visible in the density map, which is likely due to its rotational disorder in the crystal. Adapted from Hardwick *et al.*[15]

Crystal structure of a DNA duplex spin-labelled with 5-Me

Attempts to determine the crystal structure of the corresponding 5-Me labelled duplex were unsuccessful due to poorly diffracting crystals. But the structure of a similar, singly labelled sequence was determined at 2.9-Å resolution (Figure 4.4, Table 4.1). This duplex also crystallised in the B-form (see Table C.2 for conditions). The asymmetric unit contains a single duplex (space group $P2_12_12_1$), and the electron density is well defined for all residues except the 5-Me spin label—for which only density of the triazole group can be seen (Figure 4.4*c*, Figure C.1*b*). Its absence from the electron density map is probably due to rotational disorder arising from the flexibility of the linker.

Comparison of spin-labelled duplexes with their unmodified counterparts

The global structures of the 6-Me and 5-Me spin-labelled duplexes are similar to their unmodified counterparts (PDBs 1s2r and 1d98, Figure 4.5);[13,24] all atom RMSDs are

Table 4.1. Data collection and refinement statistics of PDBs 6qjs and 6qjr. Each dataset was collected from a single crystal; values in parentheses are for the highest-resolution shell.

	6qjs (6-Me)	**6qjr (5-Me)**
Data collection		
Space group	$P3_221$	$P2_12_12_1$
Cell dimensions:		
a, *b*, *c* (Å)	43.0, 43.0, 97.5	25.3, 41.2, 67.0
α, β, γ (°)	90, 90, 120	90, 90, 90
Resolution (Å)	37.2–1.80 (1.84–1.80)	35.1–2.90 (3.10–2.90)
R_{meas} (%)	9.2 (73.9)	5.7 (18.4)
$I/\sigma(I)$	9.3 (2.6)	10.3 (2.7)
$CC_{1/2}$	0.994 (0.684)	0.999 (0.998)
Completeness (%)	97.6 (95.1)	99.0 (99.6)
Redundancy	3.8 (3.5)	3.3 (3.6)
Refinement		
Resolution (Å)	37.2–1.8	35.1–2.9
Reflections	9.365	1,621
R_{work}/R_{free} (%)	21.3 / 25.1	18.8 / 24.4
Atoms:		
DNA	484	490
$[Co(NH_3)]^{3+}$	4	0
Water	7	0
B factors:		
DNA	35.5	79.0
Water	30.9	–
RMS deviations:		
Bond lengths (Å)	0.014	0.008
Bond angles (°)	1.67	1.16

1.08 Å (402 atoms) and 0.70 Å (403 atoms), respectively.* Because of differences in crystal packing between the spin-labelled duplexes and their unmodified analogues, the effects from crystallisation cannot be entirely separated from those of spin labelling. However, their RMSDs lie within the range of identical duplexes in different crystalline environments (see Chapter 2).[25] This indicates that, at least in the crystalline state, these spin labels have only a minor effect on global structure.

There is, however, a notable difference in minor-groove width around position 9, where one of the 6-Me spin labels is located in structure 6qjs. Here, the minor-groove width of the 6-Me structure is 3.8 Å greater than that of its unmodified counterpart (Figures 4.5*a* and C.4). This could have arisen either from the presence of the spin label or from differences in crystal packing. However, since the AAATTT tract in the sequence gives rise to an unusually narrow minor groove,[13] any such effect of minor-groove widening would probably be less significant for other sequences.

Some perturbation to local structure was also observed at the 6-Me- and 5-Me-labelled nucleotides. For example, the sugar puckers and glycosyl torsion angles more closely resembled those of A-DNA than of B-DNA (Figure 4.5*c*–*f*, Tables C.4–C.6). That the values of all spin-labelled uridines were similar—in spite of their differing crystalline environments—indicates that they reflect a major conformer instead of crystallisation artefacts.

Perturbation of the modified sugar puckers and glycosyl torsion angles by the spin labels was not surprising, since even single-atom modifications can exert a large effect on puckering.[27] At the time of writing, there are two other B-DNA structures in the PDB containing 2′-substituted *ribo*nucleotides. Between them, the structures contain 10 symmetrically unique modified residues, all of which adopt a narrow range of A-DNA-like puckers (C3′-*endo*, P = 4–14°, Figure 4.6). By contrast, 2′-substituted *arabino*nucleotides are more characteristic of B-DNA, clustering into two groups of P

*In both cases, the terminal base pairs were omitted to minimise crystal-packing artefacts.

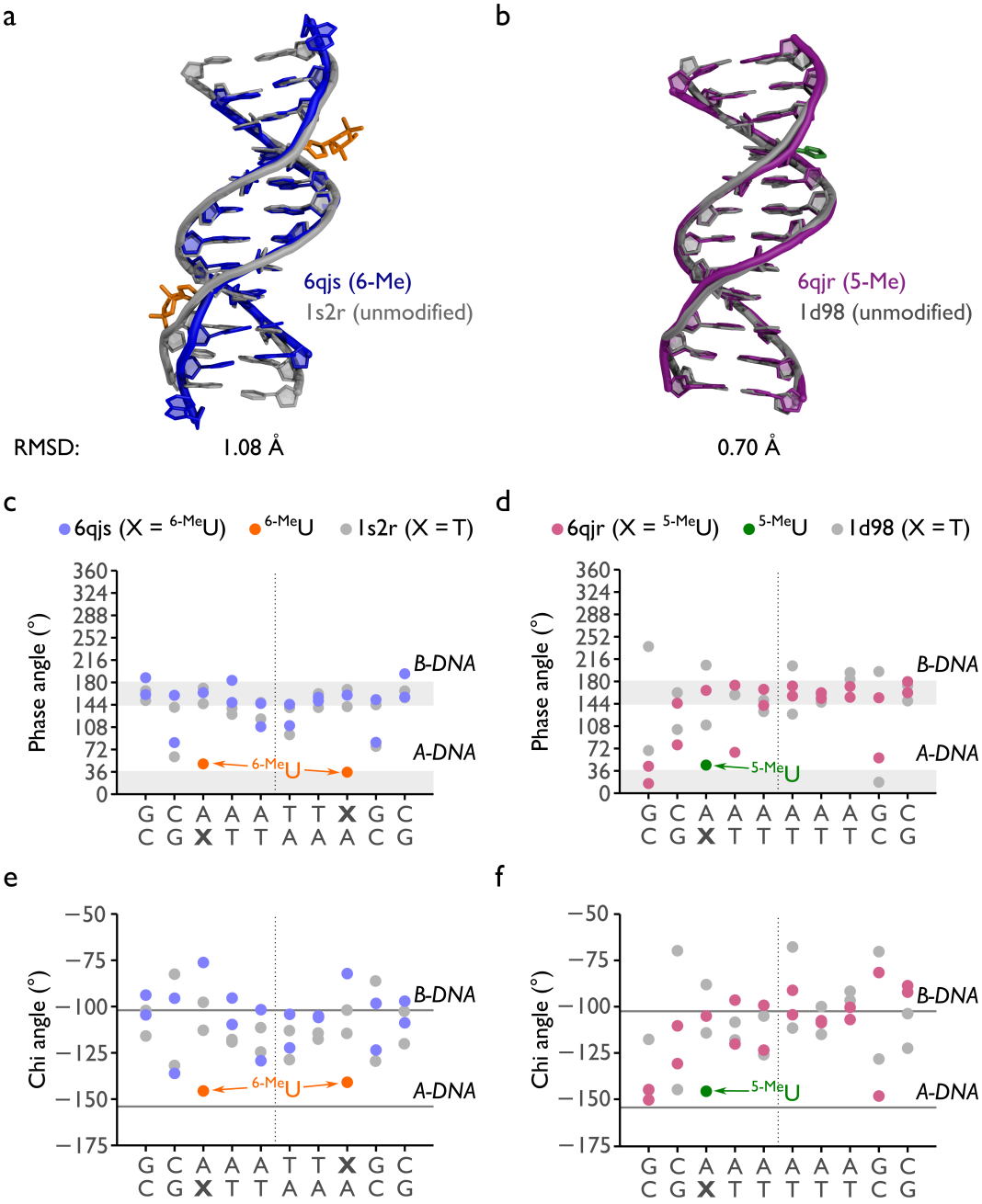


Figure 4.5. Crystal structures of spin-labelled duplexes overlaid with their unmodified counterparts. (*a*) 6-Me-labelled structure (blue, PDB 6qjs) and its unmodified variant (grey, PDB 1s2r). The 6-Me labels are shown in orange. (*b*) The 5-Me-labelled structure (purple, PDB 6qjr) and its unmodified analogue (grey, PDB 1d98). The triazole group of the 5-Me spin label is shown in green. (*c,d*) Pseudorotation phase angles and glycosyl torsion (chi) angles (*e,f*) of all non-terminal nucleotides in the four structures, shown alongside typical A- and B-DNA values.[26] The values of the 3 modified residues are similar and are characteristic of A-DNA. The plots illustrate that these perturbations do not propagate along the duplexes and are comparable to those arising from sequence-dependent structural variation or crystal-packing effects. *a* and *b* adapted from Hardwick *et al.*[15]

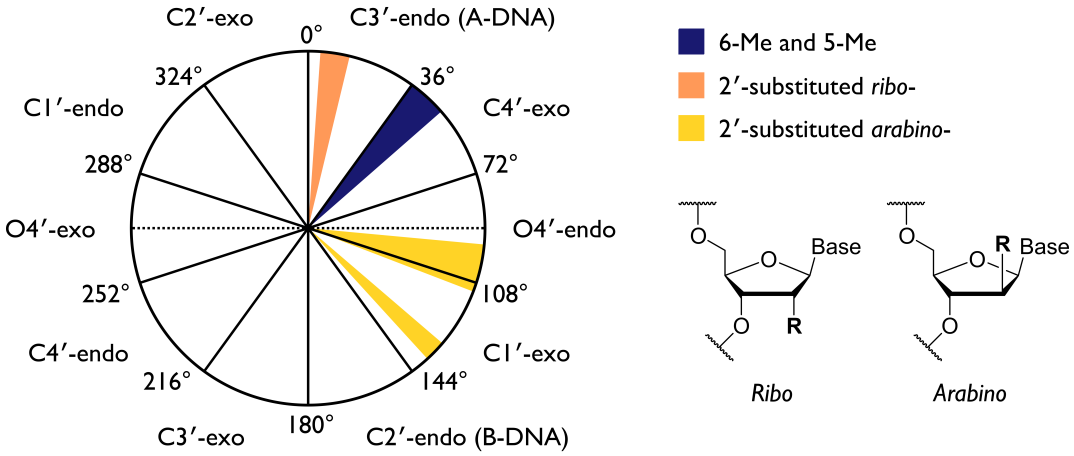


Figure 4.6. A comparison of the puckering of the 2′-spin-labelled uridines (PDBs 6qjs, 6qjr) with 2′-substituted ribo- and arabinonucleotides (10 and 20 unique nucleotides, respectively, from 11 crystal structures). Adapted from Hardwick *et al.*[15]

= 95–111 and 131–138° (O4′-*endo* and C1′-*exo*, 9 structures, 20 unique modified residues, Figure 4.6).* As expected, the values of the spin-labelled uridines are closer to those of other 2′-substitued ribonucleotides than to those of the arabino configuration. Notably, these local deviations from typical B-DNA geometry are confined to the modified sites and their nearest neighbours—they do not propagate throughout the B-form duplex.

Altogether, the crystal structures indicate that this spin-labelling strategy is well-suited to EPR-based studies of B-DNA: in the crystalline state, neither spin label causes large changes to global structure. Although local effects were observed, these are confined to the modified residues and their nearest neighbours.

4.2.4 Molecular dynamics simulations

The X-ray crystal structures represent conformational snapshots in the crystalline state. To probe the behaviour and dynamics of spin-labelled DNA in solution, duplexes containing

*PDB codes of the crystal structures containing the modified ribonucleotides are 3ey3 and 2dpc.[28,29] Those of the arabinonucleotides are 4kwo, 2fih, 461d, 436d, 1dpn, 478d, 460d, 389d, 388d.[27,30–34]

Table 4.2. Most probable interspin distances derived from MD simulations, X-ray crystallography, and EPR.

	Interspin distance, *r* (nm)	
	6-Me	5-Me
MD 'antiparallel'	3.33	2.78
MD 'minor groove'	2.88	2.93
X-ray structure	2.87	–
EPR	2.99	3.12

the 6-Me and 5-Me labels were studied using molecular dynamics (MD) simulations.

MD simulations were performed on the sequence d(CGCAAATT^{X}UGCG)$_2$, where X = 6-Me or 5-Me, for a duration of 1,000 ns. For both spin labels, two major duplex conformations were observed: one in which the spin labels point into the minor groove ('minor groove'), and another in which the spin labels align antiparallel to the DNA ('antiparallel'). The most probable interspin distances, *r*, for the two conformers are given in Table 4.2.

MD simulations of the 6-Me duplex

The *r*-values of the 'minor groove' and 'antiparallel' conformers differ significantly (Table 4.2, Δr = 0.45 nm). The *r*-value of the 6-Me duplex obtained by EPR (2.99 nm) agrees well with that of the 'minor groove' configuration. By contrast, there appears to be no significant contribution from the 'antiparallel' conformer, suggesting that it is not a major component of the real ensemble. This is also supported by X-ray crystallographic studies, which provide no evidence of the 'antiparallel' state.

In contrast to the 'antiparallel' duplex, the MD-derived *r*-value of the 6-Me 'minor groove' conformer was in excellent agreement with the crystal structure (Table 4.2), differing by only 0.1 Å, suggesting that they are conformationally similar. This was

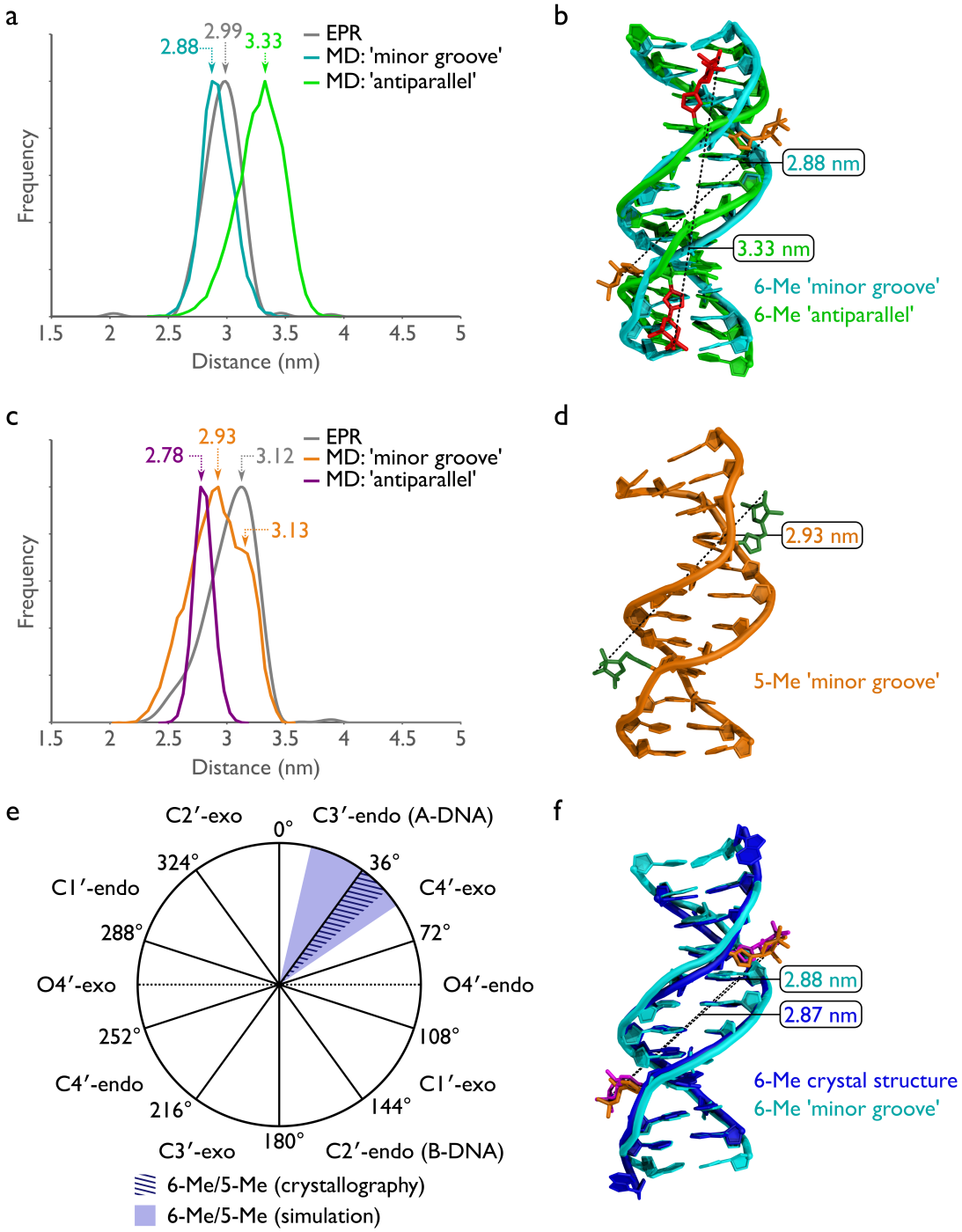


Figure 4.7. Molecular dynamics simulations of spin-labelled duplexes. (*a*) Simulated distance distributions of the 'minor groove' (turquoise) and 'antiparallel' (green) states of the 6-Me-labelled duplex, an overlay of which is shown in (*b*). (*c*) Corresponding distributions of the 5-Me-labelled duplex and a representative structure of the 'minor groove' state. (*e*) Pseudorotation phase angles of the spin-labelled sugars derived from simulations (solid blue) and X-ray crystallography (dark blue lines). (*f*) Overlay of a simulated snapshot of the 6-Me 'minor groove' conformation and crystal structure PDB 6qjs. Adapted from Hardwick *et al.*[15]

confirmed by superimposing a snapshot from the simulation with the crystal structure (Figure 4.7, RMSD of 1.08 Å, 402 atoms*). Local structural features are also in good agreement. For example, the puckers of the spin-labelled nucleotides in the 'minor groove' snapshot are C4′-*exo* and C3′-*endo*, matching exactly those observed in the crystal structure. Glycosyl torsion angles also agree well (molecular dynamics: −145 and −145°; X-ray crystallography: −141 and −146°). And as observed by crystallography, these perturbations are local to the modification sites—the average puckering of unmodified residues throughout the ensemble is C2′-*endo*, typical of B-DNA.[26]

MD simulations of the 5-Me duplex

As with 6-Me, 'minor groove' and 'antiparallel' conformations were also observed for the 5-Me spin label, though in this case the most probable interspin distances of the two states were more similar (Δr = 0.15 nm, Table 4.2). And, like 6-Me, the 'antiparallel' state did not appear to contribute to the real system, with the 'minor groove' state more closely resembling the EPR-derived distance distribution (Figure 4.7, Table 4.2). However, the simulated 'minor groove' distribution of 5-Me was much broader than that of 6-Me and contained several shoulders, reflecting the 5-Me label's greater flexibility (Figure 4.7). This echoes the broader and less symmetric distribution of the 5-Me labels observed by EPR (FWHM 0.48 and 0.37 nm for 5-Me and 6-Me, respectively).

Because the crystal structure of the 5-Me duplex contains a single spin label, an interspin distance could not be measured. Furthermore, the differing sequences of the simulated and crystallised constructs prevent a meaningful comparison of the dodecamer duplexes' global structure. However, the effects of spin labelling on local structural features determined by the two techniques could be compared, and were in good agreement. Modified residues from a simulated snapshot of the 'minor groove' conformer adopted the C3′-*endo* pucker (P = 19 and 29°, Figure 4.7*e*), similar to that of the crystal structure

*The terminal nucleotides and spin labels were excluded.

(C4′-*exo*, P = 45°). The simulated glycosyl torsion angles (−139 and −144°) are also close to that of the crystal structure (−146°).

Overall, MD simulations were in good agreement with EPR and crystallographic data in studying the behaviour of the spin labels, particularly for the more rigid 6-Me label—where the most probable distance determined by all three techniques agreed to within 1.2 Å. The simulated effects of spin labelling on local DNA structure were also in excellent agreement with results from X-ray crystallography, bolstering confidence that both techniques accurately represent the real system.

4.2.5 NMR spectroscopy

To determine experimentally the effects of spin labelling on the conformational ensemble, spin-labelled duplexes were characterised using solution-phase NMR spectroscopy. Nitroxide spin labels cause severe line broadening in NMR spectra for nuclei within a radius of ~12 Å (~3.5 bp).[35] Consequently, the doubly labelled dodecamer sequence studied by EPR and X-ray crystallography would exhibit few NMR signals. An alternative, singly labelled 14-bp duplex was therefore chosen, whose unmodified form had previously been well characterised by NMR.[14]

NMR spectra of the 5-Me and 6-Me labelled duplexes were acquired in both their paramagnetic, and diamagnetic states—following reduction of the nitroxides to hydroxylamines. This enabled the resonances of all nucleotides to be assigned (Figure 4.8, resonances given in Tables C.7 and C.8). Differences between the measurable shifts of the paramagnetic states and those of the diamagnetic state are negligible for both spin-labelled duplexes (Tables C.9 and C.10), as are the scalar coupling and NOE cross-peak patterns, indicating that the two states are conformationally identical across the measurable region (C7–G14 and C15–C23).

Differences in chemical shift are also negligible between the 5-Me and 6-Me duplexes, indicating that the two conformations are highly similar. This is true also for the spin-

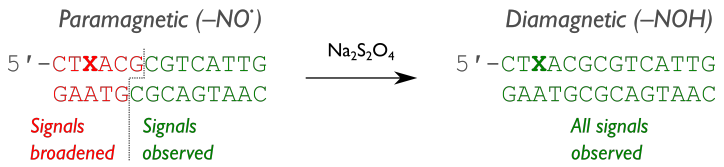


Figure 4.8. Sequence of the spin-labelled duplexes characterised by NMR, where X = either $^{6\text{-Me}}$U or $^{5\text{-Me}}$U. To observe all signals, the nitroxide groups of the spin labels were reduced to hydroxylamines.

labelled and unmodified duplexes, except for the modified base pairs and their nearest neighbours. This suggests that the effects of spin labelling on conformation are local and do not propagate throughout the duplex—consistent with the crystal structures and MD simulations. These observations were confirmed by NOESY and COSY experiments, which showed that the interproton distances and coupling constants throughout the duplex were not altered significantly by the presence of either spin label (Figure 4.9). They also showed minimal differences in conformation between the entire 6-Me and 5-Me duplexes.

In the unmodified duplex, the puckers are all primarily C2′-*endo*, with a minor admixture of C3′-*endo*.[14] The same puckering is adopted by all unmodified nucleotides in the spin-labelled duplexes. The modified uridine puckers, however, were more A-DNA-like in character (Figure 4.9*a*). Similarly, the estimated glycosyl torsion angles of the modified residues were between −140 and −150°. Thus the NMR results are in good agreement with those from X-ray crystallography and MD simulations. Furthermore, NMR spectra of both 6-Me and 5-Me duplexes also support the labels' residing in the minor groove—resonances of the spin labels' methyl groups were sharp and exhibited few NOEs to DNA.

Collectively, NMR experiments indicate that the 5-Me and 6-Me spin labels cause only local perturbations to structure, confined to the modified residues and their nearest

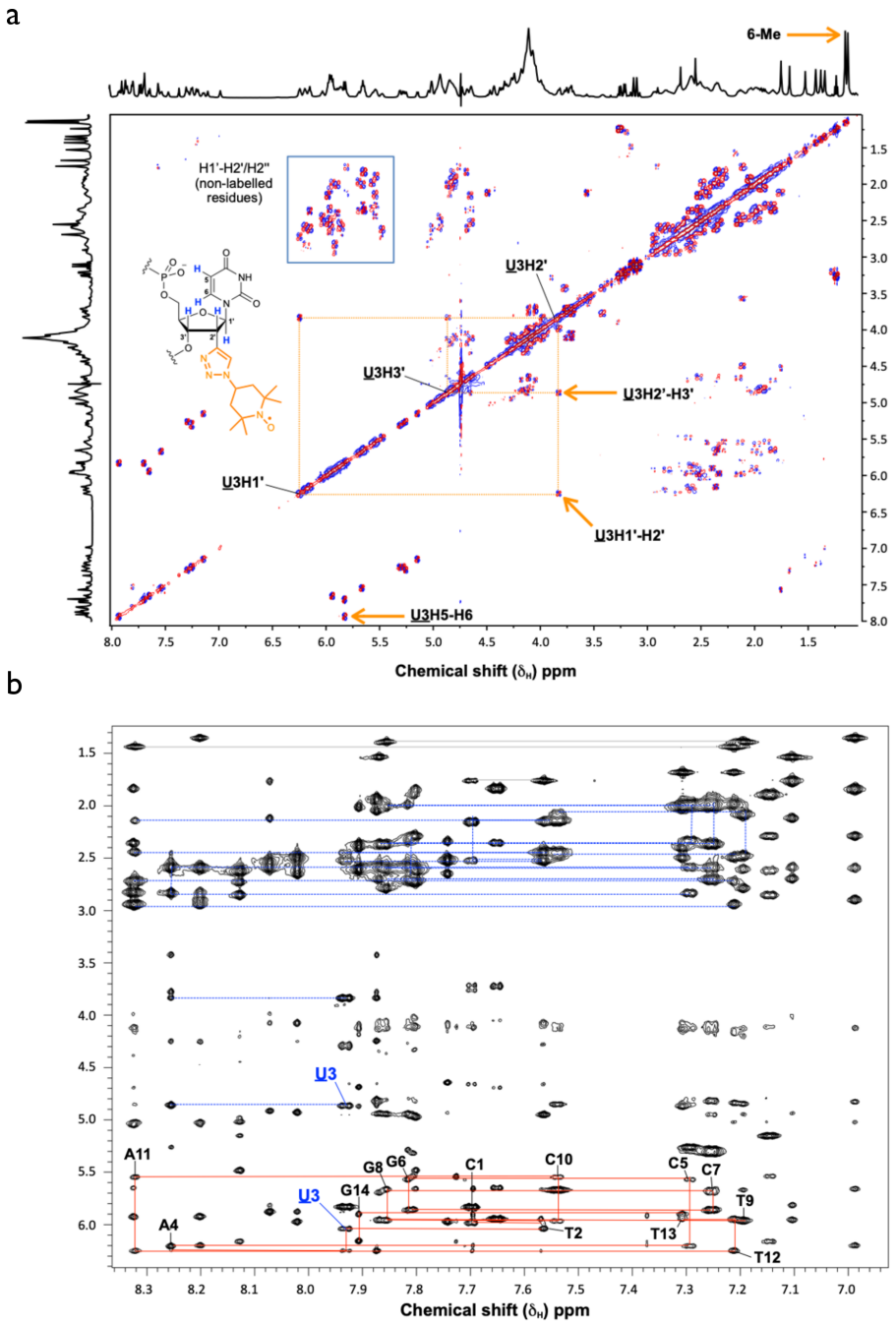


Figure 4.9. Double quantum filtered (DQF)-COSY and NOESY spectra of the reduced 6-Me-labelled duplex. (*a*) DQF-COSY spectrum, recorded at 30 °C, 14.1 T, with an acquisition time of 1 s in t_2. Cross peaks between H1′, H2′/H2″, H2′-H3′ and H3′-H4′ are consistent with B-form sugar puckers (S-type) for all nucleotides except the spin-labelled uridine U3. (*b*) ^{1}H–^{1}H NOESY spectrum, recorded at 30 °C, 14.1 T, with an acquisition time of 1 s in t_2 and a mixing time of 200 ms. Dashed lines denote H2′/H2″ to H8/H6 connectivities. The relative cross-peak intensities are characteristic of a B-DNA, apart from the spin-labelled uridine. Reproduced from Hardwick *et al.*[15]

neighbours. This is consistent with results from both X-ray crystallography and MD simulations.

