

The High-Resolution Crystal Structure of a Parallel-Stranded Guanine Tetraplex

by

Gerard Michael Laughlan

A Thesis Submitted for the Degree of

Doctor of Philosophy

Department of Chemistry

University of Glasgow

April 1995

ProQuest Number: 13834037

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 13834037

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Declaration

This thesis is less than 100,000 words in length, exclusive of tables, maps, figures and references. The work reported in this thesis was performed entirely by myself unless otherwise cited or acknowledged. Its contents have not been submitted for any other degree.

Signed

Gerard Michael Laughlan

1st April 1995

It is a long trip. We are the only riders. The work contained within this thesis is dedicated to many: *R.A.A.M.*; my mother, father and sister; The Master Therion; the lost memory of Charlotte Thompson; and all the casualties of the endocolonial War Machine – The *Vague* Generation.

My lips are now one angstrom unit from her lips, which is one ten-billionth of a meter. I begin to kiss her but she turns her head away. Don't good little boys who finish all their vegetables get dessert? I ask. I can't kiss you, we're monozygotic replicants – we share 100% of our genetic material. My head spins. You are the beautiful day, I exclaim, your breath is a zephyr of eucalyptus that does a pas de bourrée across the Sea of Galilee. Thanks, she says, but we can't go back to my house and make love because monozygotic incest is forbidden by the elders. What if I said I could change all that . . . What if I said that I had a miniature shotgun that blasts gene fragments into the cells of living organisms, altering their genetic matrices so that a monozygotic replicant would no longer be a monozygotic replicant and she could then make love to a muscleman without transgressing the incest taboo, I say, opening my shirt and exposing the device which I had stuck in the waistband of my black jeans. How'd you get that thing? she gasps, ogling its thick fiber-reinforced plastic barrel and the Uzi-Biotech logo embossed on the magazine which held two cartridges of gelated recombinant DNA. I got it for Christmas.

Mark Leyner, *I Was An Infinitely Hot and Dense Dot*

The third dimension is no longer the measure of expansion; relief, no longer the reality. From now on the latter is concealed in the flatness of pictures, the transferred representations. It conditions the return to the house's state of siege, to the cadaver-like inertia of the interactive dwelling, this residential cell that has left the extension of the habitat behind it and whose most important piece of furniture is the *seat*[siege], the ergonomic armchair of the handicapped's motor, and – who knows? – the *bed*, a canopy bed for the infirm voyeur, a divan for being dreamt of without dreaming, a bench for being circulated without circulating.

Paul Virilio, *The Last Vehicle*

In addition to the growth hormone extracted from the glands of human corpses, I was using anabolic steroids, tissue regeneration compounds, granulocyte macrophage colony stimulating factor (GMCSF) – a substance used to stimulate growth of certain vital blood cells in radiation victims – and a nasal spray of neuropeptides that accelerates the release of pituitary hormones and I was getting larger and larger and my food bills were becoming enormous. So I went on a TV game show in the hope of raising cash. This was my question, for \$250,000 in cash and prizes: If the Pacific Ocean were filled with gin, what would be, in terms of proportionate volume, the proper lake of vermouth necessary to achieve a dry martini? I said Lake Ontario – but the answer was the Caspian Sea, which is called a sea but is a lake by definition. I had failed.

Mark Leyner, *I Was an Infinitely Hot and Dense Dot*

I started making maps when I was small showing place, resources, where the enemy and where love lay. I did not know time adds to land. Events drift continually down, effacing landmarks, raising the level, like snow. I have grown up. My maps are out of date. The land lies over me now. I cannot move. It is time to go. Goodbye.

Alasdair Gray, *Lanark, A Life in Four Books*

Summary

It's a sick picture, B. J.

DNA and RNA containing runs of consecutive guanine bases may adopt four-stranded conformations based on the hydrogen-bonded guanine tetrad, or G quartet. The character of the unusual conformations adopted by polymers containing guanosine as well as guanosine mononucleotides were first investigated in the 1960s. In more recent years interest has been rekindled by the discovery that telomeric sequences in *Oxytricha* and *Tetrahymena*, which typically contain repeat tracts of guanine bases, adopt higher order structures *in vivo* and unusual conformations *in vitro*. Nucleic acid sequences containing guanine repeats have been located in a number of biological areas including chromosomal telomeres, immunoglobulin switch regions and the human immunodeficiency virus (HIV) genome, as well as in a variety of recombination sites. The oligonucleotide d(T₂G₄T₂) has recently been demonstrated to be a potent inhibitor of HIV viral-mediated cell fusion. Probing of the structures adopted by these sequences *in vitro* has revealed a capacity for several non-canonical structures, all of which involve the formation of four-stranded DNA based upon a guanine tetrad with a highly distinctive hydrogen-bonding arrangement.

Structural investigations have indicated at least two major classes of non-canonical conformation assumed by guanine-rich DNA deploying the guanine tetrad. The first class involves the folding back of a tandem sequence, or the adoption of a hairpin dimer. This results in the anti-parallel alignment of strands and the consequent formation of both *syn* and *anti* glycosyl conformations for the guanine residues within a strand. The second class are defined by the association of four independent strands in a parallel formation, and in this case it is anticipated, on the basis of results from fibre diffraction studies, that the glycosyl linkages of the guanines all adopt an *anti* conformation. Both structural classes require the presence of group I metal ions, in particular sodium and potassium, and a number of hypotheses have been promoted to account for the role of these ions in stabilising the structures adopted. Recent high-resolution NMR and X-ray crystallographic studies have succeeded in determining the key structural features of the structures of the foldback form. In the last couple of years the isolation of proteins that bind to and promote the formation of both structural classes has lent weight to their functional importance in living systems.

This thesis describes my work in determining the three dimensional crystal structure of the parallel-stranded tetraplex formed by the oligonucleotide d(TG₄T) in the presence of sodium cations. Samples of the oligonucleotide were synthesised by solid-state methods and the integrity of the final products checked by MALDI-TOF mass spectrometry. Single crystals suitable for X-ray diffraction studies were obtained using the hanging drop vapour diffusion method. Characterisation of the crystals revealed the space group to be P1 with unit cell dimensions $a=28.76\text{\AA}$, $b=35.47\text{\AA}$, $c=56.77\text{\AA}$, $\alpha=74.39^\circ$, $\beta=77.64^\circ$, $\gamma=89.73^\circ$.

X-ray diffraction data were collected using a Rigaku IIC phosphorimager area detector at the University of York. Native data were obtained from two crystals with monochromated radiation from a Cu(1.543Å) source and from one crystal with a Mo(0.711Å) source. The final processed native data-set contained a total of 211,674 reflections of which 65,264 were unique to an upper resolution limit of 1.20Å. The native data-set had a final R_{merge} of 0.056 and was 97.7% complete to 1.20Å resolution.

Initial attempts at phase determination concentrated on the production of heavy-atom derivatives. Synthetic oligonucleotides were produced by solid-state methods and the syntheses monitored by mass spectrometry. All synthetic derivatives failed to yield crystals suitable for X-ray diffraction. A number of heavy-atom soaks and co-crystallisation were attempted and X-ray data-sets collected. Analysis of the heavy-atom X-ray data-sets showed the derivatives to have been unsuccessful.

Further attempts to determine phases utilised the molecular replacement technique. Analysis of the native Patterson and self-rotation functions provided evidence that the oligonucleotide formed a tetraplex within the crystal, and suggested an approximate packing model. Search models for molecular replacement were generated from data derived from NMR studies of the d(TG₄T) parallel-stranded tetraplex in solution. Molecular replacement solutions were determined which agreed with the packing model expected from analysis of the native data.

Initial refinement of the molecular replacement solutions employed rigid body and simulated annealing techniques. An unequivocal model for the contents of the unit cell could not be deduced due to a lack of information regarding the stereochemistry of the thymine residues within the model. The structure solution was unambiguously

determined using an automated refinement protocol, coupled with least-squares minimization, which operates by iteratively determining and adjusting the water structure surrounding the macromolecule. The final stage in the refinement procedure involved conjugate-gradient and block-matrix least-squares refinement. At this stage hydrogen atoms were incorporated in a riding model and the atomic temperature factors were anisotropically refined. The final model, which has an R-index of 12.4% for all the data from 10.0–1.2Å, contains 2,453 non-hydrogen atoms, consisting of 4 parallel-stranded tetraplexes (i.e. 16 strands of d(TG₄T)), 9 calcium ions, 14 sodium ions and 514 water molecules.

Finally, the crystal structure of the parallel-stranded tetraplex formed by the oligonucleotide d(TG₄T) in the presence of sodium ions is reported. The topology of the tetraplex is in agreement with that expected for B-form DNA. Sharply resolved sodium cations are found both between and within planes of hydrogen-bonded guanine quartets. An ordered groove hydration is observed, and calcium ions are located with a well-defined hydration sphere which mediates contact with the phosphate groups of the DNA. Thymine bases are exclusively involved in making extensive lattice contacts. This tetraplex structure has been determined to one of the highest resolutions ever achieved for a nucleic acid.

Acknowledgements

Are you listening B. J.?

Needless to say this student is indebted to many who participated in the hazardous series of events culminating in the completion of this project. First of all I would like to thank Dr. Ben Luisi for his supervision and encouragement throughout the course of the work, and Professor Neil Isaacs for his reading of the manuscript as it materialized. Gratitude is also extended to Professor David Lilley of the University of Dundee who initiated the project and maintained a keen and exuberant interest in its progress. Special thanks are also due to Professor John Subak-Sharpe for kindly allowing me residence in the Institute of Virology during my studentship, and to Dr. Howard Marsden for graciously permitting me to continue to utilize the institutes computing facilities following Professor Subak-Sharpes departure.

Certain key individuals provided invaluable support and expertise. My warmest appreciation is extended to Dr. Alastair Murchie and Dr. David Norman of the University of Dundee for their generous provision of material and virtual DNA, respectively. The members of the structural biology laboratory of the University of York made me very welcome during my many working visits there. Consequently I would like to

thank Madeleine Moore, Peter Moody, Eleanor and Guy Dodson, Giles Wilson, Jeremy Tame, Camilla Smith and Stuart Priest, as well as the many other people, too numerous to include here, who lent me their wisdom and assistance at some time or another. I feel particularly indebted to Dr. Madeleine Moore and Dr. Peter Moody for their patience and support during all the stages of my encounter with the crystallographic process, and thank Madeleine again for her enthusiastic and understanding critical readings of my manuscript after all the *real* work had been accomplished.

I am grateful to Jorge Navaza for assistance with his molecular replacement package *AMoRe*; Zbyszek Otwinowski for advice and direction with his processing program *DENZO*; Uwe Pieleles for mass spectrometry analysis; and Simon Cook and Gordon Seeley for the generous use of their PCs in the generation of figures and illustrations. For miscellaneous advice, assistance and discussions beyond the call of duty I would also like to mention Fareed Aboul-ela, Gerard Bricogne, Chris Edwards, Paul Elmsley, Chris Frampton, Chris Gilmore, Adrian Laphorn, Stuart MacKay and Paul Mallinson.

Last, but most certainly not least, I wish to acknowledge my profound debt to the family, friends and colleagues who, in simply being there, have made lighter the load of the more troublesome moments encountered along the way: my mother and father for their love and financial support, especially during the fiscal desert of the last eighteen months; in the MRC Virology Unit — Marie Anderson, Dina Fortinu, Nicola Arbuckle, Susan Graham, June Harland, Graham Hope and Kathryn Phillips: in body or in byte — David, Davina, Linzi, Rod MacRae, Chris Mitchell, Linda Oliver, Ken Shankland, Norman Shankland, Norman Wolff and the Drum 'N' Bass massive. Finally, thank you to Keith McKellar for more in the way of friendship, understanding and support

that mere words could ever repay.

This project received financial support from a University of Glasgow Postgraduate Scholarship, the Medical Research Council and the Cancer Research Campaign.

Abbreviations

Nothing here now but the recordings ...

Abbreviations

- Å – Angstrom
- A – adenine
- ARP – automated refinement protocol
- a.u. – atomic mass unit(s)
- B – temperature factor
- bp – base-pair(s)
- C – cytosine, constant immunoglobulin domain
- CCP4 – computational project in protein crystallography
- CGLS – conjugate-gradient least-squares
- D – accessory immunoglobulin chain
- Da – Dalton(s)
- dH₂O – sterile distilled water
- d(· · · · ·) – deoxyribose-sugar oligonucleotide sequence

- DLS – dimer linkage structure
- DMS – dimethylsulphate
- DNA – deoxyribonucleic acid
- *E. coli* – Escherichia Coli
- EDD – electron deformation density
- ESD – estimated standard deviation
- FFT – fast Fourier transform
- fs – femtosecond
- G – guanine
- GooF – goodness-of-fit
- ^1H – nuclear magnetic resonance spectroscopy isotope hydrogen-1
- HEPES – N-2 hydroxyethyl piperazine-N'-2-ethanesulphonic acid
- HIV – human immunodeficiency virus
- HIV-1 – human immunodeficiency virus type one
- HPLC – high-performance liquid chromatography
- Hrs – hours
- I – intensity

- Ig – immunoglobulin
- J – accessory immunoglobulin chain
- K - absolute temperature, Kelvin
- kDa – kiloDalton(s)
- L – long oligonucleotide sequence
- LP – Lorentz-polarization
- ln – natural logarithm
- mer – single oligonucleotide unit
- MIR – multiple isomorphous replacement
- mg – milligram
- ml – millilitre
- mm – millimetre
- mM – millimolar
- M_r – average molecular weight
- MR – molecular replacement
- nm – nanometre
- MALDI-TOF MS – matrix assisted laser desorption ionization time of flight mass spectrometry

- MPD – 2-methyl 2,4-pentanediol
- MS – mass spectrometry
- NMR – nuclear magnetic resonance
- ° – angular degrees
- °C – temperature, degrees Celsius
- O.D. – optical density
- ^{31}P – nuclear magnetic resonance spectroscopy isotope phosphorous-31
- PEG – polyethyleneglycol
- pH – $-\log_{10}[\text{H}^+]$
- Poly(G) – poly-guanylic acid
- ps – picosecond
- $r(\cdots\cdots)$ – ribose-sugar oligonucleotide sequence
- R – reliability index
- rms – root-mean-square
- RNA – ribonucleic acid
- S – tandem repeat of short oligonucleotide sequence in immunoglobulin genes,
short oligonucleotide sequence
- $\text{Sigma}(\sigma)$ – standard deviation

- SIR – single isomorphous replacement
- T – thymine
- T_m – melting temperature
- Tris – 2-amino-2-(hydroxymethyl)-1,3-propanediol
- tRNA – transfer ribonucleic acid
- tRNA^{phe} – phenylalanine transfer ribonucleic acid
- U – uracil
- UV – ultra-violet
- V – variable immunoglobulin domain
- v/v – volume/volume
- μl – microlitre
- μm – micrometre
- μmol – micromole

Contents

Summary	ii
Acknowledgements	vi
Abbreviations	ix
1 Guanine Tetrad DNA Structures and Biological Function	3
1.1 Structural Organization of DNA	3
1.1.1 The Molecular Components of DNA and RNA	3
1.1.2 DNA Crystallography and Base-Pairing	10
1.2 Guanine Self-Association	13
1.2.1 The Formation of Guanosine Gels	13
1.2.2 Guanine-Guanine Base-Pairing	14
1.2.3 The Guanine Tetrad	15
1.2.4 Ion Specificity In Guanine Tetrad Association	17
1.2.5 Chelation Cages and Ion Selectivity	17
1.3 Proposed Biological Relevance for G4 Tetrads	21
1.3.1 Telomeres	21

1.3.2	Immunoglobulin Class Switching	24
1.3.3	Retroviral RNA Dimerization	26
1.4	Biochemical and Biophysical Evidence	28
1.4.1	Telomeric DNA	28
1.4.2	Immunoglobulin Switch Regions	30
1.4.3	Dimerization of the HIV-1 Genome	32
1.5	Proposed Models for G4 Structures	33
1.5.1	Anti-Parallel Quadruple Helical Structures	33
1.5.2	Parallel-Stranded Quadruplexes	37
1.6	X-Ray and NMR High-Resolution Studies	39
1.6.1	X-Ray Structure of d(GGGGTTTTGGGG)	39
1.6.2	NMR Structure of d(GGGGTTTTGGGG)	40
1.6.3	NMR Studies of r(UGGGGU) and d(TTGGGGT)	42
1.7	Key Structural Features of G4 DNA/RNA	43
1.7.1	syn or anti Glycosyl Conformation	43
1.7.2	Structural Isomerism	44
1.7.3	Loop Configurations	44
1.7.4	Ion Specificity, Structure and Stability	45
1.8	Biological Relevance of Guanine Tetrad Structures	47
2	Crystallization of d(TGGGGT)	51
2.1	Tetraplex Formation by d(TGGGGT)	51
2.2	Solid-State Synthesis of Oligonucleotides	54

2.3	Crystallization of Nucleic Acids	55
2.3.1	Precipitating Agents	55
2.3.2	Growth Temperature	56
2.3.3	Buffering and pH	56
2.3.4	Addition of Polyamines	57
2.3.5	Addition of Divalent Cations	59
2.3.6	Addition of Monovalent Cations	60
2.4	Preparation of d(TGGGGT)	61
2.5	Mass Spectrometry Analysis of d(TGGGGT)	62
2.6	Crystallization	64
2.7	Crystal Characterization	70
2.8	Preliminary Crystallization Trials of r(UGGGGU)	71
3	Preparation of Heavy-Atom Derivatives	74
3.1	The Isomorphous Replacement Method	74
3.2	Selection of Derivatives	77
3.3	Crystallization Trials of Synthetic Oligonucleotides	80
3.3.1	d(U _{Br} GGGGT)	81
3.3.2	d(TGGGGT-O-PO ₂ =S)	84
3.4	Heavy-Atom Soaking Experiments	86
4	Data Collection and Processing	93
4.1	The Collection of Macromolecular X-Ray Data	93
4.2	The Rotation Method	97

4.3	The R–AXIS IIC Imaging Plate Area Detector	103
4.4	Data Processing of R–AXIS Data Using DENZO	105
4.4.1	Data Processing With DENZO	105
4.4.2	Data Reduction and Scaling Using ROTAVATA/AGROVATA .	108
4.4.3	Native Data	109
4.4.4	K_2PtCl_4 Derivative	112
4.4.5	$\text{Pb}(\text{CH}_3\text{CO}_2)_2$ Derivative	114
4.4.6	Methyl Iodine (Me–I) Derivative	116
4.4.7	Molecular Iodine (I_2) Derivative	116
4.4.8	$\text{UO}_2(\text{CH}_3\text{CO}_2)_2$ Derivative	119
4.4.9	Other Derivatives	119
4.5	Comparison of Native Data–Sets	121
5	Determination of Phases	130
5.1	Phasing Strategies for d(TGGGGT)	130
5.2	Phasing by Isomorphous Replacement	131
5.3	Determination of Heavy–Atom Positions	135
5.3.1	The Patterson Function	135
5.3.2	The Patterson Method	136
5.4	Scaling of Data and Patterson Analysis	139
5.5	The Molecular Replacement Method	144
5.5.1	The Rotation Function	147
5.5.2	The Translation Function	150

5.5.3	Correlation Searches	151
5.5.4	Packing Function Analysis	152
5.6	Native Patterson Function Analysis	154
5.7	Self-Rotation Function Analysis	158
5.8	Cross-Rotation Function Analysis	168
5.8.1	Generation of Search Models for Molecular Replacement . . .	168
5.8.2	The AMoRe Package	169
5.8.3	Manipulation of Search Model Coordinates	171
5.8.4	Cross-Rotation Function Calculation using ROTING	172
5.9	Translation Function Analysis	175
5.9.1	The Programs TRAIING and FITING	175
5.9.2	Translation Function Solutions for d(TGGGGT)	176
5.9.3	Progression from the Molecular Replacement Solution	180
6	Structure Refinement	186
6.1	Refinement in Macromolecular Crystallography	186
6.1.1	Definition of Refinement	187
6.1.2	The Method of Least-Squares Refinement	189
6.1.3	Calculation of Electron Density	192
6.1.4	Resolution of Macromolecular Structures	195
6.1.5	Resolution and Determinacy in Refinement	196
6.1.6	Restraints and Constraints	198
6.1.7	Correctness of the Refined Structure	200

6.2	Refinement Strategy for the d(TGGGGT) Tetraplex	203
6.2.1	Overview of the Refinement Protocol	203
6.2.2	Rigid Body Minimization	206
6.2.3	Simulated Annealing	207
6.2.4	ARP Refinement	215
6.2.5	SHELXL-93 Refinement	226
6.3	Refinement of the d(TGGGGT) Tetraplex Structure	229
6.3.1	Rigid Body Minimization	231
6.3.2	Simulated Annealing	235
6.3.3	ARP Refinement	238
6.3.4	Refinement Using SHELXL-93	247
6.4	Quality of the Final Structure Solution	256
7	The Structure of the d(TGGGGT) Parallel-Stranded Tetraplex	277
7.1	Introduction	277
7.2	Packing of Tetraplexes in the Cell	280
7.3	The Guanine Tetrad	289
7.4	Sugar-Puckering Modes of the Tetraplex	293
7.5	Phosphate-Phosphate Separations	304
7.6	Conformation of the Sugar-Phosphate Backbone	310
7.7	RMS Comparison of the Tetraplexes	321
7.8	Location of the Sodium Cations	325
7.9	Hydration of the Tetraplex	341

7.10	Stacking of the Guanine Tetrad Planes	345
7.11	Arrangements of the Thymine Nucleotides	351
7.12	Location and Environment of the Calcium Cations	362
8	Discussion	369
8.1	Overview of the Thesis Contents	369
8.2	Future Work	387
8.2.1	Higher Resolution X-Ray Structural Studies	387
8.2.2	Electron Deformation Density Studies	390
8.2.3	Structural Studies of Ribonucleicacid Tetraplexes	393
8.2.4	The Phosphorothioate Tetraplex	394
8.2.5	Tetraplex-Protein Complexes	397
8.3	Conclusion	398
	References	402

List of Figures

1.1	Illustration of Atomic Numbering Scheme for DNA	4
1.2	Atomic Numbering and Definition of Torsion Angles for DNA	5
1.3	Envelope(E) and Twist(T) Sugar-Puckers	7
1.4	The <i>syn</i> and <i>anti</i> β -glycosyl Conformations	8
1.5	P...P Distances Commonly Observed in A- and B-type Helices	9
1.6	The Watson-Crick Base-Pairing Schemes for A-T and G-C	11
1.7	The Hoogsteen Base-Pairing Schemes for A-T and G-C	12
1.8	Structures of the Four Theoretically Possible G:G Base Pairs	14
1.9	The Cyclic Guanine Base-Tetrad	16
1.10	Schematic Illustration of Axial Metal-Binding by Guanine O ₆ Ligands	18
1.11	Structures Proposed to Form in G-Rich DNA Sequences	35
1.12	Illustration of Experimentally Determined G-Quartet Cores	40
2.1	MALDI-TOF MS Analysis of 5'-TpGpGpGpGpT-3'	63
2.2	Experimental Set-Up for Crystallizations	65
2.3	A Single Crystal of the Oligonucleotide d(TGGGGT)	67
2.4	A Twinned Crystal of the Oligonucleotide d(TGGGGT)	69
2.5	Microcrystals of the RNA Analogue of d(TGGGGT), r(UGGGGU)	72

3.1	Synthetic Heavy–Atom Derivatives of Uracil(Thymine) and Cytosine	78
3.2	The Phosphothiol Chemical Group	79
3.3	MALDI–TOF MS Analysis of 5′–U _{Br} pGpGpGpT–3′	82
3.4	Microcrystals of the Heavy–Atom Derivative d(U _{Br} GGGGT)	84
3.5	MALDI–TOF MS Analysis of 5′–TpGpGpGpGpT–O–PO ₂ =S–3′	85
3.6	Experimental Set–Up for Heavy–Atom Soaking Experiments	87
4.1	The Geometry of the Oscillation Method	99
4.2	Partially Observed Reflections in the Oscillation Method	100
4.3	Data Collection in the Symmorphic Space Group P1	101
4.4	Wilson Plot for the <i>DENZO</i> Processed Data	124
4.5	Wilson Plot for the <i>MOSFLM</i> Processed Data	125
4.6	Plot of $\bar{I}/\sigma(\bar{I})$ Against Resolution for the <i>DENZO</i> Processed Data	126
4.7	Plot of $\bar{I}/\sigma(\bar{I})$ Against Resolution for the <i>MOSFLM</i> Processed Data	127
5.1	The Ideal Situation for the Single Isomorphous Replacement Method	132
5.2	Illustration of Lack of Closure in the SIR Method	133
5.3	2–D Depiction of the Rotation And Translation Operations	145
5.4	Native Patterson Function uv Section Calculated At the Cell Origin	155
5.5	Native Patterson Function uw Section Calculated At the Cell Origin	157
5.6	Definitions of the Eulerian and Polar Coordinate Systems	162
5.7	Stereographic Projection of the Self–Rotation $\kappa=180^\circ$ Section	165
5.8	Stereographic Projection of the Self–Rotation $\kappa=90^\circ$ Section	166
5.9	The Tp0 Model as used in Molecular Replacement Studies	170

5.10	Packing Arrangement of the Tp4 Tetraplexes in the Unit Cell	179
5.11	The Model Tp4-T as used in Molecular Replacement Studies	182
5.12	Packing Arrangement of the Tp4-T Tetraplexes in the Unit Cell	183
6.1	Illustrative Overview of the Refinement Procedure	188
6.2	Illustration of the Imposition of Restraints and Constraints	199
6.3	Overview of the d(TGGGGT) Structure Refinement Procedure	204
6.4	The Unrestrained <i>ARP</i> Refinement Protocol	218
6.5	The Restrained <i>ARP</i> Refinement Protocol	222
6.6	Alternative Cell–Packing Arrangement of the Tp4-T Tetraplexes	232
6.7	Luzzati Plot for the Refined d(TGGGGT) Tetraplex Structure	254
7.1	Schematic of Strand Orientation and Packing in the Cell	281
7.2	The Unit Cell Contents	283
7.3	View of the Unit Cell Contents Along the Helical Axis	288
7.4	The Guanine Tetrad	290
7.5	Deoxyribose Sugar in a Southern Conformation	302
7.6	Deoxyribose Sugar in a Northern Conformation	303
7.7	The P···P Separation Distances of the Tetraplex	307
7.8	Guanine Nucleotide in the <i>anti</i> Configuration	319
7.9	Guanine Nucleotide in the Classical <i>anti</i> Configuration	320
7.10	RMS Superposition of the Coaxial Pair (A, B)	324
7.11	The Inter–Tetrad Sodium Cations	327
7.12	The Intra–Tetrad Sodium Cations	331

7.13	The Longitudinal, Interdigitating 3' Thymine Nucleotides	337
7.14	Hydration of the Tetraplex	343
7.15	Guanine Tetrad Stacking Within a Tetraplex	347
7.16	Stacking at the Interface Between Tetraplexes	349
7.17	The Lateral Stacking of the 5' Thymine Nucleotides	352
7.18	Stacking Patterns of the Six-Membered Thymine Stack	355
7.19	Stacking Patterns of the Five-Membered Thymine Stack	357
7.20	Other Intermolecular Thymine Contacts – I	359
7.21	Other Intermolecular Thymine Contacts – II	360
7.22	Location and Environment of the Ca^{2+} Cations	364

List of Tables

2.1	Examples of Polyamines	58
2.2	Reagents Used in Crystallization Trials	66
3.1	Reagents Used in Crystallization Trials of Synthetic Derivatives . . .	83
3.2	Mass Spectrometry of the Phosphothiol Derivative of d(TGGGGT) . .	86
3.3	Heavy-Atom Soaking Experiments	88
3.4	Heavy-Atom Screening Experiments	90
3.5	Cell Dimensions as Determined by <i>REFIX</i>	91
4.1	Scaling Statistics for <i>DENZO</i> Processed CuK α Data	110
4.2	Scaling Statistics for <i>DENZO</i> Processed CuK α & MoK α Data	111
4.3	Scaling Statistics for <i>DENZO</i> Processed K ₂ PtCl ₄ Derivative	113
4.4	Scaling Statistics for <i>DENZO</i> Processed Pb(CH ₃ CO ₂) ₂ Derivative . .	115
4.5	Scaling Statistics for <i>DENZO</i> Processed Methyl Iodine Derivative . .	117
4.6	Scaling Statistics for <i>DENZO</i> Processed I ₂ Derivative	118
4.7	Scaling Statistics for <i>DENZO</i> Processed UO ₂ (CH ₃ CO ₂) ₂ Derivative .	120
4.8	Scaling Statistics for <i>MOSFLM</i> Processed CuK α & MoK α Data . . .	122
5.1	Scaling Statistics for Native to Derivative Datasets	141

5.2	Self-Rotation Peaks For 10.0–4.00Å Data	160
5.3	<i>AMoRe</i> Cross-Rotation Peaks, 10.0–3.00Å Data	173
5.4	<i>AMoRe</i> Translation Peaks, 10.0–3.00Å Data	177
6.1	Refinement Statistics for the Restrained <i>ARP</i> Refinement Procedure	246
6.2	<i>SHELXL-93</i> Distance Restraints for Thymine Heterocycles	248
6.3	<i>SHELXL-93</i> Distance Restraints for Guanine Heterocycles	249
6.4	<i>SHELXL-93</i> Similarity Restraints for Deoxyriboses	250
6.5	<i>SHELXL-93</i> Similarity Restraints for Phosphates	251
6.6	<i>SHELXL-93</i> Similarity Restraints for Base–Deoxyribose Linkages	251
6.7	Summary of Final Refinement Statistics from <i>SHELXL-93</i>	255
6.8	Summary of Average Isotropic Temperature Factors	260
6.9	Bond Lengths and Estimated Standard Deviations – Residue 5'T(1)	264
6.10	Bond Lengths and Estimated Standard Deviations – Residue G(2)	265
6.11	Bond Lengths and Estimated Standard Deviations – Residue G(3)	266
6.12	Bond Lengths and Estimated Standard Deviations – Residue G(4)	267
6.13	Bond Lengths and Estimated Standard Deviations – Residue G(5)	268
6.14	Bond Lengths and Estimated Standard Deviations – Residue 3'T(6)	269
6.15	Bond Angles and Estimated Standard Deviations – Residue 5'T(1)	270
6.16	Bond Angles and Estimated Standard Deviations – Residue G(2)	271
6.17	Bond Angles and Estimated Standard Deviations – Residue G(3)	272
6.18	Bond Angles and Estimated Standard Deviations – Residue G(4)	273
6.19	Bond Angles and Estimated Standard Deviations – Residue G(5)	274

6.20	Bond Angles and Estimated Standard Deviations – Residue 3'T(6) . .	275
7.1	Local Helical Parameters for Tetraplexes A, B, C, and D	285
7.2	Sugar Torsion Angles(°) and Furanose Conformations for Tetraplex A	296
7.3	Sugar Torsion Angles(°) and Furanose Conformations for Tetraplex B	297
7.4	Sugar Torsion Angles(°) and Furanose Conformations for Tetraplex C	298
7.5	Sugar Torsion Angles(°) and Furanose Conformations for Tetraplex D	299
7.6	Adjacent P . . . P Distances for Tetraplexes A, B, C, and D	305
7.7	Backbone Torsion Angles(°) for Tetraplex A	311
7.8	Backbone Torsion Angles(°) for Tetraplex B	312
7.9	Backbone Torsion Angles(°) for Tetraplex C	313
7.10	Backbone Torsion Angles(°) for Tetraplex D	314
7.11	Na ⁺ . . . O ₆ Distances for Axial Sodium Ions Between Planes G2 & G-2	328
7.12	Na ⁺ . . . O ₆ Distances for Axial Sodium Ions Between Planes G2 & G3	329
7.13	Na ⁺ . . . O ₆ Distances for Axial Sodium Ions Between Planes G3 & G4	330
7.14	Na ⁺ . . . O ₆ Distances for G5 Planar Axial Sodium Ions	332
7.15	Na ⁺ . . . Na ⁺ Distances and Temperature Factors for Axial Sodium Ions	333
7.16	Ca ²⁺ Location, Temperature Factor and Ca ²⁺ . . . O(H ₂ O) Distances . .	365

We also now appreciate that molecular biology is not a trivial aspect of biological systems. It is at the heart of the matter. Almost all aspects of life are engineered at the molecular level, and without understanding molecules we can only have a very sketchy understanding of life itself. All approaches at a higher level are suspect until confirmed at the molecular level.

Francis Crick, *What Mad Pursuit*

An anomaly in the genetic code of slipper animalcules is announced. A hard blow for the champions of the code. But if the code is no longer universal, isn't this because it is no longer fashionable today to aspire to universality? The idea of genetic code is also going through its agonizing phase of revision, adjusting itself to a new scientific 'look'. The only winner now is the code of fashion, which lays no claim to universality.

A more vertiginous hypothesis: DNA *was* a universal and an invariant, but, once unmasked, so to speak, it has begun to change in order to throw the researchers off its track; it is deuniversalizing itself to muddy the waters. Out of spite or malice, or simply by adaptation, in the way bacteria adapt to antibiotics. We musn't underestimate the capacity 'objective' processes have for playing hide-and-seek with science. You understand nothing of science, nor of its failures, if you don't take into account this ultimately quite natural evil genius. If all human and animal species change their behaviour under observation, why shouldn't the same thing happen at the level of molecular species?