4.2.6 Further discussion

The accurate measurement of distances between specific sites in DNA remains an important goal in structural biology, particularly for conformational ensembles.[11,36] Spin labelling coupled with EPR spectroscopy has shown promise in addressing this, but the method's routine use in structural biology is hindered by a lack of knowledge of spin-label location and how spin labelling affects the structure of DNA itself.

The crystal structures determined in this work—which are the first reported examples of spin-labelled B-DNA—pinpoint the position of the 6-Me spin label and localise the 5-Me label to within a small radius of rotation about the flexible linker. Furthermore, they provide detailed information on the structural consequences of spin labelling, demonstrating that the strategy of Haugland *et al*[3] is compatible with B-DNA and does not cause large changes to its structure.

The X-ray crystal structure and simulated 'minor groove' conformer of the 6-Me duplex—obtained independently using the same sequence—are in excellent agreement. The most probable interspin distances differed by just 0.1 Å, and they also compared favourably with the EPR-derived value (within 1.2 Å). A crystal structure of the corresponding 5-Me variant could not be obtained, preventing direct comparison with MD simulations and EPR data. But all three techniques indicate that the 5-Me spin label is considerably more flexible than is 6-Me, which may explain the greater discrepancy between the simulated 'minor groove' and EPR-derived interspin distance distributions (1.9 Å). Thus, the 6-Me label yields distance measurements with higher precision.

For both spin labels, good agreement was found between X-ray crystallography, MD simulations, and NMR regarding the consequences of spin labelling on DNA structure. All three techniques indicate that neither spin label is strongly perturbing, imparting

A-DNA-like characteristics only at the modified sites and their nearest neighbours. These characteristics probably arise from placing a substituent at the 2′ position; similar effects on conformation arise even from single-atom substitutions.[27]

4.3 Conclusions & future work

The structure and dynamics of DNA functionalised with 2′-triazole-linked 6-Me and 5-Me spin labels were characterised in unprecedented detail by combining X-ray crystallography, MD simulations, and NMR spectroscopy. Results from these three methods are in good agreement, pinpointing the positions of the spin labels and showing that structural perturbation resulting from the modifications is modest and local. Interspin distances determined by these methods compare favourably with those derived by EPR. Results from this work will aid the reliable interpretation of EPR-derived distance distributions of DNA containing 2′-triazole-linked spin labels, helping to address a significant challenge in structural biology.

Unlike other probe-based methods (*eg* FRET and X-ray scattering interferometry),[36] the angle between two spin labels can be determined in addition to their separation—provided that the labels are sufficiently rigid.[37] The ability to measure the relative orientation between two probes would be hugely advantageous for the study of DNA structure, since both helical twist and bending can significantly alter the inter-probe distance, making the effects hard—or even impossible—to separate.[36] Future work will investigate whether angular information can be reliably obtained from EPR data. Additionally, this technique would be a valuable method to investigate the effects of cytosine modifications on the structure of DNA and its complexes with protein, as detailed in Part I.

BIBLIOGRAPHY

1. Haugland, M. M., Anderson, E. A. & Lovett, J. E. in *Electron Paramagn. Reson.* 1–34 (Royal Society of Chemistry, 2017).

2. Haugland, M. M., Lovett, J. E. & Anderson, E. A. Advances in the synthesis of nitroxide radicals for use in biomolecule spin labelling. *Chem. Soc. Rev.* **47,** 668–680 (2018).

3. Haugland, M. M. *et al.* 2'-Alkynylnucleotides: A Sequence- and Spin Label-Flexible Strategy for EPR Spectroscopy in DNA. *J. Am. Chem. Soc.* **138,** 9069–9072 (2016).

4. Cai, Q. *et al.* Site-directed spin labeling measurements of nanometer distances in nucleic acids using a sequence-independent nitroxide probe. *Nucleic Acids Res.* **34,** 4722–4730 (2006).

5. Qin, P. Z. *et al.* Measuring nanometer distances in nucleic acids using a sequence-independent nitroxide probe. *Nat. Protoc.* **2,** 2354–2365 (2007).

6. Flaender, M. *et al.* Site-specific insertion of nitroxide-spin labels into DNA probes by click chemistry for structural analyses by ELDOR spectroscopy. *Nucleic Acids Symp. Ser.* **52,** 147–148 (2008).

7. Schiemann, O. *et al.* A PELDOR-Based Nanometer Distance Ruler for Oligonucleotides. *J. Am. Chem. Soc.* **126,** 5722–5729 (2004).

8. Spaltenstein, A., Robinson, B. H. & Hopkins, P. B. A rigid and nonperturbing probe for duplex DNA motion. *J. Am. Chem. Soc.* **110,** 1299–1301 (1988).

9. Sowa, G. Z. & Qin, P. Z. in *Prog Nucleic Acid Res Mol Biol* 147–197 (Academic Press, 2008).

10. Barhate, N., Cekan, P., Massey, A. P. & Sigurdsson, S. T. A nucleoside that contains a rigid nitroxide spin label: A fluorophore in disguise. *Angew. Chemie - Int. Ed.* **46,** 2655–2658 (2007).

11. Mathew-Fenn, R. S., Das, R., Silverman, J. A., Walker, P. A. & Harbury, P. A. B. A Molecular Ruler for Measuring Quantitative Distance Distributions. *PLoS One* **3,** e3229 (2008).

12. Edwards, T. E. *et al.* Crystal structure of a DNA containing the planar, phenoxazine-derived bi-functional spectroscopic probe. *Nucleic Acids Res.* **39,** 4419–4426 (2011).

13. Woods, K. K. *et al.* High-Resolution Structure of an Extended A-Tract. *J. Am. Chem. Soc.* **126,** 15330–15331 (2004).

14. Chernatynskaya, A. V., Deleeuw, L., Trent, J. O., Brown, T. & Lane, A. N. Structural analysis of the DNA target site and its interaction with Mbp1. *Org. Biomol. Chem.* **7,** 4981–4991 (2009).

15. Hardwick, J. S. *et al.* 2'-Alkynyl spin-labelling is a minimally perturbing tool for DNA structural analysis. *Nucleic Acids Res.* (2020).

16. Seth, P. P. *et al.* Synthesis and Antisense Properties of Fluoro Cyclohexenyl Nucleic Acid (F-CeNA), a Nuclease Stable Mimic of 2'-Fluoro RNA. *J. Org. Chem.* **77,** 5074–5085 (2012).

17. Robeyns, K., Herdewijn, P. & Van Meervelt, L. Direct observation of two cyclohexenyl (CeNA) ring conformations in duplex DNA. *Artif. DNA PNA XNA* **1,** 2–8 (2010).

18. Valls, N., Steiner, R. A., Wright, G., Murshudov, G. N. & Subirana, J. A. Variable role of ions in two drug intercalation complexes of DNA. *J. Biol. Inorg. Chem.* **10,** 476–482 (2005).

19. Valls, N., Wright, G., Steiner, R. A., Murshudov, G. N. & Subirana, J. A. DNA variability in five crystal structures of d(CGCAATTGCG). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 680–685 (2004).

20. Yang, X.-l., Robinson, H., Gao, Y.-G. & Wang, A. H.-J. Binding of a Macrocyclic Bisacridine and Ametantrone to CGTACG Involves Similar Unusual Intercalation Platforms. *Biochemistry* **39,** 10950–10957 (2000).

21. Adams, A., Guss, J. M., Collyer, C. A., Denny, W. A. & Wakelin, L. P. A novel form of intercalation involving four DNA duplexes in an acridine-4-carboxamide complex of d(CGTACG)(2). *Nucleic Acids Res.* **28,** 4244–53 (2000).

22. Thorpe, J. H. *et al.* Guanine-Specific Binding at a DNA Junction Formed by d[CG(5-BrU)ACG]2 with a Topoisomerase Poison in the Presence of Co2+ Ions. *Biochemistry* **39,** 15055–15061 (2000).

23. Canals, A. *et al.* Intercalative DNA binding of the marine anticancer drug variolin B. *Sci. Rep.* **7,** 39680 (2017).

24. Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. The structure of an oligo(dA)·oligo(dT) tract and its biological implications. *Nature* **330,** 221 (1987).

25. DiGabriele, A. D., Sanderson, M. R. & Steitz, T. A. Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract. eng. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 1816–1820 (1989).

26. Blackburn, M. G., Gait, M. J., Loakes, D. & Williams, D. M. *Nucleic Acids in Chemistry and Biology* (Royal Society of Chemistry, Cambridge, 2007).

27. Li, F. *et al.* 2'-Fluoroarabino- and Arabinonucleic Acid Show Different Conformations, Resulting in Deviating RNA Affinities and Processing of Their Heteroduplexes with RNA by RNase H. *Biochemistry* **45,** 4141–4152 (2006).

28. Pallan, P. S. *et al.* A conformational transition in the structure of a 2'-thiomethyl-modified DNA visualized at high resolution. *Chem. Commun.* 2017–2019 (2009).

29. Juan, E. C. M. *et al.* Crystal structures of DNA:DNA and DNA:RNA duplexes containing 5-(N-aminohexyl)carbamoyl-modified uracils reveal the basis for properties as antigene and antisense molecules. *Nucleic Acids Res.* **35,** 1969–1977 (2007).

30. Tereshko, V., Minasov, G. & Egli, M. A "Hydrat-Ion" Spine in a B-DNA Minor Groove. *J. Am. Chem. Soc.* **121,** 3590–3595 (1999).

31. Tereshko, V., Minasov, G. & Egli, M. The Dickerson-Drew B-DNA Dodecamer Revisited at Atomic Resolution. *J. Am. Chem. Soc.* **121,** 470–471 (1999).

32. Egli, M. *et al.* X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution. *Biopolymers* **48,** 234–252 (1998).

33. Minasov, G., Tereshko, V. & Egli, M. Atomic-resolution crystal structures of B-DNA reveal specific influences of divalent metal ions on conformation and packing. *J. Mol. Biol.* **291,** 83–99 (1999).

34. Berger, I., Tereshko, V., Ikeda, H., Marquez, V. E. & Egli, M. Crystal structures of B-DNA with incorporated 2'-deoxy-2'-fluoro-arabino-furanosyl thymines: Implications of conformational preorganization for duplex stability. *Nucleic Acids Res.* **26,** 2473–2480 (1998).

35. Sjodt, M. & Clubb, R. Nitroxide Labeling of Proteins and the Determination of Paramagnetic Relaxation Derived Distance Restraints for NMR Studies. *Bio-Protocol* **7,** e2207 (2017).

36. Shi, X., Herschlag, D. & Harbury, P. A. B. Structural ensemble and microscopic elasticity of freely diffusing DNA by direct measurement of fluctuations. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E1444–51 (2013).

37. Schiemann, O., Cekan, P., Margraf, D., Prisner, T. F. & Sigurdsson, S. T. Relative orientation of rigid nitroxides by peldor: Beyond distance measurements in nucleic acids. *Angew. Chemie - Int. Ed.* **48,** 3292–3295 (2009).

5

EXPERIMENTAL

5.1 Oligonucleotide Preparation, Purification and Analysis

Oligonucleotide synthesis

Oligonucleotides were synthesised using an automated DNA/RNA synthesiser (Applied Biosystems 394) at a 1-μmol scale phosphoramidite cycle of acid-catalysed detritylation, coupling, capping, and iodine oxidation. Prior to use, all β-cyanotheyl phosphoramidite monomers were dissolved in anhydrous acetonitrile (0.1 M concentration). When modified nucleoside phosphoramidites were used, the coupling time was extended to 10 minutes.

Standard oligonucleotide deprotection

Resin-bound oligonucleotides were cleaved from the solid support and deprotected by immersion in concentrated aqueous ammonia (1 hour at room temperature, followed by 5 hours at 55 °C). The ammonia was then removed under reduced pressure and the

oligonucleotide freeze-dried.

Deprotection of oligonucleotides containing 5-formylcytosine

Oligonucleotides containing the 5-formylcytosine precursor 5-formyl-2′-deoxycytidine-I-cyanoethyl phosphoramidite (Glen Research) are incompatible with ammonia-mediated deprotection, because of the formation of a side product. Resin-bound oligonucleotides containing ^{f}C were thus treated with diethylamine in acetonitrile (2 mL, 10% v/v) for 3 minutes. The resin-bound oligonucleotides were subsequently washed with acetonitrile, dried under a stream of argon, and then immersed in methanolic sodium hydroxide (0.4 M, 1 mL, 4:1 methanol/water (v/v)) for 17 hours at room temperature. Samples were then neutralised with glacial acetic acid (28 μL). Next, the resin was resuspended by brief sonication, and the supernatant collected. The resin was rinsed with water (2 ×250 μL), which was combined with the methanolic solution. The combined mixture was then concentrated to dryness under reduced pressure. The use of isobutyryl-protected dG and acetyl-protected dC was required, due to the deprotection conditions used. To cleave the diol and generate the 5-formyl group, oligonucleotides were resuspended in chilled, aqueous sodium periodate (25 mM, 500 μL). The solution was vortexed and left to react for 30 minutes at 4 °C, after which it was desalted (Nap-10 column, GE Healthcare) and freeze-dried.

The preparation of ^{f}C-containing oligonucleotides using the above monomer was consistently successful and clean. However, the monomer is expensive (US$3,225 for 250 mg at the time of writing) and so an alternative, cheaper monomer from Glen Research was also trialed: the acetal-protected 5-formyl-dC-III-CE phosphoramidite. These oligonucleotides were deprotected in aqueous ammonia (17 hours, room temperature). Owing to incomplete guanine deprotection, samples were then heated in aqueous ammonia (2 hours, 55 °C), achieving full deprotection with no adverse effects. Samples were then concentrated to dryness under reduced pressure and resuspended in acetic acid/water

(80% v/v, 6 hours, 20 °C), then cooled on ice and neutralised by the dropwise addition of ice-cold aqueous ammonia. The samples were then desalted (Nap-25 column, GE Healthcare) and freeze-dried. Several issues were consistently encountered when using this monomer. First, MS analysis of crude samples following ammonia-deprotection indicated that a significant amount of acetal deprotection occurred during oligonucleotide synthesis, which may have resulted in side-reactions. Second, MS analysis indicated that the coupling efficiency of the monomer was relatively poor. Third, for oligonucleotides containing multiple additions of the monomer, complete acetal-deprotection was not observed, even after doubling the duration of treatment with acetic acid. Finally, significant depurination of the samples was also observed. The competing depurination and acetal-hydrolysis reactions consistently gave a mixture of the ^{f}C-containing target, various depurination products, and inseparable acetal-protected oligonucleotides. Thus, the 5-formyl-2′-deoxycytidine-I-cyanoethyl phosphoramidite produced far superior results.

Deprotection of oligonucleotides containing 5-carboxylcytosine

Oligonucleotides containing 5-carboxylcytosine were deprotected according to the procedure for 5-formyl dC CE phosphoramidite, except that the diethylamine treatment was omitted.

Deprotection of oligonucleotides containing 5-hydroxymethylcytosine

Oligonucleotides containing ^{hm}C were prepared using the 5-hydroxymethyl-2′-dC-I-CE phosphoramidite (Glen Research) and deprotected by heating in concentrated aqueous ammonia (17 hours, 75 °C). Following removal of the ammonia under reduced pressure, mass spectromety indicated that full cleavage of the cyanoethyl group from the base had not occurred. Complete deprotection was achieved by further treatment of the oligonucleotides with ammonia (4 hours, 55 °C).

Spin-labelling of oligonucleotides

The procedure of Haugland *et al* was followed.[1] Briefly, to the oligonucleotide containing 2′-ethynyl-modified dU (620 nmol in 225 μL acetonitrile) was added copper(I) iodide (124 μL, 50 mM) and aqueous tris(3-hydroxypropyltriazolylmethyl)amine (24.8 μL, 250 mM), followed by the spin label (124 μL, 500 mM). All solvents were degassed prior to use. The reactions were vortexed and left at room temperature for 3 hours, after which they were diluted to a final volume of 1 mL and desalted (2 × Nap-25, GE Healthcare).

Oligonucleotide purification by HPLC

Following deprotection, oligonucleotides were purified using a reversed-phase high-performance liquid chromatography (RP-HPLC) system (Gilson) with a Luna 10-μm C8 100-Å-pore Phenomenex 10 × 250 mm column with a gradient of acetonitrile in triethylammonium bicarbonate (TEAB) buffer. Buffer A was a 0.1-M aqueous solution of TEAB, prepared by dilution of a 1-M stock solution. The solutions pH was lowered to pH 7.5 by bubbling through CO_2 gas, prepared by the sublimation of dry ice. Buffer B consisted of 50:50 acetonitrile/TEAB (0.2 M).

Oligonucleotide purification by polyacrylamide gel electrophoresis (PAGE)

Denaturing polyacrylamide gels were prepared and run according to the procedure of Section 5.2. Typically, 0.5–2 OD_{260} of material was loaded per lane (when using a 15-tooth comb), premixed with an equal volume of formamide.

After electrophoresis, the gel was visualised by UV (not necessary for fluorescently labelled samples) and the desired bands were excised using a sterile scalpel and placed in 0.5-mL Eppendorf tubes, whose base had been thrice pierced by a needle. Each tube was placed into a 1.5-mL Eppendorf tube and briefly centrifuged (17,000 *g*, VWR Micro Star 17R) to crush the gel by forcing it through the holes. Tris·HCl (10 mM, 1.2 mL, pH 7.5) was then added to the crushed gel, and the sample was frozen at -80 °C for 30 minutes,

before being shaken (37 °C, 20 hours, 1200 rpm). Each sample was then concentrated to dryness (Labconco CentriVap), resuspended in 1.5 mL water, twice desalted (Nap-25, GE Healthcare), and freeze-dried.

Ethanol precipitation of oligonucleotides

To the DNA solution (45 μL) was added sodium acetate (5 μL, 3 M, pH 5.5) and 2 μL of glycogen (10 mg mL^{-1}). The solution was vortexed before 300 μL of ice-cold ethanol (96%) was added, followed by thorough vortexing. The samples were then cooled (at least 10 minutes at −80 °C or 30 minutes at −20 °C), then centrifuged (17,000 *g*, 15 mins, 4 °C). If a pellet was observed, the supernatant was carefully removed and the pellet washed with ice-cold ethanol (75%, 300 μL); the pellet was agitated to dislodge the pellet from the side of the tube (flicking the bottom of the tube while tilted horizontally was most effective). The sample was centrifuged (17,000 *g*, 5 minutes) and the supernatant was again removed. The washing and centrifugation steps repeated. If a pellet was not observed following the initial addition of ethanol and cooling, a further 2 μL glycogen was added to the ethanolic solution, followed by vortexing and cooling. After removing as much supernatant as possible from the second ethanol wash, the sample was dried (CentriVap) and resuspended in the desired buffer.

Preparation of the sodium salt of oligonucleotides using Dowex

A stock of the Na^+ form of Dowex resin was prepared according to the following procedure. 50 mL of Dowex 50 WX8 100–200 resin (Sigma-Aldrich) was washed with HCl (2 × 50 mL, 0.1 M) in a sintered glass funnel. The mixture was stirred thoroughly each time prior to the vacuum-assisted removal of the HCl. The resin was then treated with sterile, deionised water in the same way, followed by sodium hydroxide (2 × 50 mL, 0.2 M). The resin was then rinsed with water (2 × 50 mL) and the pH of the solution tested to ensure that it was not acidic. After further rinsing (6 × 50 mL), the resin (now

in the Na^+ form) was transferred to a glass jar and topped with sterile water to form a stiff slurry. The resin was stored at 4 °C and used within 1 year.

The sodium salts of oligonucleotides were prepared according to the following procedure. A small ball of glass wool was placed in the neck of a sterile Pasteur pipette, onto which ~1.5 mL of the Dowex slurry (Na^+ form) was loaded. After rinsing the slurry with 5 mL water, the oligonucleotide (1 mL solution) was added. Following elution of the oligonucleotide, a further 2 × 1 mL of water was added and combined with the oligonucleotide solution. This solution was again passed through the Dowex resin and the resin rinsed with water (2 × 1.5 mL). The two solutions were combined and concentrated under reduced pressure to ~0.5 mL and freeze-dried.

Quantification of oligonucleotides

Oligonucleotides were quantified by measuring their absorbance at 260 nm, using either a Varian 50 Bio UV-vis spectrophotometer or a Nanodrop 2000 (Thermo Scientific).

Mass spectrometry

Oligonucleotides were characterized by ultra-high-performance mass spectrometry (UPLC-MS) using an Acquity UPLC system with a BEH C18 1.7-μm column (Waters) in conjunction with a Bruker micrOTOF or a Waters Xevo G2-XS mass spectrometer (ESI^- mode). For UPLC, buffer A contained 8.6 mM triethylamine, 200 mM (1,1,1,3,3,3)-hexafluoro-2-propanol in 5% methanol/water (v/v); buffer B contained 20% buffer A in methanol. Buffer B was increased from 0-70% over 8 min, at a flow rate of 0.2 mL min^{-1}. Raw data acquired using the Bruker micrOTOF were processed and deconvoluted with the Bruker Daltonics Compass 1.3 software package. Data acquired using the Waters Xevo were deconvoluted using the MassLynx package (v4.1).

Annealing of oligonucleotide duplexes

All oligonucleotide duplexes were annealed using a thermal cycler (Bio Rad T100). Samples were heated to 95 °C for 5 minutes (lid temperature 105 °C) then cooled to 25 °C a rate of 1 °C per minute.

5.2 Polyacrylamide Gel Electrophoresis

Gel preparation

For non-denaturing gels, the desired buffer (7 mL, 10×) was combined with the appropriate volume of a 40% solution of acrylamide (29:1 acrylamide:bis-acrylamide, Fisher). The solution was diluted with deionised water to a total volume of 70 mL. To polymerise the acrylamide, aqueous ammonium persulphate (560 μL, 10% w/v) was added, followed by tetramethylethylenediamine (Sigma-Aldrich), and the solution thoroughly stirred. The gel was then cast immediately. For denaturing gels, the above procedure was repeated, but with the addition of urea (29.4 g), dissolved by sonication.

Electrophoresis and imaging

Gels were electrophoresed using a Hoefer SE 400 system with a Bio Rad PowerPac HV. The same plate size (18 × 0.2 ×24.4 cm) was used for all experiments. Gels were pre-run in the appropriate buffer for at least 30 minutes prior to loading. Gels were imaged using a Syngene G:Box Chemi XT. Quantitative analysis of gel images was performed using ImageJ.[2]

5.3 Circular Dichroism (CD) Spectroscopy

CD spectra were acquired on a Chirascan Plus spectropolarimeter (Applied Photophysics), using a quartz cuvette with a 10-mm path length. The sodium salt of each DNA sample

was obtained by ion exchange (Dowex 50, Na^+ form). Unless otherwise stated, DNA samples were suspended in phosphate-buffered saline (PBS, Oxoid, 137 mM sodium chloride, 3 mM potassium chloride, 8 mM disodium hydrogen phosphate, and 1.5 mM potassium dihydrogen phosphate, pH 7.3). Data were collected at 20 °C, over the range of 220–350 nm. Each spectrum was produced by averaging four scans (0.5-nm step size, 1.5-s time-per-point, 2-nm bandwidth). A sample containing only buffer was treated identically and subtracted from the collected data. Spectra were then smoothed (Savitzky-Golay, Origin) with a polynomial order of 3 and a smoothing window of 20 points, after which they were baseline-corrected (offset at 350 nm).

5.4 UV Absorption Spectroscopy

UV absorption spectra were collected on a Varian 50 Bio UV-vis spectrophotometer, using a quartz cuvette (10-mm path length). DNA solutions were prepared at a final duplex concentration of 1.8 μM in PBS. Data were collected at 20 °C over the range 220–350 nm and a scan rate of 60 nm min^{-1}.

5.5 X-Ray Crystallography

Oligonucleotide crystallisation

Crystals were obtained using the sitting-drop vapour-diffusion method. Trials were performed using the Natrix HT (Hampton Research) and Nucleix (Qiagen) screens, which were dispensed into 96-well sitting-drop crystallisation plates (CrystalQuick, Greiner) using a high-throughput dispenser (Hydra 96, Robbins Scientific). In each drop, 100 nL of the corresponding well solution was was combined with 100 nL of a 1 mM aqueous DNA solution using a crystallisation robot (Cartesian Microsys). The plates were immediately sealed and incubated at 21 °C in a Rock Imager 1000 (Formulatrix), enabling crystal growth to be monitored remotely. Typically, crystals appeared

within 3 days and were collected within 2 weeks. Conditions producing crystals were then manually optimised using 24-well sitting-drop crystallisation plates (Chryschem, Hampton Research), in which 1 μL of the 250 μL well solution was combined with 1 μL of DNA solution. The crystallisation conditions of the structures determined in this study are given in the appropriate appendix. Typically, the precipitant concentration was varied on one axis, while DNA concentration was varied on the other. The plates were immediately sealed and incubated at 20 °C. Collected crystals were flash-cooled and stored in liquid nitrogen until data collection. Further cryoprotection was unnecessary since MPD or 2-propanol was used as a precipitant in all cases.

Collection of diffraction data

Diffraction data were collected at 100 K at either the Diamond Light Source synchrotron science facility, Harwell (beamlines I02–I04, Pilatus 6M hybrid pixel array detectors). 1,800 images were typically collected for each crystal (0.1° oscillation, 100-ms exposure). X-ray wavelengths of 0.976 Å (PDB 5mvk, 5mvp, 5mvq, 5MVU), 1.060 Å (5mvt), 0.920 and 1.240 Å (5mvl), 0.979 Å (6qjs), and 0.916 Å (6qjr) were used.

Diffraction data processing, phase determination, model building and refinement

Data were indexed and scaled with XDS[3] and AIMLESS[4]. The brominated structure (5mvl) was solved experimentally by multiwavelength anomalous dispersion from the bromine atoms using the SHELX package.[5] All other structures were solved by molecular replacement (Phaser):[6] PDBs 5mvk, 5mvp, 5mvq, and 5mvu using 5mvl as the search model, and PDBs 6qjs and 6qjr using 1s2r and 1d98, respectively. To prevent model bias, any atoms that differed between the search model and the structure being solved were removed from the search model prior to molecular replacement. Following an initial round of refinement, using either REFMAC[7] or PHENIX,[8] missing atoms were added manually in COOT,[9] followed by further rounds of refinement.