Jean Baudrillard, *Cool Memories: 1980-1985*

Paul at Damascus and Christopher Columbus
Only found what they knew
Close your eyes and you may find it too
RNA and DNA
In a post-ethical age
Nothing is true
We move like shadows across the stage

Momus, *Cibachrome Blue*

There came a day when Nikoli saw his first Mechanist. The man was a diplomat and commercial agent, stationed by his faction in Nikoli's habitat. Nikoli and some children from his creche were playing in the corridor when the diplomat stalked by. One of the Mechanist's legs was malfunctioning and it went *click-whirr, click-whirr*. Nikoli's friend, Alex, mimicked the man's limp. Suddenly the man turned on them, his plastic eyes dilating. "Genelines," the Mechanist snarled. "I can buy you, grow you, sell you, cut you into bits. Your screams: my music."

Bruce Sterling, *Twenty Evocations*

Above all, it is in the environment of the politics of the body that a progressive sense of social individualism can be posed as an alternative to the dominant laissez-faire conception of the atomized individual. Prevailing conservative definitions of the individual appeal to a limited sphere of "privacy", where our bodies are either policed by puritan morality or else sanctified by capitalism as a temple of consumer choice. By contrast, an expanded conception of social individualism involves positive, liberatory rights of freedom and choice, drawn from the guarantees of democratic citizenship rather than from the *magna carta* of consumerism. Only through attention to individual rights can we build a radical democracy, guaranteeing respect both for the differences among individuals and for the environments – physical, cultural and social – that we inhabit. In this way, getting the future we deserve might still involve just desserts for all.

Andrew Ross, *Strange Weather*

Taking responsibility for the social relations of science and technology means refusing an anti-science metaphysics, a demonology of technology, and so means embracing the skilful task of reconstructing the boundaries of daily life, in partial connection with others, in communication with all of our parts. It is not just that science and technology are possible means of great human satisfaction, as well as a matrix of complex dominations. Cyborg imagery can suggest a way out of the maze of dualisms in which we have explained our bodies and our tools to ourselves . . . I'd rather be a cyborg than a goddess.

Donna Haraway, *A Manifesto for Cyborgs*

Chapter 1

Guanine Tetrad DNA Structures and Biological Function

1.1 Structural Organization of DNA

1.1.1 The Molecular Components of DNA and RNA

Deoxyribonucleic acid (DNA) is the hereditary material. It is a linear polymer consisting of monomeric units called nucleotides. Such units have miscellaneous and important biological functions in their own right.

A nucleotide is built up of three molecular components: sugar, heterocycle and phosphate. These components are illustrated as part of the DNA chain in Figure 1.1 (adapted from (Saenger (1984))), along with their organisation as a nucleotide unit. Figure 1.1 also illustrates the atomic numbering conventions for the DNA chain, the definition of the 5' to 3' chain polarity and the chemical composition of the DNA bases adenine(A), guanine(G), cytosine(C) and thymine(T). The nucleic acid sugar

is in a cyclic, furanoside form (β -D-ribose in ribonucleic acid (RNA) and β -D-2'-deoxyribose in DNA), which is phosphorylated in the 5' position and substituted at C_{1'} by one of four different heterocycles attached by a β -glycosyl C_{1'}-N linkage. The heterocycles are, as defined in Figure 1.1, the purine bases adenine and guanine and the pyrimidine bases cytosine and uracil. The latter is only found in RNA and is replaced by the functionally equivalent thymine (5-methyluracil) in DNA. The latter is only found in RNA and is replaced by the functionally equivalent thymine (5-methyluracil) in DNA.

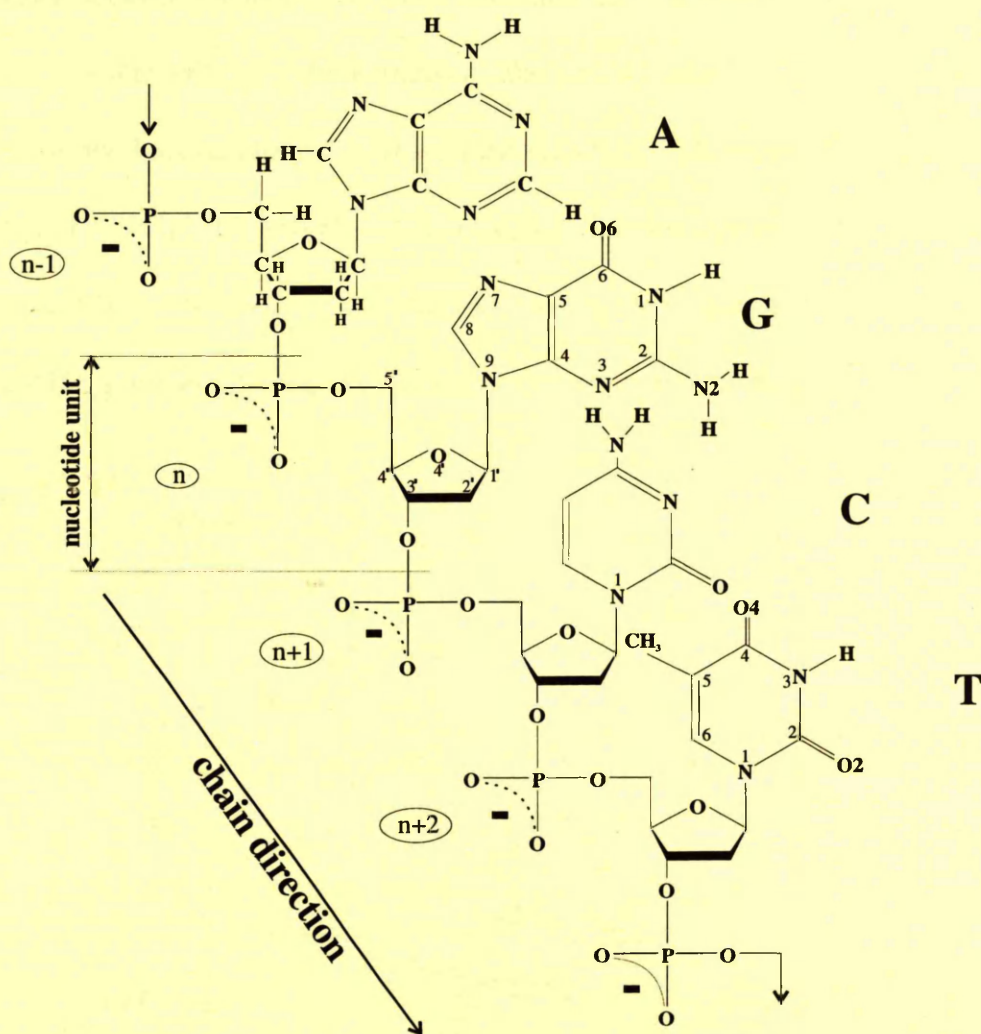


Figure 1.1: Illustration of Atomic Numbering Scheme for DNA

It is useful at the outset to provide a brief overview of the key structural elements

often invoked in describing the various details of a nucleic acid structure. All of the figures discussed in this section have been adapted from Saenger (1984). Figure 1.2 illustrates a polyribonucleotide chain. The atomic numbering scheme outlined in Figure 1.1 is included here in Figure 1.2, with the atomic element labels included along with the numbering scheme adopted. As in Figure 1.1, the counting of the nucleotides is defined from the top to the bottom of the figure, running in the 5' to 3' direction as drawn between the atoms $O_{5'}$ and $O_{3'}$ within one nucleotide unit. Hydrogens at $C_{5'}$ and oxygens at P are differentiated by the labels 1 and 2. This follows from the rule whereby, looking along $O_{5'} \rightarrow C_{5'}$ in the direction of the main chain, a clockwise counting gives $C_{4'}$, $H_{5'1}$ and $H_{5'2}$. Likewise, looking along $O_{3'} \rightarrow P \rightarrow O_{5'}$, a clockwise counting defines O_1 and O_2 at the phosphorous. In the polydeoxyribose chain, $O_{2'}$ is replaced $H_{2'1}$ and $H_{2'2}$ which are defined by an analogous convention.

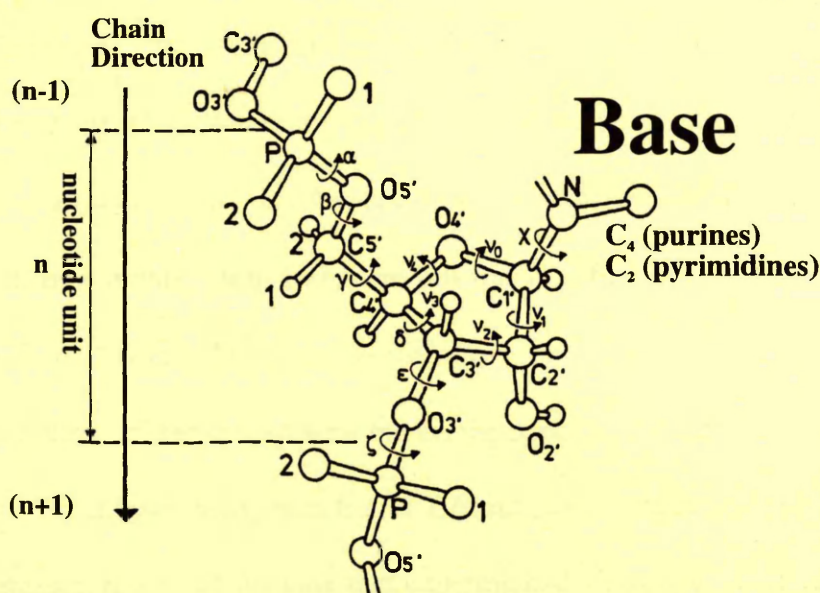


Figure 1.2: Atomic Numbering and Definition of Torsion Angles for DNA

Figure 1.2 also shows the definitions of the torsion angles of the nucleic acid backbone. Following the sequence of atoms $P \rightarrow O_{5'} \rightarrow C_{5'} \rightarrow C_{4'}$ etc, the backbone is defined by the angles (and atoms) $\alpha(O_{n-1}O_{3'}-P-O_{5'}-C_{5'})$, $\beta(P-O_{5'}-C_{5'}-C_{4'})$, $\gamma(O_{5'}-C_{5'}-C_{4'}-C_{3'})$, $\delta(C_{5'}-C_{4'}-C_{3'}-O_{3'})$, $\epsilon(C_{4'}-C_{3'}-O_{3'}-P)$ and $\zeta(C_{3'}-O_{3'}-P-O_{5'(n+1)})$. The endocyclic sugar torsion angles ($\nu_0, \nu_1, \nu_2, \nu_3, \nu_4$) collectively define the conformation of the nucleic acid sugar and are determined by the atoms $C_{4'}-O_{4'}-C_{1'}-C_{2'}$ (ν_0), $O_{4'}-C_{1'}-C_{2'}-C_{3'}$ (ν_1), $C_{1'}-C_{2'}-C_{3'}-C_{4'}$ (ν_2), $C_{2'}-C_{3'}-C_{4'}-O_{4'}$ (ν_3) and $C_{3'}-C_{4'}-O_{4'}-C_{1'}$ (ν_4). Finally, the orientation of the base relative to the sugar is described by the torsion angle χ , the β -glycosyl $C_{1'}-N$ linkage torsion angle. This torsion angle is described by the atoms $O_{4'}-C_{1'}-N_1-C_2$ in pyrimidines (thymine, uracil and cytosine) and $O_{4'}-C_{1'}-N_9-C_4$ in purines (guanine and adenine).

The five-membered furanose ring in the nucleic acid backbone is generally non-planar and is found in one of two major forms (Saenger (1984)). It can be puckered in an envelope conformation (**E**) containing four atoms in a plane with the fifth atom displaced by about 0.5\AA . Alternatively, the ring may adopt a twist conformation (**T**) in which two adjacent atoms are displaced on opposite sides of a plane constituted by the other three atoms. Both forms are illustrated in Figure 1.3. Atoms displaced from these 3- or 4-atom planes and on the same side as the $C_{5'}$ atom are designated *endo*. Those atoms likewise displaced but on the opposite side of the sugar plane are termed *exo*. The transition between **E** and **T** forms is practically negligible in energy terms. Therefore, it is often the case that experimentally observed 4-atom planes are not perfectly coplanar, whilst deviations from 3-atom planes are seldom equivalent on both sides. The largest observed deviation from planarity is termed major-puckering,

whilst the lesser is referred to as minor-puckering.

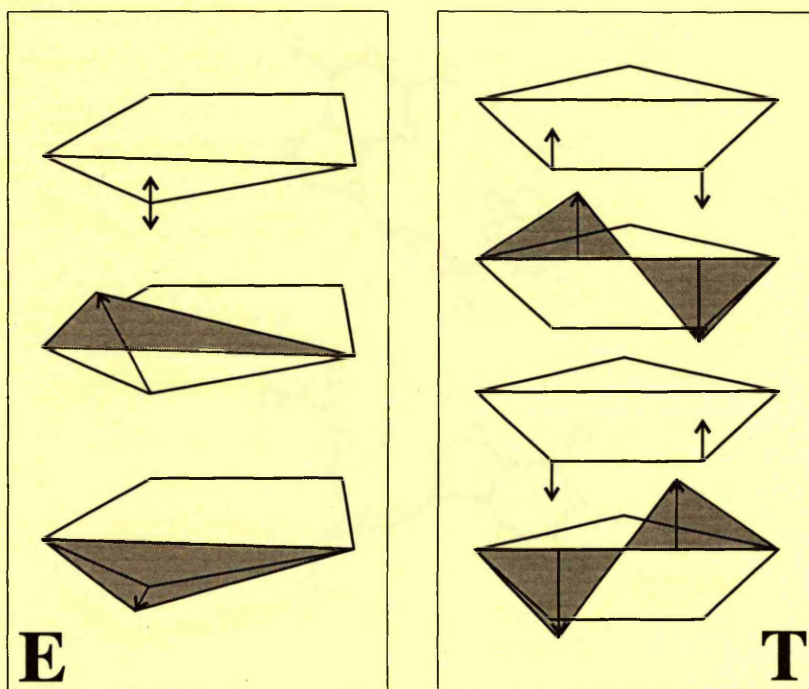


Figure 1.3: Envelope(E) and Twist(T) Sugar-Puckers

In relation to the furanose sugar of the nucleic acid backbone the heterocyclic base can adopt two main orientations about the β -glycosyl C_1' -N linkage (Saenger (1984)). These are termed the *anti* and *syn* conformations and are described by the torsion angle χ depicted in Figure 1.2.

The *anti* and *syn* conformations are illustrated for the guanine nucleotide in Figure 1.4. In the *anti* conformation the heterocycle is orientated such that it is pointing away from the sugar, whilst in the *syn* conformation it is orientated towards or above the sugar. In the pyrimidine heterocycles (uracil, thymine and cytosine) the exocyclic O_2 atom points towards the sugar for the *syn* conformation and away for the *anti*. A variety of conventions exists to describe the conformation of the β -glycosyl C_1' -N linkage,

many of which are summarised in Saenger (1984).

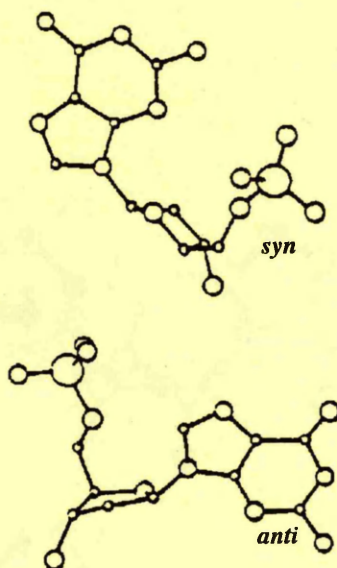


Figure 1.4: The *syn* and *anti* β -glycosyl Conformations

As a final point in this general introductory discussion it is interesting to note that DNA structures exhibit a considerable degree of polymorphism in the features outlined above (Saenger (1984)). In general, right-handed DNA helical structures can be broadly classified into two generically different families, namely A- and B-DNA. Since the B-form is adopted by the molecule in the fibrous form, under conditions of high humidity, as well as within aqueous solutions of DNA, it is generally believed to represent the native DNA structural form occurring in living organisms (Saenger (1984)).

The essential distinction between the A- and B-type polynucleotide helices lies in the mode of the deoxyribose sugar-puckering. In A-DNA families the $C_{3'}$ -*endo* envelope (**E**) form predominates, whilst in the B-DNA families the $C_{3'}$ -*exo* or $C_{2'}$ -*endo*

envelope (**E**) forms are observed. This leads to a fundamentally important variation in the distance between adjacent phosphates in the nucleic acid chain, illustrated in Figure 1.5.

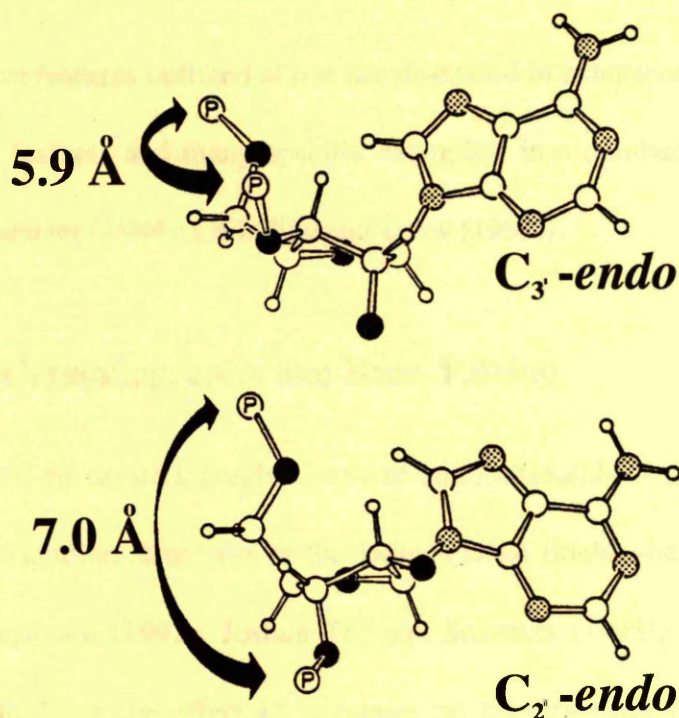


Figure 1.5: P...P Distances Commonly Observed in A- and B-type Helices

In the A-DNA family the P...P separation tends to be in the region of 5.9Å for the C_3' -endo sugar-pucker, whilst in the B-DNA family this distance can become as large as 7.0Å for the C_2' -endo sugar conformation. A direct consequence of this variance is that the helical geometry of the A- and B-DNA families are macroscopically different (discussed in great detail in Saenger (1984)). By contrast, structures of RNA helices are more conservative, exhibiting characteristics of the A-type family of helices but

not of the B-type. In solution, DNA of random sequence adopts an intermediate conformation between the ideal A and B forms, with a helical repeat of 10.5 base-pairs per turn. This is in contrast to 10 and 11 base-pairs per turn for the B- and A-forms, respectively.

The structural features outlined above are discussed in extensive detail, alongside other structural features and many specific examples, in a number of textbooks (in particular see Saenger (1984); Calladine and Drew (1992)).

1.1.2 DNA Crystallography and Base-Pairing

The early years of the crystallographic study of oligonucleotides were concerned with deriving high-resolution structures of the various DNA double-helical families (for reviews see Dickerson (1992); Joshua-Tor and Sussman (1993)). The focus soon shifted to the study of the effect of sequence on the fine structure of DNA. The importance of these structures was somewhat obscured by the fact that it was found that packing effects played an influential role in the structures of the oligonucleotides studied. However, a number of structures were resolved in different crystal forms and temperatures so that packing and sequence effects could be separated in some cases (for a review see Shakked (1991)). Many early studies investigated DNA single-base mismatches (Brown and Kennard (1992)) and DNA-drug complexes (Kennard and Hunter (1991); Wang (1992)), often utilising templates that proved successful in prior crystallographic studies.

Such double helical structures involve the formation of base-pairs. There are two rudimentary ways of doing this. The first, and the type seen almost exclusively in

double helical structures, is the Watson–Crick arrangement for A–T and G–C base-pairs (Figure 1.6).

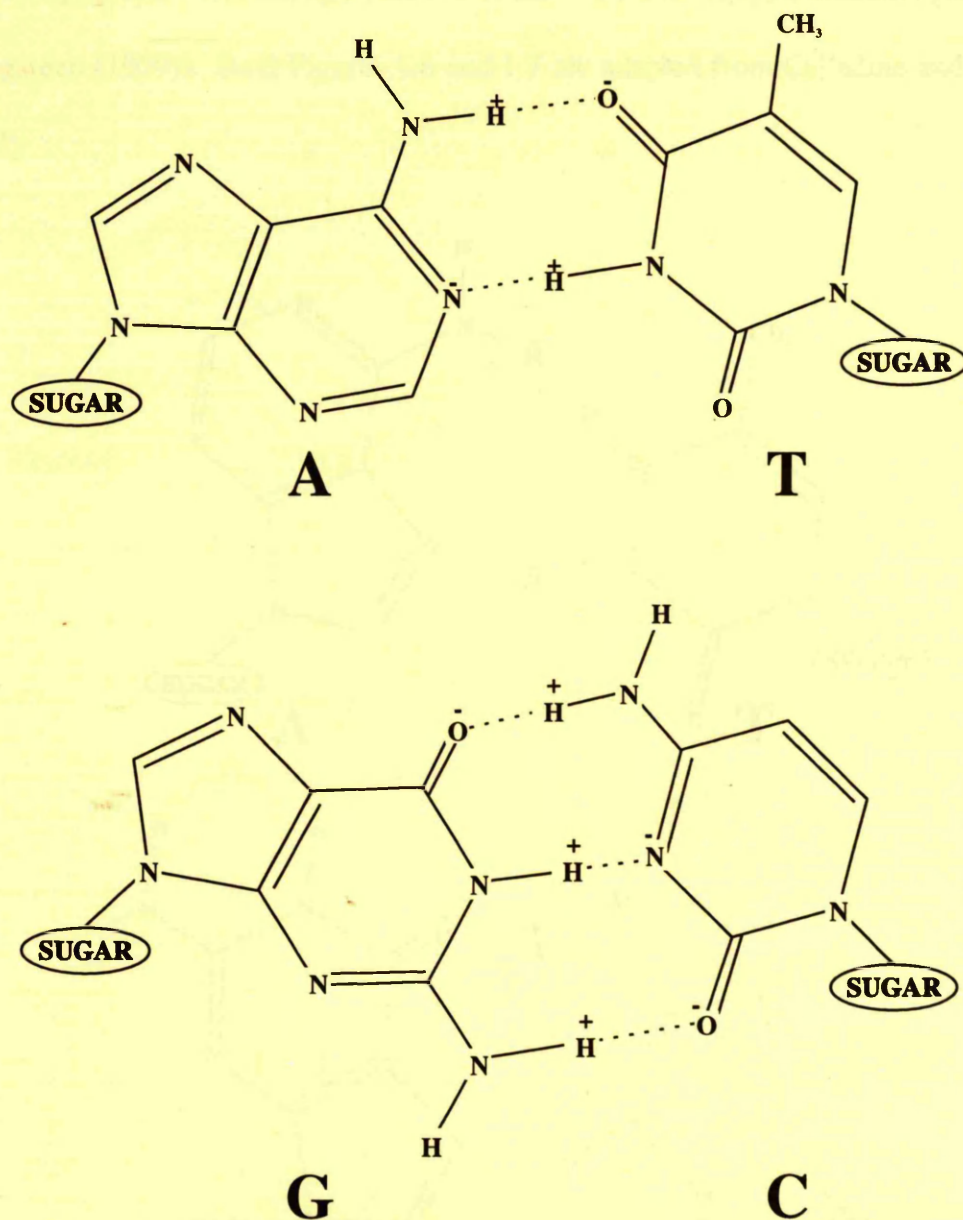


Figure 1.6: The Watson–Crick Base–Pairing Schemes for A–T and G–C

These arrangements were derived by Watson and Crick as part of their search for the double–helical structure of DNA by using paper cut–outs of bases (Watson and Crick (1953)). The second, and by far the more exotic, is the Hoogsteen base–

pair arrangements for A–T and G–C (Figure 1.7), which were discovered by Karst Hoogsteen as he attempted to verify the Watson–Crick pairs for adenine and thymine by heating up and then cooling a solution of the two bases for crystallization purposes (Hoogsteen (1959)). Both Figures 1.6 and 1.7 are adapted from Calladine and Drew (1992)

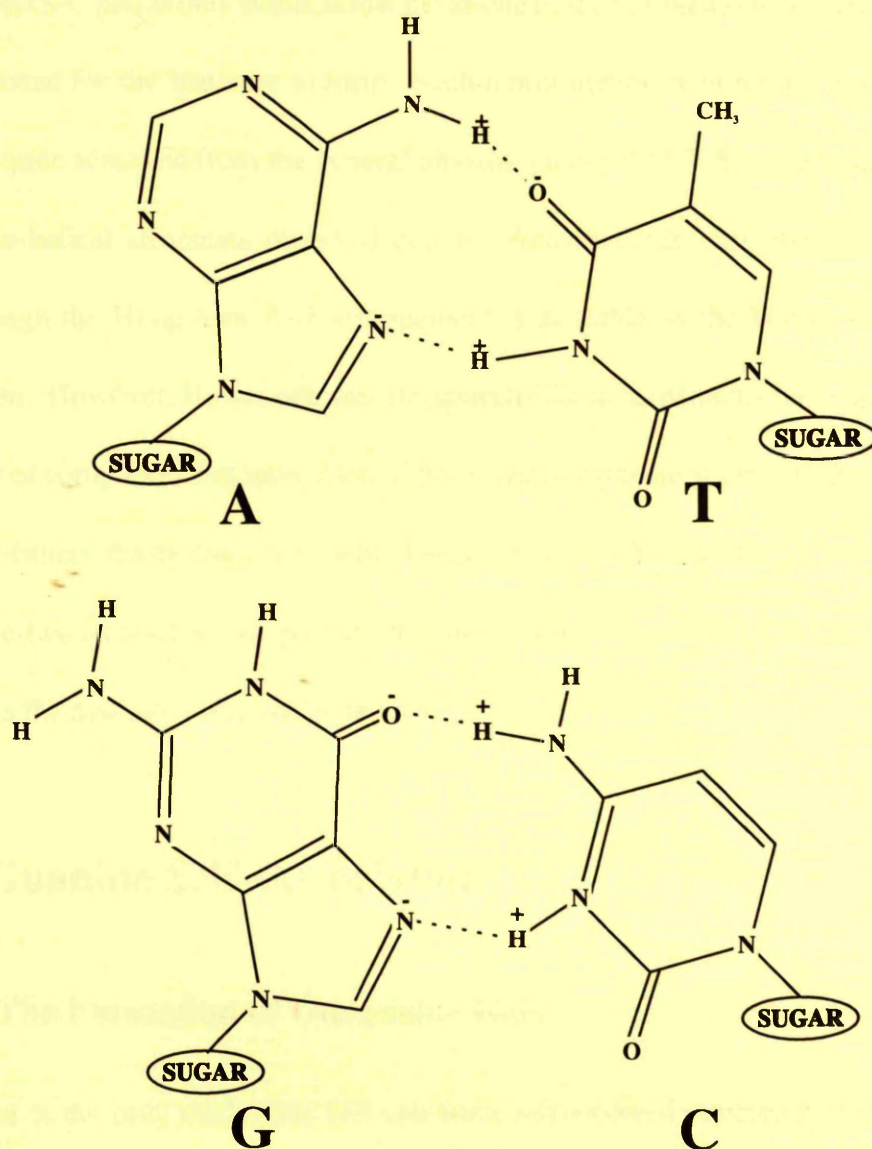


Figure 1.7: The Hoogsteen Base–Pairing Schemes for A–T and G–C

The A–T pairs are of equivalent stability in each case as both contain two hydrogen–

bonds. The two forms are interchanged by the rotation of the adenine by 180° about the glycosyl linkage.

This is not the case however with the G–C pairs. The Watson–Crick arrangement is the more stable for a number of reasons. Firstly, the Watson–Crick arrangement has three hydrogen–bonds, whilst the Hoogsteen arrangement has only two. Secondly, the Hoogsteen G–C pair is only stable at low pH as one of the nitrogens on the cytosine must be protonated for the structure to form. Such a protonation requires a pH of about 5, which is quite removed from the general physiological pH of 7–8. As a result, almost all double-helical structures observed contain Watson–Crick base–pairing schemes. Even though the Hoogsteen A–T arrangement is as stable as the Watson–Crick it is rarely seen. However, Hoogsteen and Hoogsteen-like arrangements have appeared in a number of complexes that have risen to prominence in recent years. DNA complexes with anti-cancer drugs and triple helical structures, in which a third strand joins the others, are two examples. But perhaps the most fascinating are those structures which adhere via the assembly of guanine tetrads.

1.2 Guanine Self–Association

1.2.1 The Formation of Guanosine Gels

Guanosine is the only nucleotide that can form self–ordered structures in solution, a phenomenon that was first observed in 1910 (Bang (1910)). This ability of guanosine (and its derivatives) to self–associate reflects several of the unique properties of the guanine base. In particular, guanine can participate in strong stacking interactions and

has multiple hydrogen-bonding and metal-binding sites which provide for a variety of different interactions.

1.2.2 Guanine-Guanine Base-Pairing

Four distinct G-G base-pairs may be constructed by the major tautomeric form of guanine at neutral pH, as shown in Figure 1.8 adapted from Donohue (1956).

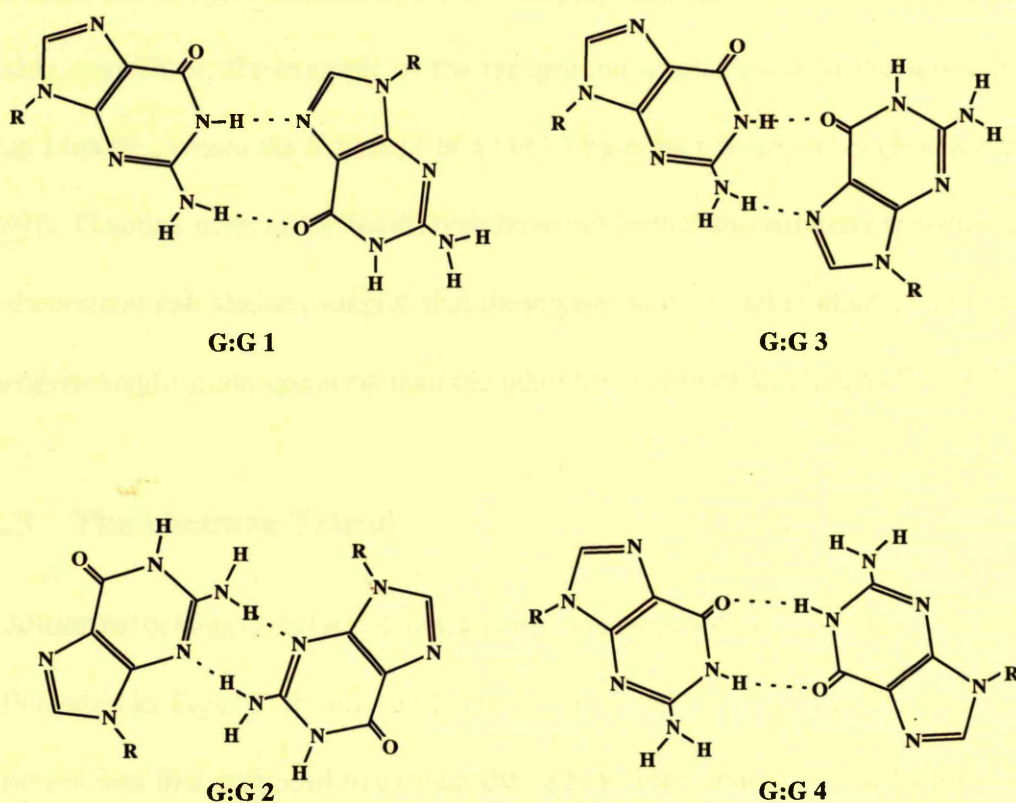


Figure 1.8: Structures of the Four Theoretically Possible G:G Base Pairs

Base-pair G:G 1 has been observed crystallographically in the structure of tRNA^{phe} (Kim *et al.*, (1974); Robertus *et al.*, (1974)), as well as in the crystallographic structures of guanine, guanosine and cyclic diguanylic acid (Bugg *et al.*, (1968); Egli *et al.*, (1990)). Chemical probing experiments indicate that G:G 1 base-pairs also form

in triple helical DNA structures of the purine–purine–pyrimidine type (Kohwi and Kohwi–Shigematsu (1988)). Guanine base–pairs of the second type have also been observed crystallographically in the intermolecular packing of B–DNA (Wing *et al.*, (1980)), cyclic diguanylic acid (Egli *et al.*, (1990)) and derivatives of tryptophan 7–methylguanine (Ishida *et al.*, (1988)).

Guanine base–pairs of the third and fourth types (G:G 3 and G:G 4) have not yet been observed in crystal structures. It has been proposed that such pairs will form under suitable conditions; for example in the recognition of guanosine in the self-splicing group I introns, where the existence of a G:G 3 base–pair is suspected (Michel *et al.*, (1989)). Guanine pairs of the fourth type have not been demonstrated experimentally, but theoretical calculations suggest that these pairs may in fact contain a more stable hydrogen–bonding configuration than the other three (Hobza and Sandorfy (1987)).