Quantitative analysis of X-ray crystal structures

Quantitative analysis of all crystal structures was performed using the 3DNA package,[10] except for groove parameters, which were determined with the Curves+ program.[11] Inter-spin distances given in Chapter 4 are between the centre of each nitroxide N–O· bond, since the unpaired electron is distributed mainly across this bond.[12,13] The distance between the bond centres was determined by halving each bond length using UCSF Chimera[14] and then measuring the distance between the oxygen atoms.

DNA model building

Models of ideal A- and B-DNA were generated using w3DNA.[15] To model the junctions between crystal structures and ideal A- and B-DNA, w3DNA input files were created manually by combining the parameters obtained from the crystal structures and ideal models. This enabled the generation of preliminary coordinate files of the modelled junctions in w3DNA. Finally, to reintroduce the non-idealised backbone geometry of the crystal structures, the modelled atomic coordinates were manually replaced with those of the crystal structures following alignment (PyMOL).

5.6 Gel Mobility

5.6.1 Polyacrylamide gel electrophoresis

10% Non-denaturing polyacrylamide gels were prepared according to 5.2. Unless otherwise stated, the 5× loading buffer consisted of Ficoll-400 (12.5% w/v, 30 mM Tris·HCl pH 7.5, and either 50 or 0 mM magnesium acetate). Following electrophoresis (90 V, 20 h), gels were stained with 1× SyBR Gold (Thermo Fisher, 100 mL, 20 mins).

5.6.2 Analysis of gel images

Using ImageJ (v1.52a),[2] gel images were rotated 90° anticlockwise, so that each lane was horizontal. All images were background subtracted using a rolling-ball radius of 50 pixels. One-dimensional plots of intensity (averaged across the lane) versus distance were generated (using boxes of the same size and lateral position). Using *R*, the plots were smoothed and the centre of each band was determined automatically with the custom script below. The following commands were used for an example output file from ImageJ, 'Lane1.csv':

```
Lane1 <- read.csv("Lane1.csv")
x <- Lane1$X
y <- Lane1$Y
plot(x, y, type='l')
library(signal)
ysmooth <- sgolayfilt(y, 3, 9)
localmax <- which(diff(sign(diff(ysmooth)))==-2)
library(ggplot2)
sp <- qplot(x, ysmooth, geom=c("line"))
sp + geom_vline(xintercept = localmax) +
scale_x_continuous(breaks = round(seq(min(0), max(1000), by = 20), 1))
write.csv(localmax, file = "Lane1_localmaxima.csv")
```

Smoothing of the densitogram reduced, but did not eliminate, false peak maxima. Although increasing the extent of smoothing eliminated these false maxima, this also reduced the number of discernable peaks corresponding to larger fragments. Thus, milder smoothing and manual removal of unwanted peaks offered superior results. Following manual removal of false values, the peak maxima were plotted against the original, unsmoothed data to ensure their validity.

```
plot(x, y, type='l')
Lane1_localmaxima2 <- read.csv(file = "Lane1_localmaxima2.csv")
sp2 <- qplot(x,y, geom=c("line"))
sp2 + geom_vline(xintercept = Lane1_localmaxima2$x) +
scale_x_continuous(breaks = round(seq(min(0), max(1000), by = 20), 1))
```

Investigating the presence of cyclised constructs

To detect any cyclised ligation products, which could introduce artefacts in the analysis of electrophoretic mobility, two-dimensional polyacrylamide gels were prepared. The procedure was adapted from Ulanovsky *et al*.[16] First, a 4% non-denaturing polyacrylamide gel was prepared and pre-run in TBE (30 mins, 140 V). The ligation product of a single reaction was then loaded onto one lane of the gel and electrophoresed (3 hours, 140 V). The lane containing the ligation products was then cut from the gel with a sterile scalpel, rotated by 90°, placed at the bottom edge between two plates, and a 10% polyacrylamide gel (containing 50 μg mL^{-1}) was then cast around it. This gel was then electrophoresed (16 hours, 140 V) in TBE (also containing 50 μg mL^{-1}). Because the ligation fragments were located at the bottom-end of the plates, the positive and negative electrodes were switched to reverse the direction of electrophoresis.

Calculating relative mobility

The observed electrophoretic mobility, μ_{obs}, is defined as:

$$\mu_{obs} = \frac{v}{E} \tag{5.1}$$

where v is the velocity of the DNA fragment and E is the applied electric field strength.[17] Thus, μ_{obs} can be calculated using the following equation:

$$\mu_{obs} = \frac{d}{tE} \tag{5.2}$$

where d is the distance migrated by the fragment, t is the migration time, and E is the applied electric field strength.[18]

Anomalous mobility can then be quantified for each of the target fragments by calculating mobility ratios, $\mu_R = \mu_{curved}/\mu_{normal}$, using equation 5.2. As shown in Equation 5.3, if the two fragments are electrophoresed on the same gel (ie, at the same electric field strength, for the same time), then the relative mobility of the two fragments can be determined by simply taking the ratio of their migration distances:

$$\mu_R = \frac{\mu_2}{\mu_1} = \frac{d_2(\frac{1}{tE})}{d_1(\frac{1}{tE})} = \frac{d_2}{d_1} \tag{5.3}$$

5.7 DNase I Cleavage

Construct design

A 32-nt random sequence was generated, into which the motif 5′-ACGCGCA was centrally inserted, using the Python script below.

```
>>> import random, string
>>> def sequence(length):
alpha = 'acgt'
return ''.join(random.choice(alpha) for i in range(length))
>>> sequence(32)
```

Preparation of the constructs

Oligonucleotides were synthesised and deprotected following standard protocols (see Section 5.1). Following purification by both RP-HPLC and PAGE, 40-μM stock samples were prepared (in 50 mM Tris·HCl pH 7.5, stored at −20 °C). Each duplex was prepared by combining 10 μL of the two desired stocks with 5 μL of 5 × 'DNase I reaction buffer'

(500 mM Tris·HCl pH 7.5 and 50 mM $MgCl_2$). The samples were then annealed and stored at −20 °C until required.

Preparation of the DNase I stock

DNase I (Sigma, D4263) was resuspended in 'DNase I Storage Buffer' (50% glycerol, 135 mM NaCl, 15 mM sodium acetate pH 6.5, 10 Kunitz units μL^{-1}) and stored at −20 °C. The desired concentration of DNase I was then achieved by serial dilution of the stock with DNase I Storage Buffer.

DNase I digestion and PAGE

To a 4 μL stock of each duplex was added 1 μL of the diluted DNase I stock, such that its final concentration was 0.005 Kunitz units μL^{-1}. After vortexing, the samples were left to react (15 min, 20 °C) and then quenched with 5 μL 'DNase I Stop Buffer' (80% formamide, 0.1% bromophenol blue, 10 mM EDTA). These conditions were found to yield sufficient cleavage without degrading more than 30% of the full-length construct.

The samples were then heated (95 °C, 3 min) then rapidly cooled to −20 °C, after which they were either loaded directly onto a pre-run 17.5% denaturing polyacrylamide gel, or stored at −20 °C until required. After loading onto a gel, samples were electrophoresed until the bromophenol blue dye had migrated to 1/3–1/4 of the plate height from the bottom of the gel (20 W, TBE buffer, ~2 hours). The gel was visualised using excitation and emission wavelengths of 460 and 516–600 nm, respectively. Each gel was imaged with both short and long exposure times. At short-exposure time (typically <3 s), all bands were not saturated and could be reliably quantified. This enabled verification that at least 70% of the duplex remained uncleaved. At long-exposure time (typically >15 s), the signal-to-noise ratio of the bands corresponding to the cleavage fragments was enhanced. The latter was used for the analysis.

Analysis of cleavage products

Densitograms of the gel images were obtained using ImageJ (v1.52a).[2] Images were processed as described in Section 5.6.

The peak corresponding to the product of cleavage 5′ of the modified site was normalised by dividing its integral by that of several combined reference peaks (see Figure B.2). This minimised errors between experiments and enabled lanes from different gels to be directly compared. Integrals were determined using OriginPro 2017, without further background subtraction. Mean values of these normalised cleavage rates were then determined (experiments performed in duplicate). Densitograms of the unmodified duplex (performed in quadruplicate) were treated in exactly the same way. Dividing the normalised value of each modified duplex by that of the unmodified duplex then gave the cleavage rate at the modified site, relative to that of its unmodified counterpart.

To measure the effects of cytosine modification on the opposite strand, the method was adjusted to reflect the need for different controls. When determining the effect of hemi-modified sites—where the modification was incorporated into the unlabelled strand—the duplex containing unmodified cytosine in both strands was used as before. But for measuring the effect of doubly modified sites, *eg* symmetric methylation, the control duplex contained the same modification only in the labelled strand; the unlabelled strand contained unmodified cytosine.

5.8 Maxam–Gilbert Sequencing

'G+A' reaction using formic acid

The following procedure was adapted from the following:
https://www.nationaldiagnostics.com/electrophoresis/article/maxam-gilbert-sequencing.
To 10 μL DNA solution in a 1.5-mL Eppendorf tube was added 25 μL formic acid. The mixture was vortexed and incubated (25 °C, 4 minutes). 200 μL of a solution contain-

ing aqueous sodium acetate (0.3 M μL, pH 7), EDTA (0.1 mM) and glycogen (0.4 mg mL^{-1}) was then added and the mixture vortexed. Ethanol (96%, 750 μL) was added, and after vortexing, the mixture was cooled at −80 °C (10 minutes) or −20 °C (60 minutes). After centrifugation (17,000 *g*, 4 °C, 15 minutes), the supernatant was removed from the resulting pellets and discarded. Pellets were rinsed with ice-cold ethanol (75%, 500 μL). If the pellet remained lodged to the wall of the tube, the tube was tilted horizontally and flicked until the pellet was dislodged. Samples were re-centrifuged (17,000 *g*, 4 °C, 5 minutes) and the supernatant removed. The previous rinsing step was repeated, and, following removal of the supernatant, pellets were dried (either left to air-dry, or dried on a speed-vac under low heat). Next, pellets were resuspended in aqueous piperidine (10 %), transferred to screw-top tubes, and heated (90 °C, 30 minutes). Note that a fresh solution of piperidine was prepared each time to minimise its oxidation, and that screw-top tubes are essential to maintain a tight seal, ensuring that the reaction goes to completion. Tubes were then cooled, centrifuged to collect any condensate, then concentrated to dryness (speed-vac). Pellets were redissolved in water (40 μL) and dried again. To ensure complete removal of the piperidine, the rehydration and drying step was repeated. Samples were then ready for electrophoretic analysis.

G+A reaction using heat

The depurination reaction described above can reportedly be achieved for radiolabelled DNA by simply heating a sample under certain conditions.[19] This method was appealing, since it would negate the use of formic acid and piperidine, as well as ethanol precipitation.

1.5 μL of labelled DNA was combined with 20 μL water and 5 μL of a solution of EDTA (10 mM), formamide (80%), bromophenol blue (0.1 %), xylene cyanol (0.1%), and sodium hydroxide (1 mM). The tube was placed in a heat block (95 °C, 20 minutes) with the lid off, to allow partial evaporation of the sample. Following this procedure, it is reported that the sample would be ready for electrophoretic analysis. This method was

attempted for fluorescently labelled DNA and the products analysed by both PAGE and capillary electrophoresis. However, the formic acid method yielded far superior results.

'C' reaction using hydrazine

This reaction was performed only for the hydroxyl radical experiments (see below). The procedure was adapted from Gaastra[20] and the following: https://www.nationaldiagnostics.com/electrophoresis/article/maxam-gilbert-sequencing.

In a 1.5-mL Eppendorf tube, DNA solution (5 μL) was combined with sodium chloride (15 μL, 5 M) and hydrazine (30 μL), and incubated (20 °C, 10 minutes). 200 μL of a mixture of aqueous sodium acetate (0.3 M, pH 5.5) and glycogen (0.4 mg mL^{-1}) was added and mixed with a pipette. Immediately afterwards, ice-cold ethanol (96%, 750 μL) was added. The mixture was vortexed and cooled at −80 °C for 10 minutes and then centrifuged (17,000 *g*, 15 minutes). The supernatant was removed and the pellet washed with ice-cold ethanol (75%, 750 μL) and centrifuged (17,000 *g*, 5 minutes). The supernatant was removed and the ethanol washing step repeated. The pellet was then resuspended in 300 μL of a solution of sodium acetate (0.3 M, pH 5.5) and glycogen (0.4 mg mL^{-1}) and vortexed. Ice-cold ethanol (96%, 900 μL) was added and the mixture vortexed again. After centrifugation (17,000 *g*, 15 minutes) and removal of the supernatant, the pellet was washed with ice-cold ethanol (75%, 900 μL) and re-centrifuged (17,000 *g*, 5 minutes). The supernatant was removed and the washing step repeated. After removal of the supernatant, the pellets were left to air-dry. Piperidine treatment was carried out as described above. Note that owing to the extreme toxicity of hydrazine, supernatant, pipette tips and vials that contained or had come into contact with the chemical were quenched in aqueous iron(III) chloride (3 M).

5.9 Hydroxyl Radical Cleavage

These experiments are ongoing, so their results have not been presented in this thesis. However, the procedures used are included here, as they may be useful to those continuing the work.

Capillary electrophoresis

Capillary electrophoresis was performed on a SeqStudio-3200 Genetic Analyzer (Applied Biosystems), with a typical running voltage of 12,000 V.

Cleavage reaction

The following was adapted from Jain *et al.*[21] Fresh, aqueous solutions of hydrogen peroxide (3%), sodium ascorbate (5 mM) and a 1:1 mixture of EDTA (4 mM) and ferrous ammonium sulphate (2 mM) were prepared by serial dilution. 1 uL of a FAM-labelled DNA duplex (8 μM) was then combined with 4.5 μL deionised water and the desired buffer (0.9 uL, 10 ×) in a 1.5-mL Eppendorf tube. 1.2 μL each of sodium ascorbate, hydrogen peroxide and the iron-EDTA solution was added as separate drops to the tube walls to avoid premature formation of hydroxyl radical. To start the reaction, the tube was tapped firmly on the bench and vortexed, then allowed to react at room temperature for 2 minutes. The reaction was then quenched by the addition of thiourea (2 μL). The DNA was collected by ethanol precipitation (see Section 5.1) and the pellets washed with 75% ethanol (2 × 65 μL) before being resuspended in water (16 μL). 1 μL of each sample was combined with 20 μL formamide ('Hi-Di', Thermo Fisher) and the desired standard in a separate well of a 96-well plate. The wells were sealed with adhesive film and the plate heated at 95 °C for 5 minutes. The plate was centrifuged (1,200 rpm, 2 minutes), the film replaced with septa, and the plate loaded into the electrophoresis instrument.

BIBLIOGRAPHY

1. Haugland, M. M. *et al.* 2'-Alkynylnucleotides: A Sequence- and Spin Label-Flexible Strategy for EPR Spectroscopy in DNA. *J. Am. Chem. Soc.* **138,** 9069–9072 (2016).
2. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).
3. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 133–144 (2010).
4. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69,** 1204–1214 (2013).
5. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. Sect. A Found. Crystallogr.* **64,** 112–122 (2008).
6. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
7. Murshudov, G. N. *et al.* REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 355–367 (2011).
8. Adams, P. D. *et al.* PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

9. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

10. Lu, X.-J. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31,** 5108–5121 (2003).

11. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **37,** 5917–5929 (2009).

12. Haugland, M. M., Lovett, J. E. & Anderson, E. A. Advances in the synthesis of nitroxide radicals for use in biomolecule spin labelling. *Chem. Soc. Rev.* **47,** 668–680 (2018).

13. Azarkh, M. *et al.* Site-directed spin-labeling of nucleotides and the use of in-cell EPR to determine long-range distances in a biologically relevant environment. *Nat. Protoc.* **8,** 131–147 (2013).

14. Pettersen, E. F. *et al.* UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

15. Zheng, G., jun Lu, X. & Olson, W. K. Web 3DNA - A web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.* **37,** 240–246 (2009).

16. Ulanovsky, L., Bodner, M., Trifonov, E. N. & Choder, M. Curved DNA: design, synthesis, and circularization. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 862–6 (1986).

17. Luan, B. & Aksimentiev, A. Electric and electrophoretic inversion of the DNA charge in multivalent electrolytes. *Soft Matter* **6,** 243–246 (2010).

18. Stellwagen, E., Lu, Y. & Stellwagen, N. C. Curved DNA molecules migrate anomalously slowly in free solution. *Nucleic Acids Res.* **33,** 4425–4432 (2005).

19. Hampshire, A. J., Rusling, D. A., Broughton-Head, V. J. & Fox, K. R. Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. *Methods* **42,** 128–40 (2007).

20. Gaastra, W. in *Nucleic Acids* 333–342 (Humana Press, New Jersey, 1984).

21. Jain, S. S. & Tullius, T. D. Footprinting protein–DNA complexes using the hydroxyl radical. *Nat. Protoc.* **3,** 1092–1100 (2008).

Part III.

Appendix

A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

A.1 Supplementary figures

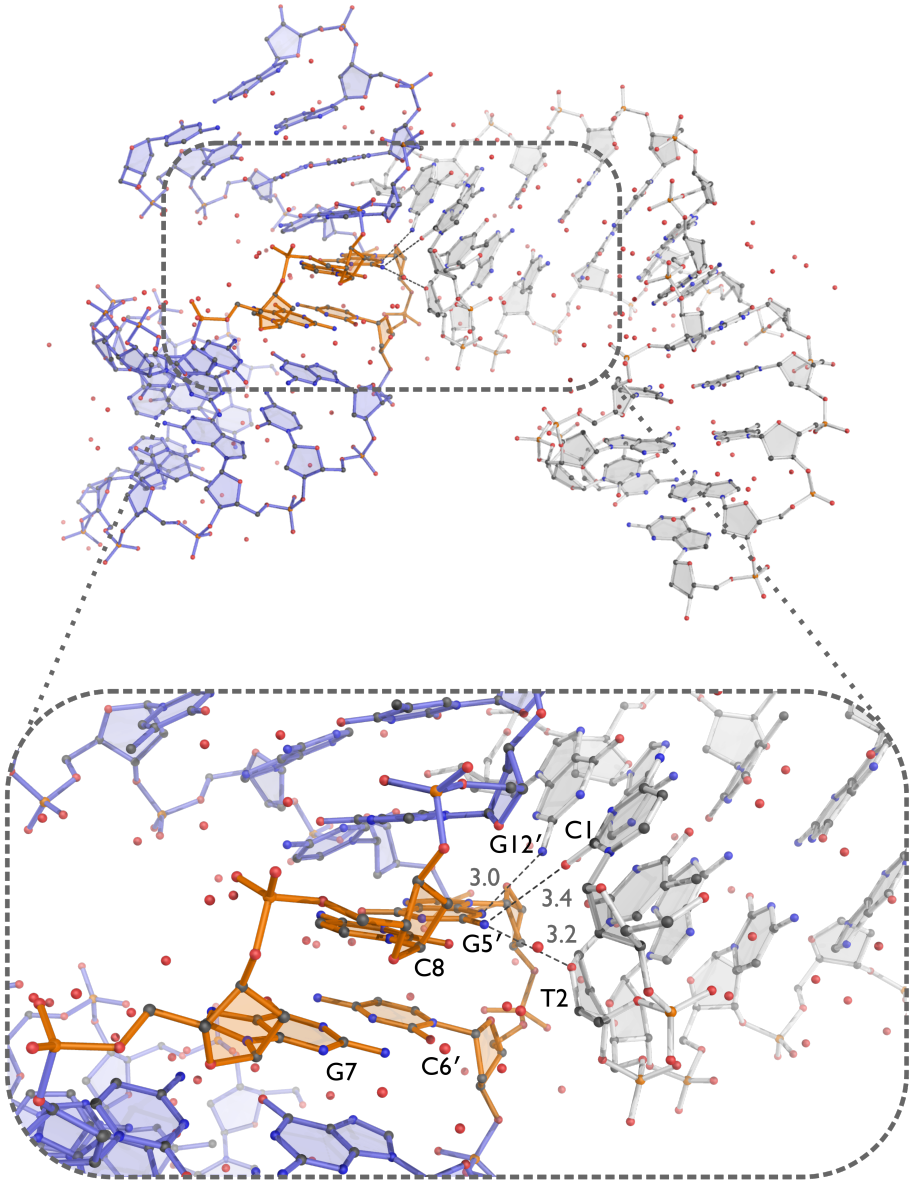


Figure A.1. Crystal contacts between step 7 (orange) of 5mvk (blue) and a symmetry mate (grey).

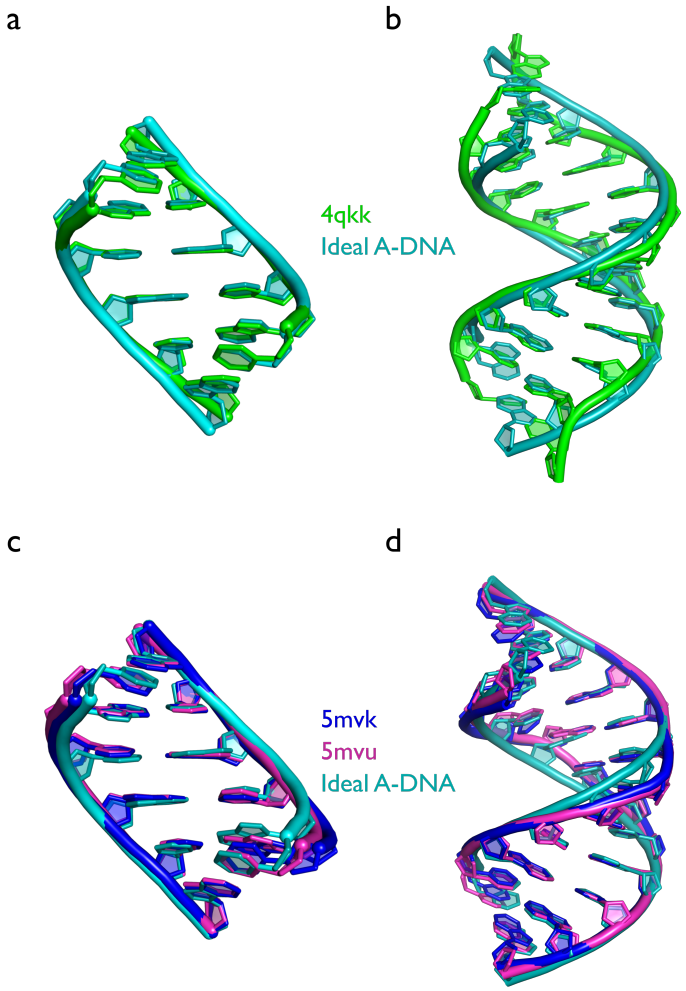


Figure A.2. Crystal structures containing either 5-formylcytosine or cytosine overlaid with ideal A-DNA of the same sequence. (*a*) The formylated $(CpG)_2$ tract of the high-salt fC structure (green, PDB 4qkk) overlaid with ideal A-DNA (cyan). The two structures overlap well—consistent with the formylated region adopting typical A-form geometry. (*b*) Overlay of the entire dodecamer duplex of 4qkk. As observed in the overlay of 4qkk and its unmodified crystal structure analogue 5mvk (Figure 2.6), overlap is poorest at the duplex ends, which are structurally anomalous in 4qkk. (*c*) Overlay of the $(CpG)_2$ tract of the unmodified (blue, PDB 5mvk) and low-salt fC (magenta, PDB 5mvu) crystal structures with ideal A-DNA. (*d*) Overlay of the entire dodecamer duplexes. All three structures are similar, though the crystal structures deviate somewhat from the ideal model in the central region—presumably due to effects from the sequence or crystal packing. Adapted from Hardwick *et al.*[1]

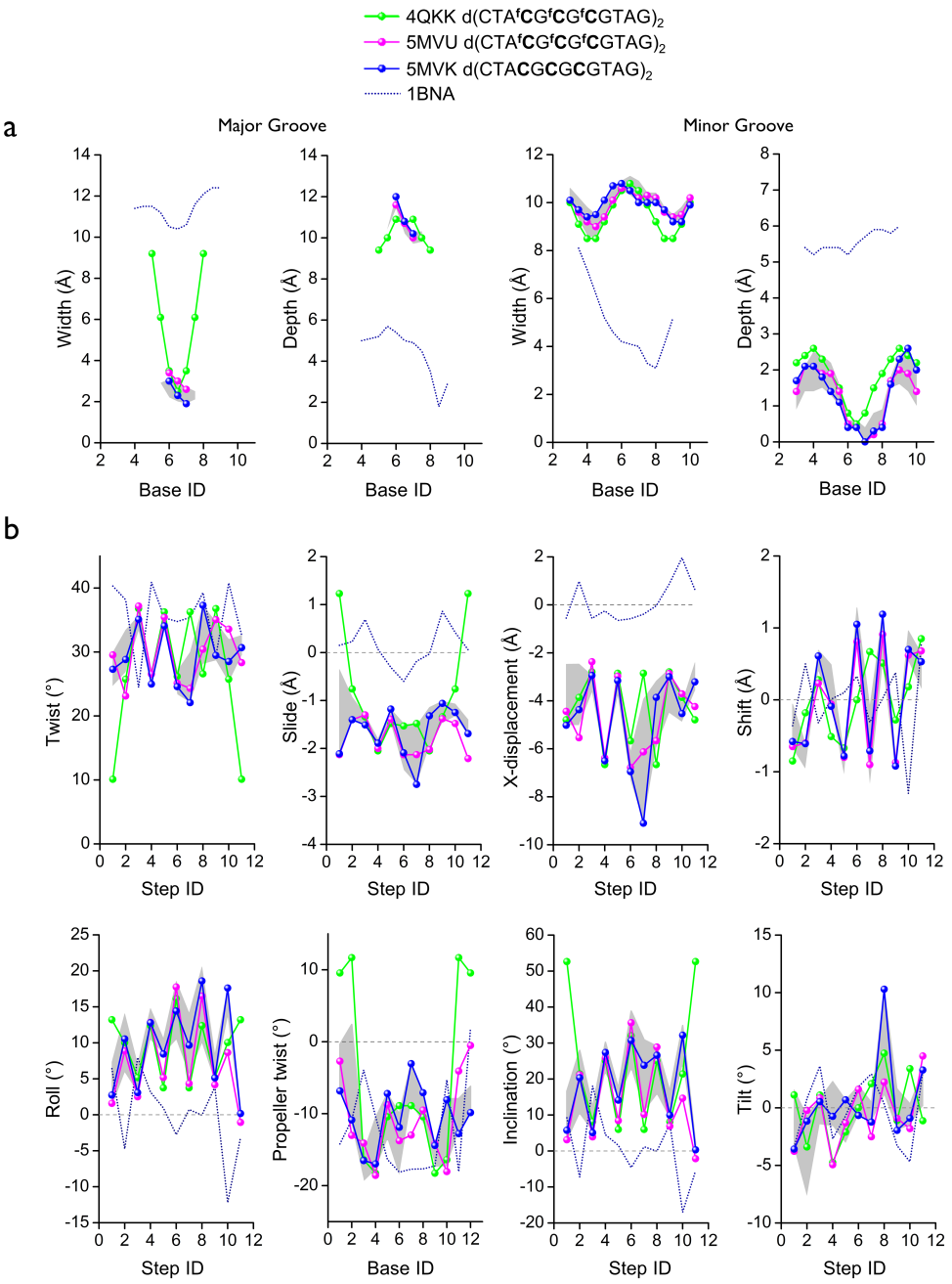


Figure A.3. Structural parameters of the low-salt (magenta, PDB 5mvu, space group $P3_221$) and high-salt (green, PDB 4qkk, space group $P4_32_12$) crystal structures of the formylated sequence and their unmodified counterpart (blue, PDB 5mvu, space group $P3_221$). Shaded regions indicate the range of values adopted by unmodified variants of the sequence (PDBs 5mvq, 5mvt, and 5mvp), which crystallised in the same space group ($P3_221$) as the low-salt fC structure and its unmodified counterpart. Thus, they provide an indication of the magnitude of sequence-dependent structural changes, against which the structural consequences of formylation can be compared. The values of the Dickerson-Drew dodecamer B-DNA structure (blue dots, PDB 1BNA) are plotted alongside. (*a*) Groove dimensions. (*b*) Local structural parameters. Reproduced from Hardwick *et al.*[1]

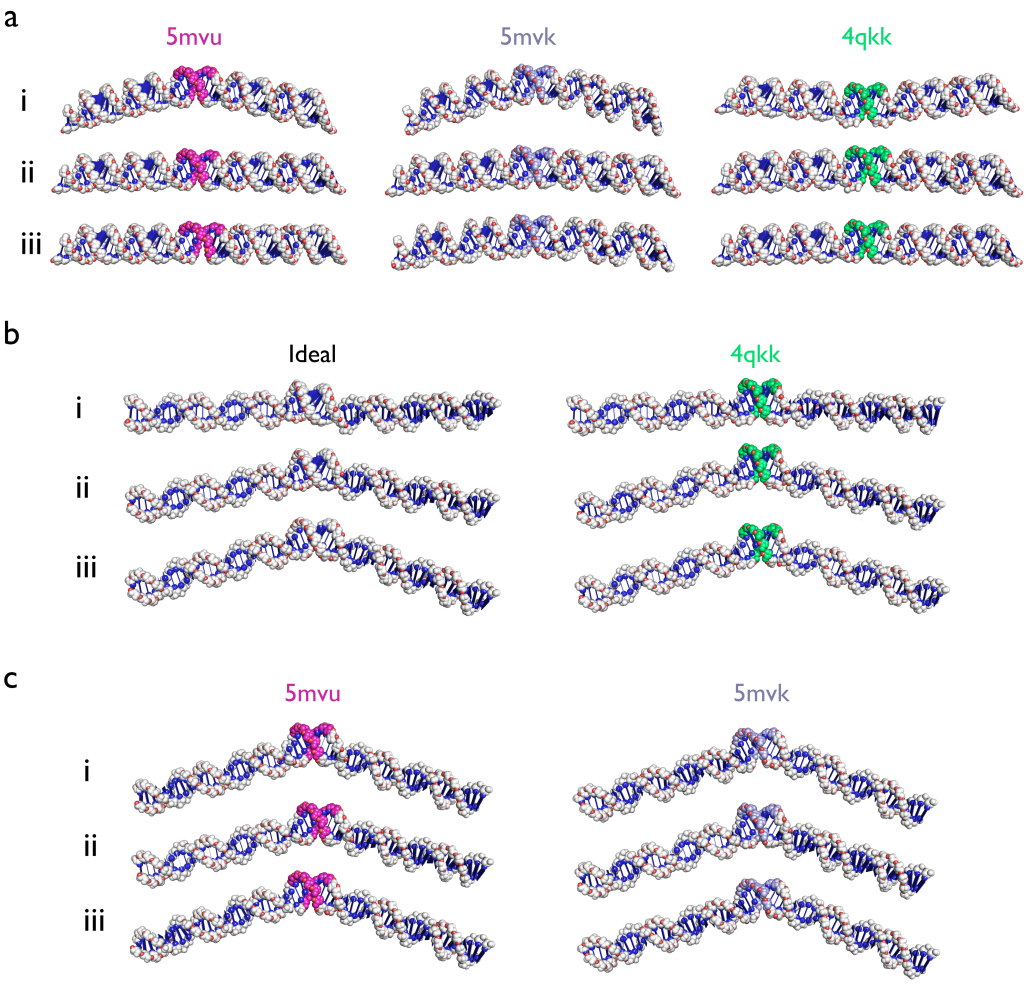


Figure A.4. 60-bp models of junctions between ideal A- or B-DNA and crystal structures of the same sequence containing either 5-formylcytosine or cytosine. The fC-containing structures 5mvu (magenta) and 4qkk (green), and unmodified structure 5mvk (blue) are flanked by either ideal A-DNA (*a*) or B-DNA (*b,c*). To illustrate how crystal packing disproportionately affects the terminal residues, three modelled variants of each crystal structure are shown: (*i*) the entire dodecamer d(CTAXGXGXGTAG)$_2$; (*ii*) the decamer core d(TAXGXGXGTA)$_2$; (*iii*) the octamer core d(AXGXGXGT)$_2$, where X = C (5mvk) or fC (5mvu, 4qkk). Reproduced from Hardwick *et al.*[1]

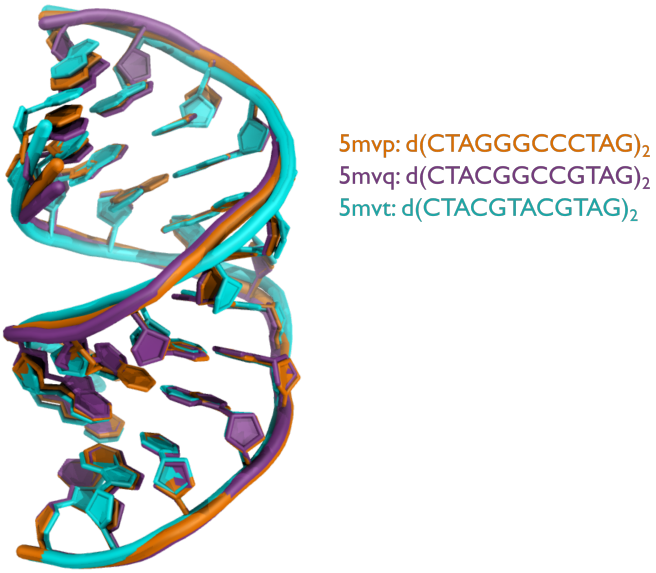


Figure A.5. Overlay of unmodified crystal structures 5mvp, 5mvq, and 5mvt. All duplexes crystallised in the same space group ($P3_221$) as the formylated duplex under low-salt conditions (5mvu) and its unmodified counterpart (5mvk). Conformational differences between 5mvu and 5mvk fall within the range exhibited by 5mvp, 5mvq, and 5mvt.

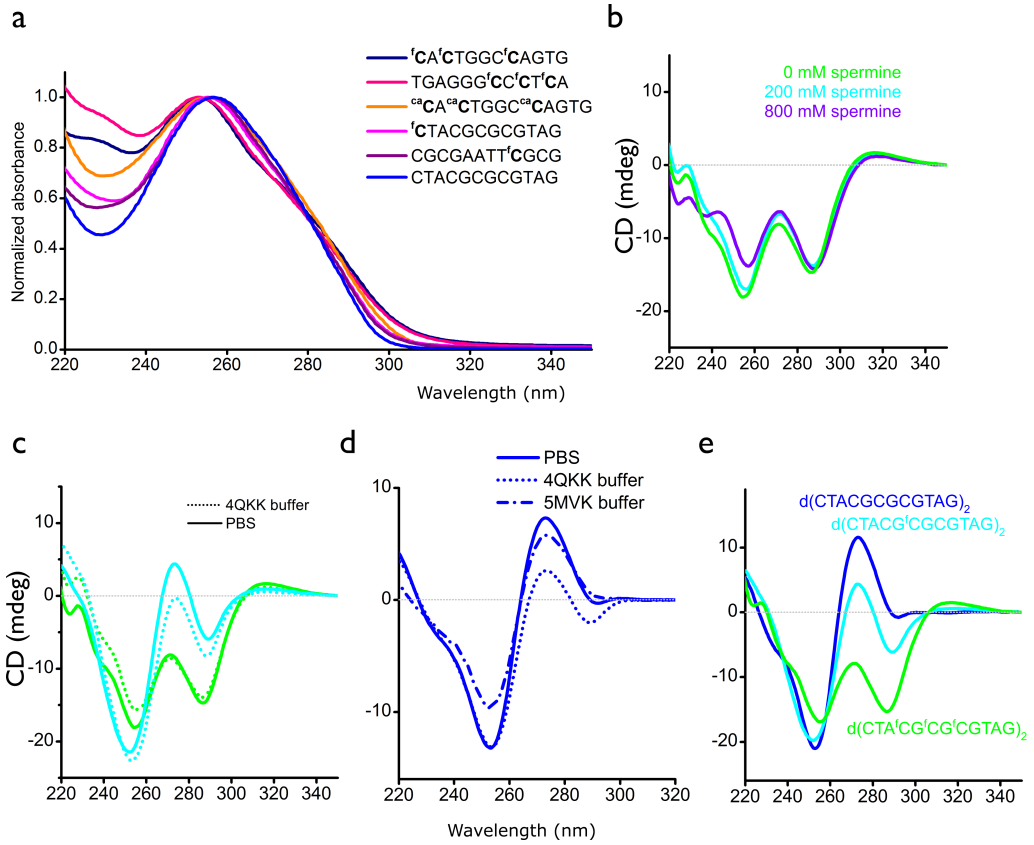


Figure A.6. Supplementary UV and CD spectra. (*a*) In all sequences tested, the addition of fC or caC to DNA duplexes caused perturbations to their UV spectra, suggesting that the unusual features of their CD spectra could be due to electronic, rather than conformational, changes. (*b*) The addition of ≥ 200 mM spermine to the formylated duplex d(CTA^{f}CG^{f}CG^{f}CGTAG)$_2$ (in PBS) did not lead to a transition to a B-type CD spectrum, contrary to the findings of Raiber *et al.* (*c*) CD spectra of the duplexes d(CTACG^{f}CGCGTAG)$_2$ (cyan) and d(CTA^{f}CG^{f}CG^{f}CGTAG)$_2$ (green) in the crystallisation buffer of 4qkk (dotted line) and PBS (solid line). (*d*) CD spectra of the unmodified duplex d(CTACGCGCGTAG)$_2$ in PBS (solid line), its crystallisation buffer (dashed line), and the crystallisation buffer of 4qkk (dotted line). (*e*) CD spectra of the duplexes analysed by NMR, in the NMR buffer (80 mM KCl, 22 mM NaPi, pH 7.3). Reproduced from Hardwick *et al.*[1]

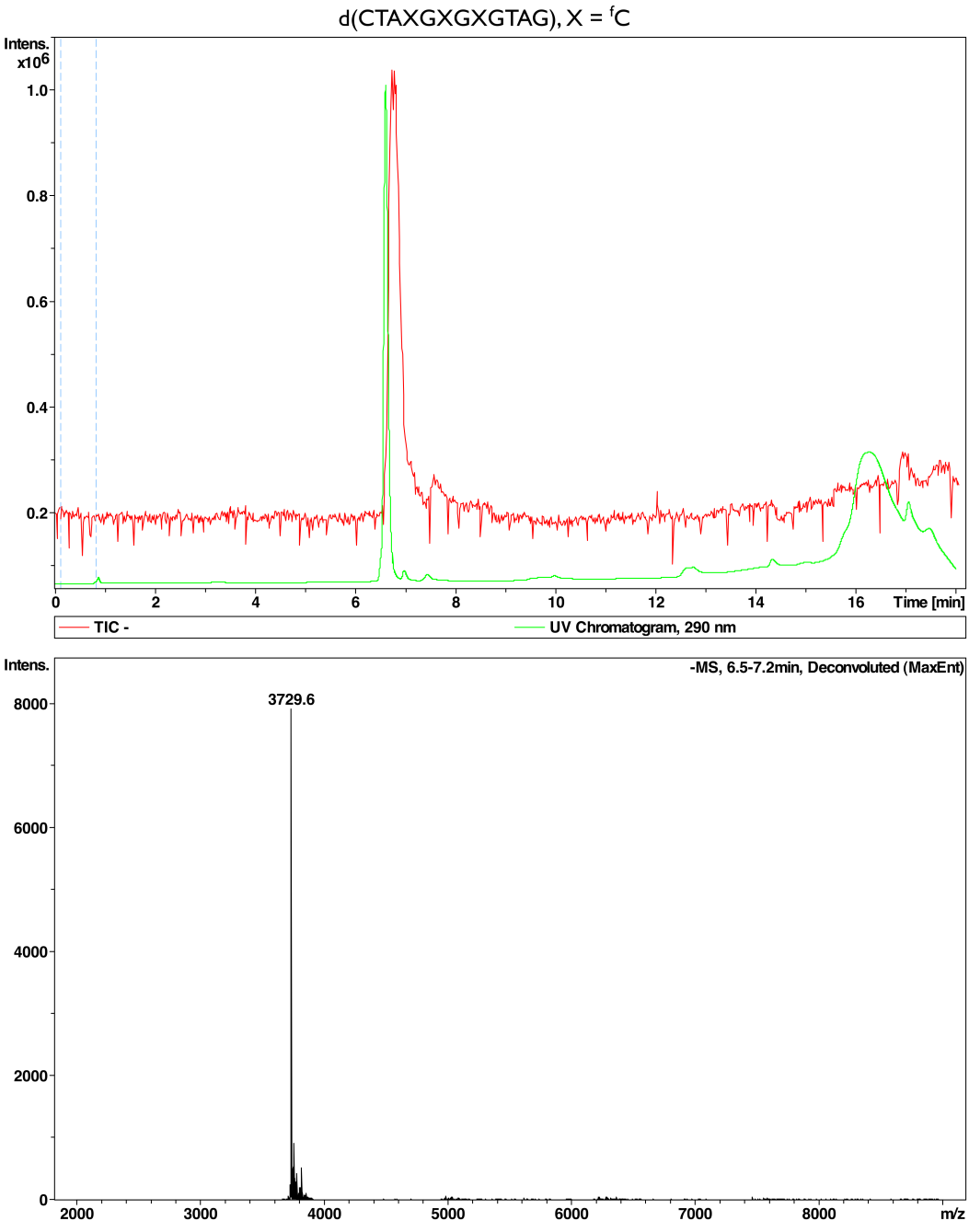


Figure A.7. Chromatograms and deconvoluted mass spectrum of d(CTA^{f}CG^{f}CG^{f}CGTAG).

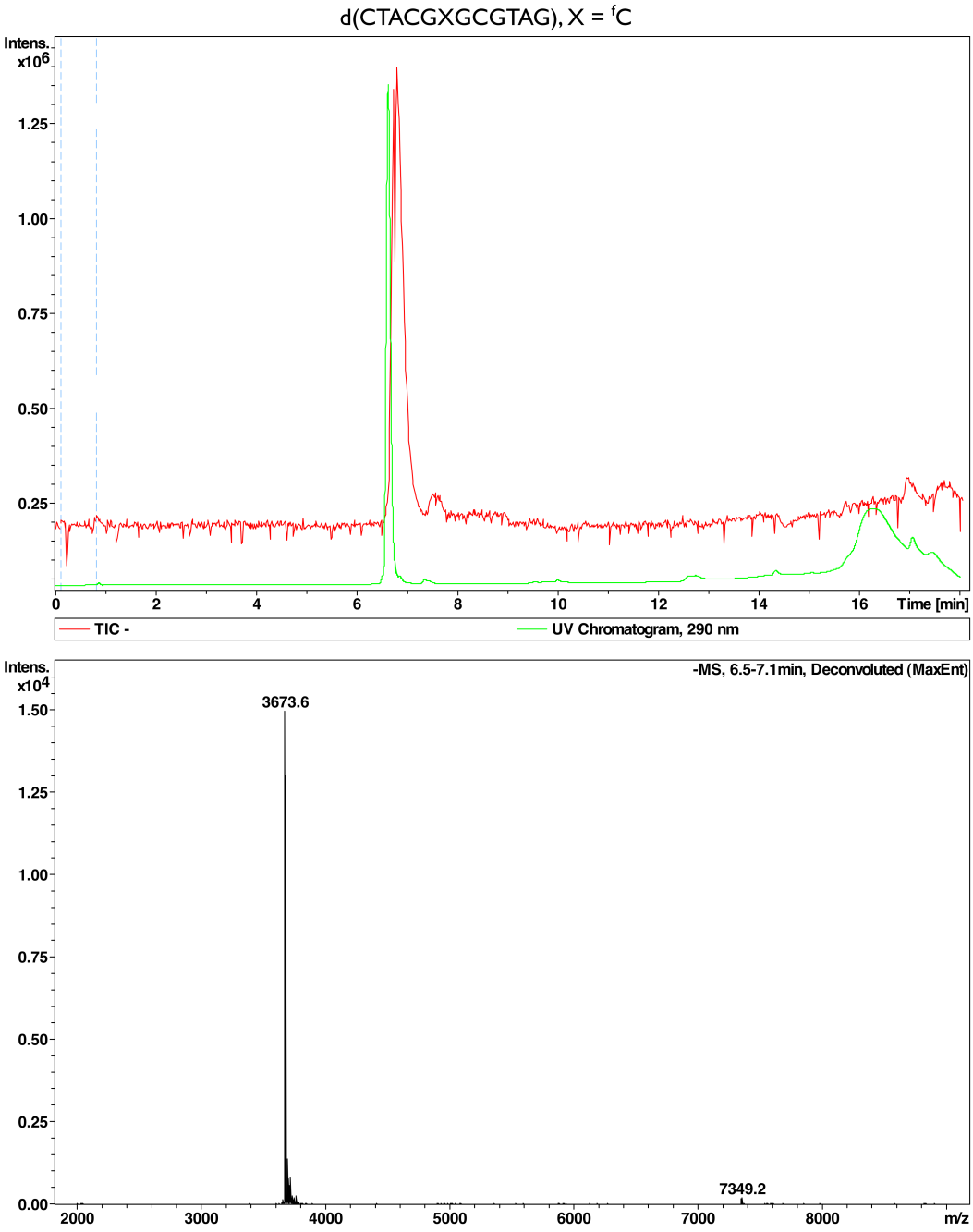


Figure A.8. Chromatograms and deconvoluted mass spectrum of d(CTACG^{f}CGCGTAG).

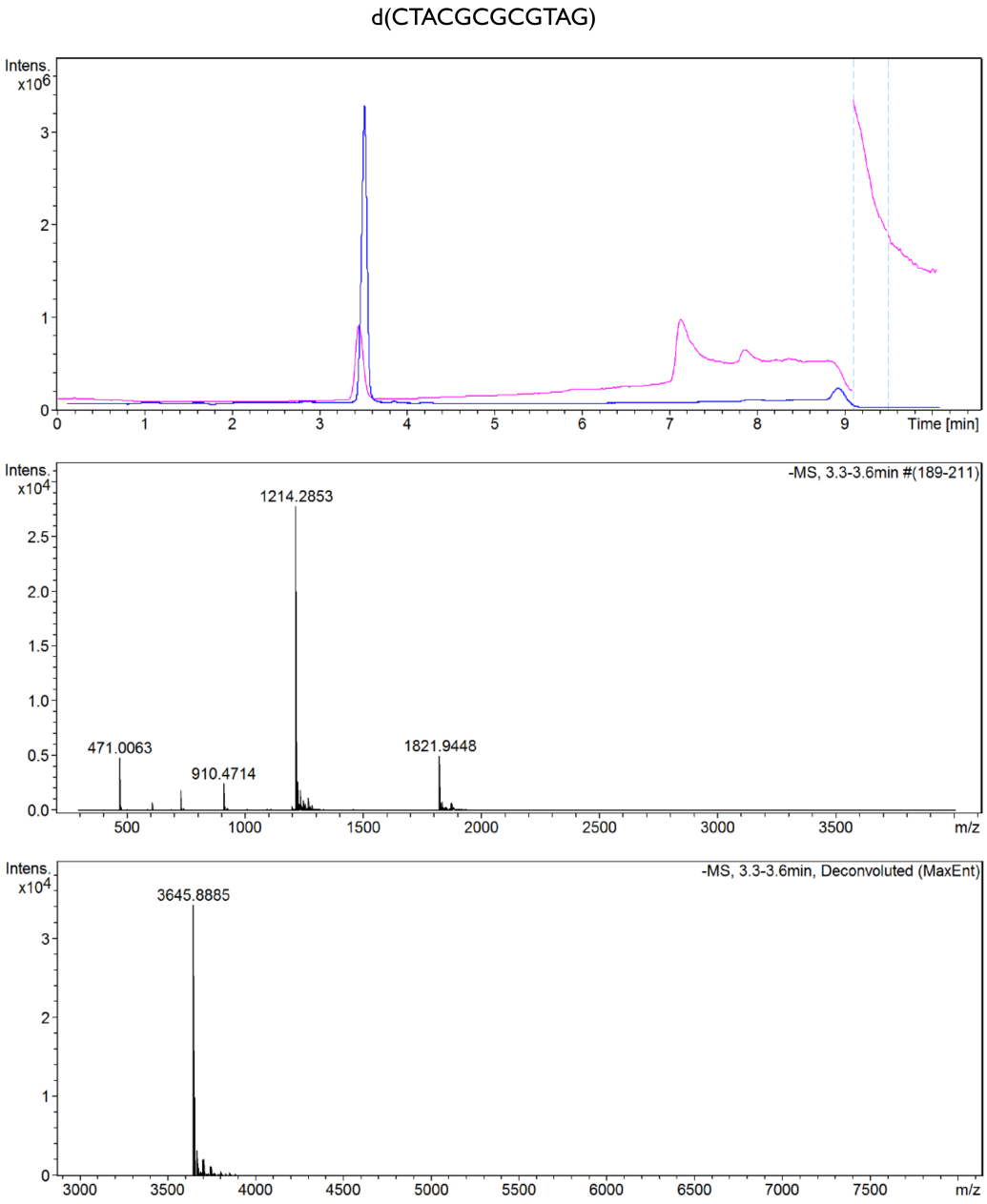


Figure A.9. Chromatograms, mass spectrum, and deconvoluted mass spectrum of d(CTACGCGCGTAG).

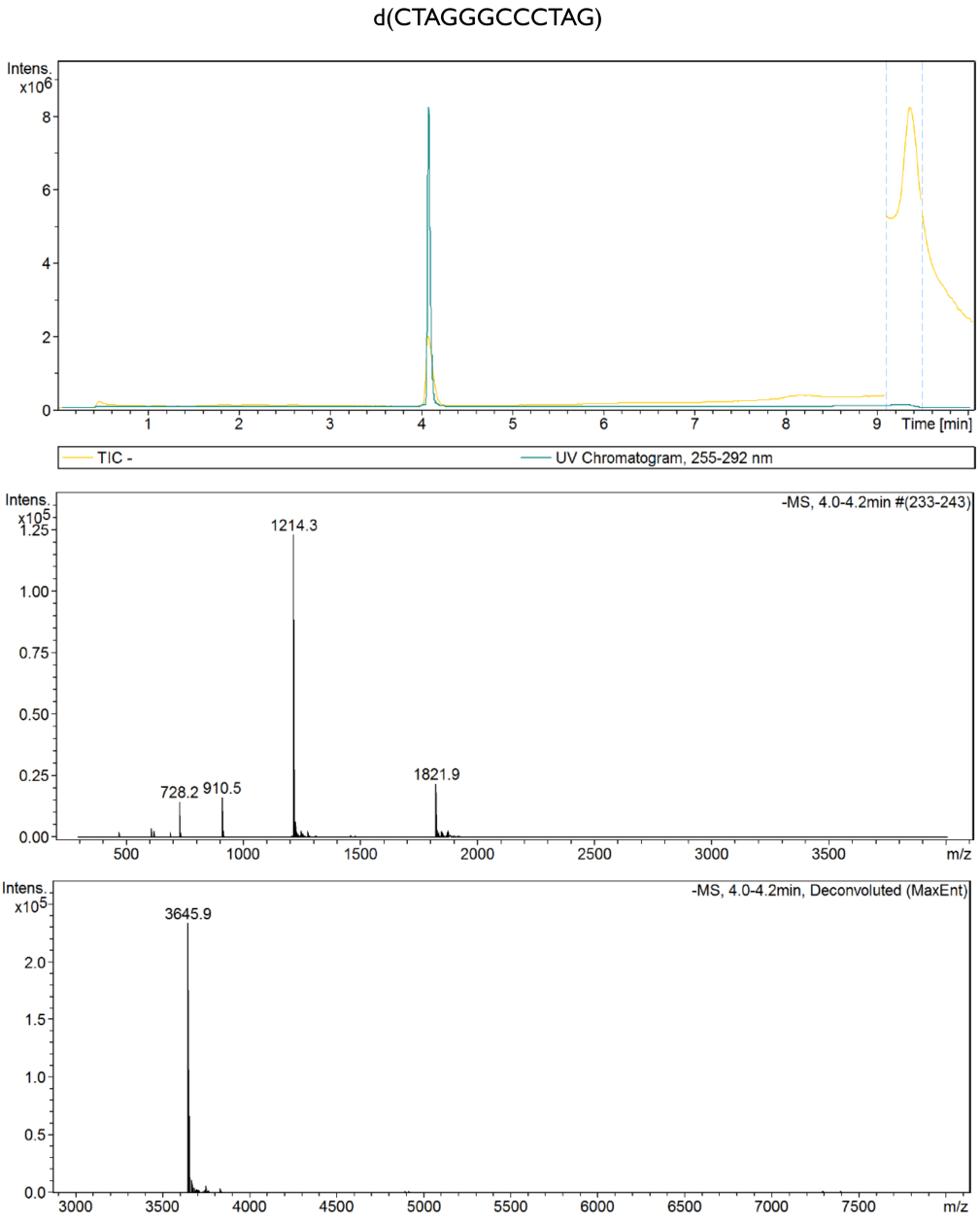


Figure A.10. Chromatograms, mass spectrum, and deconvoluted mass spectrum of d(CTAGGGCCCTAG).

d(CTACGTACGTAG)

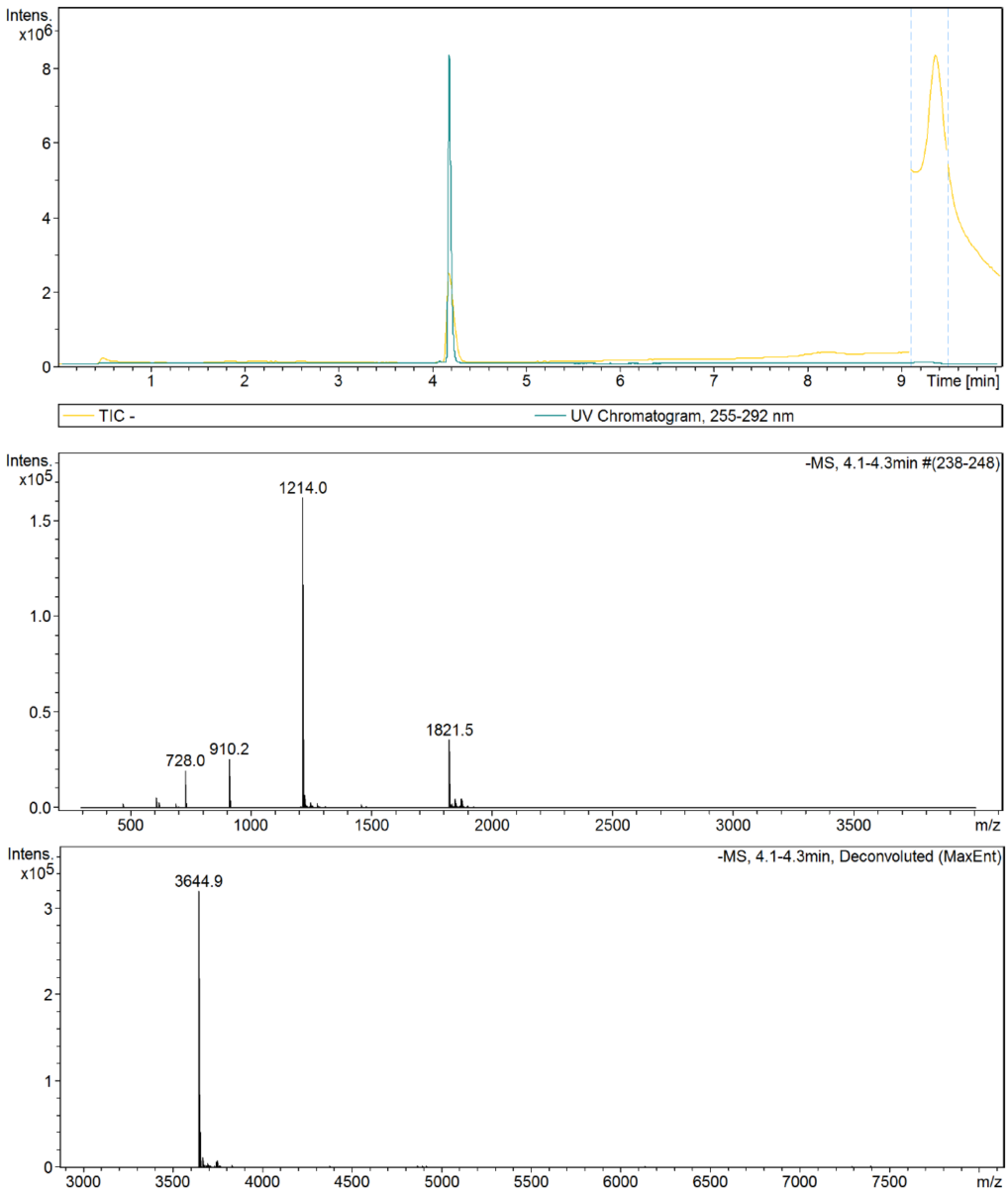


Figure A.11. Chromatograms, mass spectrum, and deconvoluted mass spectrum of d(CTACGTACGTAG).