1.2.3 The Guanine Tetrad

In addition to forming G–G base–pairs, guanine has the propensity to form cyclic tetrads as illustrated in Figure 1.9, adapted from Gellert *et al.*, (1962). Guanine base–tetrad formation was first proposed to explain the self–ordered helical structures formed by guanylic acid (Gellert *et al.*, (1962)). Infinite stacking of guanylic acid tetrads was demonstrated to generate a quadruple helical structure which was consistent with fibre diffraction data (Gellert *et al.*, (1962)). Subsequent experiments have demonstrated that most other guanosine derivatives and polymers can associate into similar quadruple helices (Tougard *et al.*, (1973); Arnott *et al.*, (1974); Zimmerman *et al.*, (1975); Zimmerman (1976); Mariani *et al.*, (1989)).

On the basis of fibre diffraction data and model building studies, two very similar models were proposed for quadruple helices of poly(G) (Arnott *et al.*, (1974); Zimmerman *et al.*, (1975)).

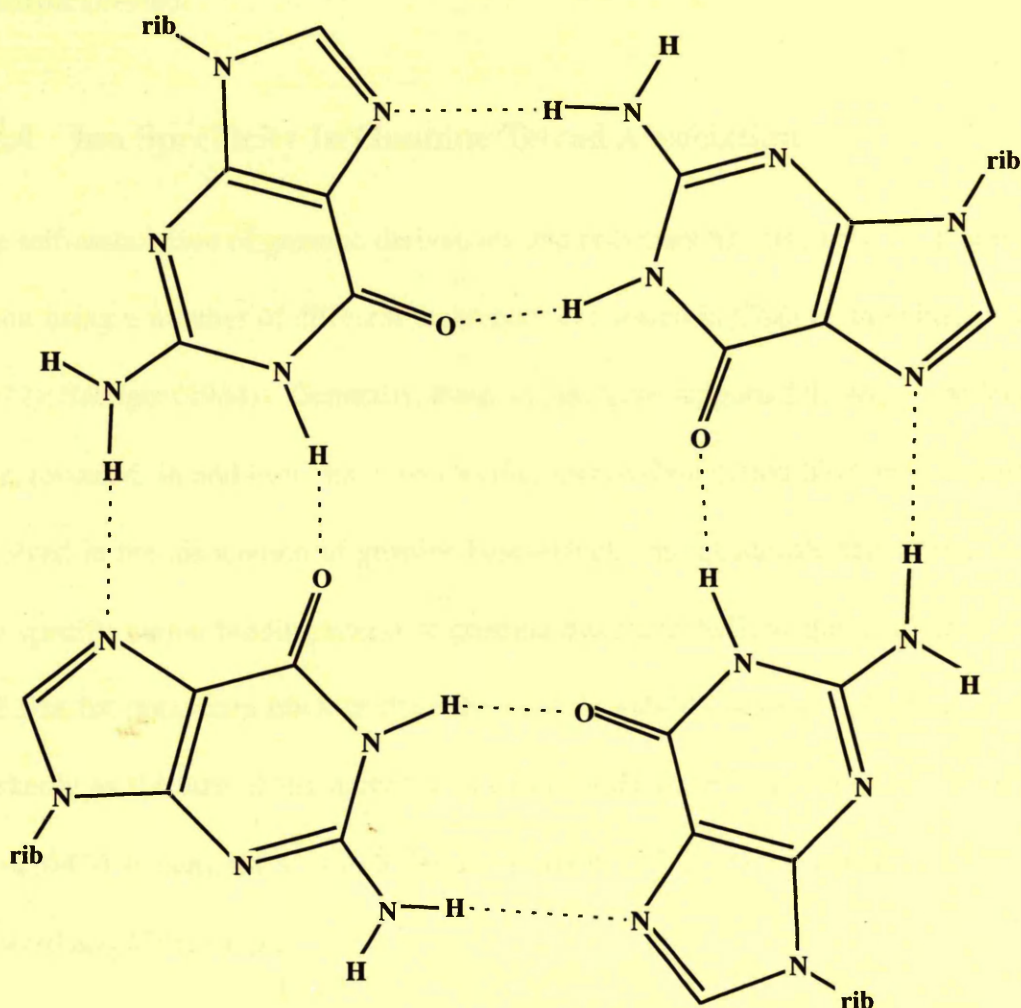


Figure 1.9: The Cyclic Guanine Base-Tetrad

In both models, poly(G) forms a right-handed quadruple helix in which each helical step contains a guanine base-tetrad. Tetrad steps are stacked with a separation of 3.4 Å with a rotation per tetrad of 31°. This produces a helix of 11.5 bases and a pitch of 39.2°. The four poly(G) strands in this structure have the same orientation and all

guanine glycosyl linkages adopt the *anti* conformation, hence avoiding steric hindrance between sugar and base. This is in opposition to the *syn* conformation in which the bulky part of the base is located over the sugar, giving rise to close interatomic contacts (Saenger (1984)).

1.2.4 Ion Specificity In Guanine Tetrad Association

The self-association of guanine derivatives and polymers has also been studied in solution using a number of different techniques (reviewed in Chantot and Guschlbauer (1972); Saenger (1984)). Generally, these studies have supported the tetrad model and have revealed, in addition, that size-specific, monovalent cation binding is integrally involved in the association of guanine base-tetrads into quadruple helical structures. The specific cation binding site(s) of guanine quadruple helices appear to be of optimal size for potassium binding since the relative stability of quadruplexes decreases markedly as the size of the monovalent cation is decreased ($K > Na > Li$) or increased ($K > Rb > Cs$) as demonstrated in Miles and Frazier (1978), Pinnavaia *et al.*, (1978) and Howard and Miles (1982).

1.2.5 Chelation Cages and Ion Selectivity

Potassium

It has been postulated that the cation-specific stability of quadruple helical structures arises because potassium can optimally bind in chelation cages between adjacent tetrads in the axial channel of the quadruplex (Miles and Frazier (1978)). In this model the

eight O_6 ligands which define the chelation cage are donated by guanines in the tetrads above and below the metal-binding site. These eight-coordinate axial cages have an ideal size and ligand stoichiometry for potassium chelation, based on calculations from the fibre diffraction model of poly(G) (Zimmerman *et al.*, (1975)). The potassium binding site within a guanine quadruplex helix is illustrated as a schematic in Figure 1.10.

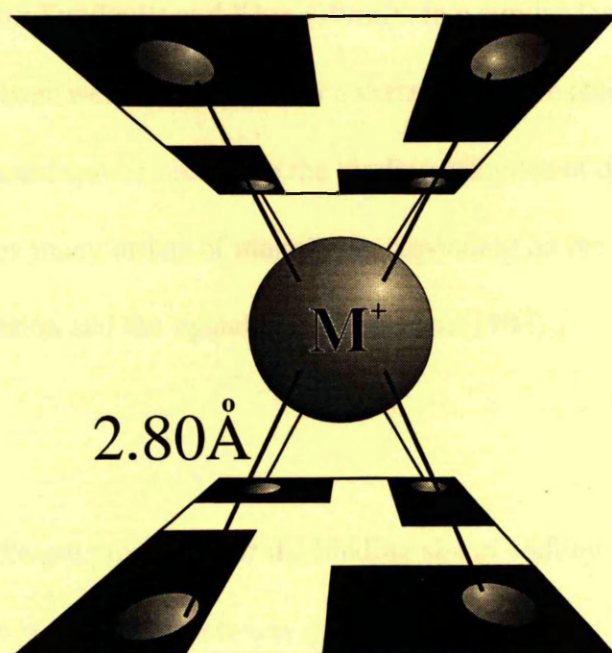


Figure 1.10: Schematic Illustration of Axial Metal-Binding by Guanine O_6 Ligands

In Figure 1.10, the eight-coordinate chelation cage created within the axial channel of the quadruplex helix by stacking of the adjacent G-tetrads is shown (adapted from Miles and Frazier (1978)). Each guanine surrounding the cage contributes a single O_6 ligand. The former are drawn as black planes, and the latter as small, shaded spheres. For clarity, the stacked tetrads are depicted without helical twist. The average $M \cdots O$ distance (2.80\AA) within the cage agrees well with the expected $K \cdots O$ bondlength

(2.78Å) as calculated from the sum of the relevant van der Waals radii. The M···O distance of 2.80Å is derived from a fiber diffraction model of the quadruple helical structure of poly(G) (Zimmerman *et al.*, (1975)). Since the fibre study produces an averaged model, individual chelation cages may be expected to vary in dimension.

Other monovalent cations, such as sodium, would be expected to fit less well into a chelation cage of this size and hence stabilize these structures to a lesser degree (Pinnaia *et al.*, (1978); Sundquist and Klug (1989)). In a similar fashion, size-selective cation binding has been well documented for a variety of crown ether structures, as well as for other ion-ligand systems in which the binding affinities of different monovalent cations can differ by many orders of magnitude, depending on the complementarity of size between the cation and the ligand cage (Pedersen (1988)).

Sodium

There exist two different proposals for the binding site of sodium in quadruple helical structures. As the axial cages between guanine tetrads appear to be too large for the optimal binding of sodium, it has been postulated that sodium binds instead within tetrad planes (Howard and Miles (1982); Williamson *et al.*, (1989)). The Na···O bonds calculated from this model are 2.3Å. Whilst this distance is acceptable in itself, sodium binding in the tetrad plane would require four coordinate, square planar coordination which has not been typically observed in sodium model complexes. Indeed, eight coordinate ligand cages which preferentially bind potassium usually also bind sodium in the same site, although the affinity is smaller and, in a number of cases, involves a reduction in the overall coordination (Gandour *et al.*, (1986)). Precedents exist in

model complexes for the idea that sodium can bind and “rattle around” in a potassium-specific cage of set size (Cram *et al.*, (1986)). Sundquist and Klug (1989) have therefore proposed that both sodium and potassium bind in the axial cages between tetrads in such quadruplexes, but that the potassium binds with increased affinity and thereby stabilizes these structures considerably better. However, to a greater extent, binding sites exclusive to sodium are feasible at the ends of quadruple helical structures where water molecules could function as additional coordinating ligands (Brown-Bouhoutsos (1982)). It is not only the size of the ion which is of importance. Another important factor is the free-energy cost of stripping away the hydration shell of the metal, necessary for the transition from solution to the “crown ether” configuration of the tetraplex (Williamson (1993)).

Thus, overall, guanine quadruple helical structures are remarkably stable despite the electrostatic and steric repulsions that might be anticipated in such a nucleic acid arrangement. This stability must arise from energetically favourable interactions that result in the highly organized helical form. In particular, stabilization is conferred by the presence of four interstrand hydrogen-bonds to each guanine and multiple cation coordination sites, as well as stacking interactions between the tetrads of a very stable nature. The slow kinetics of poly(G) association (Howard *et al.*, (1977)) would seem to reflect this. In addition, whilst the self-assembly of 5'-guanosine monophosphate in the presence of sodium is strongly thermodynamically driven (Borzo *et al.*, (1980)), the self-association of other purines is less so (Saenger (1984)).

1.3 Proposed Biological Relevance for G4 Tetrads

Interest in the structures formed by guanine self-association has been revived in recent years. This has come about due to the hypothesis that certain biological systems may utilise the guanine quadruplex in some way. The vast majority of research interest has been focused on the organisation of chromosomal telomeric DNA (Sundquist (1991)). However, attention has also been directed towards recombination events in immunoglobulin switch regions (Shimizu and Honjo (1984); Sen and Gilbert (1988)) and the RNA genome of the human immunodeficiency retrovirus (HIV) (Sundquist and Heaphy (1993)).

1.3.1 Telomeres

Linear eukaryotic chromosomes terminate with specialized sequences which are essential for complete DNA replication. These telomeric sequences, and their affiliated proteins, also appear to function in chromosome stability, organization and association, and are reviewed generally in Blackburn and Szostak (1984); Zakian (1989); Blackburn (1990); Greider (1990); and Zakian *et al.*, (1990). All known eukaryotic telomeres consist of simple repeating sequences in which one DNA strand of the double helix contains short tracts of guanines alternating with short tracts of A/T-rich sequences. Representative sequences are **TTGGGG** from *Tetrahymena thermophila* (Blackburn and Gall (1978)), **TTTTGGGG** from *Oxytricha nova* (Klobutcher *et al.*, (1981)), **TTTAGGG** from *Arabidopsis thaliana* (Richards and Ausubel (1988)), **TTAGGG** from *Homo sapiens* (Moyzis *et al.*, (1988)), **AG**₁₋₈ from *Dictyostelium discoideum* (Emery

and Weiner (1981)) and TG₁₋₃ from *Saccharomyces cerevisiae* (Shampay *et al.*, (1984); Walmsley *et al.*, (1984)). More extensive lists are to be found in Zakian (1989) and Blackburn (1990). In each case the guanine-rich strand is oriented 5' to 3' towards the end of the chromosome. This strand extends to produce a 3' single-stranded overhang of about two repeating units in species where the precise telomeric terminus is known (Klobutcher *et al.*, (1981); Henderson and Blackburn (1989)). Although telomeric sequences from different organisms are not identical, the conservation of the G-tract sequence motif suggests that the structures and functions of telomeres are strongly conserved across the eukaryotes.

A great deal of what is known about telomeres and related processes has been obtained from studies on lower eukaryotes, in particular the ciliated protozoa. Many ciliates are highly enriched in telomeres and their associated proteins, owing to the developmentally regulated fragmentation and amplification of their chromosomes into multiple short linear DNA fragments which have telomeres at both ends (Gall (1986)). In the most extreme examples, these DNA fragments may be gene-sized and may be present in thousands of copies per cell, as is the case for the hypotrichous ciliate *Oxytricha*.

Novel mechanisms must exist for complete replication of the ends of linear DNA since Okazaki fragment-primed DNA replication would result in incomplete replication of 3' terminal template sequences upon removal of the RNA primer (Watson (1972)). The discovery, in *Tetrahymena*, of telomerase, a protein which adds G-rich telomeric repeats to the 3' ends of chromosomes in a template independent manner (Greider and Blackburn (1985, 1987)), provided a solution to this problem. The ac-

tivity of telomerase was later observed in other lower eukaryotes (Zahler and Prescott (1988); Shippen-Lentz and Blackburn (1989)) and in human HeLa cell extracts (Morin (1989)). All of the telomerases studied have an essential RNA component which contains sequence tracts which are complementary to the telomeric repeat (Greider and Blackburn (1989); Shippen-Lentz and Blackburn (1990)). As such, this RNA sequence has been demonstrated to serve *in vivo* as the template for the addition of *Tetrahymena* telomeric DNA repeats (Yu *et al.*, (1990)).

In *Oxytricha*, the telomeric repeats in the macronucleus also interact with a heterodimeric protein which binds specifically at the 3' terminal telomeric overhang (Lipps *et al.*, (1982); Gottschling and Zakian (1986); Price and Cech (1989); Raghuraman and Cech (1989); Hicke *et al.*, (1990)). The exact and overall functions of the protein remain unclear, but postulated roles include extension of telomeric DNA by telomerase (Yu *et al.*, (1990)), protection of the 3' telomeric overhang (Gottschling and Zakian (1986)), as well as promotion of telomere-telomere associations (Lipps *et al.*, (1982); Raghuraman and Cech (1989)).

Other telomere related interactions are not yet well understood. Yeast telomeres can undergo recombinations that may be implicated in telomere generation as well as the preservation of telomere length (Walmsley *et al.*, (1984); Pluta and Zakian (1989)). An RNA strand complementary to guanine-rich, single-stranded telomeric DNA strands has been observed in extracts of the *Oxytricha* macronucleus (Zahler and Prescott (1989)). It has been proposed that such telomere-specific RNA molecules may be used to prime the synthesis of the cytosine-rich DNA strand of the macronuclear DNA. Often telomeres appear specifically localized in the nucleus and seem to assist

in interchromosomal associations during cell meiosis and mitosis, as reviewed in Blackburn and Szostak (1984). While such observations may indicate specific means of chromosomal organization and recombination, it is unclear if subtelomeric regions of the chromosome are involved. Hence, in summary, experimental analyses have suggested that telomeres may be fundamentally implicated in chromosomal stability (McClintock (1941)), cell meiosis (Yu *et al.*, (1990)) and senescence (Lundblad and Szostak (1989); Harley *et al.*, (1990)), although the total, necessary character required of telomeres in such processes remains unknown (Levis (1989); Biessmann *et al.*, (1990)).

1.3.2 Immunoglobulin Class Switching

The enormous diversity of immunoglobulins (Ig) depends upon two features: a large number of variable (V) regions bind to specific antigens; and these variable regions are physically linked, in the same protein, to different constant (C) regions that exert a variety of functions. The V region repertoire is shared by the Ig classes and subclasses (IgM, IgD, IgG₁, IgG_{2a}, IgG_{2b}, IgG₃, IgA and IgE) which are determined by one of eight C_H region genes in mouse (C_μ, C_δ, C_{γ1}, C_{γ2a}, C_{γ2b}, C_{γ3}, C_α and C_ε, respectively).

Somatic amplification of both types of Ig diversity is derived from DNA rearrangements of the Ig genes. Multiple DNA segments (V_L-J_L or V_H-D_H-J_H) are recombined to complete a V gene. Thus the antigen specificity of each lymphocyte is determined by random association of two or three DNA segments. The V_H-D_H-J_H recombination seems to be a key step for differentiation between T and B lymphocytes because it is the first DNA rearrangement in the B cell lineage. Common precursors to T and B

lymphocytes may not be committed to differentiate to either T or B cells until either the B cell receptor (i.e. Ig) gene or the T cell receptor gene is rearranged successfully. The rearrangement of the T and B cell receptor genes may compete in determining the lineage of lymphocytes.

B lymphocytes that have completed the correct $V_H-D_H-J_H$ and V_L-J_L recombinations produce IgM because the C_μ gene is the closest to the J_H segments. The progeny of a single lymphocyte retain essentially the same V_H region sequences, except for base replacements due to somatic mutation. However, the Ig produced among the progeny often changes from IgM to another class. Paradoxically, the constant region (C_μ) of the Ig is replaced by another C_H region while the variable region (V_H) remains constant. This phenomenon is known as class switching (Shimizu and Honjo (1984) and references therein) and it is mediated by another DNA rearrangement termed S-S recombination.

Sequence comparisons of the DNA segments and expressed γ or α genes demonstrated that after switching a complete V_H gene is directly linked to the C_γ or C_α gene. Recombination occurs in an intervening region that lies between the V_H and C_H genes. The nucleotide sequences of the regions containing frequent recombination sites are comprised of tandem repeats of short tracts that contain dispersed TGAGC and TGGGG sequences. This stretch is termed the S region.

It has been proposed that the S-S recombination might be a sort of homologous recombination recognizing scattered short homologous DNA tracts. Alternatively, it has also been suggested that class-specific recombinases might be responsible for S-S recombination. Nucleotide sequences surrounding recombination sites are typically

purine rich, with guanines making up some 50% of the total. In addition to the sequences for the S region given above, similar sequences occur in S-S recombination sites in myeloma cells (TGAG, TGGG), *E. Coli* (TGAG, TGGG, TGGT) and in the *chi* sequence that stimulates generalized recombination d(GCTGGTGG).

In their search for proteins that might bind to these immunoglobulin switch regions, Sen and Gilbert (1988) reached the conclusion that these representative sequences exhibited evidence of guanine self-association, and proposed that such association might be implicated in the necessary recognition required for specific S-S recombination events (Sen and Gilbert (1988)). Other subsequent studies have suggested that the guanine tetrad quadruplex may play a key role in other recombination events in the tRNA gene *supF* (Akman *et al.*, (1991)), as well as in chromosomal telomeres (Katinka and Bourgain (1992)).

1.3.3 Retroviral RNA Dimerization

The genomic RNA of retroviruses is organized into infectious virion particles within the infectious cycle of the virus. This, however, is a poorly understood process (reviewed in Linial and Miller (1990)). Mature retroviral virions contain two homologous copies of their single-stranded RNA genomes, stably associated within an RNA-*gag* protein complex. The retroviral *gag* genes encode polyproteins that are cleaved into at least three proteins which are present in stoichiometric amounts of about 2000 per virion. The *gag* gene products constitute the structural basis of the retroviral core or capsid in the intracellular form of the retrovirus, including the close association of the two genomic RNA strands. This close association of the two RNA strands accounts in part

for the remarkable genetic variability of retroviruses as information on both strands can be utilized upon subsequent infection (Coffin (1979); Junghans *et al.*, (1982); Stuhlmann and Berg (1992)). The most stable contact between the two genomic RNAs, termed the "dimer linkage structure" (DLS), occurs at a discrete site near the 5' end of each strand (Kung *et al.*, (1976)). The two strands emanate from the DLS with the same orientation (i.e. 5' to 5' and 3' to 3'), and this parallel alignment may facilitate subsequent recombination via a reverse transcriptase strand switching mechanism (Coffin (1979); Stuhlmann and Berg (1992)).

Strong genetic and biochemical evidence exists which suggests that genomic RNA dimerization in a number of different retroviruses is facilitated by interactions with viral gag protein(s) (Anderson *et al.*, (1992) and references therein). However, the two RNA strands in the dimer appear to associate primarily via direct RNA-RNA interactions, since genomic RNA dimers extracted from virions are quite stable under non-denaturing conditions even after all proteins are removed (Kung *et al.*, (1976)). In addition, several retroviral RNA genomes, including human immunodeficiency virus type 1 (HIV-1), have been shown to dimerize spontaneously in the absence of protein cofactors under conditions of high ionic strength (Prats *et al.*, (1988); Bieth *et al.*, (1990); Darlix *et al.*, (1990); Marquet *et al.*, (1991)). HIV-1 sequences sufficient to support spontaneous dimerization have been mapped to a 113-nucleotide domain, encompassing the 5' end of the gag gene (Darlix *et al.*, (1990)). The structural motif mediating the association of the two identical single strands has remained unknown.

It has been observed that retroviral dimerization domains are purine rich. Furthermore, the kinetics and ionic strength dependencies of dimerization are consistent

with the formation of a guanine quadruplex (Marquet *et al.*, (1991)). Investigations by Sundquist and Heaphy (1993) led them to suggest that the dimerization of the HIV genome *in vitro* may be mediated by the formation of an interstrand quadruple helix structure bases upon the guanine tetrad. Furthermore, the authors speculated that this structure may also mediate the association of genomic HIV-1 RNA *in vivo*, revealing how RNA itself can achieve the self-recognition required for subsequent genetic recombination (Sundquist and Heaphy (1993)).

1.4 Biochemical and Biophysical Evidence

1.4.1 Telomeric DNA

The first indication that telomeric DNA sequences, as defined in Section 1.3.1, could adopt a variety of different structures came from the observation that crude extracts of macronuclear DNA from hypotrichous ciliates could associate into higher order structures *in vitro* (Lipps (1980); Lipps *et al.*, (1982)). Electron microscopic and biochemical studies indicated that this association involved the pairing of the telomeric ends of macronuclear DNA, with the appearance of “cusps” in which the two double helical DNA fragments projected from the side of the cohered telomeres. These telomeric complexes were highly resistant to treatment with nuclease, although they could be dissociated by high temperature, high pH, low ionic strength and protease. This suggested, with regards to observations with protease, that the observed association of telomeric sequences was mediated by unidentified intracellular protein(s) (Lipps *et al.*, (1982)).

It was later shown that macronuclear telomeric DNA could, when purified, slowly associate *in vitro* under high salt conditions without any protein present (Oka and Thomas (1987)). Furthermore, it was demonstrated by nuclease digestion and oligonucleotide competition studies that this association was telomere specific, involving the 3' telomeric overhang. Telomeric complexes themselves were highly resistant to exonuclease digestion, but were cleaved specifically by T7 gene 3 endonuclease and endonuclease S1, indicating some single-stranded character near or within the telomeric structure. Preformed telomeric complexes showed a marked increase in thermal stability of more than 25° in the presence of potassium versus sodium. This corresponded with previous observations, noted earlier, for quadruple helical structures formed by guanine derivatives and their polymers.

Another type of telomeric DNA structure was indicated for single stranded guanine-rich oligonucleotides containing a minimum of two telomeric repeats. Under low ionic strength conditions, these oligonucleotides folded into *intramolecular* complexes that exhibited anomalously rapid migration through native polyacrylamide gels (Henderson *et al.*, (1987)). A number of experimental observations suggested that these complexes were stabilized by G-G base interactions. In particular, it was striking that folded structures were observed for different telomeric sequences whose only apparent similarity was the potential to form G-G base-pairs, e.g. d(T₂G₄)₂ from *Tetrahymena* (Blackburn and Gall (1978)) and AG₇AGAG₆AG₆ from *Dictyostelium* (Emery and Weiner (1981)).

Additionally, the study of Henderson *et al.*, (1987) proposed that oligonucleotides containing four telomeric repeats could fold into especially stable and compact in-

tramolecular structures. The four-fold repeat oligonucleotide d(T₂G₄)₄ was analysed in detail. Visible spectroscopic studies demonstrated that this oligonucleotide folded reversibly through a multistep pathway with a T_m of approximately 40°C in 20 mM Na⁺ buffer (Henderson *et al.*, (1987)). ³¹P and ¹H NMR studies also suggested a discrete, ordered structure at low temperature which unfolded into a random coil configuration at 40°C (Henderson *et al.*, (1987)). Nearly all of the amino protons in the folded structure were observable in the ¹H NMR spectra in H₂O, indicating that these protons are either stabilized by direct hydrogen-bonds or are otherwise inaccessible to solvent. The amino protons are not commonly observable in H₂O solvent as a result of rapid exchange with the hydrogens of the solvent itself.

Such evidence strongly suggested that telomeric DNA sequences had the potential to fold into a variety of unusual structures *in vitro*. Despite the evident differences between intermolecular complexes between telomeres and the intramolecular structures formed by telomeric oligonucleotides, it was apparent that these structures were stabilized by guanine-guanine interactions. This suggested that common structural principles were involved.

1.4.2 Immunoglobulin Switch Regions

In studies of proteins that bound to immunoglobulin switch region DNA sequences, Sen and Gilbert (1988) documented evidence for another type of DNA quadruplex formation. In particular, they observed that single-stranded, guanine-rich oligonucleotides could tetramerize to form parallel-stranded quadruple helical structures. This observation suggested structures analogous to models proposed from fibre diffraction studies

of polymers of guanine and its derivatives.

The tetrameric stoichiometry of these oligonucleotide complexes was demonstrated by annealing complementary strands to the complex and observing the stepwise formation of four new complexes of reduced electrophoretic mobility (Sen and Gilbert (1988)). This represented the formation of complexes containing one, two, three and four complementary strands respectively. All guanines within the putative quadruple helical regions of these complexes were protected from N₇ methylation by dimethylsulphate (DMS) (Sen and Gilbert (1988)), a result which is consistent with each guanine acting as a hydrogen-bond acceptor within a guanine base-tetrad, as illustrated in Figure 1.9.

That such heteromorphous, guanine-rich oligonucleotides formed tetrameric complexes demonstrated that discontinuities could be incorporated as bulges or mismatches within quadruple helical structures. This result suggested that higher ordered, tetrameric DNA structures might be implicated in recombination events in immunoglobulin switch regions. But Sen and Gilbert (1988) went on to suggest that the association which they observed may also have a function in the adhesion of chromatids during cell meiosis, and that this may involve the joining of chromosomal telomeres. While such an observation does not relate to the pairwise association observed in the macronuclear DNA from the hypotrichous ciliates, it does suggest another intermolecular association in which the association is not dimeric but tetrameric.

1.4.3 Dimerization of the HIV-1 Genome

In their studies of the dimerization of the HIV-1 genome in the absence of any protein co-factor, Sundquist and Heaphy (1993) presented strong evidence that quadruple helical structures were implicated in the dimerization of the HIV-1 RNA genome. In this study they took a 127-nucleotide RNA fragment from the genome of HIV-1 which encompassed the 5' end of the *gag* gene. This fragment dimerized spontaneously in the absence of any protein co-factor. Remarkable stabilization was observed for the potassium ion, with significantly lesser thermal stability for lithium, sodium and cesium ions. Deletion analyses of the 3' end of the fragment revealed that an RNA fragment as short as 94-nucleotides was capable of spontaneous dimerization, but that the further elimination of the 9-base, purine-rich sequence GGGGGAGAA completely eliminated dimerization.

It was thus proposed that the sequence GGGGGAGAA constitutes one of two tracts required for dimerization within a single-strand, required in order to allow two strands to come together in a cation dependent manner. However, it was additionally suggested that a quadruple helix need not necessarily be composed of nine nucleotides, and that the possibility exists for quadruple helix formation by the mixed purine sequence (AAG), with the presumption that the A₄ tetrad is isostructural to the G₄ (Sundquist and Heaphy (1993)). This is feasible on the basis of studies of poly(GGA) (Lee (1990)), although such sequences have not been characterised fully and demonstrate a tendency to form double-helical structures with no monovalent cation dependency (Rippe *et al.*, (1992)).

This study presents evidence for perhaps the best characterised biological function

for a quadruple helical structure stabilized by guanine tetrads. In addition, it suggests the possibility that isostructural adenine tetrads may also play a interesting structural and functional role. However, recent studies of oligonucleotides of the human telomere $d(T_2AG_3)_n$ and the *Chlamydomonas* telomere $d(T_4AG_3)_n$ indicate that the adenines are formally part of single-stranded loops where they probably interact with thymines (Murchie and Lilley (1994)).

1.5 Proposed Models for G4 Structures

1.5.1 Anti-Parallel Quadruple Helical Structures

Intermolecular Quadruplexes

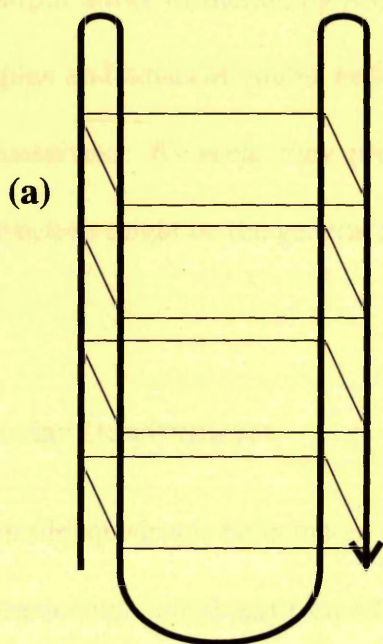
The complications involved in studying macronuclear DNA, where there are two telomere strands at each end (Oka and Thomas (1987)), were overcome by using short double helical oligonucleotides which contained an *Oxytricha* or *Tetrahymena* telomere sequence and a simple blunt duplex at the other end (Sundquist and Klug (1989)). As was observed with macronuclear DNA, these oligonucleotides were found to slowly associate to form intermolecular complexes when incubated in buffers containing high concentrations of sodium chloride (Sundquist and Klug (1989)).

Constructs were analysed that contained four $d(T_2G_4)$ double helical repeats followed by the 3'-terminal overhang $[d(T_2G_4)_2]$ of the *Tetrahymena* telomeres. Intermolecular complexes formed by oligonucleotides of this type were shown to be dimeric by co-annealing two constructs of different overall lengths and observing the formation

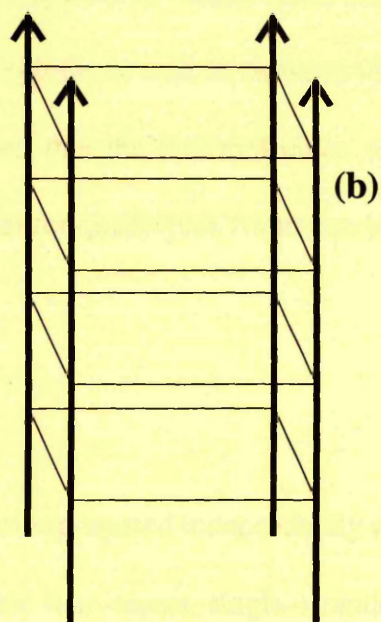
of unique heterodimeric complexes. Chemical probing of the purified dimeric complexes revealed that the N₇ position of every guanine within each telomeric overhang was protected from chemical alkylation, while the chemical accessibility of the guanines within the double helical regions remained unchanged, and thymines were still oxidized by OsO₄. This result implied that dimerization was mediated completely by the telomeric overhang. It has also been observed that random sequence duplexes will dimerize if they end with a d(T₂G₄)₂ overhang, as will the oligonucleotide d(T₂G₄)₂ itself (Sundquist and Klug (1989); Williamson *et al.*, (1989)).

A schematic model illustrating a dimeric association of telomeric overhangs, based on the results of Sundquist and Klug (1989) and Williamson *et al.*, (1989), is shown in Figure 1.11(c). This model has been drawn without helical twist for clarity. In the model proposed, the single-stranded telomeric overhangs form hairpin loops which proceed to dimerize, resulting in complexes that are stabilized by an anti-parallel, quadruple helical core. The formation of guanine base-tetrads in the centre of the quadruple helix is sufficient to ensure that all 16 guanine N₇ positions are inaccessible chemically. The significant difference between this and previous models for guanine quadruplexes, derived from guanine polymer and derivative studies, is that these quadruplexes must have an anti-parallel orientation due to the fact that the telomeric overhangs fold back on themselves.