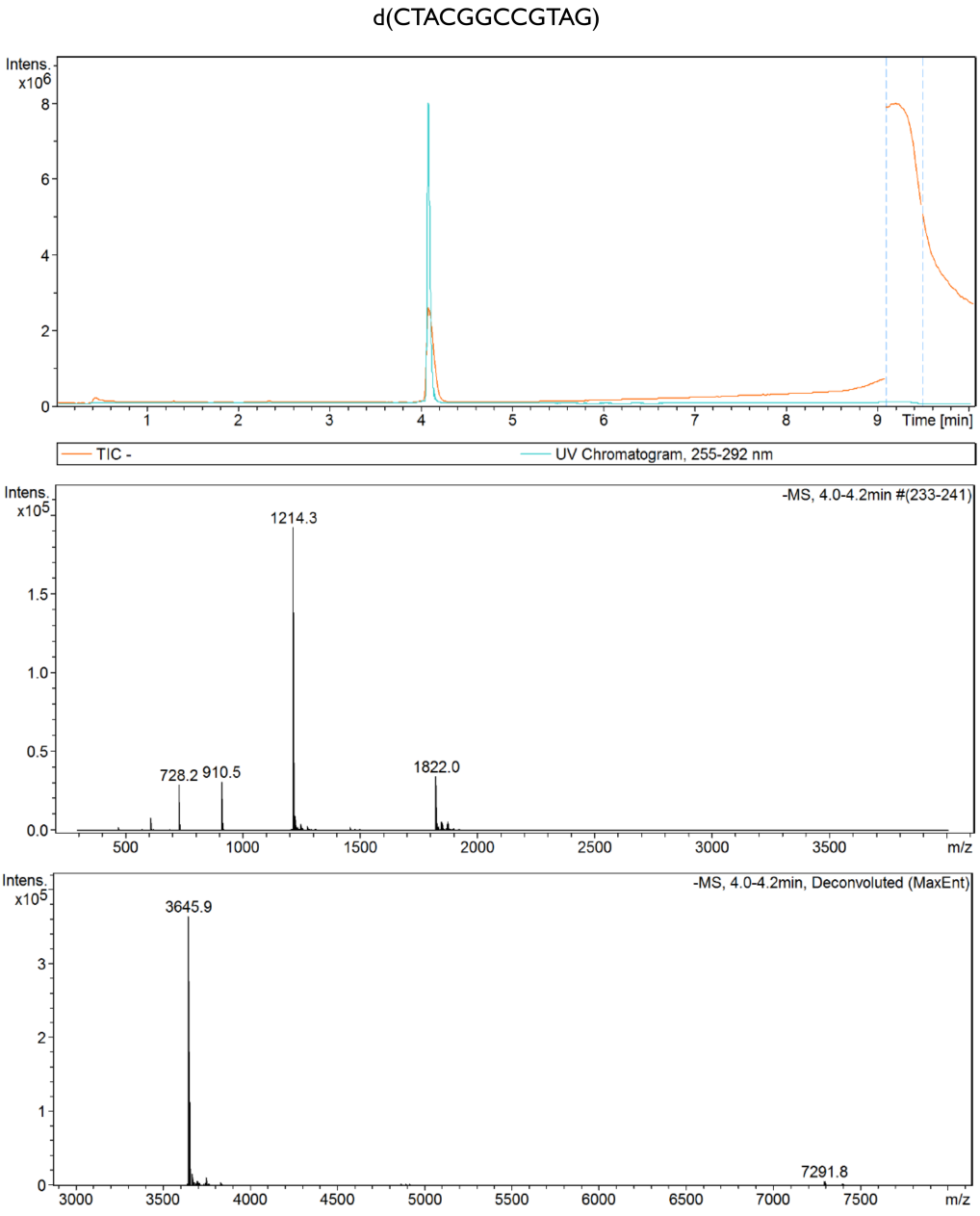


Figure A.12. Chromatograms, mass spectrum, and deconvoluted mass spectrum of d(CTACGGCCGTAG).

A.2 Supplementary tables

Table A.1. Data collection and refinement statistics of PDB 5mvl.

Data collection		
Space group	$P3_221$	
Cell dimensions:		
a, *b*, *c* (Å)	43.5, 43.5, 60.8	
α, β, γ (°)	90, 90, 120	
	Peak	*Remote*
Resolution (Å)	30.4–1.6 (1.63–1.60)[a]	30.4–1.4 (1.43–1.40)
R_{meas}	5.5 (52.9)	4.6 (117)
$I/\sigma(I)$	18.8 (4.1)	23.4 (2.0)
$CC_{1/2}$	0.999 (0.898)	0.999 (0.984)
Completeness (%)	99.1 (100)	99.9 (100)
Redundancy	7.9 (8.0)	9.1 (8.3)
Refinement		
Resolution (Å)		23.7–1.4
Reflections		13,446
R_{work}/R_{free}		15.7 / 18.2
Atoms:		
DNA		486
Mg^{2+}		1
Water		76
B factors:		
DNA		23.9
Water		33.9
RMS deviations:		
Bond lengths (Å)		0.009
Bond angles (°)		1.07

Both datasets were collected from a single crystal.

[a] *Values in parentheses are for the highest-resolution shell.*

Table A.2. Phasing statistics of 5mvl (MAD).

	Peak	*Remote*
X-ray wavelength (Å)	0.9202	1.2400
Mid-slope of anomalous normal probability	1.098	0.864
Correlation coefficient (CC_{all}/CC_{weak})	39.1/30.0	
Figure of merit	0.601	
No. of bromine sites	2	

Table A.3. Twinning statistics of 5mvu

Twin domain	Twin operator	Twin fraction
1	h, k, l	0.632
2	$-h, -k, l$	0.368

Table A.4. Crystallisation conditions of the six structures determined in Chapter 2

	5mvk	**5mvl**	**5mvp**	**5mvt**	**5mvq**	**5mvu**
Oligonucleotide (mM)	0.70	0.60	0.70	0.70	0.75	0.50
NaCl (mM)	12	80	80	12	80	0
KCl (mM)	80	12	12	80	0	0
$MgCl_2$ (mM)	0	20	20	40	20	18
$[Co(NH_3)]Cl_3$ (mM)	0	0	0	2	0	0
Spermine (mM)	12	12	12	0	12	3
2-Propanol (% v/v)	0	0	0	0	0	10
MPD (% v/v)	40	40	45	45	35	0
Sodium cacodylate (mM)	40	40	40	40	40	50
pH	7.0	6.0	7.0	5.5	6.0	6.5

Table A.5. Sequences and masses of oligonucleotides used in this study.

Oligo code	**Sequence**	**Molecular weight** / Da. Observed (calculated)
ODN1	CTACGCGCGTAG	3646 (3646)
ODN2	C^{b}UACGCGCGTAG	3711 (3711)
ODN3	CTAGGGCCCTAG	3646 (3646)
ODN4	CTACGTACGTAG	3645 (3645)
ODN5	CTACGGCCGTAG	3646 (3646)
ODN6	CTA^{f}CG^{f}CG^{f}CGTAG	3730 (3730)
ODN7	CTACG^{f}CGCGTAG	3674 (3674)
ODN8	^{f}CTACGCGCGTAG	3674 (3674)
ODN9	TGAGGG^{f}CC^{f}CT^{f}CA	3730 (3730)
ODN10	^{f}CA^{f}CTGGC^{f}CAGTG	3730 (3730)
ODN11	^{ca}CA^{ca}CTGGC^{ca}CAGTG	3775 (3775)

Table A.6. All-atom RMS deviation of the six crystal structures determined in this study. The number of atoms of each duplex included in the calculation is given in parentheses. Where the duplex sequences differ, all atoms except such differing nucleobases were included, while where only the modification differs, all atoms except those comprising the modification were included.

	RMSD (Å)					
	5mvk	5mvu	5mvl	5mvp	5mvt	5mvq
5mvk	–	0.65 (486)	0.36 (484)	0.64 (410)	1.13 (448)	0.66 (448)
5mvu	0.65 (486)	–	0.59 (484)	0.60 (410)	0.84 (448)	0.82 (448)
5mvl	0.36 (484)	0.59 (484)	–	0.60 (408)	1.13 (446)	0.57 (446)
5mvp	0.64 (410)	0.60 (410)	0.60 (408)	–	0.98 (410)	0.56 (448)
5mvt	1.13 (448)	0.84 (448)	1.13 (446)	0.98 (410)	–	1.27 (448)
Ideal A-DNA	1.52 (486)	1.27 (486)	1.45 (484)	1.25 (486)	1.27 (486)	1.32 (486)
Ideal B-DNA	6.87 (486)	6.54 (486)	6.75 (484)	6.61 (486)	6.37 (486)	6.71 (486)

Table A.7. Sugar puckering and glycosyl (χ) torsion angles of 4qkk, 5mvu (X = fC), and 5mvk (X = C). Values anomalous to A-DNA are shown in bold.

Puckering						
	4qkk		**5mvu**		**5mvk**	
Base pair	Strand I	Strand II	Strand I	Strand II	Strand I	Strand II
C·G	**C3′-*exo***	**C4′-*exo***	**C2′-*exo***	**C4′-*exo***	C3′-*endo*	**C4′-*exo***
T·A	**C1′-*exo***	**C1′-*exo***	C3′-*endo*	**C2′-*exo***	C3′-*endo*	**C2′-*exo***
A·T	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
X·G	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
G·**X**	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
X·G	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
G·**X**	C3′-*endo*	C3′-*endo*	C3′-*endo*	**C2′-*exo***	C3′-*endo*	C3′-*endo*
X·G	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
G·**X**	C3′-*endo*	C3′-*endo*	C3′-*endo*	**C2′-*exo***	C3′-*endo*	**C1′-*exo***
T·A	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	**C2′-*endo***
A·T	**C1′-*exo***	**C1′-*exo***	**C2′-*exo***	C3′-*endo*	C3′-*endo*	C3′-*endo*
G·C	**C4′-*exo***	**C3′-*exo***	C3′-*endo*	**C2′-*exo***	C3′-*endo*	C3′-*endo*
***χ*-Torsion angle** (°)						
	4qkk		**5mvu**		**5mvk**	
Base pair	Strand I	Strand II	Strand I	Strand II	Strand I	Strand II
C·G	−170.4	**−112.3**	−176.3	−152.9	−168.4	−161.4
T·A	**−75.8**	−128.1	−160.1	−174.8	−156.5	−163.3
A·T	−157.1	−153.8	−166.4	−149.3	−162.5	−155.0
X·G	−156.2	−177.3	−158.6	−174.6	−158.5	−171.5
G·**X**	−167.9	−157.4	−168.9	−156.4	−157.8	−159.1
X·G	−154.1	−163.4	−154.4	−163.7	−151.1	−156.8
G·**X**	−163.4	−154.1	−171.0	−171.3	−170.3	−170.1
X·G	−157.4	−167.9	−162.4	−158.1	−167.4	−161.3
G·**X**	−177.3	−156.2	−170.6	−159.9	−165.3	**−109.7**
T·A	−153.8	−157.1	−151.5	−159.9	−149.9	**−125.3**
A·T	−128.1	**−75.8**	−175.1	−164.5	−168.0	−157.5
G·C	**−112.3**	−170.4	−153.2	−177.7	−155.2	−171.5

Table A.8. α- and β-torsion angles of 4qkk, 5mvu (X = fC), and 5mvk (X = C).

α-*Torsion angle* (°)						
	4qkk		**5mvu**		**5mvk**	
Base pair	Strand I	Strand II	Strand I	Strand II	Strand I	Strand II
C·G	—	72.1	—	-48.6	—	-79.6
T·A	-65.2	-68.5	-44.9	-137.6	-71.9	128.5
A·T	68.6	-68.1	-101.6	-45.2	-69.6	-66.7
X·G	-68.4	150.4	-45.4	-103.7	-62.2	-64.8
G·**X**	-69.1	-69.9	-71.6	-54.8	-63.6	-73.5
X·G	-76.6	-77.6	-53.6	-51.1	-65.7	-61.1
G·**X**	-77.6	-76.6	124.1	-151.3	124.7	132.1
X·G	-69.9	-69.1	-63.8	-82.9	-61.1	-93.8
G·**X**	150.4	-68.4	-68.2	-84.1	-68.0	-73.1
T·A	-68.1	68.6	-62.8	-71.3	-68.5	-59.5
A·T	-68.5	-65.2	155.9	-62.1	127.3	-63.0
G·C	72.1	—	-47.2	—	-63.6	—

β-*Torsion angle* (°)						
	4qkk		**5mvu**		**5mvk**	
Base pair	Strand I	Strand II	Strand I	Strand II	Strand I	Strand II
C·G	—	-126.3	—	165.6	—	169.8
T·A	145.6	-178.0	164.7	128.8	-171.3	-169.1
A·T	-118.1	-177.1	162.1	173.1	161.7	-176.5
X·G	164.1	-168.3	157.3	163.2	167.5	165.6
G·**X**	176.8	170.6	-178.5	169.9	172.0	173.7
X·G	179.6	168.6	-176.5	171.1	176.8	-174.6
G·**X**	168.6	179.6	-169.2	116.7	-171.9	-174.5
X·G	170.6	176.8	178.3	-157.6	174.5	173.2
G·**X**	-168.3	164.1	166.8	167.9	168.7	172.6
T·A	-177.1	-118.1	-178.5	169.6	-173.5	-172.5
A·T	-178.0	145.6	-158.5	178.3	-163.4	179.7
G·C	-126.3	—	160.4	—	168.5	—

Table A.9. γ- and δ-torsion angles of 4qkk, 5mvu (X = fC), and 5mvk (X = C).

γ-Torsion angle (°)						
	4qkk		**5mvu**		**5mvk**	
Base pair	Strand I	Strand II	Strand I	Strand II	Strand I	Strand II
C·G	178.6	−151.1	−174.4	53.6	43.4	75.3
T·A	45.2	52.6	44.1	140.8	56.0	−161.5
A·T	−152.6	55.3	90.7	36.9	62.7	52.5
X·G	54.7	−178.9	41.6	104.8	52.9	67.0
G·**X**	60.6	62.2	62.0	45.1	57.6	54.5
X·G	55.6	61.4	36.2	46.0	48.1	44.3
G·**X**	61.4	55.6	−163.8	149.5	−165.1	−157.1
X·G	62.2	60.6	49.0	51.3	53.8	46.2
G·**X**	−178.9	54.7	67.8	68.8	60.4	52.5
T·A	55.3	−152.6	45.3	60.0	49.8	52.2
A·T	52.6	45.2	170.7	59.6	−165.8	50.2
G·C	−151.1	178.6	54.1	162.8	62.1	47.0
δ-Torsion angle (°)						
	4qkk		**5mvu**		**5mvk**	
Base pair	Strand I	Strand II	Strand I	Strand II	Strand I	Strand II
C·G	147.1	87.0	92.9	77.0	82.3	76.1
T·A	124.5	117.3	76.1	80.3	84.9	94.2
A·T	82.5	92.1	77.7	82.3	78.3	84.4
X·G	79.3	90.4	75.6	84.2	79.3	83.3
G·**X**	84.5	79.6	76.3	81.3	83.9	78.0
X·G	81.5	76.3	92.6	82.6	84.4	83.1
G·**X**	76.3	81.5	77.4	76.6	81.0	80.8
X·G	79.6	84.5	83.7	83.1	76.6	79.1
G·**X**	90.4	79.3	78.6	85.5	83.0	127.1
T·A	92.1	82.5	75.1	78.4	82.9	135.3
A·T	117.3	124.5	83.5	82.3	86.6	85.2
G·C	87.0	147.1	75.9	86.8	80.8	86.5

Table A.10. ϵ- and ζ-torsion angles of 4qkk, 5mvu (X = fC), and 5mvk (X = C).

ϵ-Torsion angle (°)						
	4qkk		**5mvu**		**5mvk**	
Base pair	Strand I	Strand II	Strand I	Strand II	Strand I	Strand II
C·G	−97.2	—	−131.4	—	−169.2	—
T·A	−149.3	−102.8	−149.3	−126.7	−146.7	−148.0
A·T	−144.4	−153.1	−137.1	−150.3	−148.6	−155.1
X·G	−159.9	−146.6	−162.1	−137.5	−151.0	−158.4
G·**X**	−155.8	−166.8	−149.8	−158.1	−156.7	−158.9
X·G	−147.8	−159.8	−147.3	−141.5	−150.5	−149.1
G·**X**	−159.8	−147.8	−150.3	−138.2	−149.4	−155.6
X·G	−166.8	−155.8	−143.0	−150.4	−144.3	−157.1
G·**X**	−146.6	−159.9	−156.4	−169.3	−164.3	−157.2
T·A	−153.1	−144.4	−160.1	−153.2	−150.9	−166.4
A·T	−102.8	−149.3	−127.2	−148.6	−141.8	−156.4
G·C	—	−97.2	—	−152.0	—	−157.0
ζ-Torsion angle (°)						
	4qkk		**5mvu**		**5mvk**	
Base pair	Strand I	Strand II	Strand I	Strand II	Strand I	Strand II
C·G	−58.9	—	−78.3	—	−61.1	—
T·A	−160.4	−103.3	−65.9	−75.0	−74.6	−54.0
A·T	−72.3	−62.8	−90.2	−41.7	−71.0	−75.5
X·G	−73.0	−67.3	−62.8	−84.1	−73.9	−74.0
G·**X**	−70.4	−70.1	−77.1	−70.5	−80.6	−72.5
X·G	−71.8	−65.3	−69.8	−75.8	−68.8	−65.4
G·**X**	−65.3	−71.8	−74.3	−84.6	−75.1	−72.3
X·G	−70.1	−70.4	−71.5	−27.8	−74.9	−68.6
G·**X**	−67.3	−73.0	−76.9	−68.5	−70.7	−127.8
T·A	−62.8	−72.3	−74.4	−68.4	−78.0	−80.8
A·T	−103.3	−160.4	−80.3	−70.3	−69.9	−71.1
G·C	—	−58.9	—	−66.6	—	−70.4

Table A.11. Shift, slide, roll, and twist values of PDBs 4qkk, 5mvu (X = fC), and 5mvk (X = C).

	Shift (Å)			**Slide** (Å)		
Base step	4qkk	5mvu	5mvk	4qkk	5mvu	5mvk
CT/AG	−0.85	−0.65	−0.58	1.23	−2.13	−2.11
TA/TA	−0.18	−0.60	−0.61	−0.76	−1.40	−1.40
A**X**/GT	0.28	0.23	0.61	−1.34	−1.30	−1.51
XG/**X**G	−0.51	−0.06	−0.09	−2.05	−2.00	−1.89
G**X**/G**X**	−0.67	−0.80	−0.78	−1.53	−2.13	−2.09
G**X**/G**X**	0.67	−0.90	−0.71	−1.48	−2.13	−2.75
XG/**X**G	0.51	0.91	1.19	−2.05	−2.02	−1.32
GT/A**X**	−0.28	−0.87	−0.92	−1.34	−1.38	−1.06
TA/TA	0.18	0.62	0.70	−0.76	−1.48	−1.25
AG/CT	0.85	0.68	0.53	1.23	−2.21	−1.69
	Roll (°)			**Twist** (°)		
Base step	4qkk	5mvu	5mvk	4qkk	5mvu	5mvk
CT/AG	13.21	1.59	2.74	10.11	29.53	27.30
TA/TA	10.03	8.88	10.54	25.74	23.12	28.82
A**X**/GT	5.16	2.51	3.08	36.78	37.18	35.10
XG/**X**G	12.39	12.84	12.79	26.58	26.55	24.99
G**X**/G**X**	3.75	5.17	8.44	36.27	35.43	34.08
XG/**X**G	16.15	17.76	14.42	26.18	25.14	24.58
G**X**/G**X**	3.75	4.34	9.67	36.27	24.30	22.08
XG/**X**G	12.39	16.50	18.59	26.58	30.42	37.30
GT/A**X**	5.16	4.17	5.13	36.78	35.05	29.47
TA/TA	10.03	8.64	17.61	25.74	33.56	28.51
AG/CT	13.21	−1.05	0.18	10.11	28.33	30.68

Table A.12. Rise and tilt values of values of PDBs 4qkk, 5mvu (X = fC), and 5mvk (X = C).

	Rise (Å)			**Tilt** (°)		
Base step	4qkk	5mvu	5mvk	4qkk	5mvu	5mvk
CT/AG	2.74	2.99	3.05	1.12	−3.77	−3.58
TA/TA	3.22	3.00	2.93	−3.40	−0.22	−1.16
A**X**/GT	3.32	3.28	3.26	1.11	0.86	0.53
XG/**X**G	3.70	3.49	3.24	−4.74	−4.94	−0.74
G**X**/G**X**	3.13	3.20	3.26	−2.10	−1.31	0.69
XG/**X**G	3.27	3.23	3.23	−0.00	1.61	−0.66
G**X**/G**X**	3.13	3.05	3.19	2.10	−2.51	−1.24
XG/**X**G	3.70	3.36	3.71	4.74	2.24	10.30
GT/A**X**	3.32	3.19	3.00	−1.11	−0.92	−1.96
TA/TA	3.22	3.21	2.98	3.40	−1.80	−0.90
AG/CT	2.74	2.87	3.04	−1.12	4.50	3.28

Table A.13. Shear, strech, stagger, and buckle values of PDBs 4qkk, 5mvu (X = fC), and 5mvk (X = C).

	Shear (Å)			**Stretch** (Å)		
Base pair	4qkk	5mvu	5mvk	4qkk	5mvu	5mvk
C·G	0.04	0.11	0.07	−0.03	−0.05	−0.17
T·A	−0.12	0.25	−0.09	−0.18	−0.03	−0.20
A·T	0.07	−0.27	−0.02	−0.07	−0.04	−0.17
X·G	0.26	−0.09	0.17	−0.09	0.06	−0.06
G·**X**	−0.24	−0.60	−0.24	−0.09	−0.12	−0.13
X·G	0.10	−0.14	0.12	−0.11	−0.02	−0.17
G·**X**	−0.10	−0.16	−0.06	−0.11	−0.04	−0.12
X·G	0.24	0.05	0.13	−0.09	−0.04	−0.20
G·**X**	−0.26	−0.38	−0.21	−0.09	−0.12	−0.05
T·A	−0.07	−0.66	−0.18	−0.07	−0.18	−0.16
A·T	0.12	0.42	0.17	−0.18	−0.00	−0.17
G·C	−0.04	0.35	−0.13	−0.03	0.16	−0.09
	Stagger (Å)			**Buckle** (Å)		
Base pair	4qkk	5mvu	5mvk	4qkk	5mvu	5mvk
C·G	0.36	0.39	0.18	−27.50	−7.70	−4.64
T·A	0.08	0.49	0.38	12.13	3.43	2.25
A·T	0.13	0.20	0.22	17.71	7.54	7.05
X·G	−0.20	−0.13	0.11	15.74	9.12	6.56
G·**X**	−0.01	0.11	0.14	−4.95	−6.08	0.70
X·G	0.02	0.12	−0.23	1.07	−2.34	2.94
G·**X**	0.02	0.07	0.08	−1.07	−2.57	1.17
X·G	−0.01	0.00	−0.13	4.95	8.65	5.36
G·**X**	−0.20	0.08	−0.43	−15.74	−4.91	−20.97
T·A	0.13	0.10	−0.06	−17.71	−2.77	−8.61
A·T	0.08	0.60	0.38	−12.13	−3.57	−2.74
G·C	0.36	0.53	0.49	27.50	11.58	6.99

Table A.14. Propeller twist and base-pair opening values of PDBs 4qkk, 5mvu (X = fC), and 5mvk (X = C).

	Propeller (°)			**Opening** (°)		
Base pair	4qkk	5mvu	5mvk	4qkk	5mvu	5mvk
C·G	9.56	−2.71	−6.84	1.77	1.55	0.30
T·A	11.69	−13.01	−10.87	−1.37	1.36	1.62
A·T	−16.39	−14.09	−16.51	1.33	5.00	−0.42
X·G	−18.27	−18.57	−16.98	1.84	2.05	2.19
G·**X**	−10.43	−8.76	−7.20	1.69	3.14	3.83
X·G	−8.89	−13.76	−11.90	−0.42	−0.36	0.86
G·**X**	−8.89	−12.98	−3.05	−0.42	−3.50	0.20
X·G	−10.43	−9.51	−7.09	1.69	−1.48	−0.01
G·**X**	−18.27	−14.14	−14.41	1.84	1.10	1.92
T·A	−16.39	−18.04	−8.07	1.33	−6.39	−0.08
A·T	11.69	−4.08	−12.78	−1.37	−3.73	−2.04
G·C	9.56	−0.51	−9.85	1.77	1.61	−1.65

Table A.15. X-displacement, Y-displacement, helical rise, and inclination values of 4qkk, 5mvu (X = fC), and 5mvk (X = C).