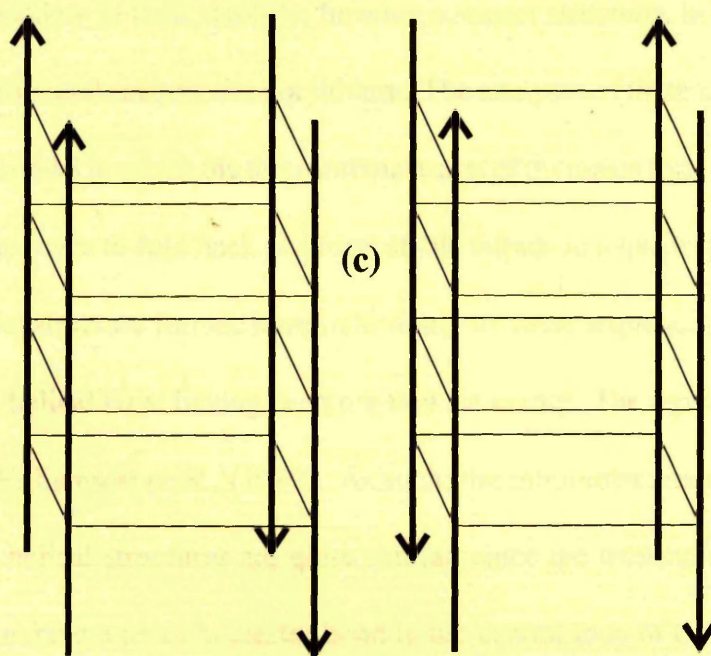
It has also been observed that the oligonucleotides d(T₂G₄)₂ and d(T₄G₄)₂ fold to form hairpins under conditions of moderate ionic strength with no anomalous cation dependence (Henderson *et al.*, (1987); Williamson *et al.*, (1989)). These experimental results led Sundquist and Klug (1989) to propose that the 5' thymine and adenine



Foldback Structure



**Parallel
Quadruplex**



Antiparallel Quadruplexes

Figure 1.11: Structures Proposed to Form in G-Rich DNA Sequences

tracts, found typically to interrupt guanine stretches in telomeric sequences, might also assist in hairpin dimer formation by acting as spacers to reduce steric clash between the quadruplex and adjacent double helical regions, as well as between the telomeric strands themselves. As such, they proposed that the intermolecular, anti-parallel foldback structure might be the general model for quadruplex formation by telomeric sequences.

Intramolecular Quadruplexes

The anti-parallel quadruple helix model was also proposed independently with regards to the intramolecular complexes formed by the four-repeat, single-stranded oligonucleotides $d(T_2G_4)_4$ and $d(T_4G_4)_4$ (Williamson *et al.*, (1989)). These complexes were also cation dependent in their stability, forming compact structures in the presence of sodium, potassium and cesium, but not lithium. The analyses of these oligonucleotides substantiated a model in which the three internal tracts of thymine form loops, allowing the four guanine tracts to fold back and form stable tetrads in a quadruple helical core. The anti-parallel structure formed intramolecularly by these sequences is illustrated in Figure 1.11(a), helical twist having been omitted for clarity. The figure is based upon the results of Williamson *et al.*, (1989). As such, the intramolecular and intermolecular quadruple helical structures are quite similar, since the intermolecular model is produced by breaking a phosphodiester bond in the central loop of the intramolecular model; conversely, the intramolecular model simply requires the fusing of two of the strands of the intermolecular model.

The verification of the formation of guanine tetrads in four-fold telomeric repeat

structures resulted from base substitution, methylation protection and methylation interference studies (Williamson *et al.*, (1989); Henderson *et al.*, (1990); Panyutin *et al.*, (1990); Venczel and Sen (1993)). In addition, UV-induced cross-links are easily formed between thymines in different tracts of the sequences $d(T_2G_4)_4$ and $d(T_4G_4)_4$ (Williamson *et al.*, (1989)). Thus, the intramolecular folding of the oligonucleotides brings thymines from distant tracts into close proximity at the ends of the core of the quadruplex. Recent findings have also suggested that cytosine–cytosine⁺ (C–C⁺) base-pairing may also play a role in further stabilising DNA quadruplex structures (Hardin *et al.*, (1993)).

1.5.2 Parallel-Stranded Quadruplexes

Guanine-rich sequences from telomeres and immunoglobulin switch regions have been shown to tetramerize to form parallel-stranded structures stabilized by guanine tetrads (Sen and Gilbert (1988, 1990)). In parallel-stranded structures, as in the anti-parallel examples, all guanines are protected from dimethylsulphate methylation. This parallel-stranded association is illustrated in Figure 1.11(b), again shown without helical twist for clarity. The figure is based upon the results of Sen and Gilbert (1988).

Experiments were carried out using long(L) and short(S) oligonucleotides. Coannealing of these oligonucleotides led to the formation of three heterotetrameric structures (LSSS, LLSS, LLLS) (Sen and Gilbert (1988)) which co-migrate in gel electrophoresis. This result verified the formation of parallel stranded tetraplexes, since the formation of anti-parallel complexes would have resulted in species of very different electrophoretic mobilities.

In principle the species formed by tetrameric association of strands could be parallel or anti-parallel. The adoption of a parallel orientation is presumed to be due to the steric favourability of the four-fold symmetric parallel structure, in which the phosphate backbones are evenly and maximally separated and the guanines adopt the sterically favourable *anti* conformation. Despite this stability, parallel-stranded complexes form very slowly (Sen and Gilbert (1990)). As a result they favour high salt, temperature and DNA concentration. Still, recent NMR studies have demonstrated the stable formation of parallel-stranded tetraplex structures by a number of oligonucleotides: d(TG₄T) (Aboul-ela *et al.*, (1992)); d(T₂AG₃) and d(T₂G₄) (Wang and Patel (1992)); d(TG₃T) and d(TG₃T₂G₃T) (Jin *et al.*, (1992)); d(G₄T₄) (Sarma *et al.*, (1992)); d(CGCG₃GCG) (Hardin *et al.*, (1992)); d(TG₄) and d(T₄G₄T) (Lu *et al.*, (1992)). Several other studies have offered similar conclusions on the basis of the mobility of such sequences in non-denaturing polyacrylamide gels (Balagurumoorthy *et al.*, (1992); Gaffney *et al.*, (1992); Guo *et al.*, (1992a, 1992b); Hammond-Kosack and Docherty (1992); Murchie and Lilley (1992)). Formation of parallel-stranded tetraplexes was observed, in the above examples, in the presence of sodium and potassium. All of the glycosyl linkages were determined to be *anti*, with resonances indicating a four-fold, and hence parallel-stranded, structure. Furthermore, the formation of an extremely stable parallel-stranded tetraplex was observed for the sequence 5'-GCCGAUGGUAGUGUGGGGU-3' from the 5S RNA of *E. coli* (Kim *et al.*, (1991)). Tetraplex formation was dependent on the guanines at the 3' end.

1.6 X-Ray and NMR High-Resolution Studies

The first high-resolution experimental models of quadruple helical structures formed by oligonucleotides were reported in 1992. These reports consisted of two different structures for the oligonucleotide d(G₄T₄G₄), which contains 1.5 repeats of the telomeric DNA sequence from the ciliated protozoan *Oxytricha* (Kang *et al.*, (1992); Smith and Feigon (1992)). Kang *et al.*, (1992) determined the structure by X-ray crystallography, while Smith and Feigon (1992) carried out their study by NMR. The structures were remarkable for the fact that, though they were very similar, they contained a number of striking differences. The results of these studies are shown in Figure 1.12, which is adapted from Williamson (1993). In each case the models have been drawn without helical twist for clarity. The *syn* and *anti* labels for the glycosyl linkages, situated in the tetrad planes of the models, refer to the residues 1–4 of the strand in the right of the foreground of the model in each of the three cases.

1.6.1 X-Ray Structure of d(GGGGTTTTGGGG)

The X-ray structure determination was carried out using potassium as the stabilizing cation, and was solved to a resolution of 2.5 Å using molecular replacement (Kang *et al.*, (1992)). The resultant structural solution is illustrated schematically in Figure 1.12(a). Four guanine tetrads stack on top of each other, forming a right-handed quadruple helix. The first tract of four guanines folds back by means of the thymine linker to form hydrogen-bonds to the second tract, and the two hairpins dimerize to form an anti-parallel, intermolecular quadruplex. The thymine residues which form loops between

adjacent strings of guanines are situated at opposite ends of the quadruple helical core. There was some evidence in the electron density to indicate the presence of an axial potassium ion, bound between the central two guanine tetrads of the quadruplex.

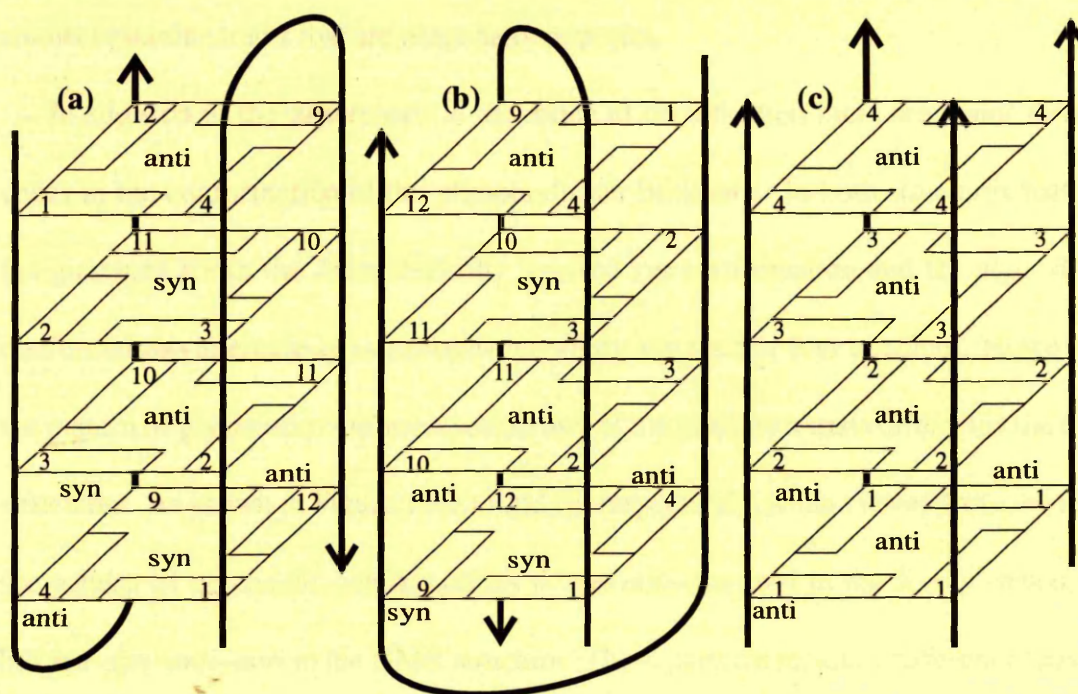


Figure 1.12: Illustration of Experimentally Determined G-Quartet Cores

1.6.2 NMR Structure of d(GGGGTTTTGGGG)

In their NMR study of the oligonucleotide d(G₄T₄G₄), Smith and Feigon (1992) used sodium as the stabilizing cation. The resultant structural solution from their studies is shown in Figure 1.12(b). In a manner similar to the X-ray structure, the four guanine tetrads are stacked on top of each other to form an anti-parallel quadruple helix via the dimerization of two hairpin oligonucleotides, as previously suggested by

preliminary NMR study of the similar oligonucleotides d(GGTTTTTCGG) (Wang *et al.*, (1991a)) and d(GGTTTTTTGG) (Wang *et al.*, (1991b)). However, it is the nature of the dimerization which serves to distinguish the NMR and X-ray structures of d(G₄T₄G₄). In the NMR structure the thymine loops do not attach adjacent guanine tracts but rather connect guanine tracts that are diagonally opposite.

In addition to the differences in the mode of dimerization there are major differences in the conformation of the phosphodiester backbone. In both structures half of the guanines are in the more sterically stressed *syn* conformation and the glycosidic conformations alternate *syn-anti-syn-anti* along any tract of four guanines. However, the pattern of glycosidic conformations in any of the guanine tetrads differs for the two structures. As shown in Figure 1.12(a) and (b) respectively, when viewed from the top; the pattern of glycosidic conformations is *syn-anti-syn-anti* in the X-ray structure, but *syn-syn-anti-anti* in the NMR structure. These patterns result in different grooves along the four sides of the quadruple helical core. In the X-ray structure there are two wide and two narrow grooves, while in the NMR structure there are two medium grooves, one wide groove and one narrow groove. There is also a different pattern of exposed substituents for the two structures.

Smith and Feigon (1992) also observed the formation of an intramolecular quadruplex with the sequence d(G₄T₄G₄T₄G₄T₄G₄), as illustrated schematically in Figure 1.11(a). This arrangement of grooves, observed in the NMR structure in the presence of sodium, was also later observed in the high-resolution NMR determination of the anti-parallel helical quadruplex formed intramolecularly by the human telomeric repeat d(AG₃(T₂AG₃)₃), also determined using sodium as the stabilizing cation (Wang

and Patel (1993*b*)). Another example is the recent high-resolution NMR study of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG) (Macaya *et al.*, (1993)), selected by amplification methods from a randomly generated population of sequences.

The NMR study of the guanine quadruplex dimer formed by the *Oxytricha* repeat oligonucleotide d(G₄T₄G₄) has been refined further in a couple of later studies (Smith and Feigon (1993); Schultze *et al.*, (1994)).

1.6.3 NMR Studies of r(UGGGGU) and d(TTGGGGT)

High-resolution structural characterization of a parallel stranded tetraplex came slightly later with the NMR analysis of the oligoribonucleotide r(UGGGGU) (Cheong and Moore (1992)). The structure resulting from this study is illustrated in Figure 1.12(c), with the omission of helical twist for clarity. As observed in the anti-parallel foldback structures, the guanines form tetrads in a helical stack that is right-handed. In this case, all the glycosidic arrangements are in the sterically preferred *anti* conformation, with the phosphate backbones parallel. This is presumably due to the absence of the steric stress induced as a result of the presence of hairpin folds, and makes the parallel-structure sterically favourable.

Another interesting feature of this RNA tetraplex was evidence which suggested that the uracil bases also formed a highly stable tetrad at one of the ends of the tetraplex. This feature is not illustrated in Figure 1.12(c). This interaction would be completely novel since there is no prior evidence for such a thymine or uracil tetrad. NMR characterizations of parallel-stranded DNA tetraplexes, cited above, presented no evidence for the analogous formation of thymine tetrads. A recent high-resolution

NMR study of the *Tetrahymena* telomeric sequence d(TTGGGGT) showed a similar parallel-stranded tetraplex to that seen for RNA (Wang and Patel (1993a)). However, the thymines were seen to sample multiple conformations and to remain undefined in the solution structure.

1.7 Key Structural Features of G4 DNA/RNA

The high-resolution structural studies which have been detailed above summarize the key structural features of the complexes that can be formed by oligonucleotides utilising guanine tetrads. In addition, they also underline what are some of the remaining key questions concerning the structural organization of these complexes. These features are summarized below.

1.7.1 *syn* or *anti* Glycosyl Conformation

The glycosidic conformation angle for a guanine within a tetrad structure can adopt either the *anti* or *syn* conformation, the *anti* being the sterically preferred. Geometric constraints require that two adjacent guanines in a guanine tetrad have the same glycosidic conformation, but must have opposite glycosidic conformations if they are on anti-parallel strands. There is a strong correlation between the glycosidic conformation and the relative strand direction in all cases where both are known (Jin *et al.*, (1992)). The glycosidic conformations in tetrameric guanine tetrad structures with parallel strands are all *anti* (Aboul-ela *et al.*, (1992); Cheong and Moore (1992); Sarma *et al.*, (1992); Wang and Patel (1992, 1993a)). In guanine tetrad structures containing

anti-parallel strands, the observed glycosidic conformation angles are half *syn* and half *anti*, alternating *syn-anti-syn-anti* along the guanine tract (Kang *et al.*, (1992); Smith and Feigon (1992); Wang and Patel (1993b)).

1.7.2 Structural Isomerism

The NMR and X-ray crystallographic studies of the *Oxytricha* telomere d(G₄T₄G₄) demonstrated that oligonucleotides encoding telomeric sequences could fold into structurally inequivalent configurations that were still consistent with the formation of anti-parallel helical quadruplexes (Kang *et al.*, (1992); Smith and Feigon (1992)). It has previously been observed that the oligonucleotide d(T₄G₄)₄ folded into different intramolecular quadruplexes that could be distinguished by gel electrophoresis (Williamson *et al.*, (1989)). In addition, different electrophoretic mobilities were observed for four hairpin dimers formed by the sequence d(T₂G₄)₂ (Sundquist and Klug (1992)). Model building studies demonstrate the possibility of forming a variety of structural isomers, all of which have anti-parallel quadruple helical configurations (reviewed in Sundquist (1991)).

1.7.3 Loop Configurations

It has been shown that differing loop lengths can markedly affect the stability of Watson-Crick hairpin loops (Hilbers *et al.*, (1985); Groebe and Uhlenbeck (1988)). Biophysical evidence has also suggested that different loop lengths may affect the stability and isomeric distribution of telomeric DNA structures (Williamson *et al.*, (1989)). The X-ray crystallographic and NMR studies of d(G₄T₄G₄) have illustrated

two distinct loop arrangements for two loops of four thymine bases (Figure 1.12(a) and (b) respectively). The third possibility, a “head-to-head” arrangement, has not been observed for quadruplexes formed by intermolecular dimers, although it has been observed for the intramolecular case (Smith and Feigon (1992); Wang and Patel (1993b)), and is shown in Figure 1.11(a).

1.7.4 Ion Specificity, Structure and Stability

Recent studies of the formation of quadruple helical structures by oligonucleotides has demonstrated the ability of such structures to discriminate amongst different ions, with the potassium form generally being more stable than the sodium form (Zahler *et al.*, (1989); Hardin *et al.*, (1991); Chen (1992); Hardin *et al.*, (1992); Lu *et al.*, (1992)). This was predicted from earlier studies of fibres and guanine derivatives. The divalent strontium ion has been shown to be even better than potassium in stabilizing guanine tetrad structures (Chen (1992)).

Using ion-concentrations that are close to the physiological levels, it has been shown that while intramolecular hairpin structures form very rapidly (Henderson *et al.*, (1987)), dimerization to form anti-parallel quadruplexes in the presence of sodium can take two hours (Sundquist and Klug (1989)), while parallel tetraplexes are generally even slower (Sen and Gilbert (1990)). Generally, structures formed in the presence of potassium dissociate more slowly (Oka and Thomas (1987); Raghuraman and Cech (1989); Sundquist and Klug (1989); Sen and Gilbert (1990)), and studies of $d(T_4G_4)_4$ have shown that potassium increases the equilibrium constant to favour the folded structures by a factor of about sixty (Raghuraman and Cech (1990)), thereby demon-

strating that structures stabilised by the K^+ cation are generally favoured on kinetic and thermodynamic grounds. However, in certain cases, anti-parallel hairpin dimeric structures (Oka and Thomas (1987)) and parallel tetrameric structures (Sen and Gilbert (1990)) also form more slowly in the presence of potassium. Sundquist (1991) has suggested that the latter may be explained if rate-determining steps in the formation of these complexes involve disrupting structures anomalously stabilized by potassium. Williamson (1993) has suggested that in addition to the possibility that potassium is a better "fit" than sodium, it is also possible that the difference in hydration energies for the two ions favours potassium as the ion of choice.

The NMR and crystallographic studies of $d(G_4T_4G_4)$ illustrated the importance of the choice of ion in determining the resultant structural isomerism. It would appear that the various structural parameters such as the glycosidic conformations, strand directions, strand associations and loop configurations may be determined by the nature of the ion binding site(s). Furthermore, the nature of the site(s) occupied by sodium and potassium would appear to be very different. The only high-resolution structural information on ion sites to date is that determined from the only crystallographic study to date (Kang *et al.*, (1992)). In this case, a single, monovalent ion is observed in a central axial position. A sodium site or sites would most probably differ quite significantly. Further elucidation of the nature of ion binding site(s) may prove instructive in understanding the role monovalent cations play in determining the stability and preference of different quadruplex structures.

1.8 Biological Relevance of Guanine Tetrad Structures

There is as yet no clear evidence for a biological role for structures that form guanine tetrads. It has been proposed that such structures may play a role in recombination events involving immunoglobulin switch regions (Sen and Gilbert (1988)) and in the dimerization of the two copies of the RNA genome of HIV-1 (Sundquist and Heaphy (1993)). The latter example is perhaps one of the best characterized candidates.

Most of the work involved in characterizing the quadruplexes that might have some biological function has undertaken studies of telomere sequences (reviewed in Sundquist (1991)). The strongest argument in support of biological roles for such structures in these cases is the remarkable conservation of the guanine tract motif and the interchangeability of telomeric DNA sequences from different species (Szostak and Blackburn (1982); Pluta *et al.*, (1984); Greider and Blackburn (1985); Brown (1989); Cross *et al.*, (1989)). Such conservation and interchangeability does not however indicate the use of unusual DNA structures *in vivo*, as the telomeric sequences and the 3' overhang may be simply required for systems that interact with telomers, such as telomerase (Blackburn (1990)).

The two most characterized enzymes involved in the maintenance of telomeric 3' tracts are telomerase from *Tetrahymena* and the stoichiometric telomere binding protein from *Oxytricha*. Experiments have shown that neither of these proteins interacts directly with quadruple helical telomeric DNA structures (Henderson *et al.*, (1990); Raghuraman and Cech (1990)). While such quadruple helical structures should be able to form under physiological conditions, experiments have shown that the *Oxytricha*

telomere binding protein can prevent the formation of intramolecular quadruple helical structures (Raghuraman and Cech (1990)), suggesting that the function of such proteins might be to prevent quadruplex formation. Stable protein–telomere complexes formed *in vivo* have been shown to be inconsistent with quadruplex formation by footprinting studies (Price and Cech (1989); Raghuraman *et al.*, (1989)).

The strong conservation of telomeric DNA sequences and the fact that telomerase utilizes an RNA template in its synthesis has led to suggestions that perhaps telomeres may be ancient molecular structures, or remnants of an RNA world (Weiner (1988)). The proposition that dimeric and tetrameric structures of telomeric DNA might be involved in cell meiosis and mitosis (Sen and Gilbert (1988); Sundquist and Klug (1989); Sen and Gilbert (1990)) appeared weakened when experiments suggested that such functions may alternatively be performed by proteins, in particular the telomere binding protein from *Oxytricha* (Raghuraman and Cech (1989)).

A number of studies have presented evidence for selective binding of guanine tetraplex structures by a variety of cellular proteins (Price *et al.*, (1992); Walsh and Gilberto (1992); Fang and Cech (1993*a*, 1993*b*); Liu *et al.*, (1993); Pearson *et al.*, (1993); Weismanshomer and Fry (1993); Giraldo and Rhodes (1994); Liu and Gilbert (1994); Schierer and Henderson (1994)).

The β subunit of the *Oxytricha* telomere binding protein is of particular interest, since it has been demonstrated to promote the formation of foldback and four–stranded guanine tetrad containing conformations (Fang and Cech (1993*a*, 1993*b*)). Similarly, the yeast *RAP1* protein, essential for the maintenance of the telomeres of *Saccharomyces cerevisiae*, appears to facilitate the formation of the guanine tetraplex structure (Giraldo

and Rhodes (1994)). The yeast *KEM1* gene product has been shown to be a nuclease specific for the parallel-stranded G4 tetraplex structure (Liu *et al* (1993); Liu and Gilbert (1994)), formed by such sequences as the oligonucleotide d(TG₄T). The yeast *KEM1* protein specifically recognises the parallel-stranded tetraplex form and cleaves at the contiguous single-strand regions. A role has been suggested for this nuclease activity and guanine tetrad tetraplex DNA in meiosis (Liu and Gilbert (1994)).

That telomeric proteins act as molecular chaperones for guanine tetrad formation, alongside the observation of specific binding and nuclease activity for at least one intracellular protein, presents perhaps the strongest evidence yet that quadruple helical structures composed of guanine tetrads may exist *in vivo*, at least at the sites of chromosomal telomeres.

High-resolution X-ray crystallographic studies of such quadruplexes are likely to yield valuable information on cation coordination, tetrad interactions and other structural features that may provide insight relevant to all guanine quadruplexes, parallel and anti-parallel. Such structural detail is likely to be important in the analysis of quadruplex-protein interactions, and thus in understanding the biological function of these compelling structures.

Molecules, which are systems of quanta of, and which act as stable vehicles of physical information, zoosemiotic and cultural systems including language, constitute a continuous chain of stages, with ever more complex energy levels, in the context of a unique and universal evolution. It is therefore possible to describe both language or living systems from a unifying cybernetic point of view. For the moment, this is only a useful and provisional analogy . . . A reciprocal rapprochement between genetics, animal communication and linguistics may lead to a complete science of the dynamics of semiosis, in which science may turn out, in the final analysis, to be nothing other than a definition of life.

Thomas Sebeok, *How Animals Communicate*

The mathematically minded Leibniz saw in the mystical elegance of the binary system, where only the zero and the one count, the very image of creation. The unity of the Supreme Being, operating by means of a binary function against the nothing, was sufficient ground, he thought, from which all things could be made.

Marshall McLuhan, *Understanding Media*

After the metaphysics of being and appearance, after energy and determinacy, the metaphysics of indeterminacy and the code. Cybernetic control, generation through models, differential modulation, feedback, question/answer, etc.: this is the new *operational* configuration . . . Digitality is its metaphysical principle (Leibniz's God), and DNA is its prophet.

Jean Baudrillard, *The Orders of Simulacra*

"It's all in your genes." This slogan of Eugenics deconstructs its own resolute objective of genetic improvement when its pun implies that "it" – your sexuality, your identity – is all in your designer jeans. In the postmodern Americanization of the globe, Eugenics is transformed into Eu(jean)ics, a postmodern science that strives for hyperreal perfection by splicing panic genes with panic jeans. This post-fetishistic power field promises Eu(jean)ic perfectibility by deploying signifiers of radical politics in the service of simulated health, wealth and sexual desire.

Lee Quinby, *Panic Jeans*

Chapter 2

Crystallization of d(TGGGGT)

2.1 Tetraplex Formation by d(TGGGGT)

As indicated in Chapter 1, the hexadeoxynucleotide d(TG₄T) has been shown by NMR study to form a parallel-stranded tetraplex (Aboul-ela *et al.*, (1992)). In this study, Aboul-ela and colleagues investigated the solution structure of the complex formed by the hexadeoxynucleotide d(TG₄T) in the presence of sodium cations. Two structural forms were observed in the ¹H NMR spectrum, with one form predominating at high temperature, 70°C, and the other at low temperature, 30°C. The rate of exchange between the two forms was seen to be slow which, as discussed in Chapter 1, is consistent with quadruplex formation (Oka and Thomas (1987); Sundquist and Klug (1989)). Due to the number of peaks that were observed, the structure symmetry was either two-fold if duplex, and four-fold if tetraplex.

In the presence of 100mM K⁺, 10mM Na⁺ ions at 38°C, the observed spectrum was identical to that obtained at 30°C in the presence of 100mM Na⁺ alone. This indication

of a cation specific stability for the d(TG₄T) complex, in particular the greater stability of the potassium form in relation to the sodium form, suggested the formation of a quadruplex rather than a duplex structure, since duplex stability is generally cation independent. Furthermore, ¹H and ³¹P spectra demonstrated that the conformation of the sugar-phosphate backbone was similar to that found in canonical B-form duplex DNA (Saenger (1984)).

B-form DNA is widely believed to be the most common conformation of double-helical DNA existing in the cell. Generally speaking, the B-form of DNA is characterized by the phosphates of one polynucleotide chain being approximately 7Å apart, with the deoxyribose sugars occupying one of the equivalent C_{3'}-*exo* or C_{2'}-*endo* conformations. This results in the structure being overwound in that the rotation per nucleotide ranges from 36 to 45°, whilst the axial rise varies from 3.03 to 3.37Å. The structural details of B-form DNA and the other canonical DNA types, namely A- and Z-DNA, are discussed in detail by Saenger (1984).

Subsequently, observed resonances for d(TG₄T) indicated the formation of a right-handed helix with glycosidic torsion angles exclusively *anti* and sugars in the C_{3'}-*exo* conformation. This is in contrast to the anti-parallel fold-back structures where the glycosidic angles alternate in a *anti-syn* manner, as discussed in Chapter 1.

The spectra derived from the guanine amino resonances indicated an unusual hydrogen-bonding pattern that was shown to be consistent with the formation of a guanine tetrad (Figure 1.9), and inconsistent with any other base associations that could be devised (Aboul-ela *et al.*, (1992)). Spectral resonances did not exhibit the features characteristic of an anti-parallel association of strands. Thus it was concluded

that the hexadeoxynucleotide d(TG₄T) associates as a tetraplex with four-fold symmetry along the helical axis and with guanines involved in hydrogen-bonded tetrads. Aboul-ela *et al.*, (1992) further concluded that the most stable G₄ tetrad structure involves a parallel rather than an anti-parallel association of strands. This conclusion assumes the absence of the entropic advantage inherent in intramolecular folding which may force an anti-parallel association, as when thymine and/or adenine linkers are present between self-associating guanine tracts. Qualitatively similar results have been found in studies of the related sequences d(T₂AG₃)/d(T₂G₄) (Wang and Patel (1992)) and d(T₄G₄) (Gupta *et al.*, (1993)) in K⁺-containing buffer.

Parallel-stranded and anti-parallel foldback structures are significantly different at the nucleotide level. While these structures have not been observed *in vivo*, it is possible that such structural polymorphism may be biologically important. The intermolecular fold-back structure is the only DNA quadruplex formed by an oligonucleotide to have been characterised to high-resolution using X-ray crystallographic techniques (Kang *et al.*, (1992)). The hexadeoxynucleotide d(TG₄T) presents perhaps the simplest possible case of an oligonucleotide that has been shown to form an intermolecular quadruplex at physiological pH and monovalent cation concentration.

Simplicity, rather than explicit biological relevance, was the key consideration in the choice of this sequence for study. However, it should be noted that the sequence is very similar in composition to the telomere sequence d(TTG₄T) from *Tetrahymena*. The latter has been shown to be equivalent in complex formation to d(TG₄T) by NMR analysis (Wang and Patel (1993a)). As such it is a prime candidate for the crystallographic study of the parallel-stranded, tetrameric form of G₄ DNA. Structural

information of this kind may prove to be useful in the understanding of both parallel and anti-parallel quadruplexes.

The first step in any crystallographic analysis is the reproducible growth of diffraction quality single crystals. The growth and characterization of crystals of the oligonucleotide d(TG₄T) in the presence of the monovalent sodium cation will be discussed in this chapter. Initial attempts to obtain crystals of the RNA analogue of d(TG₄T), r(UG₄U), are outlined at the end of the chapter.

2.2 Solid-State Synthesis of Oligonucleotides

The number of high resolution X-ray crystallographic studies of DNA oligonucleotides has grown rapidly in number over the last fifteen years (for reviews see Dickerson (1992); Joshua-Tor and Sussman (1993)). This has largely been made possible because of considerable advances in synthetic oligonucleotide chemistry. The processes involved in synthesising DNA have been automated and DNA-synthesizers are commercially available. A standard 10 μ mol solid support resin can yield as much as 20mg of a 20-nucleotide sequence or 20-mer. This material can then be rapidly purified using High-Performance Liquid Chromatography (HPLC). Short RNA fragments can also be chemically synthesized, although this presents additional problems due to the need to protect the 2'-hydroxyl of the ribose. Larger RNAs, such as transfer RNA (tRNA), require cellular extraction or *in vitro* transcription systems. A variety of methods and protocols have been devised (Dock-Bergon and Moras (1992)).

2.3 Crystallization of Nucleic Acids

Data pertaining to the specific conditions that proved to be successful in the growth of diffraction quality DNA crystals are presented in a number of reviews (Kennard and Hunter (1989); Dock-Bergon and Moras (1992)). Crystals grown have used oligonucleotide concentrations in the range 0.5–4.0mM, with most in the range 1.0–2.0mM. The crystallization of an ever increasing number of DNA oligonucleotides has revealed a number of features common to the materials and methods employed.

2.3.1 Precipitating Agents

The most commonly employed precipitating agents used in the crystallization of nucleic acids are alcohols. MPD (2-methyl 2,4-pentanediol) has proven to be particularly popular since it is non-volatile and easy to handle with a boiling point of 197°C. It is used in the range of about 30% for oligonucleotides. Isopropanol and ammonium sulphate have also proved to be very successful in work with tRNAs, as has the polymer polyethyleneglycol (PEG) in the average molecular weight (M_r) range 4000–8000 a.u. Smaller PEGs (M_r 600) have shown some success with DNA oligonucleotides, in particular the Z-DNA hexamer d(CG₃) (Wang *et al.*, (1984)).

The phosphates of nucleic acids are negatively charged at neutral, physiological pH. Thus DNA and RNA may be considered to be highly negative polyelectrolytes. The mixing of salts with PEG, MPD or some other alcohol presents a more complex and interesting system, as the salts act as an electrostatic shield mediating the interactions between the nucleic acid, the precipitating agents and the solvent water. This is a

particularly important consideration in the crystallization of the d(TG₄T) tetraplex where the Na⁺ cation is necessary for stability and is introduced using NaCl.

2.3.2 Growth Temperature

Nucleic acids are stable over a wide range of temperatures. Typically crystals are grown at 4°C in a self-contained cold-room, or at a regulated ambient temperature of about 20°C. In certain cases crystals have been grown at around 35°C using an incubator or oven (Chevrier *et al.*, (1986)). The d(TG₄T) tetraplex is observed in NMR studies below temperatures of 60°C (Aboul-ela *et al.*, (1992)). The noted slow rate of formation of the tetraplex, and the predominance of the single-stranded oligonucleotide at temperatures greater than 60°C, suggests that room temperature or 4°C trials would be more likely to yield tetraplex crystals.

2.3.3 Buffering and pH

In the crystallization of proteins it has been long observed that pH often plays a very important role, affecting overall protein charge and the nature of the packing arrangement (Blundell and Johnson (1976)). However, pH appears to play a much lesser role in the crystallization of nucleic acids since, as mentioned earlier, these are negatively-charged polyelectrolytes. Cytosine bases are protonated at pH 4.0–5.0. If such cytosine nucleotides are accessible then the additional charge interaction may promote crystallization.

Interestingly, it has recently been shown by NMR that the cytosine tract d(TC₅)₄ folds at low pH to form a structure consisting of a four-stranded backbone, similar to

that seen in guanine quadruplex structures (Gehring *et al.*, (1993)). Hemi-protonated cytosine bases participate in non-Watson-Crick base-pairs, forming parallel-stranded C-C⁺ duplexes. Two such duplexes intermesh in an anti-parallel manner, with the C-C⁺ base-pairs of one duplex fully intercalated with those of the other duplex, in what has been termed an *i-motif*. Recent native gel electrophoresis studies of the cytosine-rich sequences of human telomeric DNA demonstrated that the oligonucleotide d(CCCTAA)₄ forms an intramolecular complex at low pH (Ahmed *et al.*, (1994)). Gel species were consistent with the formation of four-stranded structures and NMR analysis revealed characteristic C⁺ imino proton resonances and connectivities consistent with the *i-motif*. This raises the question of whether cytosine tracts can form *i-motifs in vivo*, perhaps in concert with complementary guanine tetrad structures. The obvious weakness in this proposal is the necessity for low pH and cytosine protonation. As yet, no crystal structures of such a species have been reported.

For crystallization of nucleic acids generally, low pH may introduce local structural artifacts, while alkaline pH will lead to degradation for RNAs. Sodium cacodylate is the buffer most often used, with an effective buffering range of 5.2–7.2, and has the advantages that it is pH stable over a large temperature range and possesses anti-bacterial properties. The buffer can also act as a mild precipitating agent with variation in concentration if low ionic strength additives such as MPD or PEG are used.

2.3.4 Addition of Polyamines

Polyamines are extensively distributed in biological systems. They are involved in DNA condensation as well as in protein synthesis (Tabor and Tabor (1976)). A few

examples of polyamines are given in Table 2.1. For crystallization, spermine is most commonly used. It is a linear molecule with four positive charges at neutral pH. Spermine played a key role in the first crystallizations of tRNAs, and was subsequently used in the high-resolution structure determination of yeast tRNA^{phe} (Ladner *et al.*, (1972)). Following on from this success, spermine was systematically employed in crystallization trials with nucleic acids. Spermidine, which contains three positive charges at neutral pH, has also been utilized in crystallization experiments but with less success (Dock-Bergon and Moras (1992)).

Putrescine	$\text{H}_2\text{N}-(\text{CH}_2)_4-\text{NH}_2$
Cadaverine	$\text{H}_2\text{N}-(\text{CH}_2)_5-\text{NH}_2$
Spermidine	$\text{H}_2\text{N}-(\text{CH}_2)_3-\text{NH}-(\text{CH}_2)_4-\text{NH}_2$
Thermine	$\text{H}_2\text{N}-(\text{CH}_2)_3-\text{NH}-(\text{CH}_2)_3-\text{NH}-(\text{CH}_2)_3-\text{NH}_2$
Spermine	$\text{H}_2\text{N}-(\text{CH}_2)_3-\text{NH}-(\text{CH}_2)_4-\text{NH}-(\text{CH}_2)_3-\text{NH}_2$

Table 2.1: Examples of Polyamines

As a consequence of its positive charge, spermine binds in the phosphate grooves of nucleic acids. The structure of yeast tRNA^{phe} was refined to 2.5Å resolution and disclosed two spermine molecules (Ladner *et al.*, (1972)). One spermine molecule is coiled in the deep major-groove of the anticodon arm of the tRNA. It is hydrogen-bonded to four different phosphates on both sides of the groove. The second spermine

molecule is visibly sandwiched between two distinct sections of the tRNA. Spermine molecules have been identified in a number of Z-DNA structures (Fujii *et al.*, (1982) and references therein), although no spermine molecules have yet been seen in structures of A-type oligonucleotides. Kang *et al.*, (1992) observed spermine molecules in their 2.5Å structure determination of the *Oxytricha* telomeric DNA sequence oligonucleotide d(G₄T₄G₄), although stereochemical details were not given. Therefore, it is likely that spermine, or an alternative polyamine, may be necessary for the successful growth of diffraction quality crystals of the d(TG₄T) tetraplex.

2.3.5 Addition of Divalent Cations

Divalent cations, and particularly Mg²⁺, have been known for a long time to be involved in the stabilization of nucleic acid structures and to play an important role in their function. The structures of tRNA gave the first insight into the structural role played by Mg²⁺ ions, where they were observed to stabilize loops and bends in non-helical regions of the structure (Ladner *et al.*, (1972)). The role of divalent cations is interesting from the standpoint of crystallization since divalent cations have also been observed to stabilize the crystal packing. This effect has been observed in tRNAs, as well as in A- and B- helical forms, and especially in structures of Z-DNA oligonucleotides at high-resolution (reviewed in Kennard and Hunter (1989, 1991); Joshua-Tor and Sussman (1993)). In these structures, one Mg²⁺ ion is co-ordinated by six oxygens, one of which is a phosphate oxygen of the backbone, the others belonging to water molecules. A number of the sites also form intermolecular contacts. The presence of such contacts appears to convey greater crystal packing stability, and

would seem to contribute to the generally high-resolution diffraction observed with Z-DNA structures.

A number of additional divalent cations have proved successful in crystallizations, used either in place of or in conjunction with Mg^{2+} . In the case of tRNAs these include manganese, calcium, cobalt, nickel, barium and mercury (Dock-Bergon and Moras (1992)). A very small number of cases have reported the complex ion cobalt hexammine ($[Co(NH_3)_6]^{3+}$) as a stabilizing agent (Dock-Bergon and Moras (1992)). In the case of Z-DNA, certain studies have shown a five-fold increase in crystal stability using $[Co(NH_3)_6]^{3+}$ in comparison with the use of Mg^{2+} (Dock-Bergon and Moras (1992)). Of particular interest to guanine tetrad oligonucleotide structures is the role played by the $[Co(NH_3)_6]^{3+}$ ion in stabilizing the tertiary structure of tRNA^{phe} (Hingerty *et al.*, (1982)). In this study, X-ray evidence demonstrated that the $[Co(NH_3)_6]^{3+}$ ion hydrogen-bonded to pairs of guanine bases in the major-groove, rather than replacing Mg^{2+} ions involved in stabilizing crystal contacts. However, there have been no comparable investigations of the role of the $[Co(NH_3)_6]^{3+}$ ion in relation to guanine tetrad quadruplexes, and the ion was not used in the X-ray structure determination of the *Oxytricha* telomere repeat d(G₄T₄G₄) (Kang *et al.*, (1992)).

2.3.6 Addition of Monovalent Cations

Generally speaking, the stabilization of intermolecular, crystal packing contacts is assisted by divalent cations such as Mg^{2+} . However, in the case of the Z-DNA structure of d(Br⁵CG)₃, crystals were grown using Na⁺ ions only (Chevrier *et al.*, (1986)). The final structure revealed Na⁺ ions occupying interstitial positions, making

bridging contacts between neighbouring molecules. The octahedral geometry of the Na^+ ions was less well-defined than that of Mg^{2+} . As detailed earlier in this chapter, the monovalent ions Na^+ or K^+ are essential for formation and stability of the parallel-stranded tetraplex formed by the oligonucleotide d(TG₄T). In the crystallization of the sodium form of the tetraplex it is possible that the sodium ion may play a role in stabilizing crystal contacts, in addition to its putative role within the four-stranded helix.

2.4 Preparation of d(TGGGGT)

The oligonucleotide 5'-TpGpGpGpGpT-3' was synthesised by Dr. Alastair Murchie of the CRC Nucleic Acid Structure Research Group of the University of Dundee. Samples were prepared on a 1 μmol scale using β -cyanoethylphosphoramidite chemistry (Sinha *et al.*, (1984)) without 5' deprotection and purified by reverse phase chromatography using NENSORBTM columns.

Lyophilized samples were resuspended in 200 μl of 5mM HEPES pH 7.0, 100mM NaCl. These were then annealed by placing the samples in an insulated waterbath at 70°C and then cooling slowly to room temperature in order to induce tetraplex formation. Cooled samples were then dialysed extensively against 50mM NaCl using a 3 kDa molecular weight cut-off (Centricon-3) by serially concentrating and diluting the sample into buffer. The samples were then ready for crystallization trials and were stored at 4°C.

$$c_{dna} = \frac{s_{abs}}{\epsilon H_y} \quad (2.1)$$

The DNA concentration of the final sample was quantified spectrophotometrically using the absorption of the sample at 260nm within a standard 1cm path cell. This concentration was calculated using equation 2.1, a variation of the Beer–Lambert Law of optical absorption (Eisenberg and Crothers (1979)), where c_{dna} is the final mMol concentration of the sample; s_{abs} is the optical absorption of the sample measured at 260nm in O.D. units; ϵ is the extinction coefficient of the oligonucleotide, calculated from base composition to be 65.6; and H_y is the hypochromic correction factor for DNA which has a value of 0.77. Base stacking within nucleic acid structures results in a reduction in the absorption of UV radiation. This phenomena is termed the hypochromic effect (Saenger (1984)).

2.5 Mass Spectrometry Analysis of d(TGGGGT)

Matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI–TOF MS) is a recently reported technique which demonstrates considerable precision and utility in the analysis and characterization of natural and modified oligonucleotides (Pieles *et al.*, (1993)). The general methodology involves embedding the sample under analysis in a crystalline matrix. The sample is then released by laser–pulsed desorption and ionization, whereupon the sample is accelerated in a constant electric field and detected at the end of a linear tube.

Previous methods, which have proved very successful in the analysis of peptides

and protein molecules, are inclined to be inefficient in dealing with oligonucleotides. This tends to be as a result of poor desorption and/or ionization behaviour of larger molecules and the high tendency of oligonucleotides to form adducts with sodium and potassium ions. The latter leads to multiple peaks and resultant broad signals which make accurate mass determination extremely difficult. Pieleś *et al.*, (1993) have employed a new crystalline matrix of 2,4,6-trihydroxy acetophenone, together with di- and tri-ammonium salts of organic and inorganic acids. This approach has proved successful in suppressing peak broadening due to the formation of multiple adducts.

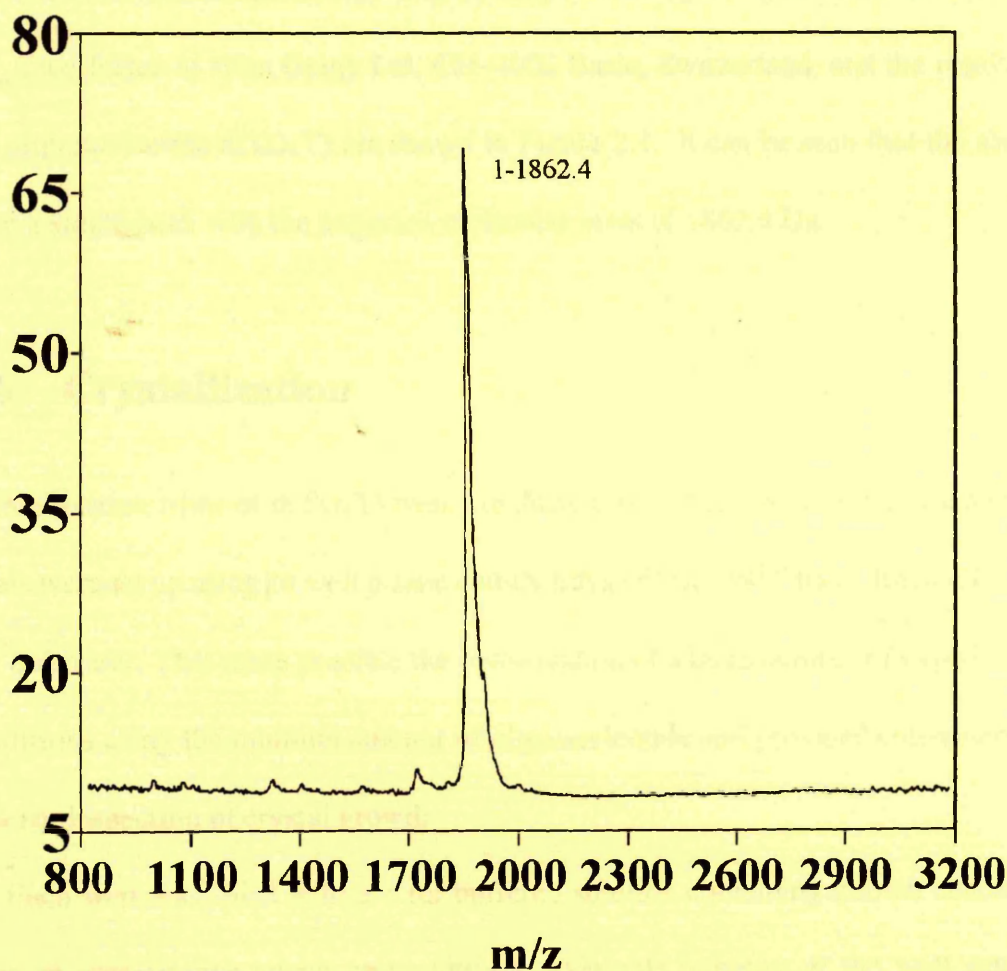


Figure 2.1: MALDI-TOF MS Analysis of 5'-TpGpGpGpGpT-3'

A number of samples of oligonucleotides were submitted for analysis in order to ascertain the success of the synthesis and the stability of the products. The results from derivatives of d(TG₄T) are discussed in Chapter 3. Samples of d(TG₄T) were prepared for MALDI-TOF MS analysis by extensive dialysis against distilled water using a 3 kDa molecular weight cut-off (Centricon-3), in order that the sample might be salt free. Concentration of the final sample was determined by UV absorption at 260nm, as detailed in Section 2.4. Samples were then lyophilized. Each sample contained approximately 200 picomoles of DNA. MALDI-TOF MS has demonstrated picomole and even femtomole sensitivity. The mass spectroscopy analysis was carried out by Dr. Uwe Pieleles of Ciba Geigy Ltd, CH-4002 Basle, Switzerland, and the results for the oligonucleotide d(TG₄T) are shown in Figure 2.1. It can be seen that the sample gave a single peak with the expected molecular mass of 1862.4 Da.

2.6 Crystallization

Crystallization trials of d(TG₄T) were conducted using the vapour diffusion method. Trials were set up using 24 well plastic culture trays (FALCONTM trays, Becton Dickinson Labware). This made possible the investigation of a large number of experimental conditions using the minimal amount of oligonucleotide and provided convenient access for inspection of crystal growth.

Each well was filled with a 1 ml buffered solution containing known concentrations of precipitating agents and additives. Variable volumes of the well solution were transferred to the drop containing the oligonucleotide in order to attain desired

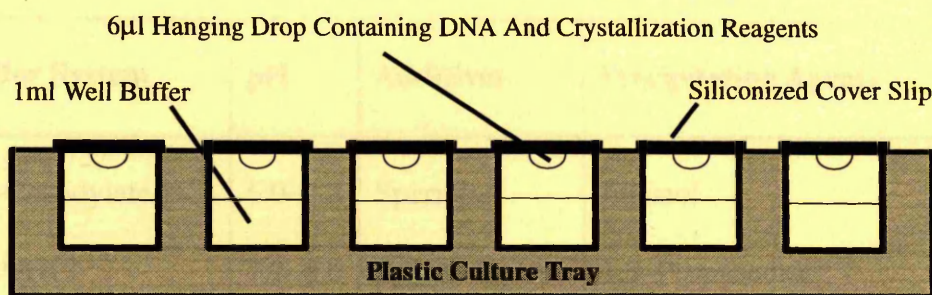


Figure 2.2: Experimental Set-Up for Crystallizations

concentrations of the various factors in the initial drop. Spermine, spermidine and $[\text{Co}(\text{NH}_3)_6]\text{Cl}_3$ were added directly to the drop and not to the well, in accordance with previous experimental procedures (Dock-Bergon and Moras (1992)). The drop was then suspended over the well using siliconized coverslips and the well sealed with vacuum grease. This set up is illustrated in Figure 2.2 and is adapted from Blundell and Johnson (1976). The crystallization of oligonucleotides is, as with proteins, largely a trial and error process. The many chemical and physical factors involved are detailed in Section 2.3. Table 2.2 provides a list of the buffers, additives and precipitants employed in attempts to crystallize d(TG₄T).

Initially, trials were restricted to room temperature. The drops generally contained an initial buffer concentration of 20 mM and an oligonucleotide concentration of 1 mM (0.25 mM tetraplex). Early experiments suggested that spermine was essential to crystal growth and that a pH range of 6.5–7.5 was optimal. Furthermore, spermidine and $[\text{Co}(\text{NH}_3)_6]\text{Cl}_3$ were demonstrably less effective in promoting crystal growth. The best initial results were obtained using MPD as the precipitating agent and an initial concentration of 6mM spermine in the starting droplet. Showers of microcrystals were observed within a few hours of setting up the drops. Fewer and larger crystals could

Buffer System	pH	Additives	Precipitating Agents
Na-Cacodylate/HCl	6.0–7.2	Spermine	Ethanol
Tris ¹ /HCl	7.2–8.0	Spermidine	1,3-Propanediol
		[Co(NH ₃) ₆]Cl ₃	1,2-Pentanediol
		MgCl ₂	DL-1,2-Hexanediol
		CaCl ₂	MPD ²
		BaCl ₂	PEG ³ 400, 2K, 4K, 6K
		ZnCl ₂	Saturated Ammonium Sulphate
		CdCl ₂	Saturated Sodium Chloride
		NaCl	Distilled Water

¹ 2-Amino-2-(Hydroxymethyl)-1,3-Propandiol

² 2-Methyl 2,4-Pentanediol

³ Polyethylene Glycol

Table 2.2: Reagents Used in Crystallization Trials

be obtained by lowering the MPD concentration in the starting droplet to about 5% v/v, regardless of the concentration in the well. Variation of spermine concentration resulted in precipitate at higher ranges and a loss of crystalline form at lower ranges. In addition, at 1 mM concentration the oligonucleotide precipitated in the presence of 6 mM spermine and 5% v/v MPD if the NaCl concentration in the initial drop was less than 50 mM. Crystals grown in this way were too small for diffraction studies and were unstable, dissolving after 1–2 weeks. Variation in oligonucleotide concentration

yielded no visible improvement. The other precipitants were largely unsuccessful with the exception of the PEGs and 1,3-propanediol. The former, however, largely resulted in oily solids, whilst the small crystals obtained with the latter were unstable and dissolved. BaCl_2 , ZnCl_2 and CdCl_2 all produced precipitate, while CaCl_2 and MgCl_2 both appeared to promote crystal growth in the range 10–20 mM in the initial drop.

When facilities for storage at 4°C became available, a number of trays previously set up at room temperature were transferred with significant results. Trials using MPD as the precipitant produced sharp, well-defined needle crystals that remained stable in the mother liquor, and appeared to require spermine and Ca^{2+} , but not Mg^{2+} , for growth. This result was reproduced with trays set up at 4°C. Again, variation of pH, spermine, MPD, oligonucleotide and Ca^{2+} concentration failed to produce larger crystals.



Figure 2.3: A Single Crystal of the Oligonucleotide d(TGGGGT)

Crystal size was dramatically improved when NaCl was introduced as a precipitating agent in the well, in addition to MPD, with a proportional increase in the NaCl concentration in the initial droplet as a result. Crystals grown from droplets where the initial NaCl concentration was approximately 50–80 mM were large enough for diffraction analysis, but were hopelessly twinned. The quality of the crystals was further improved by increasing the initial concentration of NaCl in the droplet. Diffraction quality crystals were subsequently grown from drops containing 20 mM Na–Cacodylate/HCl (pH 6.6–7.2) or 20 mM TRIS/HCl (pH 7.5), 10 mM CaCl_2 , 6 mM spermine, 130–180 mM NaCl, 5% v/v MPD and 1 mM d(TG₄T). The drops were equilibrated against wells containing 120 mM Na–Cacodylate/HCl (pH 6.6–7.2) or TRIS/HCl (pH 7.5), 70 mM CaCl_2 , 700–1000 mM NaCl and 26–32% v/v MPD. The drops were 6 μl in size.

Crystals appeared within a week and grew to their full size of about 340 x 200 x 200 μm over two to three weeks. The crystal morphology suggested a triclinic or possible monoclinic habit, as shown in Figure 2.3. The crystals exhibited a predominant tendency to grow as twins and as a result a certain number were unsuitable for X-ray work. However, a significant proportion twinned with a slender, well-defined join that could be easily severed, using a hair or finely stretched glass capillary, yielding two crystals for analysis (Figure 2.4).

The crystals appeared to be indefinitely stable in the drop liquor and were mounted in glass capillaries directly from the drop. Hydrating liquor was inserted at one end of the capillary only, to avoid wetting the crystal upon cooling. The crystals were mounted at 4°C using a custom-built dry-air crystal cooling device. Diffraction

patterns were recorded on an R–AXIS IIC phosphorimager area detector, using a Rigaku rotating anode X–ray generator and a graphite monochromator. Using $\text{CuK}\alpha$ radiation, diffraction was observed to beyond 1.6\AA . Subsequent diffraction experiments using Molybdenum radiation demonstrated diffraction to a limit of 1.0\AA . Diffraction tests at the Daresbury Synchrotron by Dr. Ben Luisi have exhibited diffraction to a limit of 0.88\AA at 4°C . It has been observed that crystallization drops like those described above, but containing 80–130 mM NaCl in the initial droplet (with 200–700 mM NaCl proportionally situated in the well) yield crystals of comparable quality over a period of several months. These crystals also appear to be indefinitely stable in the mother liquor.

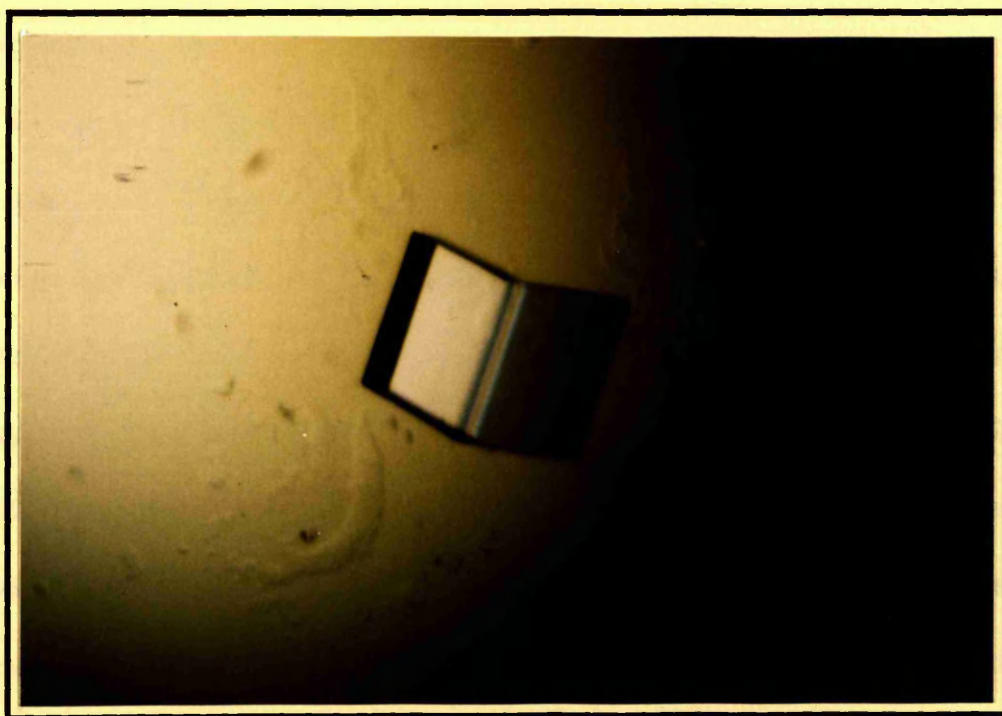


Figure 2.4: A Twinned Crystal of the Oligonucleotide d(TGGGGT)

2.7 Crystal Characterization

As noted above, crystals of d(TG₄T) grew with a triclinic, or possibly monoclinic, habit. A crystal was mounted at 4°C and aligned with the beam such that the longest axis was vertical. A number of 2° oscillation images were collected using the R-Axis IIC phosphorimager area detector, a Rigaku rotating anode X-ray generator and a graphite monochromator. The cell dimensions were measured to be $a=28.76\text{\AA}$, $b=35.47\text{\AA}$, $c=56.77\text{\AA}$, $\alpha=74.39^\circ$, $\beta=77.64^\circ$ and $\gamma=89.73^\circ$, as determined by the auto-indexing package *REFIX* (Kabsch (1988, 1993)). The diffraction pattern contained no symmetry, except for Freidel symmetry, and there were no apparent systematic absences. Such evidence suggested that the cell belonged to the symmorphic space group P1.

The presence of a near 90° angle in the cell, whilst not impossible for P1, raised the question of whether the crystal did in fact have a cell of a higher symmetry. This concern was heightened by the presence of a large non-origin peak in the native Patterson of the data processed using the cell above, which is discussed in detail in Chapter 5. The cell was subsequently checked in two ways. Firstly, the crystal was mounted at 4°C on the Rigaku AFC5R 4-circle diffractometer, using a Rigaku rotating anode generator, and an auto-indexing program was set to search for peaks and to calculate the smallest cell of the highest symmetry. The software employed utilises the Delaunay reduction procedure (AFC5R SOFTWARE PROGRAM) and constitutes a more accurate experimental method of crystal characterization. Secondly, the cell dimensions obtained above were tested using the *NRCVAX* cell reduction program

CREDUC (Gabe *et al.*, (1989)). In both cases, the cell and space group were confirmed to be correct.

2.8 Preliminary Crystallization Trials of r(UGGGGU)

The RNA analogue of d(TG₄T), r(UG₄U), has been studied to high-resolution using NMR (Cheong and Moore (1992)). The oligonucleotide r(UG₄U) has been observed to form a parallel-stranded tetraplex that is stabilized by K⁺ cations. There is also evidence for the formation of a novel U₄ tetrad. Analogous T₄ tetrads have not been observed in studies of parallel-stranded DNA sequences, as discussed in this chapter and in Chapter 1. The crystals of d(TG₄T) diffract to very high-resolution and crystals of r(UG₄U) would form the basis of an interesting structural comparison between the RNA and DNA forms of the parallel-stranded tetraplex. The two structures would be expected to be quite different at the nucleotide level. While DNA is structurally polymorphous, adopting A-, B- and Z-forms, RNA is quite structurally conservative, adopting the A-form exclusively (Saenger (1984)). Indeed, while the d(TG₄T) tetraplex has been characterized as adopting a B-form structure (Aboul-ela *et al.*, (1992)), the r(UG₄U) tetraplex has been characterized as having an A-like backbone by Cheong and Moore (1992).

Samples of r(UG₄U) were prepared for crystallization in the same way as described for d(TG₄T) in this chapter. The conditions described above for crystals of the sodium form of the d(TG₄T) tetraplex resulted in microcrystalline aggregates of the RNA analogue. The sodium form of the r(UG₄U) tetraplex appears to have a much higher

sensitivity to spermine. The most promising results to date have emerged from experiments where 1,6-hexanediol is used analogously to MPD, in conditions similar to those reported above. Small, needle crystals of r(UG₄U) have been observed, and are shown in Figure 2.5. Work continues to obtain crystals suitable for X-ray data collection.



Figure 2.5: Microcrystals of the RNA Analogue of d(TGGGGT), r(UGGGGU)

It may be asked, of course, whether all the invariants, conservations and symmetries that make up the texture of scientific discourse are not fictions substituted for reality in order to obtain a workable image . . . A logic itself founded upon a purely abstract, perhaps 'conventional', principle of identity – a convention with which, however, human reason seems to be incapable of doing without.

Chapter 2 Jacques Monod, *Chance and Necessity*

The entities postulated by science are not *found*, and they do not constitute an 'objective' stage for all cultures and all of history. They are *shaped* by special groups, cultures, civilizations: and they are shaped from a material which, depending on its treatment, provides us with gods, spirits, a nature that is a partner of humans rather than a laboratory for their experiments, or with quarks, fields, molecules, tectonic plates. Social monotony thus implies cosmic monotony – or 'objectivity', as the latter is called today.

Paul Feyerabend, *Against Method*

Thoughtless people are in the habit of pointing out that every 'reasonable' person will be persuaded that science knows best. The comment admits a weakness of argumentation: arguments do not work on everyone, they work only on people who have been properly prepared. And this is a general feature of all ideological debates: arguments in favour of a certain world view depend on assumptions which are accepted in some cultures, rejected in others, but which because of the ignorance of their defenders are thought to have universal validity.

Paul Feyerabend, *Farewell to Reason*

The significance of language for the evolution of culture resides in the fact that in language man set a world of his own over against the other world . . . To the extent that man has for long ages believed in the concepts and names of things as in *aeternae veritates* [eternal truths] he has appropriated to himself that pride by which he raised himself above the animal: he really thought that in language he possessed knowledge of the world. The sculptor of language was not so modest as to believe that he was only giving things designations, he conceived rather that with words he was expressing supreme knowledge of things: language is, in fact, the first stage of the occupation with science . . .

Friedrich Wilhelm Nietzsche, *Human, All Too Human*

Chapter 3

Preparation of Heavy-Atom Derivatives

3.1 The Isomorphous Replacement Method

There are two main techniques of phase determination for macromolecular structures. These are the molecular replacement and isomorphous replacement methods. The molecular replacement method is suitable for a protein or oligomeric nucleic acid where a homologous structure is already known. Furthermore, in the case of nucleic acids, models may be available from preliminary fibre diffraction or NMR studies. Most of the successful structure determinations of oligomeric nucleic acids to date have employed the molecular replacement method (Kennard and Hunter (1989, 1991)), and the application of this approach in the structure solution of d(TG₄T) is discussed in Chapter 5.

The isomorphous replacement method is employed when no homologous or rudi-

mentary model is available for study. This is more often the case for protein molecules which are commonly much more structurally variant and complex than the helical duplexes, trimers and higher order multimers formed by the nucleic acids. For oligomeric nucleic acids, with the exception of one structure solved using maximum entropy methods (Miller *et al.*, (1988)), isomorphous replacement techniques have been central to the analysis of oligonucleotide structures for which no investigative model was available (Kennard and Hunter (1989, 1991)). As, initially, a three-dimensional model of the parallel-stranded tetraplex formed by d(TG₄T) was not available, experiments were undertaken to tackle the phase problem using the isomorphous replacement approach.

The isomorphous replacement method was first used in the structure determination of phthalocyanines (Robertson and Woodward (1937)), where the strategy employed involved the direct substitution of an atom in the molecule with a similar, but heavier, element. This is the strategy commonly employed when the method is applied in small molecule studies. The method was first utilized for biological macromolecules in 1954 in the structural studies of the protein haemoglobin (Green *et al.*, (1954)). It is difficult to introduce heavy-atoms directly into the structure of a protein or ribonucleic acid, such as a tRNA, except in cases where a metal ion may be replaced with one of a heavier metal. Commonly, heavy-atoms are introduced into the macromolecule crystal by addition rather than replacement. This is possible since the percentage of solvent that constitutes a macromolecule crystal is typically large, in the region of 30–70%, and this permits the diffusion of heavy-atoms along the solvent channels. Thus, the technique used is still isomorphous replacement in the sense that it is solvent within the crystal that is replaced. In the case of oligonucleotides prepared by solid-

state synthesis techniques, atoms may be replaced in the manner described for small molecule studies. This commonly involves the formation of a covalent Br-C or I-C bond, usually involving the C₅ atom of the pyrimidine bases cytosine, thymine or uracil. The resultant derivatives must then be grown in crystals isomorphous to the native.

Crick and Magdoff (1956) have shown that the average change of intensity for acentric reflections due to the addition of heavy-atoms is given by equation 3.1, where N_H heavy-atoms of scattering factor f_H are bound to a molecule of N_P light atoms of scattering factor f_P . This indicates that the required size of the heavy-atom contribution is dependent upon the size of the macromolecule, with heavier atoms clearly being necessary for larger macromolecules. Oligonucleotides and their assemblies, which are commonly lighter than most proteins, have thus been successfully solved using bromine and iodine atom derivatives, with 35 and 46 electrons respectively.

$$\left(\frac{2N_H}{N_P}\right)^{\frac{1}{2}} \left(\frac{f_H}{f_P}\right) \quad (3.1)$$

Ideally speaking, a perfect isomorphous derivative is one in which the only change in electron density between it and the native crystal is a peak at the site of the heavy-atom. Therefore, practically speaking, in order to obtain a good isomorphous derivative the aim would be to introduce a heavy-atom into a crystal with minimal disturbance of the macromolecule conformation, orientation or crystal lattice. In addition, there should be optimally a limited number of heavy-atoms sites. All sites should have high occupancy and be tightly bound to the macromolecule. This would greatly simplify the solution of the resultant Patterson functions. In practice multiple binding sites of low

occupancy are often found and small changes in the unit cell dimensions are commonly observed.