	X-displacement (Å)			**Y-displacement** (Å)		
Base step	4qkk	5mvu	5mvk	4qkk	5mvu	5mvk
CT/AG	−4.79	−4.44	−5.02	3.69	0.55	0.43
TA/TA	−3.86	−5.54	−4.37	−0.40	1.34	0.97
A**X**/GT	−2.79	−2.37	−2.94	−0.29	−0.24	−0.93
XG/**X**G	−6.66	−6.38	−6.50	−0.02	−0.84	0.04
G**X**/G**X**	−2.85	−2.97	−3.16	0.80	1.13	1.39
XG/**X**G	−5.67	−6.78	−6.95	−0.00	−1.28	−2.26
G**X**/G**X**	−2.85	−6.13	−9.10	−0.80	1.43	1.39
XG/**X**G	−6.66	−5.66	−3.86	0.02	−1.22	−0.49
GT/A**X**	−2.79	−2.86	−3.00	0.29	1.30	1.42
TA/TA	−3.86	−3.71	−4.53	0.40	−1.29	−1.34
AG/CT	−4.79	−4.24	−3.21	−3.69	−0.46	−0.41
	Helical rise (Å)			**Inclination** (°)		
Base step	4qkk	5mvu	5mvk	4qkk	5mvu	5mvk
CT/AG	2.58	2.93	2.88	52.64	3.11	5.75
TA/TA	2.73	2.32	2.31	21.40	21.19	20.34
A**X**/GT	3.12	3.19	3.13	8.12	3.94	5.09
XG/**X**G	2.56	2.28	2.05	25.06	25.87	27.37
G**X**/G**X**	3.00	3.00	2.88	6.00	8.44	14.13
XG/**X**G	2.01	1.48	1.73	32.06	35.64	30.71
G**X**/G**X**	3.00	2.71	1.87	6.00	10.16	23.81
XG/**X**G	2.56	2.08	2.98	25.06	28.89	26.60
GT/A**X**	3.12	3.03	2.83	8.12	6.89	9.98
TA/TA	2.73	2.72	1.89	21.40	14.65	32.17
AG/CT	2.58	3.02	3.07	52.64	−2.13	0.33

Table A.16. Tip and helical twist of 4qkk, 5mvu (X = fC) & 5mvk (X = C).

	Tip (°)			**Helical twist** (°)		
Base step	4qkk	5mvu	5mvk	4qkk	5mvu	5mvk
CT/AG	-4.45	7.35	7.52	16.66	29.81	27.66
TA/TA	7.26	0.52	2.24	27.80	24.74	30.67
A**X**/GT	-1.74	-1.35	-0.88	37.14	37.27	35.23
XG/**X**G	9.60	9.95	1.59	29.65	29.84	28.03
G**X**/G**X**	3.36	2.13	-1.15	36.52	35.81	35.09
XG/**X**G	0.00	-3.24	1.40	30.69	30.74	28.45
G**X**/G**X**	-3.36	5.89	3.05	36.52	24.80	24.11
XG/**X**G	-9.60	-3.92	-14.73	29.65	34.58	42.74
GT/A**X**	1.74	1.52	3.81	37.14	35.30	29.96
TA/TA	-7.26	3.05	1.65	27.80	34.67	33.42
AG/CT	4.45	-9.11	-6.17	16.66	28.70	30.85

Table A.17. Shift, slide, roll, and twist values of PDBs 5mvq, 5mvp, and 5mvt.

	Shift (Å)			**Slide** (Å)		
Base step	5mvq	5mvp	5mvt	5mvq	5mvp	5mvt
1	−0.68	−0.21	−0.06	−2.14	−1.86	−0.35
2	−0.97	−0.66	−0.86	−1.40	−1.36	−0.93
3	0.69	0.39	0.68	−1.30	−1.52	−1.49
4	0.19	0.49	0.06	−1.80	−2.01	−2.05
5	−0.93	−1.07	−1.04	−1.43	−1.49	−1.33
6	1.28	1.15	0.86	−2.45	−2.13	−1.48
7	−1.00	−1.22	−0.11	−2.23	−2.10	−1.52
8	1.18	1.13	0.46	−1.14	−1.28	−1.95
9	−0.93	−1.03	−0.82	−1.01	−1.47	−1.36
10	0.96	0.55	0.53	−1.07	−1.24	−1.33
11	0.15	0.30	0.64	−1.43	−1.40	−1.74
	Roll (°)			**Twist** (°)		
Base step	5mvq	5mvp	5mvt	5mvq	5mvp	5mvt
1	1.09	1.22	7.31	28.96	29.92	24.68
2	14.06	12.35	5.87	26.84	30.84	33.43
3	6.76	7.05	4.98	35.65	33.22	34.74
4	14.80	11.58	10.41	25.74	25.74	27.16
5	10.61	10.31	9.20	34.42	35.49	33.99
6	10.41	10.92	18.58	25.01	27.02	23.35
7	9.66	14.12	5.56	23.86	23.88	29.96
8	20.56	9.91	13.51	37.15	36.59	28.13
9	4.59	8.41	6.62	30.84	28.66	34.65
10	17.30	16.17	13.50	25.24	30.28	31.98
11	−0.85	1.00	0.98	32.62	30.89	29.45

Table A.18. Rise and tilt values values of PDBs 5mvq, 5mvp, and 5mvt.

	Rise (Å)			**Tilt** (°)		
Base step	5mvq	5mvp	5mvt	5mvq	5mvp	5mvt
1	3.08	3.03	2.76	−0.72	−2.53	1.56
2	2.99	3.22	3.28	−1.22	−0.71	−7.79
3	3.15	3.25	3.40	0.81	−1.46	0.19
4	3.31	3.34	3.48	1.47	2.31	−1.62
5	3.41	3.33	3.07	−2.15	−3.11	−1.24
6	3.06	3.05	3.20	1.66	1.70	1.94
7	3.03	2.97	3.20	−1.29	−0.28	0.50
8	3.83	3.24	3.37	10.35	3.96	1.03
9	3.03	3.36	3.17	−1.47	1.67	0.97
10	3.01	3.16	3.20	−0.36	0.37	−1.55
11	2.91	3.00	2.92	0.11	1.03	2.49

Table A.19. X-displacement, Y-displacement, helical rise, and inclination values of PDBs 5mvq, 5mvp, 5mvt.

	X-displacement (Å)			**Y-displacement** (Å)		
Base step	5mvq	5mvp	5mvt	5mvq	5mvp	5mvt
1	−4.50	−3.81	−2.47	1.20	−0.07	0.49
2	−5.02	−4.27	−2.46	1.66	1.05	0.24
3	−2.97	−3.66	−3.22	−1.01	−0.88	−1.10
4	−6.24	−6.52	−6.24	−0.11	−0.51	−0.45
5	−3.81	−3.69	−3.42	1.21	1.28	1.56
6	−7.34	−6.19	−6.12	−2.40	−1.98	−1.33
7	−7.18	−6.92	−3.93	1.94	2.50	0.30
8	−3.81	−3.17	−5.95	−0.44	−1.25	−0.68
9	−2.67	−4.51	−3.15	1.47	2.34	1.49
10	−4.88	−4.35	−4.09	−1.88	−0.88	−1.10
11	−2.41	−2.80	−3.59	−0.26	−0.39	−0.79
	Helical rise (Å)			**Inclination** (°)		
Base step	5mvq	5mvp	5mvt	5mvq	5mvp	5mvt
1	3.02	2.96	2.54	2.17	2.35	16.61
2	2.06	2.52	3.19	27.97	22.15	9.94
3	2.88	2.86	3.16	10.91	12.15	8.28
4	2.00	2.28	2.53	30.21	24.42	21.19
5	2.91	2.89	2.66	17.41	16.45	15.39
6	1.97	2.12	1.67	22.78	22.22	38.88
7	2.04	1.53	2.88	22.22	30.90	10.64
8	3.05	2.91	2.23	29.11	15.39	25.97
9	2.89	2.77	2.85	8.56	16.53	10.99
10	1.89	2.24	2.43	34.85	28.54	23.24
11	2.94	2.96	2.91	−1.51	1.87	1.92

Table A.20. Tip and helical twist values of PDBs 5mvq, 5mvp, and 5mvt.

	Tip (°)			**Helical twist** (°)		
Base step	5mvq	5mvp	5mvt	5mvq	5mvp	5mvt
1	1.45	4.89	−3.54	28.99	30.05	25.77
2	2.42	1.27	13.19	30.27	33.17	34.78
3	−1.31	2.52	−0.31	36.28	33.96	35.09
4	−2.99	−4.88	3.31	29.67	28.28	29.09
5	3.53	4.97	2.08	36.03	37.04	35.20
6	−3.65	−3.45	−4.06	27.11	29.16	29.82
7	2.97	0.60	−0.96	25.74	27.69	30.46
8	−14.66	−6.14	−1.98	43.49	38.07	31.16
9	2.74	−3.27	−1.62	31.21	29.88	35.27
10	0.73	−0.65	2.68	30.52	34.23	34.68
11	−0.20	−1.93	−4.88	32.63	30.93	29.57

Table A.21. Sugar puckering and glycosyl (χ) torsion angles of 5mvq, 5mvp, and 5mvt.

Puckering						
	5mvq		**5mvp**		**5mvt**	
Base pair	Strand I	Strand II	Strand I	Strand II	Strand I	Strand II
1	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
2	C3′-*endo*	C3′-*endo*	C3′-*endo*	C2′-*exo*	C2′-*endo*	C3′-*endo*
3	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
4	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
5	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
6	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
7	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
8	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
9	C3′-*endo*	C1′-*exo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
10	C3′-*endo*	C2′-*endo*	C3′-*endo*	C4′-*exo*	C3′-*endo*	C3′-*endo*
11	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
12	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*
χ-*Torsion angle* (°)						
	5mvq		**5mvp**		**5mvt**	
Base pair	Strand I	Strand II	Strand I	Strand II	Strand I	Strand II
1	−174.2	−163.4	−173.2	−145.8	−155.3	−132.6
2	−154.5	−167.3	−159.3	−166.7	−117.2	−147.3
3	−159.4	−151.5	−164.2	−150.8	−151.1	−156.0
4	−157.7	−166.0	−165.7	−163.7	−159.7	−173.4
5	−164.7	−157.7	−173.0	−157.9	−168.3	−157.5
6	−158.3	−159.2	−160.9	−159.0	−143.1	−157.3
7	−168.8	−172.2	−169.9	−169.9	−156.6	−154.4
8	−161.4	−160.1	−160.6	−155.1	−159.5	−163.0
9	−166.8	−108.4	−160.1	−166.6	−168.4	−161.1
10	−151.7	−120.9	−150.2	−149.6	−152.7	−155.7
11	−166.0	−149.6	−160.8	−156.1	−162.6	−159.7
12	−149.7	−171.6	−145.4	−169.5	−153.9	−172.9

Table A.22. Chemical shift assignments of the formylated duplexes.

^{1}H Chemcal shifts (ppm) of d(CTA^{f}CG^{f}CG^{f}CGTAG)$_2$								
Nt	H8/6	H2/5/Me	H1′	H2′	H2″	H3′	H4′	NH/NH$_2$ (288 K)
C1	7.84	5.94	5.88	2.12	2.53	4.67	4.10	–
T2	7.53	1.71	5.65	2.26	2.48	4.88	4.18	13.73
A3	8.34	7.75	6.29	2.76	3.00	5.07	4.45	–
^{f}C4	8.16	9.24	5.58	2.02	2.41	4.78	4.21	9.00/7.93
G5	7.79	–	5.94	2.55	2.80	4.95	4.34	12.56
^{f}C6	8.12	9.20	5.66	2.02	2.41	4.78	4.22	8.98/7.82
G7	7.83	–	5.97	2.61	2.83	4.96	4.35	12.58
^{f}C8	8.15	9.23	5.68	2.10	2.44	4.79	4.18	9.00/7.87
G9	7.93	–	5.96	2.59	2.78	4.96	4.35	12.4
T10	7.22	1.46	5.55	1.96	2.26	4.83	4.13	13.56
A11	8.19	7.50	6.07	2.72	2.88	5.02	4.40	–
G12	7.72	–	6.01	2.44	2.25	4.64	4.17	13.08
^{1}H Chemcal shifts (ppm) of d(CTACG^{f}CGCGTAG)$_2$								
Nt	H8/6	H2/5/Me	H1′	H2′	H2″	H3′	H4′	NH/NH$_2$ (288 K)
C1	7.85	5.94	5.88	2.15	2.54	4.68	4.10	–
T2	7.56	1.72	5.66	2.24	2.52	4.89	4.18	13.79
A3	8.33	7.72	6.24	2.74	2.89	5.04	4.44	–
C4	7.24	5.25	5.55	1.94	2.29	4.81	4.26	–
G5	7.74	–	5.94	2.56	2.81	4.95	4.34	12.76
^{f}C6	8.13	9.26	5.63	2.01	2.43	4.63	4.17	–
G7	7.91	–	5.89	2.63	2.74	4.96	4.35	12.95
C8	7.25	5.24	5.70	1.99	2.39	nd	4.16	–
G9	7.89	–	5.93	2.62	2.75	4.96	4.35	12.57
T10	7.21	1.55	5.55	1.96	2.27	4.63	4.34	13.64
A11	8.19	7.53	6.07	2.70	2.87	5.02	4.39	–
G12	7.72	–	6.01	2.45	2.28	4.62	4.18	13.10

Table A.23. Chemical shift assignments of the unmodified duplex.

	^{1}H Chemcal shift (ppm)							
Nt	H8/6	H2/5/Me	H1′	H2′	H2″	H3′	H4′	NH/NH_2 (288 K)
C1	7.85	5.93	5.88	2.15	2.55	4.66	4.10	–
T2	7.57	1.73	5.68	2.25	2.53	4.90	4.20	13.8
A3	8.34	7.58	6.23	2.75	2.90	5.05	4.44	–
C4	7.25	5.27	5.55	1.93	2.32	4.78	4.16	6.58/8.19
G5	7.83	–	5.85	2.59	2.70	4.95	4.35	12.88
C6	7.27	5.30	5.65	1.96	2.36	4.78	4.16	6.39/8.29
G7	7.86	–	5.85	2.61	2.75	4.96	4.35	12.94
C8	7.24	5.27	5.65	1.98	2.37	4.82	4.17	6.41/8.29
G9	7.89	–	5.92	2.61	2.75	4.95	4.35	12.8
T10	7.22	1.52	5.56	1.96	2.27	4.79	4.18	13.7
A11	8.19	7.53	6.06	2.70	2.87	5.01	4.41	–
G12	7.71	–	6.01	2.43	2.27	4.63	4.17	13.10

BIBLIOGRAPHY

1. Hardwick, J. S. *et al.* 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.* **24,** 544–552 (2017).

B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

B.1 Supplementary figures

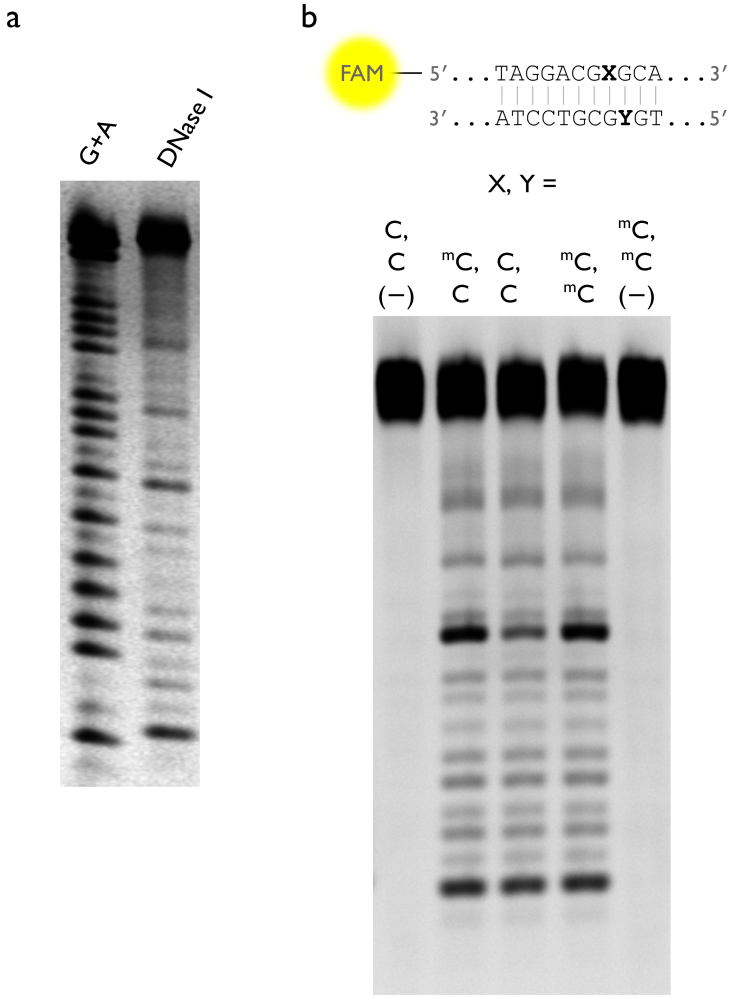


Figure B.1. Example gels showing Maxam–Gilbert sequencing and DNase I cleavage products. (*a*) Denaturing polyacrylamide gel showing the products of the 'G+A' Maxam–Gilbert sequencing reaction, alongside the DNase I cleavage products. In both cases, the duplex used was that containing unmodified cytosine. Note the difference in mobility of the 'G+A' fragments compared to the corresponding DNase I cleavage products, which are shifted down by ~0.5–1 nt. This is because the 'G+A' reaction produces a 3′-phosphate after strand cleavage, whereas the DNase I reaction leaves a 3′-OH.[1] (*b*) Denaturing polyacrylamide gel showing the DNase I cleavage products of three duplexes, whose modification states are indicated. Also shown are methylated and unmethylated duplexes treated identically, except they were not exposed to DNase I, denoted by (−).

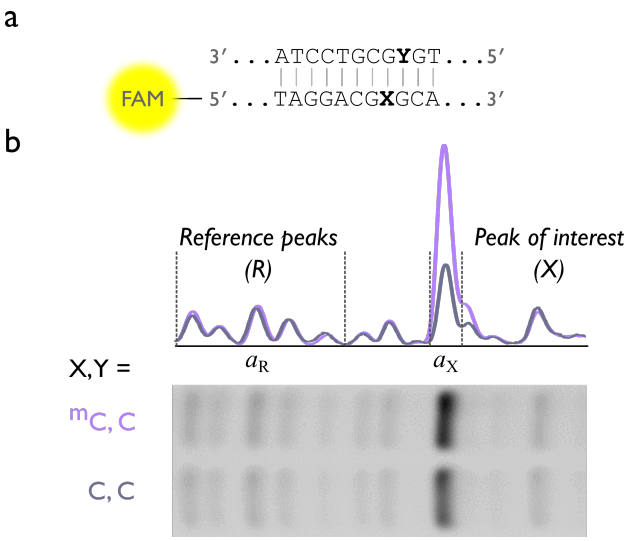


Figure B.2. The reference peaks used to determine relative DNase I cleavage rates. To normalise values, the ratio of the areas of the peak of interest and reference peaks was taken (a_X / a_R). The reference peaks were not significantly affected by the presence of cytosine modifications.

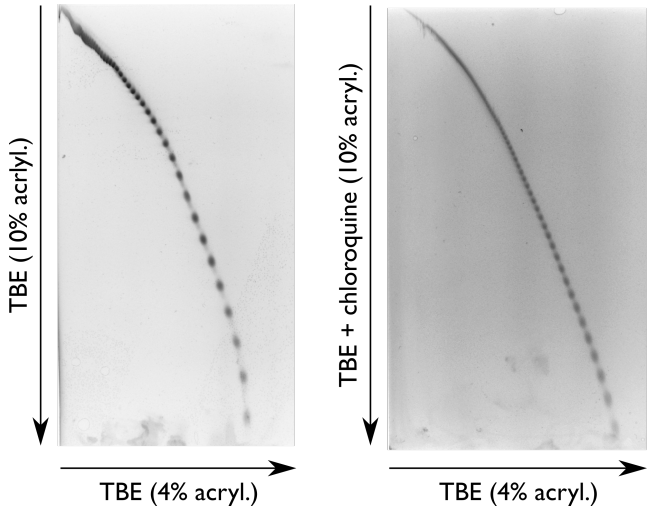


Figure B.3. 2D polyacrylamide gels to detect cyclised ligation products. Ulanovsky *et al* used 2D PAGE to detect cyclised ligation products by exploiting the differential effect of chloroquine on the mobility of circular and linear DNA fragments.[2] Ligation products were first electrophoresed on a 4% acrylamide gel without chloroquine, followed by a 10% acrylamide gel with chloroquine in the second dimension. The direction of electrophoresis is indicated by the arrows. The 2D gels obtained in this study of the unmodified A-tract fragments indicated that cyclised fragments are not present in the measurable region.

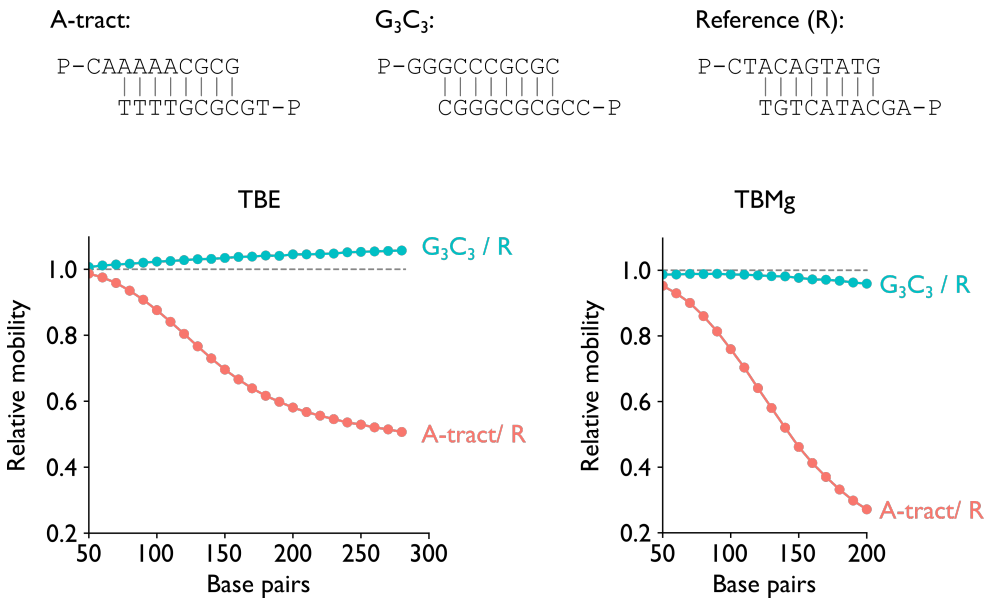


Figure B.4. The electrophoretic mobilities of the unmodified A-tract and G_3C_3 constructs, relative to an unmodified, random sequence. In the absence of divalent cations (TBE buffer), the G_3C_3-containing fragments have a higher mobility than do the fragments of random sequence. Conversely, the intrinsically curved A-tract fragments exhibited anomalously slow mobility. In the presence of 10 mM Mg^{2+} the G_3C_3 fragments migrated more slowly than did the fragments of random sequence, consistent with an increase in net curvature. However, the degree of curvature was much less pronounced than has previously been reported.[3]

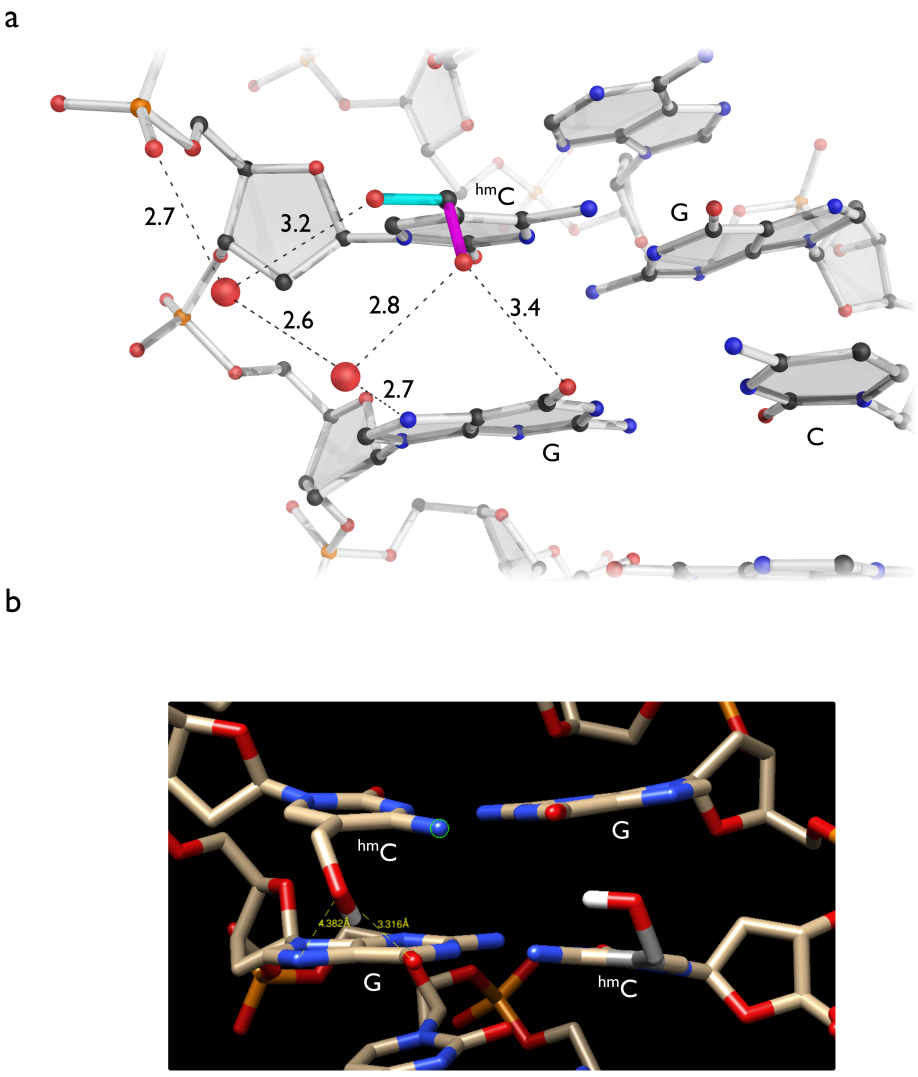


Figure B.5. Interactions of the 5-hydroxymethyl group in the major groove. (*a*) Section of a crystal structure of the Dickerson-Drew dodecamer containing hemimethylated ^{hm}CpG (PDB 4c5x).[4] The hydroxymethyl (hm) group exhibits conformational dimorphism. The hm group in the major conformer (magenta) is within direct hydrogen-bonding distance (3.4 Å) of the O6 of the 3′ guanine in the CpG site and also interacts with the guanine's N7 via a water molecule (red sphere). The hm group in the minor conformer (cyan) exhibits a water-mediated interaction with a phosphodiester oxygen of the same residue. (*b*) Model of a symmetrically hydroxymethylated CpG site, based on the hmC crystal structure PDB 4c5x (model built using UCSF Chimera).[4,5] The model was built to investigate why the presence of two hmC groups in a CpG site appears to induce less bending into the minor groove than does the presence of one. The hydroxymethyl groups are too far from each other to form direct hydrogen bonds. But they could interact via a network of water molecules in the major groove, which could reduce the region's tendency to bend into the minor groove. Alternatively, the effect could arise from the interactions described in *a*, where the interactions of the two hydroxymethyl groups reinforce each other, analogous to the chelate effect. Part *a* reproduced from Hardwick *et al.*[6]

B.2 Supplementary tables

Table B.1. Mean DNase I cleavage rates and their standard deviations directly 5′ to the modified cytosine on the labelled strand ('X'), relative to that of the duplex containing the unmodified CpG site; 'Y' denotes the opposing modification on the unlabelled strand.