Calculations have shown that a 0.5% change in the overall cell dimensions or a 0.5° rotation of the molecules in the cell can result in an average change in intensities of 15% to 3 Å resolution (Crick and Magdoff (1956)). Such changes are resolution dependent and the resultant lack of isomorphism between native and derivative crystals can be analysed by plotting fractional isomorphous differences against resolution. Lack of isomorphism is typically indicated by a rise in the ratio with increase in resolution. Isomorphous differences in the range 10–25% are considered to be experimentally acceptable. As reported in Chapter 2, the crystals of d(TG₄T) diffract to atomic resolution. Therefore, an accurate structural study to this resolution requires at least two derivatives that are strictly isomorphous to high-resolution, or one high-quality derivative with an anomalous scattering contribution.

3.2 Selection of Derivatives

The selection and preparation of heavy-atom derivatives remains largely a matter of trial and error. Any rational strategy of derivative preparation assumes the suitability of a particular site, information which ultimately will only be obtained from the solved structure itself. Generally, consideration should be given as to whether the mother liquor is appropriate for derivative preparation or not. High concentrations of salt-precipitating agents and phosphate buffers can react with many potential derivatives, and crystals may need to be transferred to less reactive medium if grown under these

conditions. The pH of the crystallization liquor, which is an important consideration in the derivative preparation of proteins, is less significant for oligonucleotides which commonly crystallize at around neutral pH.

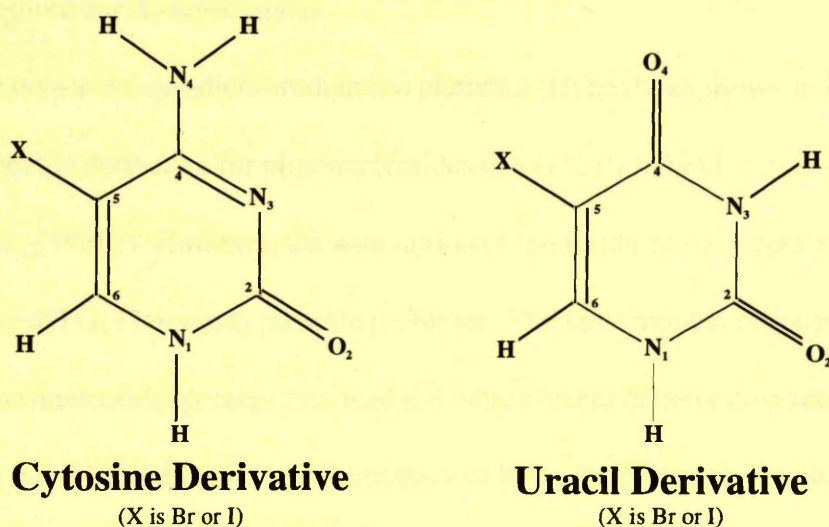


Figure 3.1: Synthetic Heavy-Atom Derivatives of Uracil(Thymine) and Cytosine

As indicated in Section 3.1, the solid-state synthesis of oligonucleotides presents an opportunity for the covalent incorporation of heavy-atoms into the sequence under study. Derivatives are formed by the attachment of a bromine or iodine atom to the C₅ carbon of the pyrimidine bases cytosine, uracil or thymine. The resultant derivatives are 5-bromocytosine and 5-iodocytosine, derived from cytosine, and 5-bromouracil and 5-iodouracil, derived from either thymine or uracil. These derivatives are illustrated in Figure 3.1. The sequence d(TG₄T), containing a 5' and 3' thymine, can therefore potentially yield two distinct derivatives by this variation of the isomorphous replacement method. Such covalent syntheses have been used in nearly all oligonucleotide structures where isomorphous replacement has been employed. This approach has the advantage that the resultant derivatives will have a specific number of sites of high

occupancy, although in cases where the asymmetric unit cell contains a considerable number of strands, the Patterson can be complex. Unfortunately, in many cases, the derivatives do not form isomorphous crystals or do not produce crystals of the size or quality required for X-ray analysis.

The heavy-metal *cis*-dichlorodiamino platinum (II) has been shown to be an effective and specific derivative for oligonucleotides in its ability to bind to guanine residues (Wing *et al.*, (1980)). However, the assumption of the formation of a parallel-stranded tetraplex by d(TG₄T) presents possible problems. The large number of roughly equivalent guanine nucleotides present may lead to a large number of derivative sites, resulting in a highly complex Patterson, low occupancy or both. Furthermore, the guanines may be inaccessible to the heavy-metal, especially as the platinum(II) metal preferentially coordinates the N₇ and O₆ atoms of the guanine base which, in the tetraplex strand arrangement, will be situated in the core of the structure. The highly symmetrical nature of the tetraplex, and the equivalence of the grooves in the parallel-stranded form, may well present a significant obstacle generally in obtaining derivatives of low site number and high occupancy by heavy-atom diffusion.

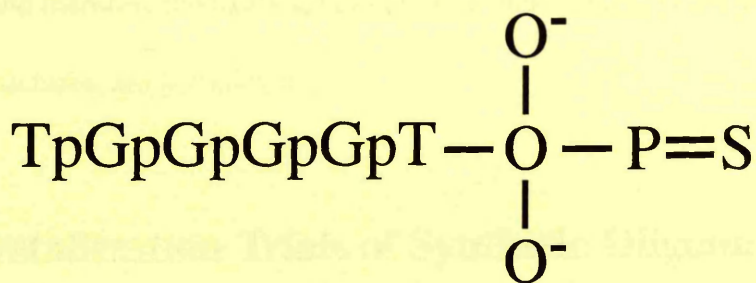


Figure 3.2: The Phosphothiol Chemical Group

It may, however, be possible to form a mercury derivative by the addition of a

phosphothiol group into the synthesised oligonucleotide. The phosphothiol group is illustrated in Figure 3.2, where it is covalently bonded to the 3' end of the d(TG₄T) oligonucleotide. The sulphur may act as an attachment site for mercury compounds, as is often observed for free cysteine residues in proteins. The attachment of the phosphothiol group to the 5' or 3' ends of the oligonucleotide again offers the choice of two potential derivatives.

It has been observed that quadruplexes containing guanine tetrads can form in the presence of cesium ions with a cation dependent stability (Williamson *et al.*, (1989); Sundquist and Heaphy (1993)). Tetraplex formation of d(TG₄T) using cesium as the stabilizing cation in place of sodium presents another possible derivative. However, the larger ion radius of cesium may influence the dimensions of the tetraplex and hence the isomorphism of any derivative. The divalent cation strontium, which has been shown to stabilise quadruplex formation more effectively than potassium (Chen (1992)), also forms a highly stable tetraplex with d(TG₄T) (David Lilley, Personal Communication). This information was not available when heavy-atom derivatives trials were carried out. The relative dimensions of the strontium, cesium and sodium tetraplexes, and therefore the likely suitability of strontium and cesium as isomorphous derivative structures, are not known.

3.3 Crystallization Trials of Synthetic Oligonucleotides

The oligonucleotides d(U_{Br}G₄T), d(U_IG₄T), and d(TG₄T-O-PO₂=S) were synthesised by Dr. Alastair Murchie of the CRC Nucleic Acid Structure Research Group of the

University of Dundee. Samples were prepared and quantified for crystallization trials in the same manner as described in Chapter 2, Section 2.4. The bromo- and iodo-derivatives were shielded from ambient light in order to prevent possible photolysis of the Br-C and I-C bonds. Crystallization trials were conducted using the vapour diffusion method as detailed in Chapter 2, section 2.6.

3.3.1 d(U_{Br}GGGGT)

Initially, crystallization trials were conducted using both the bromouracil and iodouracil derivatives. Both oligonucleotides exhibited very similar behaviour under a variety of crystallization conditions. Furthermore, the bromouracil derivative was abundantly available, whilst the iodouracil derivative was relatively scarce. Consequently, due to the availability of the bromouracil syntheses, the decision was taken to carry out further investigations using only the bromouracil derivative.

Samples of the bromouracil syntheses were submitted for mass spectrometry analysis, as described in Chapter 2, Section 2.5. The analysis was carried out by Dr. Uwe Pieleles of Ciba-Geigy Ltd, CH-4002 Basle, Switzerland, and the results for the oligonucleotide d(U_{Br}GGGGT) are shown in Figure 3.3. All of the samples submitted gave a single peak with an expected molecular mass of 1927.5 Da.

The bromouracil derivative was screened extensively. All trials were carried out, and the results monitored, under red light in order to minimize the chance of photolysis as mentioned above. Buffers, precipitants and additives employed in experimental trials are listed in Table 3.1. In a similar fashion to the native form, the bromouracil derivative of d(TG₄T) displayed a marked preference for the Ca²⁺ cation, spermine

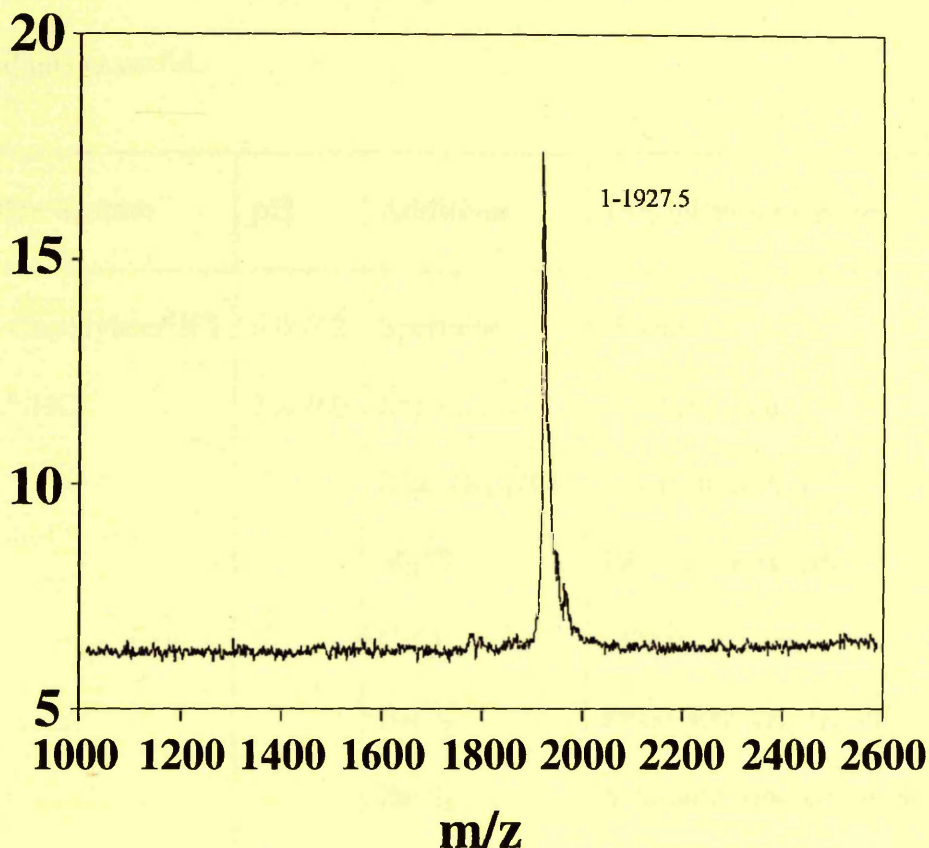


Figure 3.3: MALDI-TOF MS Analysis of 5'-U_{Br}pGpGpGpGpT-3'

and a growth temperature of 4°C, whilst exhibiting little variance with regards to pH. Crystals large enough for X-ray analysis could not be obtained despite strenuous efforts to enlarge the microcrystals that grew readily. Crystals with a maximal dimension of 60 μm were eventually grown from 6 μl drops containing 15–20% 1,6-hexanediol, 140 mM NaCl, 12 mM CaCl₂, 20 mM Na-Cacodylate/HCl pH 7.0, 1 mM spermine and 1 mM d(U_{Br}GGGGT), equilibrated against wells containing 15–20% 1,6-hexanediol, 5% v/v MPD, 700 mM NaCl, 70 mM CaCl₂ and 120 mM Na-Cacodylate/HCl pH 7.0 at 4°C. However, these crystals could not be enlarged and grew in a distinctly hexagonal

form, as shown in Figure 3.4, and not in the triclinic form previously observed for the native material. Attempts to enlarge the d(U_{Br}GGGGT) crystals by microseeding proved unsuccessful.

Buffer System	pH	Additives	Precipitating Agents
Na-Cacodylate/HCl	6.0–7.2	Spermine	Ethanol
Tris ^a /HCl	7.2–9.0	Spermidine	1,3-Propanediol
		[Co(NH ₃) ₆]Cl ₃	1,2-Pentanediol
		MgCl ₂	DL-1,2-Hexanediol
		CaCl ₂	MPD ^b
		BaCl ₂	PEG ^c 400, 2K, 4K, 6K
		ZnCl ₂	Saturated Ammonium Sulphate
		CdCl ₂	Saturated Sodium Chloride
		NaCl	Distilled Water
			1,6-Hexanediol

^a 2-Amino-2-(Hydroxymethyl)-1,3-Propandiol

^b 2-Methyl 2,4-Pentanediol

^c Polyethylene Glycol

Table 3.1: Reagents Used in Crystallization Trials of Synthetic Derivatives

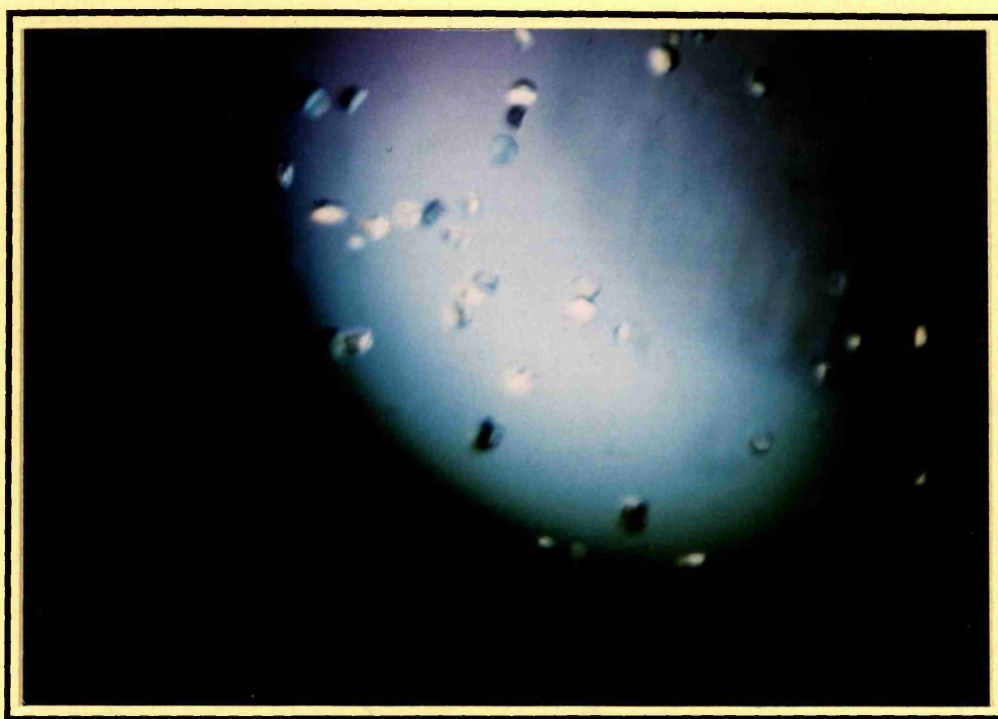


Figure 3.4: Microcrystals of the Heavy-Atom Derivative d(U_{Br}GGGGT)

3.3.2 d(TGGGGT-O-PO₂=S)

The oligonucleotide d(TGGGGT-O-PO₂=S) performed poorly in a number of crystallization screens, yielding only precipitate and amorphous “glasses” or globules. Mass spectrometry analysis of the syntheses, again conducted by Dr. Uwe Pieleles, revealed that the desired product d(TGGGGT-O-PO₂=S), whilst present, was heavily contaminated, as illustrated in Figure 3.5. Identification of the products of this analysis is presented in Table 3.2.

The synthesis of the phosphothiol was an exploratory investigation and it was suspected that the sulphur might exchange with oxygen when exposed to water (Alastair Murchie, Personal Communication). This was confirmed by the observation of the peak

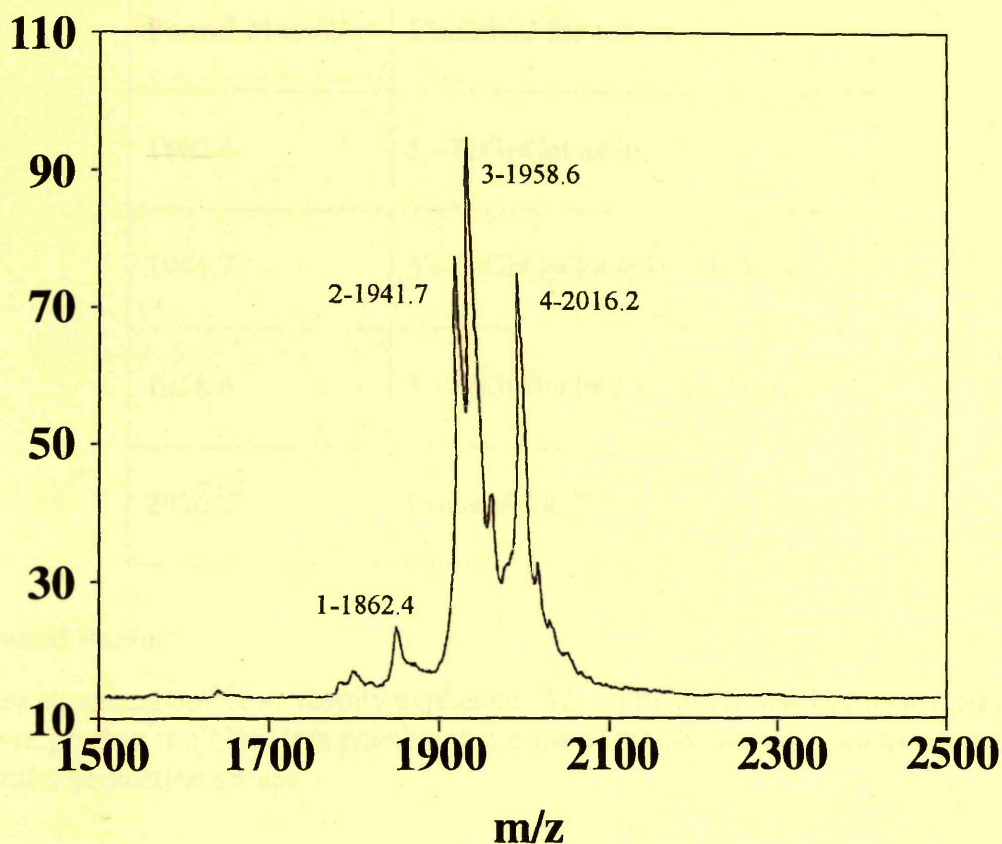


Figure 3.5: MALDI-TOF MS Analysis of 5'-TpGpGpGpGpT-O-PO₂=S-3' at 1941.7 Da. The additional degradation of the phosphothiol derivative to the native form as indicated by the peak at 1862.4 Da, and the retention of chemical blocking groups indicated by the 2016.2 Da peak, suggests that the synthesis of a suitably stable phosphothiol oligonucleotide for derivative formation may be more difficult than originally envisaged.

Found Mass(Da)	Chemical Structure
1862.4	5'-TpGpGpGpGpT-3'
1941.7	5'-TpGpGpGpGpT-O-PO ₃ -3'
1958.6	5'-TpGpGpGpGpT-O-PO ₂ =S-3' ^a
2016.2	Unidentified ^b

^a Desired Product

^b This signal cannot be so readily explained. The mass difference between it and the desired product is 57 Da. It is possible that some of the oligonucleotides have retained chemical protecting groups.

Table 3.2: Mass Spectrometry of the Phosphothiol Derivative of d(TGGGGT)

3.4 Heavy-Atom Soaking Experiments

Further attempts to find heavy-atom derivatives of the d(TG₄T) tetraplex relied upon the strategy of adding heavy-atoms to the crystal by means of soaking experiments. These trials were carried out in the mother liquor in which the crystals were grown where, as noted in Chapter 2, they were indefinitely stable. Experimentation with a number of stabilizing solutions indicated that the crystals were very sensitive to variations in NaCl concentration, particularly as the concentration in the drop was lowered below approximately 50mM.. The experimental setup is illustrated in Figure

3.6, adapted from Blundell and Johnson (1976).

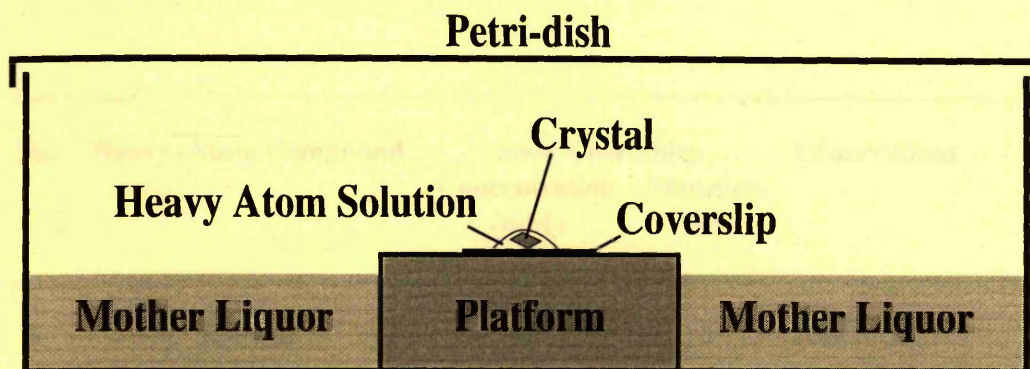


Figure 3.6: Experimental Set-Up for Heavy-Atom Soaking Experiments

Droplets approximately $2\ \mu\text{l}$ in volume, in which the crystals were grown, were transferred to a petri-dish containing 2 ml of the corresponding mother liquor. The dishes were sealed with vacuum grease and allowed to equilibrate overnight. Heavy-atom solutions were made up to the required concentrations in a solution with the composition of the mother liquor of the crystal under trial. This served to minimize any disruption to the crystal as a result of fluctuations in the liquor conditions, particularly the NaCl concentration. A calculated volume of the heavy-atom solution, usually 8–10 μl , was added to the mother liquor droplet containing the crystal and the petri-dish was resealed. The crystals were monitored during the specified duration of the soak. Details of the soaking experiments are given in Table 3.3. Once the soaking procedure was complete, integral crystals were mounted directly, without backsoaking, and placed immediately in the X-ray beam.

In the preparation of the Me-I and I₂ derivatives a different experimental procedure was employed. Native crystals were first mounted in capillary tubes ready for X-ray analysis. The capillary tubes were briefly opened at their dry end and a few

No.	Heavy-Atom Compound	Soak Conditions		Observations
		Concentration (mM)	Duration (Hrs)	
1	K ₂ PtCl ₄	2	3	Fine.
2	Pb(CH ₃ CO ₂) ₂	2	48	Fine.
3	Pb(CH ₃ CO ₂) ₂	5	48	Fine.
4	Pb(CH ₃ CO ₂) ₂	10	48	Fine.
5	UO ₂ (CH ₃ CO ₂) ₂ ^a	0.2	48	Very bad cracking. Heavy precipitate.
6	EuCl ₃ ^a	0.6	48	Surface cracks. Heavy precipitate.
7	CsCl ₂ ^b	700	3	Crystal disintegrated.
8	CsCl ₂ ^b	910	3	Slight cracking.
9	I ₂	—	2	Fine. Colour uptake.
10	Me-I	—	2	Fine.

^a In the case of the uranium and europium soaks the metals precipitated at the concentrations noted. The oligonucleotide crystallized in the presence of these metals at the same concentrations as those noted above, using the conditions described in Chapter 2, Section 2.6. Drops were visibly clear immediately after being set-up. A light precipitation was seen after a few weeks once crystals had grown. Hence, it is these co-crystallizations, and not the above soaks, that are referred to in Tables 3.4 and 3.5.

^b Crystallization experiments employing cesium as the counter-ion in place of sodium yielded only precipitate. The concentrations of cesium chloride referred to above are equivalent to the Na⁺ concentration in the mother liquor of the respective crystals used.

Table 3.3: Heavy-Atom Soaking Experiments

microlitres of volatile liquid Me-I or I₂ added. The capillary was then quickly resealed using molten wax and the volatile vapour was allowed to diffuse for a specified period. The remainder of the volatile liquid was then removed, the capillary sealed as before and the crystals placed in the X-ray beam. This procedure was carried out in a 4°C cold room, taking care to ensure that as little hot wax as possible was used in the resealing procedure.

Derivatives were screened quickly using the R-AXIS IIC phosphorimager area detector and CuK α radiation. Data-sets of 180° could be collected in 15 hours and then analysed against the native. The crystals were kept at 4°C using a dry-air, custom-built crystal cooling device. With the exception of the europium chloride co-crystallization, all of the derivatives screened in this manner survived in the beam long enough to permit the collection of complete data-sets to high-resolution. The cesium chloride soak, whilst it survived in the beam, diffracted poorly and the cell dimensions could not be determined accurately. Subsequently this data could not be processed. For the remaining derivatives it was then possible to obtain accurate cell dimensions and check for non-isomorphism. Details of the heavy-atom compounds screened in this way are given in Table 3.4. Cell dimensions, determined using the auto-indexing option of the software package *REFIX* (Kabsch (1988, 1993)), are presented in Table 3.5 along with the dimensions for the native cell. Methods relating to data collection and processing are discussed in the next chapter.

No.	Heavy-Atom Compound	Method	Comments
1	K_2PtCl_4	2° Oscillations	Good diffraction. 1.6Å data-set collected.
2	$\text{Pb}(\text{CH}_3\text{CO}_2)_2$	2° Oscillations	Good diffraction to at least 1.8Å.
3	$\text{Pb}(\text{CH}_3\text{CO}_2)_2$	2° Oscillations	Good diffraction to at least 1.8Å.
4	$\text{Pb}(\text{CH}_3\text{CO}_2)_2$	2° Oscillations	Good diffraction. 1.8Å data-set collected.
5	$\text{UO}_2(\text{CH}_3\text{CO}_2)_2$	2° Oscillations	Weak data-set collected to 1.8Å.
6	EuCl_3	2° Oscillations	Very poor diffraction. Died.
8	CsCl_2	2° Oscillations	Diffraction badly smeared. Collected 2.5Å data-set.
9	I_2	2° Oscillations	Good diffraction. 2Å data-set collected.
10	Me-I	2° Oscillations	Good diffraction. 2Å data-set collected.

Table 3.4: Heavy-Atom Screening Experiments

No.	Derivative	$a(\text{\AA})$	$b(\text{\AA})$	$c(\text{\AA})$	$\alpha(^{\circ})$	$\beta(^{\circ})$	$\gamma(^{\circ})$
1	K_2PtCl_4	28.61	35.41	56.61	74.29	77.60	90.08
4	$\text{Pb}(\text{CH}_3\text{CO}_2)_2$	28.64	35.46	56.73	74.16	77.80	89.83
5	$\text{UO}_2(\text{CH}_3\text{CO}_2)_2$	28.68	35.41	56.65	74.19	77.91	90.00
8	CsCl_2	28.60	35.30	56.95	74.14	77.33	90.02
9	I_2	28.63	35.52	56.85	74.17	78.06	89.92
10	Me-I	28.70	35.58	56.74	74.21	77.80	89.95
—	Native	28.76	35.47	56.77	74.39	77.64	89.73

Table 3.5: Cell Dimensions as Determined by *REFIX*

The grand movements of scientific and technical knowledge, the perception of an intimate state of matter, will have been only a series of procedures signalling in advance what was to follow at the heart of Occidental culture: the decay of visible markers, the loss of sensible referents, the disintegration of various "standards", excesses affecting the interpretation of phenomena and acculturation successively reaching to physics and geophysics, in order to pursue the astrophysical and cosmogenic quest for "the first moments of the universe", a retrospective inquiry that consists of climbing back through time with the aid of scientific certainties, in order to pose yet again the question of beginnings, the famous Big Bang, the explosion of causality that, according to physicists, was supposed to end tomorrow in a gigantic implosion of finality, a theoretical or meta-theoretical construction capable of saving the matter of the absence of sense, of preserving the creation of a creator, a secret desire for autonomy and for *universal automation* uniting all contemporary apocalyptic trends, this revelation of the precariousness of the human will, this face of hopelessness that is perfectly matched to the degree of ambition among the sciences, this deception in which the idea of nature from the Enlightenment blurs into – and finally becomes confused with – the idea of the real, left over from the century of the speed of light.

Paul Virilio, *The Lost Dimension*

Constant revolutionizing of production, uninterrupted disturbance of all social relations, everlasting uncertainty and agitation, distinguish the bourgeois epoch from all earlier times. All fixed, fast-frozen relationships, with their train of venerable ideas and opinions, are swept away, all new-formed ones become obsolete before they can ossify. All that is solid melts into air, all that is holy is profaned, and men at last are forced to face with sober sense the real conditions of their lives and their relations with their fellow men.

Karl Marx and Friedrich Engels, *The Communist Manifesto*

Even technology is a fable or *Sage*, a transmitted message: when seen in this light it is stripped of all its (imaginary) claims to be able to constitute a new 'strong' reality that could be taken as self-evident, or glorified as what Plato calls the *ontos on*. The myth of a dehumanizing technology, as well as the 'reality' of this myth in a wholly administered and regulated society, are metaphysical accretions which lead us to continue to read the fable as a 'truth'. An accomplished nihilism, like the Heideggerian *Ab-grund*, calls us to a fictionalized experience of reality which is also our only possibility for freedom.

Gianni Vattimo, *The End of Modernity*

Chapter 4

Data Collection and Processing

4.1 The Collection of Macromolecular X-Ray Data

X-ray diffraction data from single crystals of biological macromolecules are commonly collected using standard laboratory sources, synchrotron radiation sources or a combination of the two. The choice of the data collection method is largely dependent on empirical considerations such as the availability of crystals for experimentation, their diffraction limit, the effects of radiation damage on the crystals and the nature of the diffraction experiment undertaken.

Standard laboratory sources fall into two main categories, namely the sealed-tube and rotating-anode generators, both of which utilize the same physical principles in the generation of experimental X-rays. In these generators X-rays are produced when a beam of electrons is accelerated under high voltage and strikes a metal target. Two distinct X-ray spectra are generated. The rapid deceleration of the incident electrons produces a white radiation spectrum which is a function of the accelerating voltage

employed. Superimposed on this is a discontinuous spectra consisting of precisely defined wavelengths produced by the ionization of the target metal atoms through the ejection of electrons from the inner orbital shells. When electrons from higher energy levels move in order to fill the vacancy created, X-ray photons are emitted. This second spectra is known as the characteristic spectra and contains peaks particular to the target material employed. The characteristic lines of this type of spectrum are called K, L and M and correspond to transitions from higher energy orbitals to the K, L and M orbitals i.e. to the orbitals of principle quantum numbers $n = 1, 2$ and 3 . When the two orbitals involved in the transition are adjacent the line is called α , if they are separated by another shell, the line is called β . In the case of copper, the most intense X-ray peak produced by the target material is the $K\alpha$ line and this peak is subsequently monochromated and collimated for laboratory X-ray work. The target of the rotating-anode generator is continuously rotated, as opposed to the sealed-tube where the target is stationary, and this allows the usage of higher voltages and hence the production of a more intense experimental X-ray beam. The rotating-anode generator is preferred in laboratory biological experiments for two main reasons. First of all, macromolecular crystals of protein and/or DNA generally contain intrinsic crystalline disorder which results in weaker and lower resolution diffraction than that found for small compounds. Secondly, the crystals are commonly smaller than small molecule crystals and contain a large amount of solvent. The use of high intensity radiation may optimize the signal for both of these conditions.

One limitation of the laboratory source is that it only produces X-ray radiation of a fixed wavelength which is determined by the characteristic line of the target material.

Copper (Cu) is the target material used in most conventional laboratory generators and the resultant $\text{CuK}\alpha$ radiation has a wavelength of 1.5418\AA . Molybdenum (Mo) is also employed in laboratory sources and produces a beam of $\text{MoK}\alpha$ radiation with a shorter wavelength of 0.7107\AA . Copper radiation is frequently used for the large unit cells of macromolecular crystals since, for any given case, the diffracted intensities observed will be more separated when a longer wavelength is used. Spot overlap may become a problem with shorter wavelength radiation, such as molybdenum. The copper beam produced is also more intense than that for molybdenum. Molybdenum radiation may be used for high-resolution experiments since the separation of the diffracted intensities is less than that for copper, with the volume of reciprocal space accessible in a diffraction experiment being greater for lower X-ray wavelengths. However, the Mo beam is weaker in comparison to that produced by Cu.