X, Y:	*Rel. cleavage rate*	*Std. dev.*
C,C	1.00	0.02
mC, C	2.46	0.10
mC, mC	2.52	0.13
hmC, C	2.77	0.21
hmC, hmC	2.53	0.03
fC, C	2.20	0.26
fC, mC	2.25	0.02
fC, fC	1.78	0.07
caC, C	1.02	0.19
caC, mC	0.98	0.07
caC, caC	0.94	0.03

Table B.2. DNase I cleavage rates on the strand opposing the modified cytosine (4-nt downstream of position 'X'), relative to corresponding duplexes where position Y contained unmodified cytosine.

X, Y:	*Rel. cleavage rate*	*Std. dev.*
C, C	1.00	0.05
C, mC	0.91	0.02
mC, mC	0.91	0.07
C, hmC	0.84	0.07
hmC, hmC	0.87	0.07
C, fC	0.65	0.10
fC, fC	0.52	0.03
C, caC	0.90	0.11
caC, caC	0.95	0.14

Table B.3. Sequences and masses of oligonucleotides used in the DNase I cleavage study.

Sequence	Mol. weight / Da. *Observed* (*Calculated*)
*F*CATGCATACTAGGACGCGCAGACTGAGGTTCA*	10,395 (10,395)
*F*CATGCATACTAGGACG^{m}CGCAGACTGAGGTTCA	10,409 (10,409)
*F*CATGCATACTAGGACG^{hm}CGCAGACTGAGGTTCA	10,426 (10,425)
*F*CATGCATACTAGGACG^{f}CGCAGACTGAGGTTCA	10,425 (10,423)
*F*CATGCATACTAGGACG^{ca}CGCAGACTGAGGTTCA	10,439 (10,438)
TGAACCTCAGTCTGCGCGTCCTAGTATGCATG	9,791 (9,791)
TGAACCTCAGTCTG^{m}CGCGTCCTAGTATGCATG	9,805 (9,805)
TGAACCTCAGTCTG^{hm}CGCGTCCTAGTATGCATG	9,821 (9,821)
TGAACCTCAGTCTG^{f}CGCGTCCTAGTATGCATG	9,820 (9,819)
TGAACCTCAGTCTG^{ca}CGCGTCCTAGTATGCATG	9,834 (9,834)

**F denotes the 5′-FAM label.*

Table B.4. Sequences and masses of oligonucleotides used in the electrophoretic mobility study.

	Mol. weight / Da.
Sequence	*Observed* (*Calculated*)
*P*CAAAAACGCG	3,110 (3,108)
*P*CAAAAACGCA	3,094 (3,092)
*P*CAAAAACG^{m}CG	3,124 (3,122)
P^{m}CAAAAA^{m}CG^{m}CG	3,152 (3,150)
*P*CAAAAACG^{hm}CG	3,140 (3,138)
*P*CAAAAACG^{f}CG	3,138 (3,136)
*P*CAAAAACG^{m}A	3,108 (3,106)
*P*TGCGCGTTTT	3,105 (3,103)
*P*TGTGCGTTTT	3,120 (3,118)
*P*TG^{m}CGCGTTTT	3,119 (3,117)
*P*TG^{m}CG^{m}GTTTT	3,133 (3,131)
*P*TG^{hm}CGCGTTTT	3,135 (3,133)
*P*TG^{f}CGCGTTTT	3,133 (3,131)
*P*GGGCCCGCGC	3,110 (3,108)
*P*GGGCCCG^{m}GC	3,123 (3,122)
*P*GGGCCCG^{hm}GC	3,138 (3,138)
*P*CCGCGCGGGC	3,109 (3,108)
*P*CCG^{m}CGCGGGC	3,124 (3,122)
*P*CCG^{hm}CGCGGGC	3,140 (3,138)

**P denotes 5′-phosphate.*

BIBLIOGRAPHY

1. Leblanc, B. & Moss, T. in *Methods Mol. Biol.* 31–38 (Humana Press, Totowa, NJ, 2001).

2. Ulanovsky, L., Bodner, M., Trifonov, E. N. & Choder, M. Curved DNA: design, synthesis, and circularization. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 862–6 (1986).

3. Brukner, I., Susic, S., Dlakic, M., Savic, A. & Pongor, S. Physiological Concentration of Magnesium Ions Induces a Strong Macroscopic Curvature in GGGCCC-containing DNA. *J. Mol. Biol.* **236,** 26–32 (1994).

4. Lercher, L. *et al.* Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem. Commun.* **50,** 1794–1796 (2014).

5. Pettersen, E. F. *et al.* UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

6. Hardwick, J. S., Lane, A. N. & Brown, T. Epigenetic Modifications of Cytosine: Biophysical Properties, Regulation, and Function in Mammalian DNA. *BioEssays,* 1700199 (2018).

C

SUPPLEMENTARY INFORMATION FOR CHAPTER 5

C.1 Supplementary figures

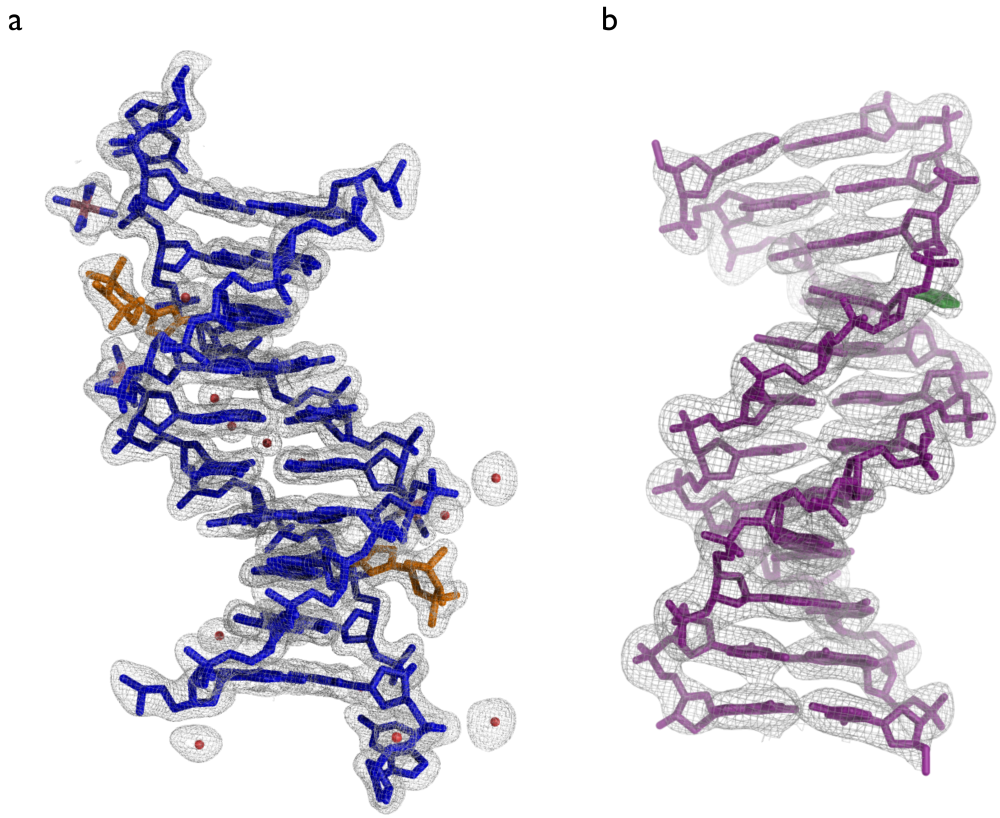


Figure C.1. $2F_o$–F_c electron density maps of the spin-labelled crystal structures, contoured at 1.0 σ. (*a*) The 6-Me-labelled structure 6qjs. (*b*) The 5-Me-labelled structure 6qjr. Reproduced from Hardwick *et al.*[1]

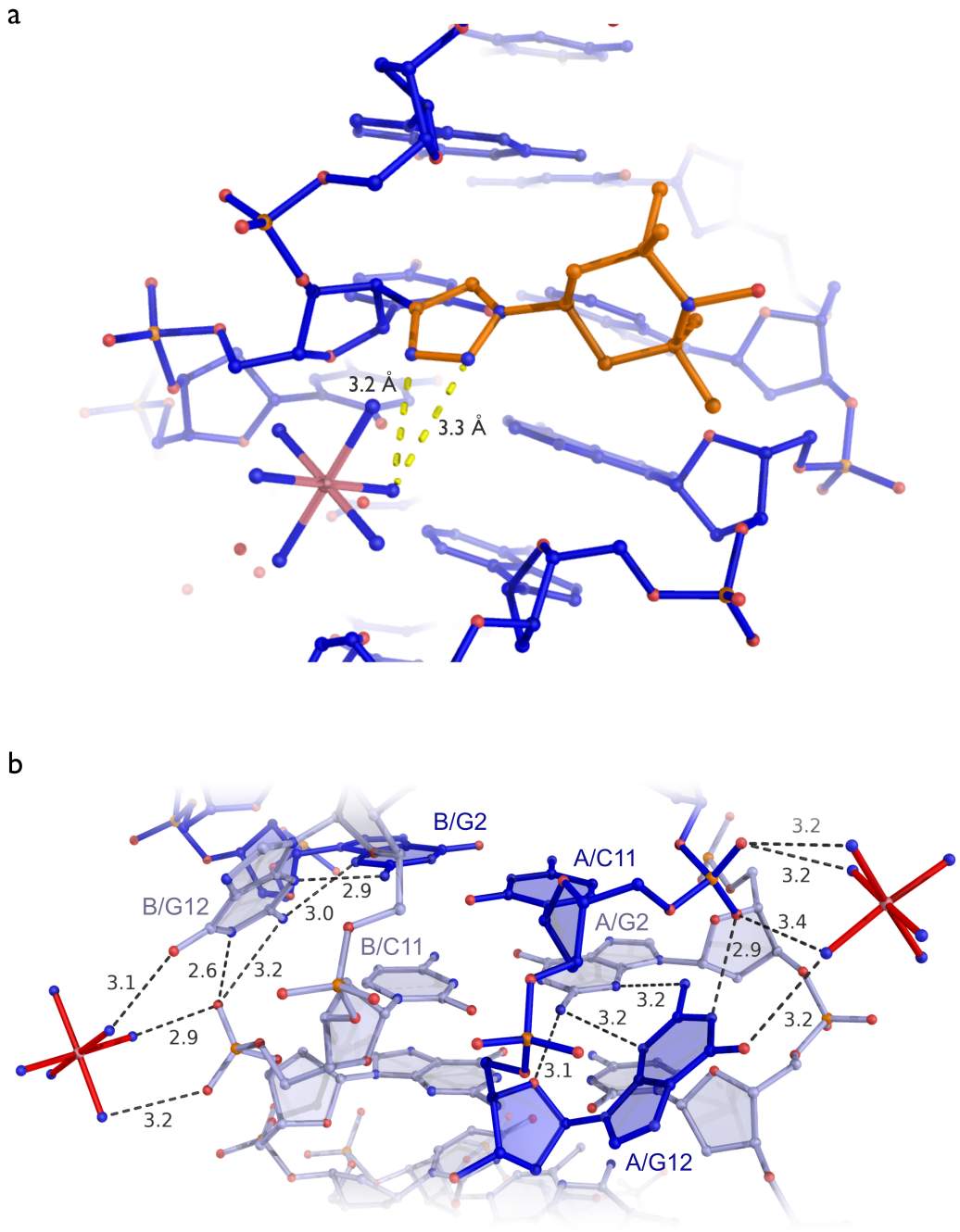


Figure C.2. Interactions in the 6-Me-labelled crystal structure 6qjs. (*a*) Interactions between the 6-Me spin label and the cobalt(III) hexammine cation. (*b*) Interactions between terminal base pairs and cobalt(III) hexammine cations. Reproduced from Hardwick *et al.*[1]

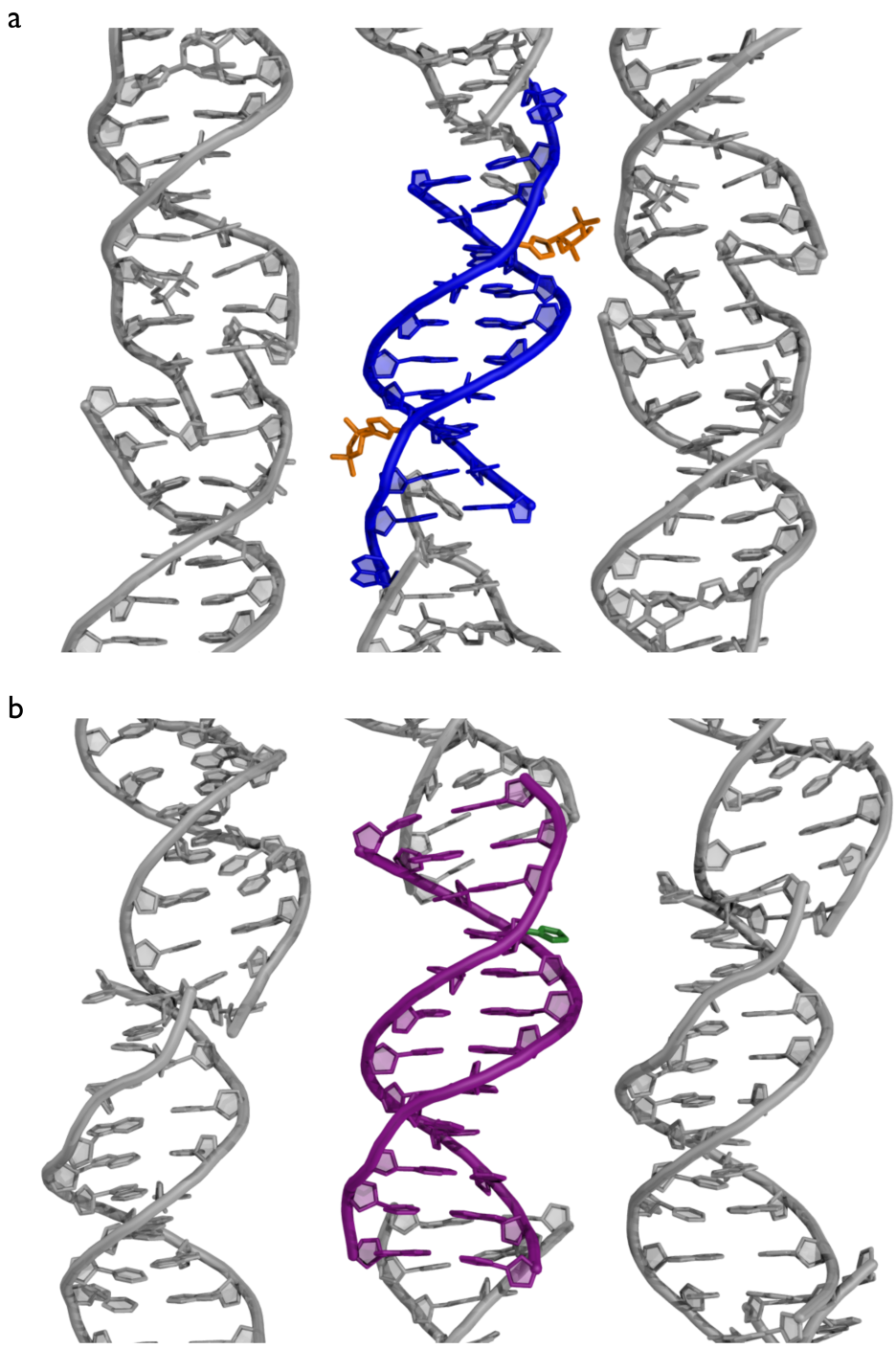


Figure C.3. Crystal packing of the spin-labelled duplexes. (*a*) The 6-Me labelled structure PDB 6qjs. (*b*) The 5-Me labelled structure PDB 6qjr.

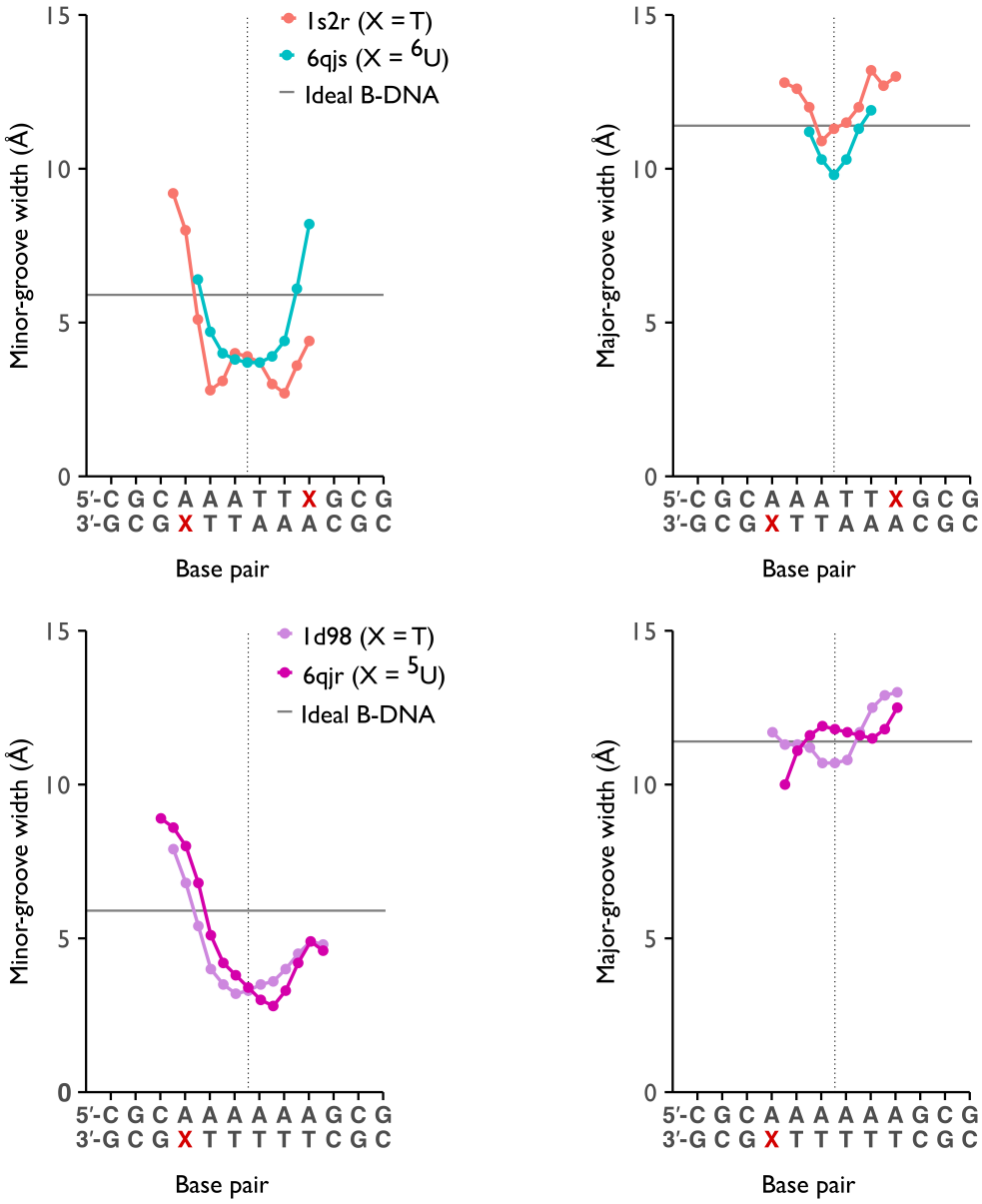


Figure C.4. Groove widths of the 6-Me and 5-Me spin-labelled structures (PDBs 6qjs and 6qjr) and their unmodified counterparts (PDBs 1s2r and 1d98). Adapted from Hardwick *et al.*[1]

C.1.1 Mass spectra of spin-labelled oligonucleotides

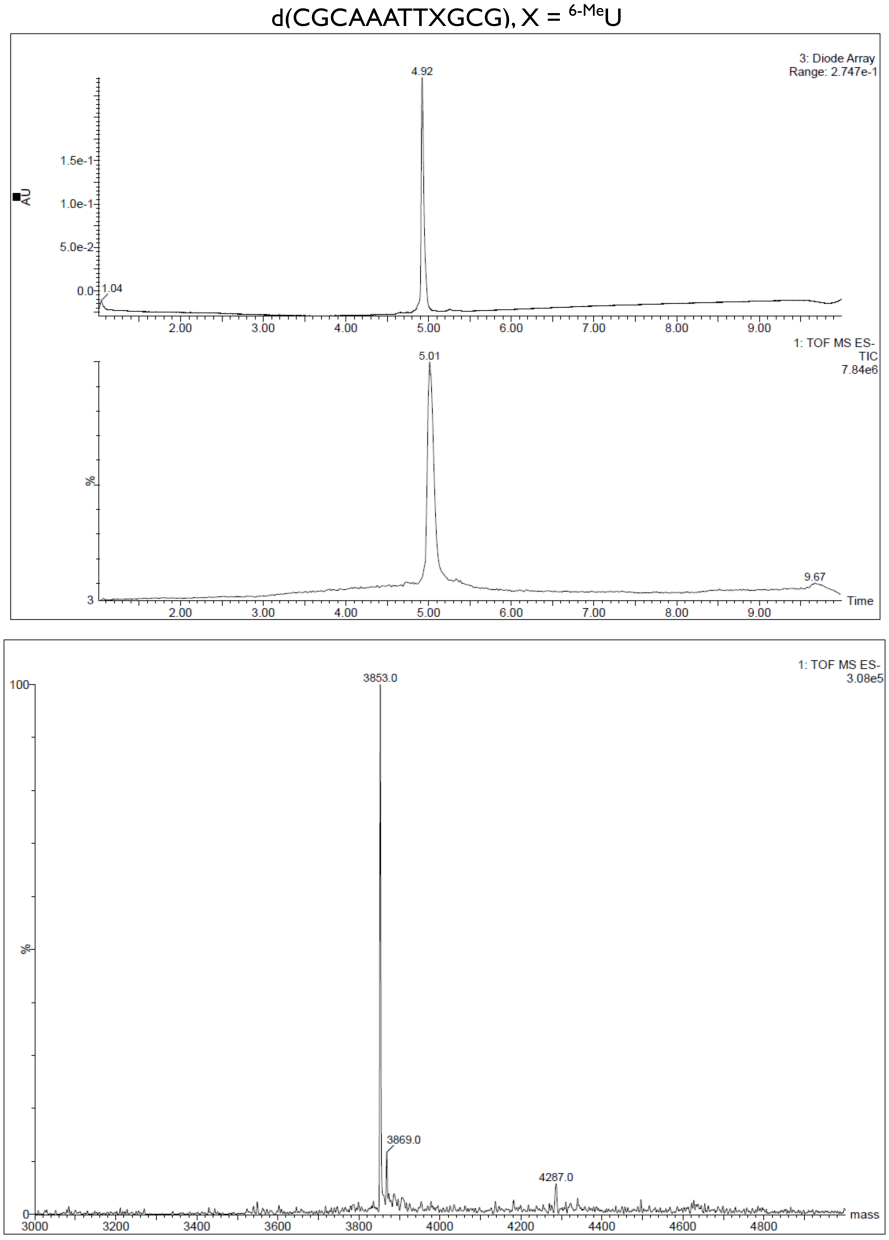


Figure C.5. Chromatograms and deconvoluted mass spectrum of d(CGCAAATTXGCG), where X denotes $^{6\text{-Me}}$U.