Whilst the crystals may survive radiation damage for a sufficient or longer time using a laboratory source, it may be the case that either the observed diffraction is of too low a resolution or that a specific X-ray wavelength is required to exploit the presence of a particular atom in the crystal, as in anomalous dispersion experiments. Despite the greater availability and accessibility of laboratory generators, in such cases a synchrotron source may be well be necessary. Synchrotron radiation facilities are large installations in which electrons are accelerated at relativistic velocities along orbits of hundreds of metres. The intensity of the radiation produced by a synchrotron ring is such that $I \propto 1/R$, where I is the intensity of the radiation generated and R is the radius of the ring. At the Daresbury synchrotron facility, for example, the ring has a diameter of approximately 50 metres.

Because they are moving in circular or near-circular paths, the accelerating electrons, which are moving at near light velocities, emit radiation within the range of the X-ray spectra. The spectrum emitted is very wide and, unlike the laboratory sources discussed above, can be tuned to the wavelength required for a particular experiment. In addition, the brightness of the synchrotron source can be greater than 10^{10} times that of the laboratory rotating-anode generator. Hence, a synchrotron source may permit the recording of X-ray diffraction data inaccessible in the laboratory and thus allow a high-resolution structural study to be undertaken.

Typically macromolecular crystals of biological molecules perish more quickly through radiation damage effects in the more intense synchrotron beam than they would in the milder laboratory source. As a consequence many crystals may be needed for a complete data collection at a synchrotron station. However, the use of a lower wavelength radiation at a synchrotron station may reduce radiation damage. The relative merits of the laboratory and synchrotron sources have to be weighed up in each individual case, depending on the diffraction properties observed for the crystals under study and the particular requirements placed upon the X-ray data.

Preliminary experiments investigating the diffraction quality of crystals of d(TG₄T) were undertaken at York University using the R-Axis IIC imaging plate area detector with an added Rigaku rotating-anode generator and a graphite monochromator. When the crystals were mounted at 4°C and allowed to equilibrate slowly to ambient temperature the quality and extent of the crystal diffraction quickly deteriorated. Data were therefore collected at 4°C using a custom-built crystal-cooling device designed by Dr. Peter Moody. Using CuK α radiation strong diffraction was observed beyond

a resolution of 1.6Å. The subsequent installation of a molybdenum anode revealed diffraction to a limit of 1.0Å resolution, with strong data extending to beyond 1.2Å. The crystals displayed a strong tolerance to radiation damage effects. This tolerance permitted the quick and convenient collection of complete high-resolution data-sets from a single crystal, using the R-AXIS IIC laboratory set-up in York. The collection and processing of the native and derivative data-sets obtained in this manner is detailed in this chapter.

4.2 The Rotation Method

X-ray diffraction data from crystals of d(TG₄T) were collected using the rotation method. This method was introduced in X-ray crystallography in 1926 (Bernal (1926)), where flat film was employed as the medium for the detection of diffracted intensities. The use of the method was not at first wide-spread, however, due to the difficulties involved in indexing and quantitatively measuring the diffracted intensities on the film. The introduction of computer-controlled microdensitometers for accurate film scanning and assessment of diffracted intensities, and the evolution of computational techniques for the indexing of rotation photographs, led to the reintroduction of the technique in the early 1970s in the field of macromolecular crystallography (Arndt and Wonacott (1977)). The rotation, or oscillation, method has since become perhaps the major technique for the quick, efficient and accurate collection of macromolecular X-ray diffraction data.

Data from crystals of d(TG₄T) were not collected using film but rather the R-AXIS

IIC imaging plate area detector. Area detectors have come to replace film as the preferred medium for the recording of X-ray diffraction intensities in macromolecular crystallography. The R-AXIS IIC is discussed briefly in the next section.

In the rotation method a crystal is made to oscillate back and forth through a small angle about an axis perpendicular to the X-ray beam. The resultant diffraction intensities are recorded on a flat film, or area detector, perpendicular to the direction of the incident X-rays. Small rotation angles are required to avoid the overlap of reflections. The recording of successive oscillation images permits the collection of an appropriate asymmetric volume of reciprocal space.

The geometry of the method is illustrated in Figure 4.1, adapted from Giacovazzo *et al.*, (1992). The motion of the crystal is demonstrated leaving the reciprocal lattice planes fixed at the top of the figure and rotating the Ewald sphere about the point **O**. Once the crystal has been oscillated through an angle $\Delta\Phi$, the intersection of the Ewald sphere with each reciprocal lattice plane defines two crescent-shaped lunes per plane, as shown on the bottom of the figure where O_F is the origin of the film or area detector surface. The detector is orientated flat and normal to the incident X-ray beam and remains stationary throughout the motion. Within the lunes are contained all the reciprocal lattice points that will produce diffraction intensities on the detector. All of these diffraction intensities have one index in common since they originate from the same lattice plane.

Lattice points in reciprocal space have a volume. Hence a reflection is only completely recorded when the entire volume crosses the edge of the Ewald sphere. A reflection resulting from a lattice point which has not passed completely through the

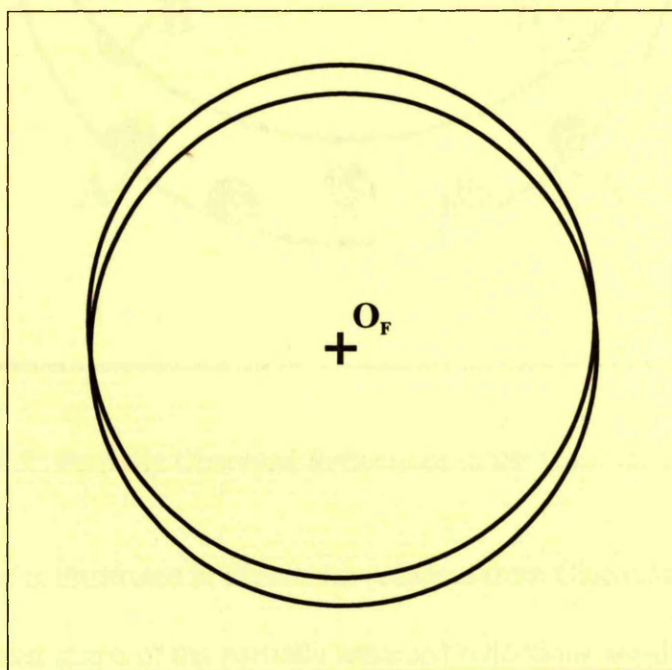
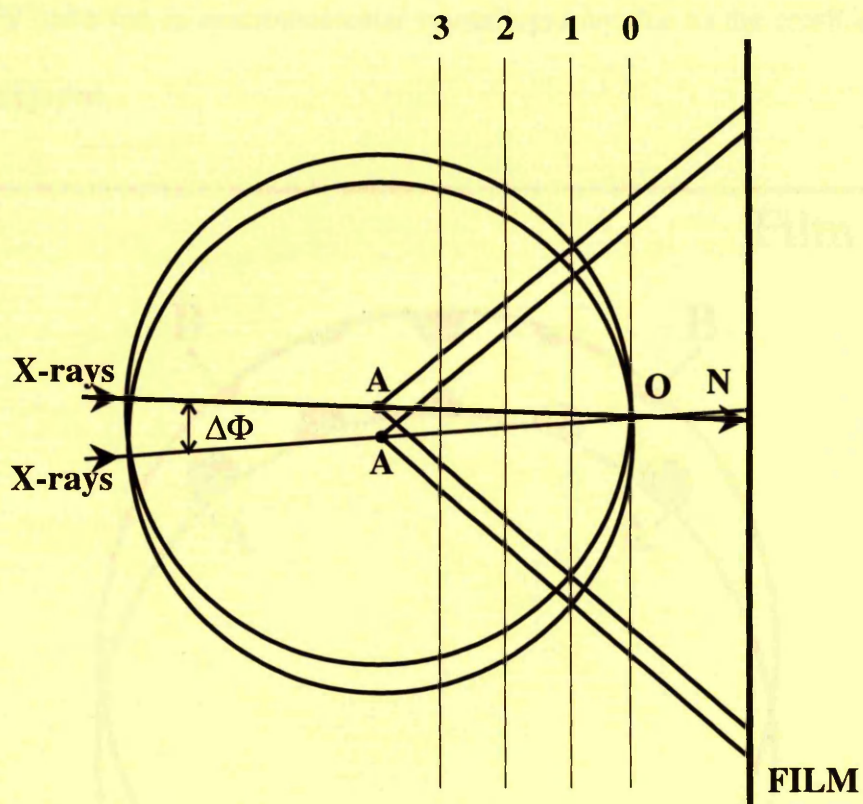


Figure 4.1: The Geometry of the Oscillation Method

Ewald sphere is termed a partially recorded reflection or partial. Such reflections are commonly observed in macromolecular crystallography due to the small oscillation angles employed.

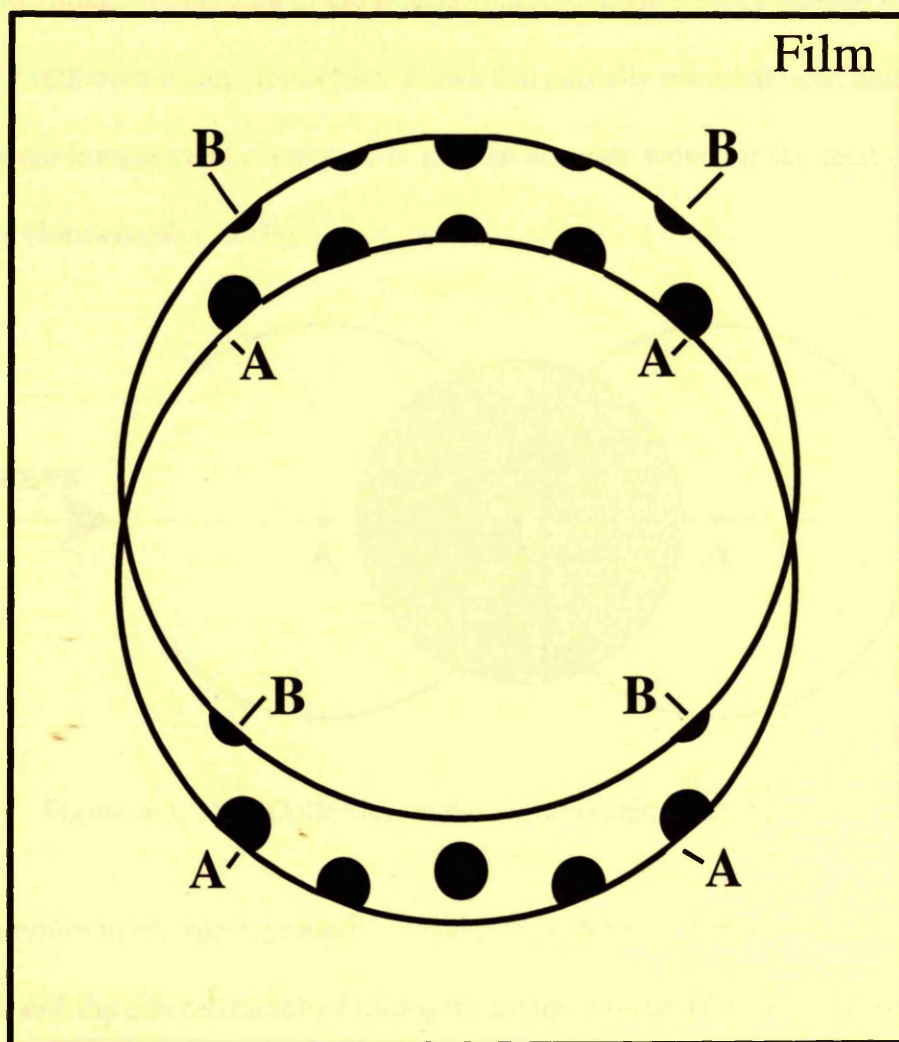


Figure 4.2: Partially Observed Reflections in the Oscillation Method

This situation is illustrated in Figure 4.2 (adapted from Giacovazzo *et al.*, (1992)), where the idealized shape of the partially recorded reflections are shown, as recorded for example on a piece of film, and are labelled **A** and **B**. The idealized shape of the partially recorded reflections is different if they are recorded at the beginning or

the end of the rotation range. The reflections labelled **A** are recorded at one end of the oscillation range and those labelled **B** at the other end. In the instance of the **A** reflections the missing part of the reflection will have been recorded on the previous oscillation image. In the case of the **B** reflections the missing part would be located on the next oscillation image. It has been shown that partially recorded reflections located on different images can be summed to give an accurate value for the total diffracted intensity (Jones *et al.*, (1977)).

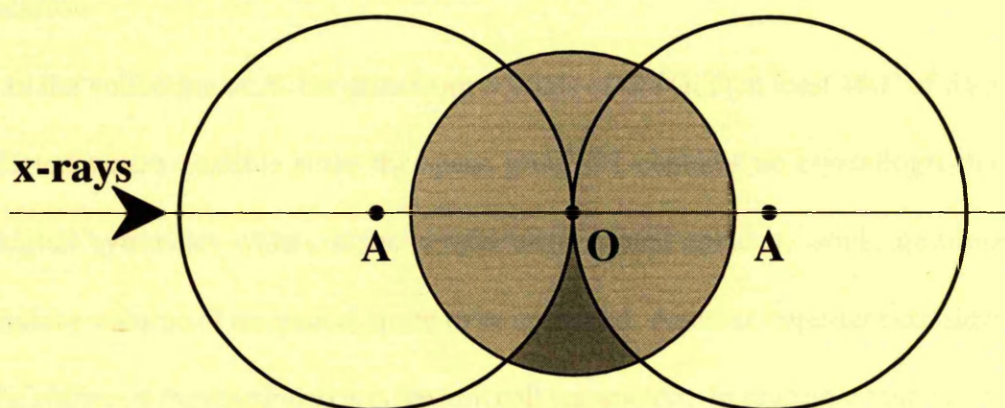


Figure 4.3: Data Collection in the Symmorphic Space Group P1

The symmorphic space group P1 contains no symmetry operations apart from the identity, with the unit cell thereby defining the asymmetric unit for which data have to be collected in order to obtain a complete data-set. Therefore, the collection of a complete data-set in P1 requires a sweep of at least 180° of reciprocal space, as illustrated in Figure 4.3, adapted from Giacovazzo *et al.*, (1992). The sphere centred on **O** represents a limiting sphere which defines the resolution limit imposed on data collection. The contingent spheres with centres **A** represent a total rotation of 180° . The lightly shaded region represents the volume of reciprocal space that has been covered by a sweep of

180°. The darkly shaded region represents the volume of reciprocal space that has not been covered in the 180° sweep. Even if the rotation axis has no symmetry it is never necessary to rotate a crystal much more than 180° in order to collect all of reciprocal space with the exception of the blind region. The blind region takes the shape of the cusp of a torus and it can be collected only by rotating the crystal about a different axis. The volume of the blind region increases the higher the resolution of the data collected. Thus the percentage of data inaccessible to a rotation experiment rises with increasing resolution.

In the collection of X-ray data from crystals of d(TG₄T) at least 180° of data were collected where possible since the space group P1 contains no crystallographic axis of higher symmetry which, if the crystal were rotated about it, would minimize the necessary volume of reciprocal space to be collected. Another important consideration in the choice of the rotation axis is the unit cell parameters. In examining the parameters which limit the maximum rotation range permitted, without incurring overlaps from different reciprocal lattice planes, Wonacott (1977) demonstrated the expression for the maximum rotation range and this is shown in equation 4.1 below.

$$\Delta\Phi_{\max} \leq |r^*|/R_{\max}^* - \Delta \quad (4.1)$$

In this equation $\Delta\Phi_{\max}$ is the maximum allowed rotation range in radians; R_{\max}^* is the maximum resolution in reciprocal space of the data to be collected; r^* is the reciprocal lattice vector pertinent to the axis of oscillation; and Δ is the reflecting range of the crystal. Hence, choosing the axis with the largest unit cell parameter as the

oscillation axis prevents it becoming a limiting factor in the choice of the maximum rotation range. Since the unit cell of the crystals of d(TG₄T) is of a moderate size, with none of the parameters being outstandingly large, data were collected about an oscillation axis with an arbitrary orientation to the unit cell axes in all cases. Generally speaking, however, the crystals were mounted in such a way that the longest physical dimension of the crystal was parallel with the capillary axis. The selection of appropriate oscillation ranges and crystal-to-detector distances was evaluated empirically by the collection of a number of sequential oscillation images. These images were then analysed using the data collection simulation option contained in the R-AXIS software package *PROCESS*. This ensured that the data collection strategy avoided overlapping reflections. Further experimental details are presented for individual data-sets later in this chapter.

4.3 The R-AXIS IIC Imaging Plate Area Detector

Area detectors are now commonly used in macromolecular crystallography. The three area detector types that are currently employed in macromolecular X-ray crystallography are the multiwire proportional counters, the television area detectors and the imaging plates. They combine a high precision photon counting efficiency with the ability to record a large fraction of the reflections which simultaneously cross the Ewald sphere when the rotation method is employed. Data-sets can therefore be collected quickly, minimising the number of crystals used and the radiation damage incurred in the collection of a high-resolution X-ray data-set from a macromolecular crystal.

The performance characteristics of the first two detector types have been studied and documented by Arndt (1986), where the newly emerging image plate technology is also discussed.

The R-AXIS IIC is an example of the imaging plate area detector type. The imaging plate is in fact a phosphorimager device. The plate is covered by a photostimulable material consisting of Ba F Br:Eu²⁺ crystals. When incident X-ray photons hit the plate some of the Eu²⁺ ions are ionized to Eu³⁺, with the freed electrons becoming trapped in Br vacancies introduced in the crystals. The resultant latent image produced by the incident X-ray diffraction intensities can be read when the plate is exposed to a He-Ne laser of wavelength 633 nm. The electrons in the Br vacancies are thus excited and return to the Eu³⁺ ions to produce excited Eu²⁺ ions. The electronic transition to the stable Eu²⁺ ion state produces light of a 390 nm wavelength in a process called photostimulated luminescence. The intensity of this emitted light is proportional to that of the incident X-rays. An image reader collects the emitted radiation and converts it into a electrical signal by means of a photomultiplier tube.

Within the body of the R-AXIS IIC there are contained two identical, interchangeable image plates. Each plate measures 200 mm by 200 mm so that the effective circular area available has a radius of 100 mm. Data can be collected in the corners of the plate but will be less complete than than collected in the disk.

Consequently one plate may be used to collect X-ray diffraction data while the other plate is being scanned by the image reader. Once a plate has been read it can be reused again as it is restored to its original state by exposure to visible radiation. This places a limitation on the time taken in collecting a data-set in so far as the

process of reading and restoring the plate for reuse requires 10 minutes in total. Hence exposures of less than 10 minutes still require a total period of ten minutes before the next exposure can begin.

4.4 Data Processing of R-AXIS Data Using DENZO

All native and derivative data-sets were collected using the R-AXIS IIC imaging plate in York University at 4°C. Derivative data-sets are those collected from crystals soaked or co-crystallized with heavy-atom compounds, as distinct from native crystals. The preparation of derivatives is described in Chapter 3. After processing, derivative data-sets were scaled relative to the native data and difference Pattersons calculated for evaluation. This procedure is described in Chapter 5. The data collection strategy has been detailed in the previous section. Parameters specific to each data-set are presented later in this section.

4.4.1 Data Processing With DENZO

That a single oscillation image contains all of the necessary information to derive the diffraction intensities from that image was first successfully demonstrated by Rossman (1979). The program *DENZO*, written by Zbyszek Otwinowski, integrates reflections from single crystal diffraction data measured on film or phosphofluorescence imaging plate detectors, collected using the oscillation, Weissenberg or precession methods (Otwinowski (1993a); Gewirth (1993)). The program takes the raw X-ray diffraction data and reduces them to a file containing the *hkl* indices and background and Lorentz-

Polarization (LP) corrected intensity of the spots on the image, along with an estimate of the error.

The four steps involved in the analysis of single crystal diffraction data using the *DENZO* package are:

1. Visualization of the original, unprocessed detector data.
2. Indexing of the diffraction pattern.
3. Refinement of the crystal and detector parameters.
4. Integration of the diffraction maxima.

DENZO requires the input of detector and X-ray parameters such as the wavelength, the crystal-to-detector distance, the precise co-ordinates of the direct beam on the imaging plate, the crystal and cassette missetting angles, the non-orthogonality of the scanner head motions and the scale of the y pixels to the x pixels of the scanned image, all of which are known prior to processing. The program also contains an autoindexing procedure to determine the exact unit cell dimensions and the precise crystal and detector orientation parameters (Otwinowski (1993a)).

The autoindexing option was not available during the processing of the d(TG₄T) data. Instead, accurate cell dimensions and orientation matrices were determined using the program *REFIX* (Kabsch (1988, 1993)), and these were then input into the program *DENZO*. In this procedure three one degree oscillations were analysed by the program *IMSTILLS* which produced an output of spot co-ordinates and intensities which were then input into *REFIX*. The quality of the solution was then evaluated by means of

the positional residual (standard deviation of spot position) which ought to be below 0.2 mm for the imaging plate detector. The correctness of the initial crystal, detector and X-ray parameters was evaluated further using the *DENZO* interactive display which permitted comparison of the X-ray diffraction oscillation image with the pattern predicted by the forementioned parameters.

The next step in the data processing procedure involved the interactive refinement of the initial crystal and detector parameters for the chosen image, as well as the definition of the spot and background size and shape and crystal mosaicity. This strategy was interactive and completely user-controlled, involving an iterative procedure whereby the resolution limits and the parameters to be refined within those limits were defined.

Parameters are fitted using a fast-converging least-squares procedure. *DENZO* permits the simultaneous refinement of crystal and detector parameters which can be unstable, particularly at low-resolution (Otwinowski (1993a)). An eigenvalue filtering procedure automatically becomes active should high correlation of parameters make the refinement unstable. Should this occur, the most correlated parameters are removed from the refinement.

Typically, the strategy employed involved the refinement of a few individual parameters at low-resolution, with the addition of more parameters at high-resolution where more could be fitted simultaneously. Once a satisfactory strategy has been derived the images were then submitted to be processed as a batch job.

For each image, *DENZO* refines all parameters using only the data from that single image. The diffraction maxima are integrated using a profile fitting procedure (Otwinowski (1993a)). A final file is produced for each image containing the *hkl*

indices, the background and LP corrected intensities and an estimate of the error.

4.4.2 Data Reduction and Scaling Using ROTAVATA/AGROVATA

Zbyszek Otwinowski has provided his own companion program *SCALEPACK* for the scaling of overlapping batches of data output by *DENZO* (Otwinowski (1993b)). However, due to the unavailability of this program at the time the d(TG₄T) data were collected, the native and derivative data-sets were scaled and reduced using the *CCP4* suite of programs *ROTAVATA* and *AGROVATA*.

The individual image files output from *DENZO* were compiled into a multirecord *MTZ* file suitable for input to *ROTAVATA* using the *CCP4* program *ROTAPREP*. Individual files were then merged and sorted using the *CCP4* program *SORTMTZ*.

ROTAVATA calculates scale factors and/or temperature factors between overlapping batches of data, using the method of Fox and Holmes (1966). In this scaling procedure, the *n*th scale factor is provided by the expression in equation 4.2.

$$K_n = C_n \exp \left[-2B_n \sin^2 \left(\frac{\theta}{\lambda} \right) \right] \quad (4.2)$$

In the expression given in equation 4.2, C_n and B_n are the two parameters that are varied for each individual batch. The temperature factor functions as a correction for radiation damage, in particular to compensate for the greater damage experienced by the high-resolution reflections. This temperature factor should not be confused with the proper use of the term as applied to atoms within molecular models, as discussed later in Chapter 6. A single image is chosen and appointed a scale factor of 1.0.

The scale factors of the other images are then calculated relative to the chosen image. Temperature and scale factors are refined simultaneously over a specified number of iterative cycles.

The final scale factors calculated for each image were then input to the program *AGROVATA* which applies the scale factors and averages the resultant X-ray diffraction intensities. Partially recorded reflections are summed and the data are monitored for the rejection of bad agreements between repeated measurements or symmetry equivalents. The final intensities are then averaged and output to an *MTZ* file. A high redundancy in the data is desirable for the purposes of the rejection criteria and also improves the quality of the final averaged data by averaging out systematic errors.

4.4.3 Native Data

The first native data collected were obtained using $\text{CuK}\alpha$ radiation at a temperature of 4°C . The crystal to film distance was set at 75 mm and 2° oscillation images with an exposure time of 10 minutes per image were collected. Two crystals were used in the collection of the data and 90 images, or 180° of data, were collected for each crystal. Strong diffraction spots were observed to 1.6\AA at the edge of the image plate. Both crystals were processed separately using *DENZO* and then scaled together using *ROTAVATA*. The final data-set contains 24,103 unique reflections with 87.0% of the reflections having $I > 3\sigma(I)$. The data is 72.0% complete to 1.58\AA with a final R_{merge} of 6.4%. Details are given in Table 4.1.

The installation of a molybdenum anode permitted the extension of the resolution of the initial data-set described above. Data were collected from one crystal using 1.7°

Dmin(Å)	Nmeas	Nuniq	% Poss.	Mult.	R _{merge}	R _{cum}	I>3σ(I)	$\bar{I}/\sigma(\bar{I})$
4.74	4,394	1,057	99.8	4.1	0.049	0.049	98.0	12.5
3.35	8,312	1,938	100.0	4.2	0.060	0.057	98.4	13.3
2.74	10,702	2,462	98.2	4.3	0.056	0.056	97.5	9.1
2.37	12,452	2,870	96.7	4.3	0.067	0.058	95.9	11.7
2.12	13,446	3,274	97.2	4.1	0.079	0.060	92.8	9.9
1.94	13,213	3,539	95.1	3.7	0.087	0.062	89.8	9.1
1.79	14,164	3,799	93.9	3.7	0.104	0.063	81.7	7.4
1.68	11,446	3,502	80.5	3.3	0.124	0.064	72.8	6.5
1.58	4,458	1,662	35.9	2.7	0.136	0.064	60.6	5.2
Total	92,587	24,103	72.0	3.8	–	0.064	87.0	8.2

Dmin – Resolution limit of shell

Nmeas – Number of measurements within resolution shell used for data reduction

Nuniq – Number of unique reflections within resolution shell

% Poss – Percentage of unique data after merging

Mult – Multiplicity of reflections

$$R_{\text{merge}} = \frac{\sum_h \sum_i |I(h)_i - \langle I(h) \rangle|}{\sum_h \sum_i | \langle I(h)_i \rangle |}$$

R_{cum} – **R_{merge}** up to this resolution range

I>3σ(I) – Percentage of reflections within resolution shell with intensity greater than 3σ

$\bar{I}/\sigma(\bar{I})$ – Ratio of the average intensity and the standard deviation of the average intensity in this resolution range

Table 4.1: Scaling Statistics for *DENZO* Processed CuKα Data

Dmin(Å)	Nmeas	Nuniq	% Poss.	Mult.	R _{merge}	R _{cum}	I>3σ(I)	$\bar{I}/\sigma(\bar{I})$
3.16	19,796	3,604	99.9	5.5	0.041	0.041	99.0	9.4
2.24	36,633	6,583	99.9	5.6	0.052	0.045	97.1	9.4
1.83	46,617	8,516	99.8	5.6	0.070	0.050	93.4	8.5
1.58	38,766	10,045	99.4	3.9	0.089	0.053	83.0	6.9
1.41	20,918	11,117	97.0	1.9	0.089	0.054	62.7	7.4
1.29	23,093	12,237	96.6	1.9	0.161	0.055	49.0	4.6
1.20	24,851	13,162	95.5	1.9	0.224	0.056	40.0	3.4
1.12	26,457	14,058	95.0	1.9	0.341	0.058	27.5	2.2
1.05	27,656	14,870	94.4	1.9	0.582	0.060	14.6	1.3
1.00	20,015	11,690	70.2	1.7	0.893	0.061	7.3	0.8
Total	285,802	105,882	92.9	2.7	–	0.061	48.5	5.0

Dmin – Resolution limit of shell

Nmeas – Number of measurements within resolution shell used for data reduction

Nuniq – Number of unique reflections within resolution shell

% Poss – Percentage of unique data after merging

Mult – Multiplicity of reflections

$$R_{\text{merge}} = \frac{\sum_h \sum_i |I(h)_i - \langle I(h) \rangle|}{\sum_h \sum_i |\langle I(h)_i \rangle|}$$

R_{cum} – **R_{merge}** up to this resolution range

I>3σ(I) – Percentage of reflections within resolution shell with intensity greater than 3σ

$\bar{I}/\sigma(\bar{I})$ – Ratio of the average intensity and the standard deviation of the average intensity in this resolution range

Table 4.2: Scaling Statistics for *DENZO* Processed CuKα & MoKα Data

oscillations with an exposure time of 40 minutes for each image. The crystal to film distance was set to 115 mm. Weak 1.0Å spots could be seen at the plate edge. A total of 107 images were collected in this manner. However, the crystal perished dramatically during frame 103 and frames 103–107 could not be processed, resulting in the collection of 173.5° of data. After processing, the integrated images from the molybdenum crystal diffraction were scaled along with the 180 images collected using the copper anode and then reduced. The final native data consists of 285,862 observations of which 64,156 are partials, resulting in 105,882 unique recorded reflections with an overall R_{merge} of 6.1% to 1.0Å. Diffraction was observed to 1.0Å, with the data 92.9% complete. Of these data measurements, 48.5% of the reflections have $I > 3\sigma(I)$ to this limit and a total of 51.5% with $I \leq 3\sigma(I)$. The 51.5% of the data with $I \leq 3\sigma(I)$ breaks down to 42.1% of the data with $I \leq 2\sigma(I)$ and 29.2% of the data with $I \leq 1\sigma(I)$. The observed data is, however, very weak beyond 1.2Å resolution. Consequently, measurements beyond 1.2Å resolution were not used in further applications of the data. Details of the complete native data are presented in Table 4.2.

4.4.4 K_2PtCl_4 Derivative

The soaking conditions for this derivative are as for Experiment no. 1 in Table 3.3. One crystal was used to collect 180 frames of data, using $\text{CuK}\alpha$ radiation, at a temperature of 4°C. The crystal to detector distance was set at 75 mm and 2° oscillation were collected with 10 minute exposure times. As with the native crystals, strong diffraction was observed to 1.6Å at the plate edge. A total of 93,680 measurements were merged, of which 18,647 were partial, to give a final R_{merge} of 4.7% and 23,420

Dmin(Å)	Nmeas	Nuniq	% Poss.	Mult.	R _{merge}	R _{cum}	R _{anom}	I>3σ(I)	$\bar{I}/\sigma(\bar{I})$
4.74	4,279	1,046	98.9	4.1	0.032	0.032	0.022	99.7	12.8
3.35	7,725	1,904	98.6	4.1	0.034	0.034	0.022	99.3	13.8
2.74	9,797	2,398	95.9	4.1	0.043	0.037	0.031	97.8	8.9
2.37	11,527	2,829	95.9	4.1	0.047	0.038	0.038	96.5	11.6
2.12	12,906	3,162	94.2	4.1	0.053	0.040	0.045	93.4	10.1
1.94	14,091	3,450	92.9	4.1	0.061	0.042	0.050	91.4	9.4
1.79	14,895	3,677	91.1	4.1	0.073	0.044	0.061	86.1	7.9
1.68	11,589	3,062	70.6	3.8	0.093	0.046	0.073	80.5	6.6
1.58	5,525	1,488	32.2	3.7	0.121	0.047	0.105	71.7	5.5
1.50	1,342	404	8.3	3.3	0.200	0.047	0.172	59.2	3.8
Total	93,676	23,420	70.1	4.0	–	0.047	0.036	89.9	8.5

Dmin – Resolution limit of shell

Nmeas – Number of measurements within resolution shell used for data reduction

Nuniq – Number of unique reflections within resolution shell

% Poss – Percentage of unique data after merging

Mult – Multiplicity of reflections

$$R_{\text{merge}} = \frac{\sum_h \sum_i |I(h)_i - \langle I(h) \rangle|}{\sum_h \sum_i |\langle I(h)_i \rangle|}$$

R_{cum} – **R_{merge}** up to this resolution range

$$R_{\text{anom}} = \frac{\sum |\langle I+ \rangle - \langle I- \rangle|}{\sum (\langle I+ \rangle + \langle I- \rangle)}$$

I>3σ(I) – Percentage of reflections within resolution shell with intensity greater than 3σ

$\bar{I}/\sigma(\bar{I})$ – Ratio of the average intensity and the standard deviation of the average intensity in this resolution range

Table 4.3: Scaling Statistics for *DENZO* Processed K₂PtCl₄ Derivative

unique measurements. Anomalous measurements were made for 23,168 reflections, with only 3,253 having anomalous differences greater than 3σ . The data-set is 70.1% complete to 1.5\AA and 89.9% of the reflections have $I > 3\sigma(I)$.

Complete details for the K_2PtCl_4 derivative are given in Table 4.3. In addition to the quantity R_{merge} , the index R_{anom} is also reported in this table. The R_{anom} index serves as an indication of the possible breakdown of Friedel's law (Blundell and Johnson (1976)) as a result of the presence of an anomalous scattering contribution from the heavy-atom(s) within the crystal. Friedel's law states that $I(hkl) = I(\bar{h}\bar{k}\bar{l})$ for a reflection hkl and its centrosymmetrically related partner $\bar{h}\bar{k}\bar{l}$, where \bar{h} , \bar{k} , \bar{l} refer to negative indices. In the expression for R_{anom} in Table 4.3, the quantities $\langle I+ \rangle$ and $\langle I- \rangle$ represent the intensities of the Friedel pairs defined above. The importance of anomalous scattering information in the derivation of phase information towards a structure solution is discussed further in Chapter 5.