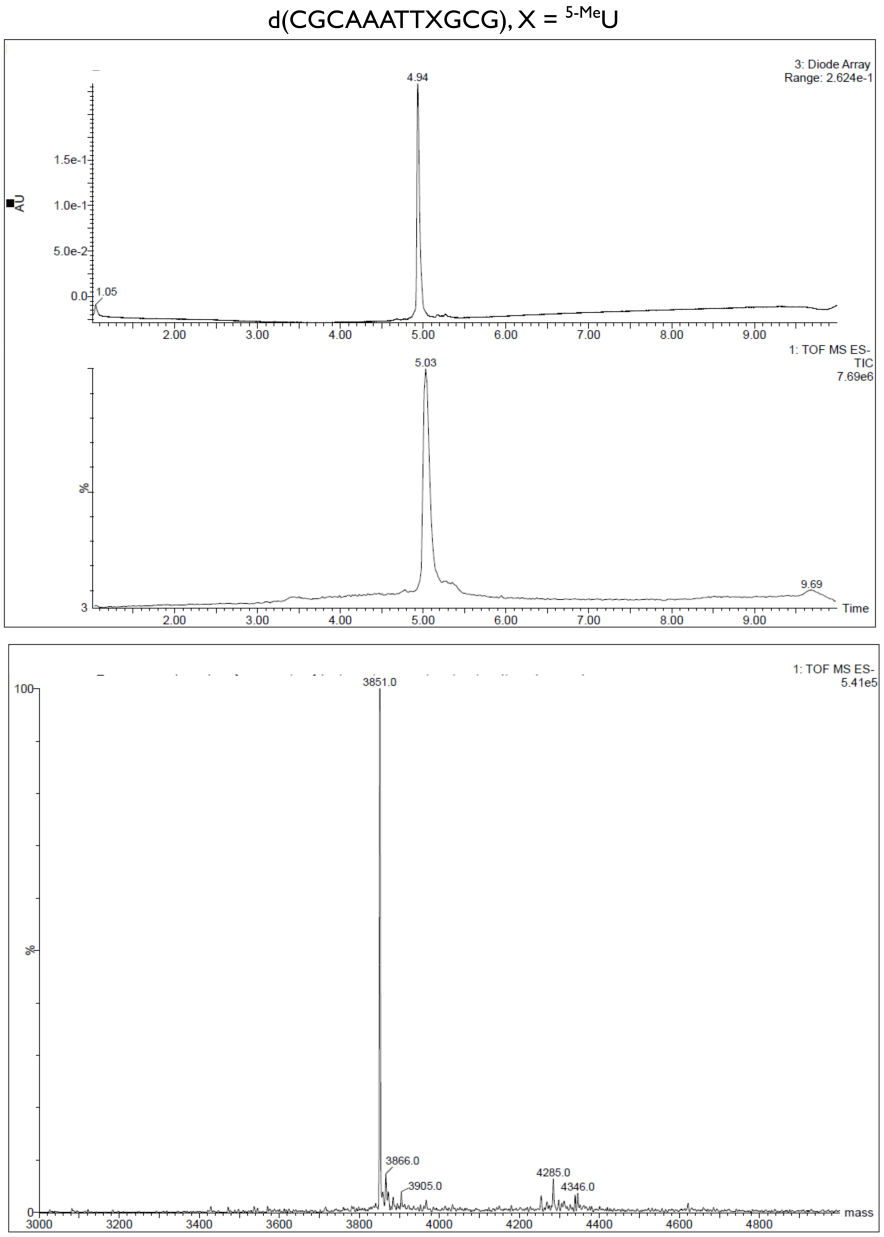


Figure C.6. Chromatograms and deconvoluted mass spectrum of d(CGCAAATTXGCG), where X denotes $^{5\text{-Me}}$U.

d(CGCTTTTTXGCG), X = $^{5\text{-Me}}$U

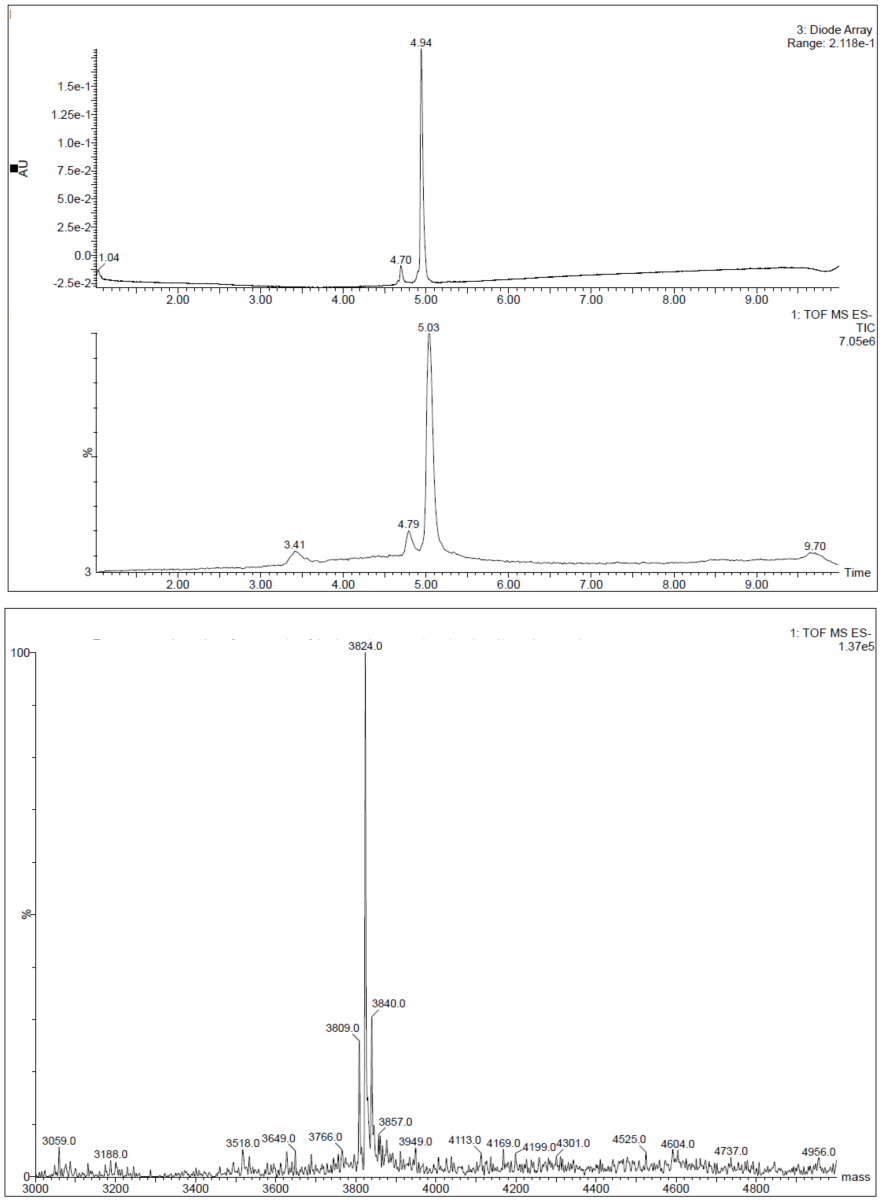


Figure C.7. Chromatograms and deconvoluted mass spectrum of d(CGCTTTTTXGCG), where X denotes $^{5\text{-Me}}$U.

d(CTXACGCGTCATTG), X = $^{6\text{-Me}}$U

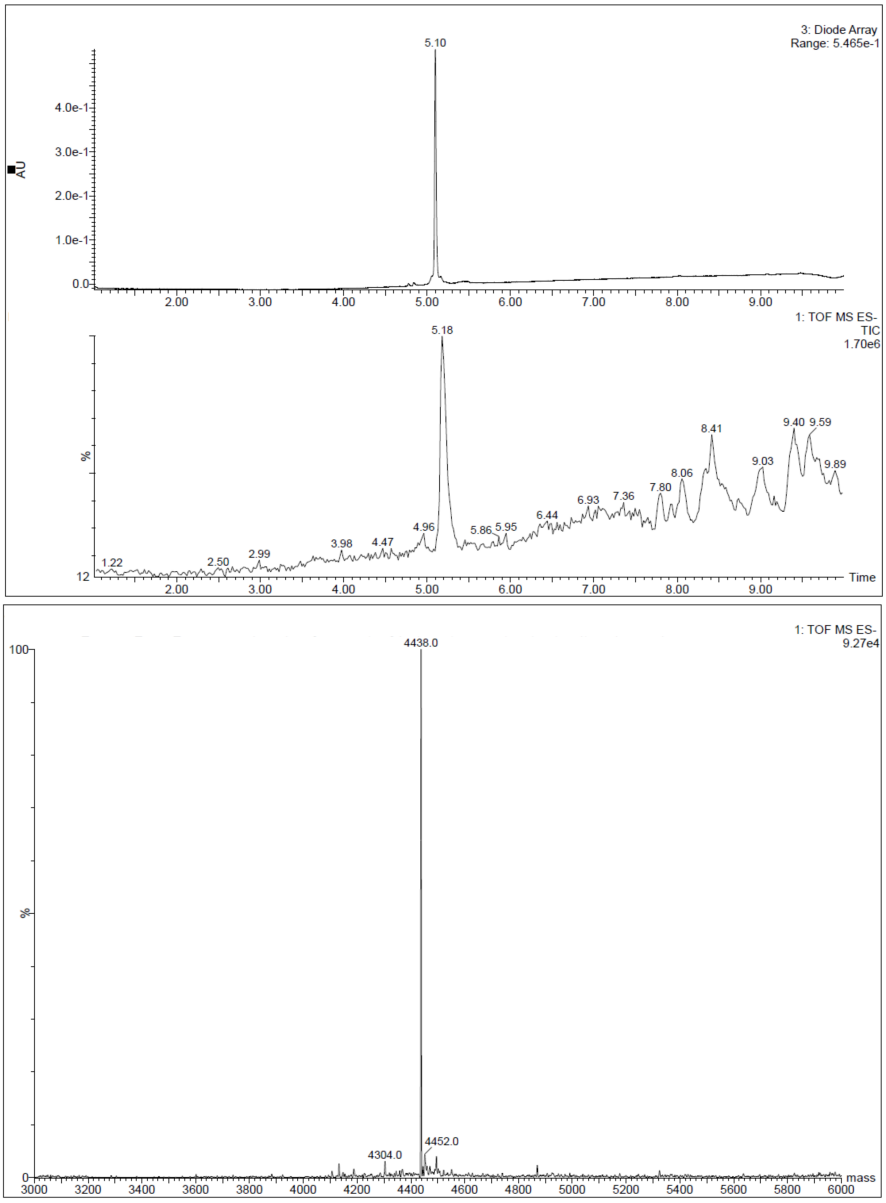


Figure C.8. Chromatograms and deconvoluted mass spectrum of d(CTXACGCGTCATTG), where X denotes $^{6\text{-Me}}$U.

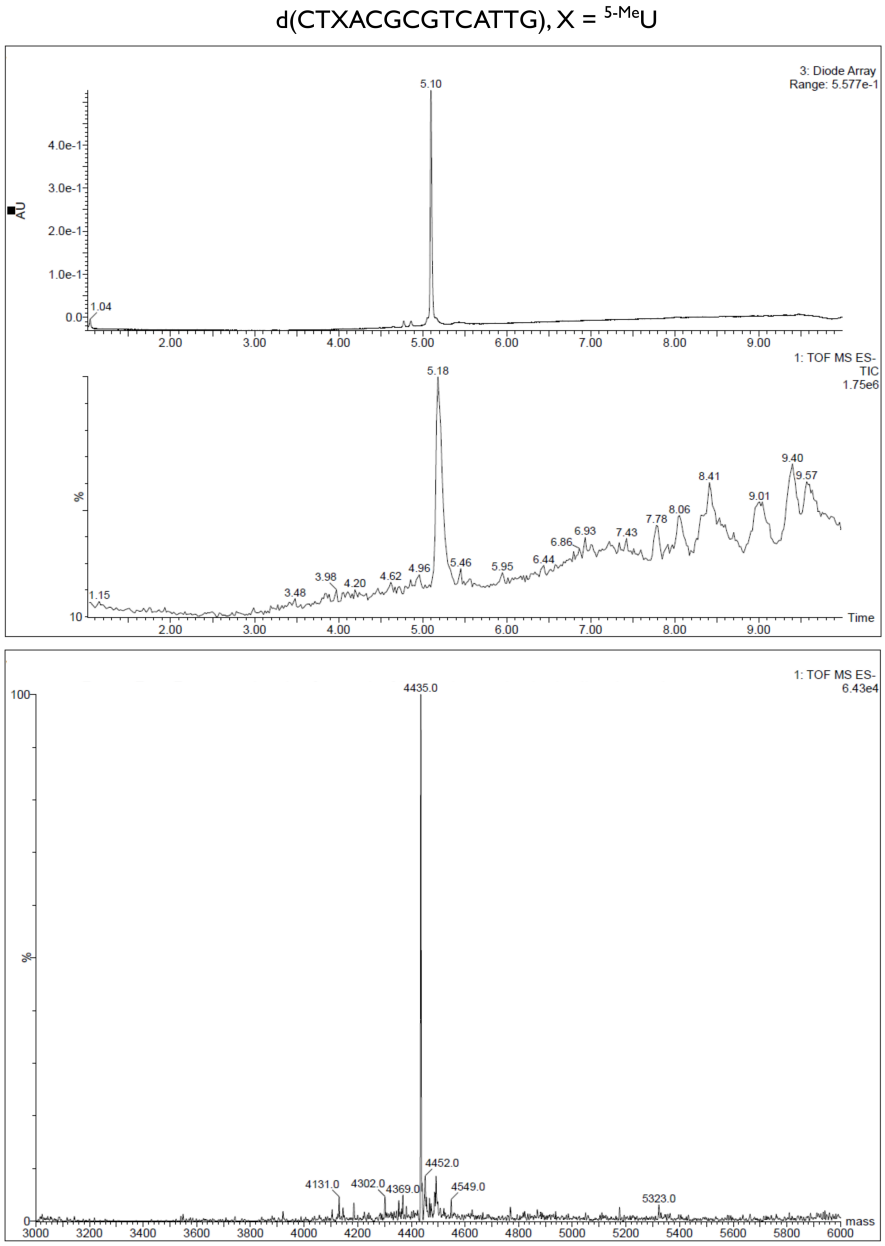


Figure C.9. Chromatograms and deconvoluted mass spectrum of d(CTXACGCGTCATTG), where X denotes $^{5\text{-Me}}$U.

C.2 Supplementary tables

Table C.1. Sequences of the spin-labelled oligonucleotides used in this study.

Oligo code	Sequence	**Molecular weight** / Da. Observed (calculated)
ODN1	CGCAAATT$^{5\text{-Me}}$UGCG	3851 (3850)
ODN2	CGCAAATT$^{6\text{-Me}}$UGCG	3853 (3852)
ODN3	CGCTTTTT$^{5\text{-Me}}$UGCG	3824 (3824)
ODN4	CT$^{5\text{-Me}}$UACGCGTCATTG	4435 (4435)
ODN5	CT$^{6\text{-Me}}$UACGCGTCATTG	4438 (4437)

Table C.2. Crystallisation conditions of the six structures determined in Chapter 2

	6qjs (6-Me)	**6qjr (5-Me)**
Oligonucleotide (mM)	1	1
NaCl (mM)	80	0
KCl (mM)	10	90
$MgCl_2$ (mM)	20	20
$[Co(NH_3)]Cl_3$ (mM)	2	0
Spermine (mM)	0	12
MPD (% v/v)	35	45
Sodium cacodylate (mM)	40	40
pH	5.5	6.0

Table C.3. Twinning statistics of PDB 6qjs

Twin domain	Twin operator	Twin fraction
1	*h*, *k*, *l*	0.481
2	−*h*, −*k*, *l*	0.519

Table C.4. Sugar puckering of the 6-Me-labelled crystal structure (PDB 6qjs, X = $^{6\text{-Me}}$U) and its unmodified counterpart (PDB 1s2r, X = T); pseudorotation phase angles are given in parentheses.

	6qjs (X = $^{6\text{-Me}}$U)		1s2r (X = T)	
Base pair	Strand I	Strand II	Strand I	Strand II
2. G·C	C3′-*exo* (187.6°)	C2′-*endo* (160.4°)	C2′-*endo* (150.9°)	C2′-*endo* (166.6°)
3. C·G	O4′-*endo* (83.1°)	C2′-*endo* (159.2°)	C4′-*exo* (60.3°)	C1′-*exo* (139.9°)
4. A·X	C2′-*endo* (163.6°)	C4′-*exo* (48.8°)	C2′-*endo* (170.8°)	C2′-*endo* (145.7°)
5. A·T	C3′-*exo* (183.3°)	C2′-*endo* (147.7°)	C1′-*exo* (128.5°)	C1′-*exo* (138.4°)
6. A·T	C2′-*endo* (146.1°)	C1′-*exo* (108.7°)	C2′-*endo* (147.6°)	C1′-*exo* (121.3°)
7. T·A	C1′-*exo* (110.4°)	C2′-*endo* (144.7°)	O4′-*endo* (95.8°)	C1′-*exo* (139.4°)
8. T·A	C2′-*endo* (150.1°)	C2′-*endo* (156.1°)	C1′-*exo* (139.6°)	C2′-*endo* (161.9°)
9. X·A	C3′-*endo* (35.5°)	C2′-*endo* (159.5°)	C1′-*exo* (141.0°)	C2′-*endo* (168.7°)
10. G·C	C2′-*endo* (152.3°)	O4′-*endo* (83.5°)	C2′-*endo* (144.0°)	O4′-*endo* (77.0°)
11. C·G	C2′-*endo* (156.3°)	C3′-*exo* (193.9°)	C2′-*endo* (166.3°)	C2′-*endo* (155.4°)

Table C.5. Sugar puckering of the 5-Me-labelled crystal structure (PDB 6qjs, X = $^{5\text{-Me}}$U) and its unmodified counterpart (PDB 1d98, X = T); pseudorotation phase angles are given in parentheses.

	6qjr (X = $^{5\text{-Me}}$U)		1d98 (X = T)	
Base pair	Strand I	Strand II	Strand I	Strand II
1. C·G	C2′-*endo* (161.9°)	C4′-*exo* (53.9°)	C2′-*endo* (176.8°)	C3′-*endo* (4.3°)
2. G·C	C3′-*endo* (15.6°)	C4′-*exo* (43.4°)	C4′-*exo* (68.8°)	C4′-*endo* (236.2°)
3. C·G	O4′-*endo* (77.9°)	C2′-*endo* (144.9°)	O4′-*endo* (102.6°)	C2′-*endo* (161.9°)
4. A·X	C2′-*endo* (165.6°)	C4′-*exo* (45.3°)	C3′-*exo* (206.2°)	C1′-*exo* (109.8°)
5. A·T	C2′-*endo* (173.9°)	C4′-*exo* (65.8°)	C2′-*endo* (158.8°)	C2′-*endo* (158.3°)
6. A·T	C2′-*endo* (167.3°)	C1′-*exo* (141.4°)	C2′-*endo* (149.7°)	C1′-*exo* (131.5°)
7. A·T	C2′-*endo* (172.6°)	C2′-*endo* (156.5°)	C1′-*exo* (127.4°)	C3′-*exo* (205.1°)
8. A·T	C2′-*endo* (161.8°)	C2′-*endo* (152.6°)	C2′-*endo* (158.5°)	C2′-*endo* (146.6°)
9. A·T	C2′-*endo* (171.9°)	C2′-*endo* (154.5°)	C3′-*exo* (183.7°)	C3′-*exo* (194.4°)
10. G·C	C2′-*endo* (153.7°)	C4′-*exo* (56.9°)	C3′-*exo* (195.9°)	C3′-*endo* (17.7°)
11. C·G	C2′-*endo* (179.7°)	C2′-*endo* (161.8°)	C2′-*endo* (173.5°)	C2′-*endo* (148.7°)
12. G·C	C3′-*endo* (28.9°)	C2′-*endo* (166.3°)	C1′-*exo* (119.7°)	C3′-*exo* (186.8°)

Table C.6. Glycosyl (χ-)torsion angles of the spin-labelled crystal structures (PDBs 6qjs and 6qjr) and their unmodified counterparts (PDBs 1s2r and 1d98).

	6qjs (°, X = $^{6\text{-Me}}$U)		1s2r (°, X = T)	
Base pair	Strand I	Strand II	Strand I	Strand II
2: G·C	−93.8	−104.5	−102.1	−115.9
3: C·G	−136.1	−95.5	−131.8	−82.5
4: A·X	−76.2	−145.6	−97.7	−112.8
5: A·T	−95.4	−109.7	−117.8	−119.1
6: A·T	−101.7	−129.3	−111.4	−124.5
7: T·A	−122.3	−104.2	−128.7	−113.0
8: T·A	−105.9	−105.1	−117.6	−114.3
9: X·A	−140.9	−82.2	−114.5	−101.9
10: G·C	−98.3	−123.5	−86.2	−129.5
11: C·G	−108.8	−97.1	−102.5	−120.1
	6qjr (°, X = $^{5\text{-Me}}$U)		1d98 (°, X = T)	
Base pair	Strand I	Strand II	Strand I	Strand II
1. C·G	−122.5	−147.6	−65.1	−136.9
2. G·C	−158.3	−144.2	−144.7	−117.2
3. C·G	−133.1	−112.4	−144.3	−69.3
4. A·X	−100.7	−144.2	−87.6	−113.7
5. A·T	−94.3	−103.3	−107.9	−117.5
6. A·T	−101.2	−145.5	−104.6	−125.5
7. A·T	−91.3	−102.0	−111.1	−67.3
8. A·T	−111.3	−127.2	−99.5	−114.5
9. A·T	−107.5	−119.2	−96.2	−91.2
10. G·C	−77.6	−165.0	−69.8	−127.8
11. C·G	−86.6	−77.1	−103.3	−122.0
12. G·C	−88.9	−117.1	−113.6	−71.5

Table C.7. Chemical shift assignments of the reduced 6-Me-labelled duplex, where X denotes the modified residue FINISH NMR TABLES

	^{1}H Chemcal shift (ppm)						
Nt	H8/6	H5/Me	H1′	H2′	H2″	H3′	H4′
C1	7.70	5.82	5.98	2.15	2.52	4.66	4.10
T2	7.56	1.76	6.03	2.44	2.52	4.94	4.27
X	7.93	5.83	6.24	3.82	–	4.86	4.65
A4	8.26	7.31	6.20	2.57	2.82	4.86	4.47
C5	7.29	5.25	5.56	2.0	2.32	4.80	4.1
G6	7.82	–	5.86	2.58	2.70	4.94	4.33
C7	7.26	5.30	5.67	2.0	2.36	4.81	4.11
G8	7.86	–	5.96	2.59	2.77	4.94	–
T9	7.19	1.39	5.96	2.07	2.47	4.83	–
C10	7.55	5.66	5.54	2.14	2.45	4.85	4.10
A11	8.32	7.57	6.25	2.71	2.94	5.02	4.42
T12	7.21	1.44	5.94	2.00	2.49	4.83	–
T13	7.30	1.68	5.89	2.01	2.37	4.86	–
G14	7.91	–	6.15	2.59	2.40	4.68	4.44
C15	7.65	5.94	5.64	1.83	2.36	4.69	4.04
A16	8.33	7.38	5.92	2.82	2.93	5.03	4.47
A17	8.20	7.70	6.19	2.59	2.89	5.02	4.47
T18	6.99	1.34	5.65	1.83	2.29	4.82	4.11
G19	7.80	–	5.48	2.61	2.72	4.97	4.43
A20	8.13	7.73	6.16	2.61	2.84	5.02	4.44
C21	7.15	5.16	5.54	1.89	2.28	4.79	4.10
G22	7.81	–	5.85	2.60	2.71	4.94	–
C23	7.25	5.31	5.68	1.99	2.36	4.82	4.11
G24	7.87	–	5.95	2.58	2.69	4.94	–
T25	7.11	1.54	5.81	1.72	2.12	–	–
A26	8.07	–	5.87	2.55	2.62	4.91	4.33
A27	8.02	–	5.97	2.52	2.65	4.92	4.40
G28	7.74	–	5.96	2.51	2.33	4.64	–

Table C.8. Chemical shift assignments of the reduced 5-Me-labelled duplex, where X denotes the modified residue

	^{1}H Chemcal shift (ppm)						
Nt	H8/6	H5/Me	H1′	H2′	H2″	H3′	H4′
C1	7.70	5.83	5.96	2.11	2.48	4.64	–
T2	7.50	1.73	5.95	2.14	2.52	4.93	4.36
X	7.94	5.83	6.33	3.92	–	4.87	–
A4	8.26	7.31	6.20	2.57	2.83	4.87	4.47
C5	7.29	5.27	5.60	2.00	2.32	4.81	–
G6	7.82	–	5.86	2.57	2.68	4.94	4.34
C7	7.25	5.30	5.65	1.99	2.36	4.81	4.11
G8	7.85	–	5.95	2.61	2.78	4.93	4.36
T9	7.19	1.39	5.97	2.08	2.47	4.84	–
C10	7.54	5.65	5.53	2.13	2.42	4.89	–
A11	8.32	7.57	6.24	2.72	2.93	5.02	4.43
T12	7.21	1.42	5.93	1.98	2.47	4.84	–
T13	7.30	1.67	5.89	2.00	2.38	4.87	–
G14	7.90	–	6.14	2.60	2.36	4.69	–
C15	7.65	5.94	5.64	1.82	2.34	4.69	–
A16	8.32	7.38	5.91	2.81	2.93	5.03	4.38
A17	8.20	7.70	6.19	2.58	2.88	5.02	4.47
T18	6.98	1.35	5.65	1.84	2.27	4.82	4.11
G19	7.80	–	5.48	2.61	2.73	4.98	–
A20	8.13	7.73	6.16	2.60	2.84	5.02	4.44
C21	7.14	5.14	5.53	1.89	2.29	4.80	4.11
G22	7.81	–	5.85	2.59	2.69	4.92	4.33
C23	7.25	5.28	5.69	1.98	2.36	4.81	–
G24	7.87	–	5.94	2.57	2.69	4.93	4.36
T25	7.10	1.54	5.82	1.76	2.11	–	–
A26	8.08	–	5.91	2.53	2.60	4.88	–
A27	8.03	–	6.01	2.51	2.63	4.93	–
G28	7.74	–	5.96	–	2.35	4.63	–

Table C.9. Chemical shift assignments of the 6-Me-labelled duplex in the paramagnetic state.

	^{1}H Chemcal shift (ppm)						
Nt	H8/6	H5/Me	H1′	H2′	H2″	H3′	H4′
C7	7.25	5.3	5.65	1.98	2.34	–	–
G8	7.85	-	5.96	2.6	2.78	4.95	–
T9	7.19	1.39	5.96	2.08	2.48	4.84	–
C10	7.54	5.67	5.55	2.14	2.44	4.85	4.14
A11	8.32	–	6.24	2.72	2.93	5.03	4.43
T12	7.22	1.44	5.95	1.99	2.44	4.83	4.37
T13	7.31	1.68	5.93	2.02	2.39	4.85	4.37
G14	7.9	–	6.15	2.57	2.4	4.68	–
C15	7.64	5.93	5.65	1.81	2.33	4.68	–
A16	8.32	–	5.9	2.81	2.94	5.03	4.47
A17	8.21	–	6.19	2.59	2.89	5.02	4.47
T18	6.99	1.35	5.66	1.85	2.28	4.82	4.14
G19	7.8	–	5.49	2.61	2.72	4.97	4.32
A20	8.13	–	6.16	2.62	2.85	5.01	4.47
C21	7.15	5.15	5.54	1.9	2.29	4.79	4.16
G22	7.81	–	5.86	2.59	2.7	4.95	4.33
C23	7.25	5.31	5.65	1.99	2.36	4.82	–
G24	–	–	5.95	2.6	2.78	4.94	4.37

Table C.10. Chemical shift assignments of the reduced 5-Me-labelled duplex.

	^{1}H Chemcal shift (ppm)						
Nt	H8/6	H5/Me	H1′	H2′	H2″	H3′	H4′
G6	–	–	–	–	–	4.96	–
C7	7.25	5.30	5.65	1.99	2.34	4.84	4.11
G8	7.85	–	5.96	2.61	2.78	4.95	4.37
T9	7.19	1.39	5.66	2.07	2.37	4.85	4.11
C10	7.54	5.66	5.55	2.13	2.44	4.85	4.14
A11	8.32	–	6.22	2.72	2.93	5.03	4.42
T12	7.21	1.43	5.95	1.98	2.47	4.83	–
T13	7.31	1.68	5.93	2.00	2.39	4.85	4.38
G14	7.91	–	6.15	2.60	2.37	–	–
C15	7.63	5.92	5.65	1.81	2.33	4.67	–
A16	8.32	–	5.91	2.81	2.94	5.02	–
A17	8.21	–	6.19	2.58	2.90	5.02	4.47
T18	6.99	1.35	5.66	1.85	2.29	4.83	4.11
G19	7.80	–	5.49	2.61	2.72	4.97	4.32
A20	8.13	–	6.16	2.62	2.85	5.01	4.43
C21	7.15	5.15	5.54	1.89	2.29	4.79	4.33
G22	7.81	–	5.86	2.59	2.70	4.95	–
C23	7.25	5.30	5.65	1.99	2.35	4.82	4.11

BIBLIOGRAPHY

1. Hardwick, J. S. *et al.* 2'-Alkynyl spin-labelling is a minimally perturbing tool for DNA structural analysis. *Nucleic Acids Res.* (2020).