4.4.5 $\text{Pb}(\text{CH}_3\text{CO}_2)_2$ Derivative

The soaking conditions for the $\text{Pb}(\text{CH}_3\text{CO}_2)_2$ derivative are as described in Table 3.3, Experiment no. 4. One crystal was used to collect 180° of data, using $\text{CuK}\alpha$ radiation at a temperature of 4°C . Images were collected using 2° oscillations and 10 minute exposures. The crystal to detector distance was set to 79 mm and strong diffraction was evident to 1.8\AA at the plate edge. The 35,312 individual measurements, of which 377 were partials, were merged resulting in an R_{merge} of 4.6%. The data-set consists of 18,036 unique reflections, with 92.9% completeness to 1.8\AA and 86.1% of the reflections having intensities greater than 3σ . A total of 12,482 reflections had

Dmin(Å)	Nmeas	Nuniq	% Poss.	Mult.	R _{merge}	R _{cum}	R _{anom}	I>3σ(I)	$\bar{I}/\sigma(\bar{I})$
5.68	1,073	543	88.0	2.0	0.028	0.028	0.028	98.7	18.1
4.02	2,161	1,105	98.4	2.0	0.029	0.029	0.036	98.5	17.2
3.29	2,770	1,417	97.5	2.0	0.036	0.033	0.037	97.9	15.1
2.85	3,247	1,650	95.8	2.0	0.051	0.038	0.050	95.7	8.6
2.55	3,637	1,839	94.2	2.0	0.041	0.038	0.062	94.1	13.0
2.32	4,010	2,033	94.2	2.0	0.048	0.040	0.061	91.3	13.8
2.15	4,310	2,202	93.8	2.0	0.049	0.041	0.073	86.2	13.1
2.01	4,518	2,313	91.7	2.0	0.057	0.042	0.089	84.6	11.6
1.90	4,745	2,433	90.6	2.0	0.073	0.045	0.108	75.8	9.1
1.80	4,711	2,501	88.1	2.0	0.085	0.046	0.141	66.1	8.1
Total	35,182	18,036	92.9	2.0	–	0.046	0.049	86.1	9.8

Dmin – Resolution limit of shell

Nmeas – Number of measurements within resolution shell used for data reduction

Nuniq – Number of unique reflections within resolution shell

% Poss – Percentage of unique data after merging

Mult – Multiplicity of reflections

$$R_{\text{merge}} = \frac{\sum_h \sum_i |I(h)_i - \langle I(h) \rangle|}{\sum_h \sum_i |I(h)_i|}$$

R_{cum} – **R_{merge}** up to this resolution range

$$R_{\text{anom}} = \frac{\sum | \langle I+ \rangle - \langle I- \rangle |}{\sum (\langle I+ \rangle + \langle I- \rangle)}$$

I>3σ(I) – Percentage of reflections within resolution shell with intensity greater than 3σ

$\bar{I}/\sigma(\bar{I})$ – Ratio of the average intensity and the standard deviation of the average intensity in this resolution range

Table 4.4: Scaling Statistics for *DENZO* Processed Pb(CH₃CO₂)₂ Derivative

anomalous measurements, with only 1,575 reflections having anomalous differences greater than 3σ . The data are described in Table 4.4.

4.4.6 Methyl Iodine (Me-I) Derivative

The preparation of this derivative is described in Table 3.3, Experiment no. 10, and in Section 3.4. The crystal to detector distance was set to 90 mm and data were collected using $\text{CuK}\alpha$ radiation and a temperature of 4°C . At this setting diffraction was observed to 2.0\AA at the edge of the image plate. Using 2° oscillations and 10 minute exposures, 360 degrees of data were collected from one crystal. The 360 frames were scaled and reduced to give 53,149 measurements, of which 11,005 were partials. The final R_{merge} is 3.3%, and the data consists of 13,543 unique reflections, with 95.1% completeness and 97.9% of reflections with $I > 3\sigma(I)$ to 2.0\AA . Anomalous measurements were recorded for 13,503 reflections, with only 1,626 having anomalous differences greater than 3σ . The data are described in Table 4.5.

4.4.7 Molecular Iodine (I_2) Derivative

The preparation of this derivative is described in Table 3.3, Experiment no. 9, and in Section 3.4. Again, data were obtained using $\text{CuK}\alpha$ radiation at a temperature of 4°C . Images were collected using 2° oscillations, 10 minute exposures and a crystal to detector distance of 90 mm. A total of 90 images, or 180° of data, were collected. The 24,323 reflections, of which 7,013 were partials, were merged to give 13,382 unique reflections with an R_{merge} of 7.2%. The data are 94.1% complete to 2.0\AA , with 94.5% of the data having intensities greater than 3σ . Anomalous measurements were recorded

Dmin(Å)	Nmeas	Nuniq	% Poss.	Mult.	R _{merge}	R _{cum}	R _{anom}	I>3σ(I)	$\bar{I}/\sigma(\bar{I})$
6.31	1,714	448	98.9	3.8	0.018	0.018	0.023	100.0	26.1
4.47	3,093	800	97.1	3.9	0.028	0.025	0.024	99.6	13.8
3.65	4,137	1,067	100.0	3.9	0.028	0.026	0.026	99.7	17.8
3.16	4,726	1,218	96.5	3.9	0.033	0.029	0.029	98.9	10.5
2.83	5,361	1,382	96.5	3.9	0.035	0.030	0.032	99.0	11.7
2.58	5,918	1,525	96.3	3.9	0.032	0.030	0.036	98.6	16.6
2.39	6,314	1,624	94.3	3.9	0.036	0.031	0.039	98.2	14.5
2.24	6,759	1,740	94.1	3.9	0.040	0.031	0.043	96.6	12.7
2.11	7,155	1,846	93.8	3.9	0.040	0.032	0.045	96.7	14.1
2.00	7,266	1,893	91.0	3.8	0.044	0.033	0.048	95.7	12.9
Total	52,443	13,543	95.1	3.9	–	0.033	0.033	97.9	10.9

Dmin – Resolution limit of shell

Nmeas – Number of measurements within resolution shell used for data reduction

Nuniq – Number of unique reflections within resolution shell

% Poss – Percentage of unique data after merging

Mult – Multiplicity of reflections

$$R_{\text{merge}} = \frac{\sum_h \sum_i |I(h)_i - \langle I(h) \rangle|}{\sum_h \sum_i |\langle I(h)_i \rangle|}$$

R_{cum} – **R_{merge}** up to this resolution range

$$R_{\text{anom}} = \frac{\sum |\langle I+ \rangle - \langle I- \rangle|}{\sum (\langle I+ \rangle + \langle I- \rangle)}$$

I>3σ(I) – Percentage of reflections within resolution shell with intensity greater than 3σ

$\bar{I}/\sigma(\bar{I})$ – Ratio of the average intensity and the standard deviation of the average intensity in this resolution range

Table 4.5: Scaling Statistics for *DENZO* Processed Methyl Iodine Derivative

Dmin(Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{merge}	R_{cum}	R_{anom}	I>3σ(I)	$\bar{I}/\sigma(\bar{I})$
6.31	814	441	97.6	1.8	0.171	0.171	0.036	99.1	2.1
4.47	1,444	792	96.2	1.8	0.047	0.066	0.037	98.5	8.7
3.65	1,870	1,038	97.4	1.8	0.124	0.099	0.044	98.2	1.7
3.16	2,180	1,213	96.2	1.8	0.057	0.082	0.050	98.8	6.6
2.83	2,472	1,360	95.1	1.8	0.093	0.083	0.050	96.8	1.3
2.58	2,704	1,519	96.0	1.8	0.049	0.077	0.050	96.0	5.6
2.39	2,865	1,592	92.6	1.8	0.060	0.075	0.055	94.8	8.9
2.24	3,098	1,723	93.3	1.8	0.057	0.073	0.061	91.6	8.0
2.11	3,318	1,832	93.2	1.8	0.062	0.072	0.068	90.7	7.0
2.00	3,332	1,872	90.1	1.8	0.064	0.072	0.072	90.0	7.2
Total	24,097	13,382	94.1	1.8	–	0.072	0.049	94.5	2.0

Dmin – Resolution limit of shell

Nmeas – Number of measurements within resolution shell used for data reduction

Nuniq – Number of unique reflections within resolution shell

% Poss – Percentage of unique data after merging

Mult – Multiplicity of reflections

$$R_{\text{merge}} = \frac{\sum_h \sum_i |I(h)_i - \langle I(h) \rangle|}{\sum_h \sum_i |\langle I(h)_i \rangle|}$$

R_{cum} – **R_{merge}** up to this resolution range

$$R_{\text{anom}} = \frac{\sum |\langle I+ \rangle - \langle I- \rangle|}{\sum (\langle I+ \rangle + \langle I- \rangle)}$$

I>3σ(I) – Percentage of reflections within resolution shell with intensity greater than 3σ

$\bar{I}/\sigma(\bar{I})$ – Ratio of the average intensity and the standard deviation of the average intensity in this resolution range

Table 4.6: Scaling Statistics for *DENZO* Processed I₂ Derivative

on 8,499 reflections, with only 1,052 having anomalous differences greater than 3σ . The data statistics are given in Table 4.6.

4.4.8 $\text{UO}_2(\text{CH}_3\text{CO}_2)_2$ Derivative

The preparation conditions for the $\text{UO}_2(\text{CH}_3\text{CO}_2)_2$ derivative are described in Table 3.3, Experiment no. 5. The crystal to detector distance was set to 79 mm and 2° oscillations were collected with an exposure time of 10 minutes. $\text{CuK}\alpha$ radiation was used along with a temperature of 4°C . The intensity of the observed diffraction was weaker than observed for the other crystals investigated due to the small size of the crystal. The first 40° of data were able to be processed but, thereafter, the quality of the diffraction rapidly deteriorated and the crystal died. A total of 7,434 measurements were recorded, of which 2,213 were partial. The available data were merged with an R_{merge} of 6.4% to give 7,114 unique reflections. The data are only 36.7% complete to 1.8\AA with 57% of the data having $I > 3\sigma(I)$. The data statistics are given in Table 4.7.

4.4.9 Other Derivatives

Two other derivatives were screened with a view to data collection. $\text{CuK}\alpha$ radiation and a temperature of 4°C were used. The EuCl_3 co-crystallization, described in Table 3.3, Experiment no. 6, exhibited very weak diffraction and died shortly after being placed in the X-ray beam. Data were collected on the CsCl_2 soaked crystal, the preparation of which is described in Table 3.3, Experiment no. 8. Diffraction was observed to 2.5\AA but was very weak beyond this limit. Using 2° oscillations, 180° of data were collected. However, the diffraction pattern was very "streaky" and the spots exhibited

Dmin(Å)	Nmeas	Nuniq	% Poss.	Mult.	R_{merge}	R_{cum}	I>3σ(I)	$\bar{I}/\sigma(\bar{I})$
5.68	237	184	29.8	1.3	0.048	0.048	91.8	9.5
4.02	457	379	33.8	1.2	0.077	0.070	91.6	4.4
3.29	572	517	35.6	1.1	0.040	0.060	89.7	11.9
2.85	662	613	35.6	1.1	0.077	0.061	79.0	7.6
2.55	768	734	37.6	1.0	0.133	0.063	75.1	3.8
2.32	839	820	38.0	1.0	0.102	0.064	66.7	6.4
2.15	900	894	38.1	1.0	0.056	0.064	57.0	12.3
2.01	955	952	37.8	1.0	0.223	0.064	45.2	2.8
1.90	1,018	1,009	37.6	1.0	0.388	0.064	34.5	1.5
1.80	1,014	1,012	35.7	1.0	0.180	0.064	20.5	4.1
Total	7,422	7,114	36.7	1.0	–	0.064	57.0	4.7

Dmin – Resolution limit of shell

Nmeas – Number of measurements within resolution shell used for data reduction

Nuniq – Number of unique reflections within resolution shell

% Poss – Percentage of unique data after merging

Mult – Multiplicity of reflections

$$R_{\text{merge}} = \frac{\sum_h \sum_i |I(h)_i - \langle I(h) \rangle|}{\sum_h \sum_i |\langle I(h)_i \rangle|}$$

R_{cum} – **R_{merge}** up to this resolution range

I>3σ(I) – Percentage of reflections within resolution shell with intensity greater than 3σ

$\bar{I}/\sigma(\bar{I})$ – Ratio of the average intensity and the standard deviation of the average intensity in this resolution range

Table 4.7: Scaling Statistics for *DENZO* Processed $\text{UO}_2(\text{CH}_3\text{CO}_2)_2$ Derivative

a high mosaicity. Calculation of the orientation matrix and the cell dimensions using *IMSTILLS* and *REFIX* produced tolerable values, although the positional residuals were very high at 1.4mm. Consequently, despite many attempts to optimise the crystal parameters in *DENZO*, the data could not be processed.

4.5 Comparison of Native Data-Sets

The native data collected using $\text{CuK}\alpha$ and $\text{MoK}\alpha$ radiation were also processed by Dr. Peter Moody of York University using the *MOSFLM* package (Leslie (1990)). The resultant integrated intensities were then scaled and merged using the *CCP4* packages *ROTAVATA* and *AGROVATA*, as previously described. The statistics for the *MOSFLM* processed native data are presented in Table 4.8.

The structure factors for both the *DENZO* and *MOSFLM* processed native data-sets were calculated using the *CCP4* program *TRUNCATE* (French and Wilson (1978)). Comparison of Table 4.8 with the *DENZO* processed native data in Table 4.2 indicates that both data-sets are of roughly equivalent quality, with individual R_{merge} values of 0.061 for the *DENZO* processed data and 0.058 for the *MOSFLM* processed data. Scaling together of the two individual data-sets using the *CCP4* program *SCALEIT* produced an R_{merge} of 0.066. The quality of the individual data-sets can be evaluated further by inspection of the Wilson plots for the two data-sets.

The construction of a Wilson plot is based upon the assumption that the atoms in a unit cell are randomly distributed. This assumption is not true for a low-resolution structure and is only partially true for a high-resolution structure, but it is still useful

Dmin(Å)	Nmeas	Nuniq	% Poss.	Mult.	R _{merge}	R _{cum}	I>3σ(I)	$\bar{I}/\sigma(\bar{I})$
3.16	17,637	3,525	99.0	5.0	0.039	0.039	99.7	10.4
2.24	33,406	6,525	100.0	5.1	0.051	0.044	98.1	9.3
1.83	40,938	8,402	99.4	4.9	0.067	0.048	94.8	8.5
1.58	34,573	9,795	97.9	3.5	0.095	0.051	85.3	6.1
1.41	19,428	10,555	93.0	1.8	0.113	0.052	65.8	6.0
1.29	20,246	11,589	92.3	1.7	0.164	0.053	52.8	4.3
1.20	21,726	12,468	91.3	1.7	0.228	0.054	43.2	3.1
1.12	23,309	13,365	91.1	1.7	0.316	0.056	30.9	2.2
1.05	24,605	14,063	90.1	1.7	0.521	0.057	17.8	1.4
1.00	10,277	6,293	38.1	1.6	0.736	0.058	10.2	1.0
Total	246,145	96,590	85.6	2.5	–	0.058	53.8	5.4

Dmin – Resolution limit of shell

Nmeas – Number of measurements within resolution shell used for data reduction

Nuniq – Number of unique reflections within resolution shell

% Poss – Percentage of unique data after merging

Mult – Multiplicity of reflections

$$R_{\text{merge}} = \frac{\sum_h \sum_i |I(h)_i - \langle I(h) \rangle|}{\sum_h \sum_i |\langle I(h)_i \rangle|}$$

R_{cum} – **R_{merge}** up to this resolution range

I>3σ(I) – Percentage of reflections within resolution shell with intensity greater than 3σ

$\bar{I}/\sigma(\bar{I})$ – Ratio of the average intensity and the standard deviation of the average intensity in this resolution range

Table 4.8: Scaling Statistics for *MOSFLM* Processed CuKα & MoKα Data

for placing the data on an absolute scale. The theory of Wilson (1949) states that the relationship expressed by equation 4.3 holds where K_p is a scale factor; $\langle I_p \rangle$ is the average values of the observed intensities within the appropriate range of $\sin^2 \theta / \lambda$; $\sum f_j^2$ is the sum of the squares of the scattering factors of the atoms in the macromolecule; and B_p is the overall isotropic temperature factor.

$$K_p \langle I_p \rangle = \sum f_j^2 \exp \left[-2B_p \left(\frac{\sin^2 \theta}{\lambda} \right) \right] \quad (4.3)$$

Thus a plot of $\ln \langle I_p \rangle / \sum f_j^2$ against $\sin^2 \theta / \lambda^2$ permits the calculation of an approximate absolute scale and temperature factor, giving amplitudes of a sensible magnitude for further calculations. This is achieved by fitting a least-squares line through the Wilson plot.

Two plots are presented, one for each of the native data-sets. The Wilson plot for the *DENZO* processed data is shown in Figure 4.4, and the corresponding plot for the *MOSFLM* processed data is shown in Figure 4.5. In both Figure 4.4 and Figure 4.5 the line labelled *Observed Reflections* uses the average $\langle I_p \rangle$ from the number of observed reflections in a hemisphere of reciprocal space. If the reflections that were not measured were all weak then $\langle I_p \rangle$ is more accurately estimated if the average is derived using the number of possible reflections. This is represented in the figures by the line labelled *Possible Reflections*.

The results of the Wilson scaling are $K=0.00967$, $B=7.66336$ for the *DENZO* processed data, and $K=0.01700$, $B=7.99331$ for the *MOSFLM* processed data. Inspection of Figures 4.4 and 4.5 indicates that the higher resolution data produces a more linear

plot that the lower resolution data, illustrating the breakdown of the Wilson assumption at lower resolutions.

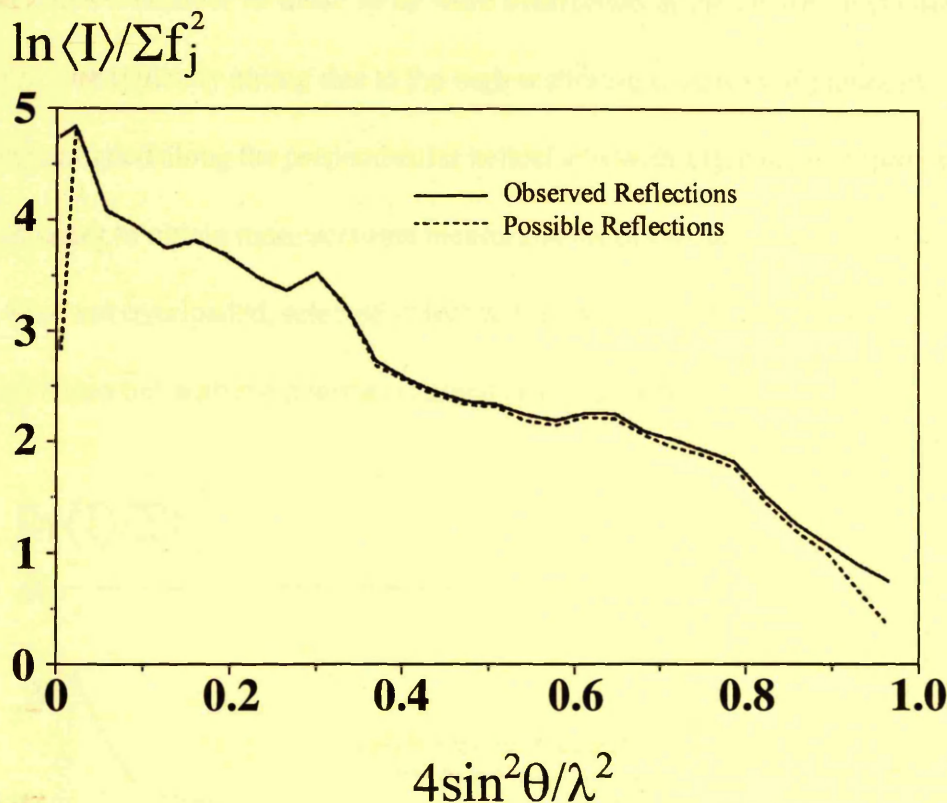


Figure 4.4: Wilson Plot for the *DENZO* Processed Data

Inspection of the two data-sets indicates that the *DENZO* processed data contain estimates for 9,292 more unique reflections than the *MOSFLM* processed data. The *DENZO* data contain estimates for 6,897 more reflections than the *MOSFLM* data in the weak high-resolution shells between 1.00 and 1.20 Å. As already indicated in Section 4.4.3, these data were judged too weak to be of practical use. However, this suggests that the *DENZO* data may contain measurements for weaker data in lower resolution shells that have perhaps been rejected by *MOSFLM*.

Visual inspection of certain strong reflections, namely the low-resolution reflections between 20\AA and approximately 50\AA and the base-stacking reflections at about 3.3\AA , indicated that a number of these spots were overloaded at the centre. Base-stacking reflections are typically strong due to the high scattering in unison of planes of nucleic acid bases aligned along the perpendicular helical axis with a spacing of approximately 3.3\AA . In order to obtain more accurate measurements of the base-stacking reflections, which were not overloaded, selected individual frames were recollected using the same exposure times but with the generator turned down to half-power.

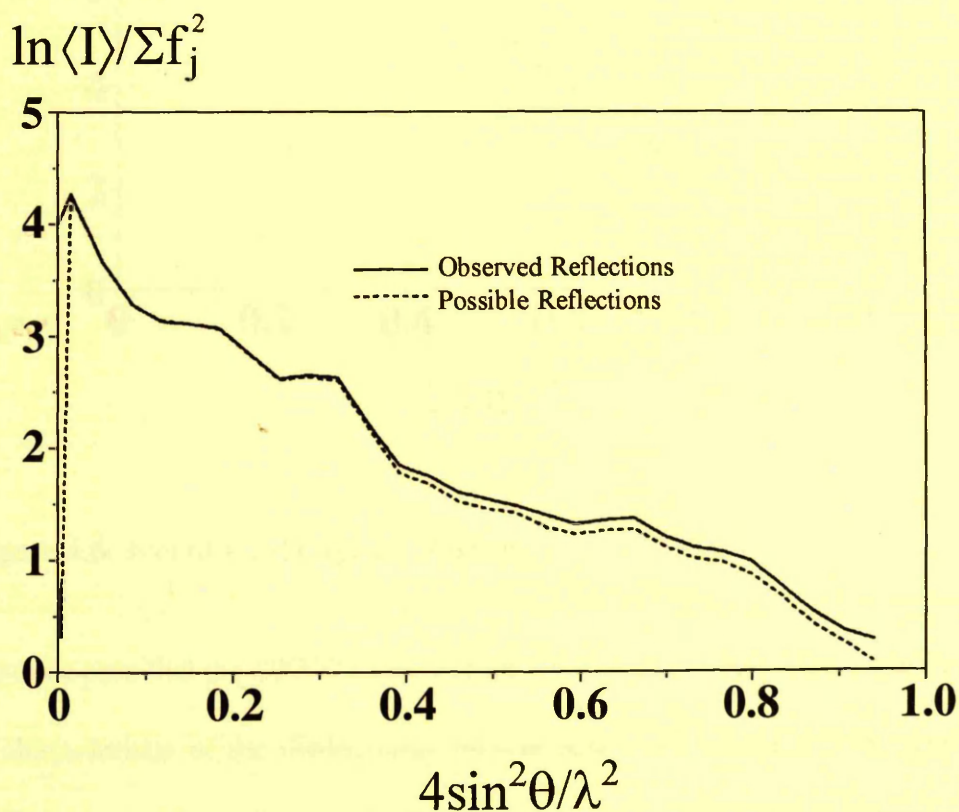


Figure 4.5: Wilson Plot for the *MOSFLM* Processed Data

Both *DENZO* and *MOSFLM* can estimate overloads using profile fitting techniques

(Otwinowski (1993a), Leslie(1990)). The *DENZO* data contain estimates for most of the very strong low-resolution reflections out to 53\AA while the *MOSFLM* data contain measurements to a limit of about 15\AA . Plots of the $\bar{I}/\sigma(\bar{I})$ distribution against resolution for the two data-sets are presented in Figures 4.6 and 4.7.

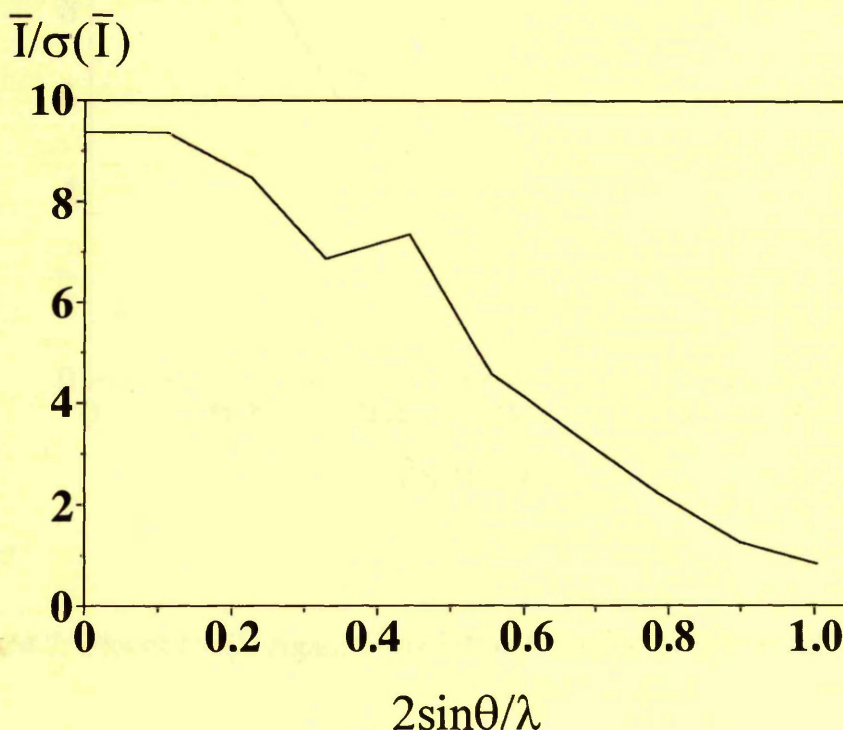


Figure 4.6: Plot of $\bar{I}/\sigma(\bar{I})$ Against Resolution for the *DENZO* Processed Data

It can be seen that the *DENZO* data, Figure 4.6, contain a “bulge” at approximately 3.3\AA , characteristic of the distinctively intense base-stacking reflections, while the *MOSFLM* data, Figure 4.7, do not. This may indicate that the *DENZO* data contain better estimates, or estimates of more of the overloaded base-stacking reflections. Such reflections are important in that they are useful in determining a model of the unit cell contents of oligonucleotides by Patterson analysis and molecular replacement, as

described in the next chapter.

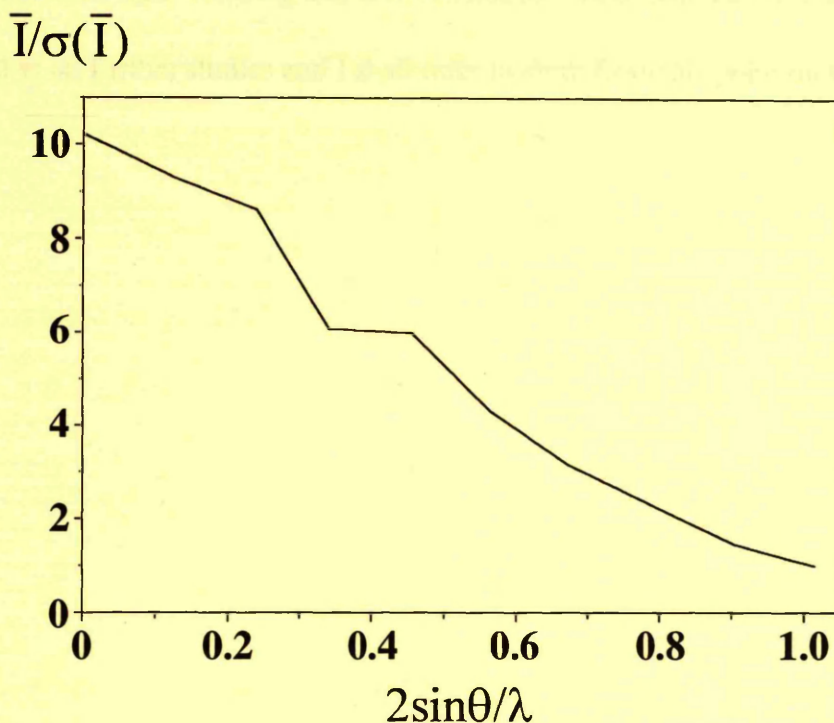


Figure 4.7: Plot of $\bar{I}/\sigma(\bar{I})$ Against Resolution for the *MOSFLM* Processed Data

In summary, the *DENZO* data are marginally more complete than the *MOSFLM* data, processed and reduced from the same R-AXIS imaging plate records, with minor differences apparent in the handling of very weak and very strong, largely overloaded, data. The resultant data-sets are however of comparable quality and the discrepancy in the number of reflections in each set is largely, as in the case of the 1.0-1.2Å data, inconsequential.

This observation in no way discriminates between the quality of the two programs themselves, as it is impossible to say to what extent either was optimised for these particular data relative to the other, since both processing procedures were carried out

independently. However, due to its marginally greater completeness and its treatment of the overloaded base-stacking and low-resolution reflections, the *DENZO* data were employed in all further studies and I shall refer to them from this point on as the native data.

And do you know what 'the World' is to me? Shall I show it to you in my mirror? This World: a monster of energy, without beginning, without end; . . . enclosed by 'nothingness' as by a boundary; not something blurry or wasted, not something endlessly extended, but set in a definite space as a definite force, and not a space that might be 'empty' here or there, but rather as force throughout, as a play of forces and waves of forces, at the same time one and many, increasing here, and at the same time decreasing there; a sea of forces flowing and rushing together, eternally changing, eternally flooding back, with tremendous years of recurrence, with an ebb and a flood of its forms; out of the simplest forms striving towards the most complex, out of the stillest, most rigid, coldest forms towards the hottest, most turbulent, most self-contradictory, and then again returning home to the simple out of this abundance, out of the play of contradictions back to the joy of concord, still affirming itself in this uniformity of its courses and its years, blessing itself as that which must return eternally, as a becoming that knows no satiety, no disgust, no weariness: this, my Dionysian world of the eternally self-creating, the eternally self-destroying, this mystery world of the twofold voluptuous delight, my 'beyond good and evil', without goal, unless the joy of the circle is itself a goal; without will, unless a ring feels good will towards itself – do you want a *name* for this world? A *solution* for all its riddles? A *light* for you, too, you best-concealed, strongest, most intrepid, most midnightly men? – *This world is the will to power – and nothing besides!* And you yourselves are also this will to power – and nothing besides!

Friedrich Wilhelm Nietzsche, *The Will to Power*

Increasingly, our concepts of past, present and future are being forced to revise themselves. Just as the past itself, in social and psychological terms, became a casualty of Hiroshima and the nuclear age (almost by definition a period where we were forced to think prospectively), so in turn the future is ceasing to exist, devoured by the all-voracious present. We have annexed the future into our own present, as merely one of those manifold alternatives open to us. Options multiply around us, we live in an almost infantile world where any demand, any possibility, whether for life-styles, travel, sexual roles and identities, can be satisfied instantly.

In addition, I feel that the balance between fiction and reality has changed significantly in the past decade. Increasingly their roles are reversed. We live in a world ruled by fictions of every kind – mass-merchandising, advertising, politics conducted as a branch of advertising, the instant translation of science and technology into popular imagery, the increased blurring and intermingling of identities within the realm of consumer goods, the pre-empting of any free or original imaginative response to experience by the television screen. We live inside an enormous novel. For the writer in particular it is less and less necessary for him to invent the fictional content of the novel. The fiction is already there. The writer's task is to invent the reality.

J. G. Ballard, Introduction to the French Edition of *Crash* (1974)

Chapter 5

Determination of Phases

5.1 Phasing Strategies for d(TGGGGT)

This chapter describes the strategies employed to determine the phases for the oligonucleotide d(TG₄T), using the various X-ray diffraction data-sets described in Chapter 4. The methods used and the results obtained break down into two distinct approaches to the phasing problem.

Initial investigations concentrated on the preparation of heavy-atom derivatives of the native crystals, in particular the possibility of growing isomorphous crystals of the synthesised derivatives d(U_{Br}G₄T) and d(U_IG₄T). Whilst crystals of the forementioned oligonucleotides could not be obtained in a form sufficient for X-ray analysis, a number of putative derivatives were prepared by co-crystallization and soaking experiments, as described in Chapter 3. Consequently, a number of X-ray diffraction data-sets were collected for these putative derivatives. The analysis of these derivatives and the results obtained are discussed initially in this chapter.

The second strategy, which was investigated concurrently with isomorphous replacement techniques, is that of molecular replacement. Native Patterson function and self-rotation function analyses facilitated the proposal of a possible packing model for the native crystals in the P1 cell. The initial characterisation of the d(TG₄T) tetraplex by NMR (Aboul-ela *et al.*, (1992)) allowed the generation of a family of search models for cross-rotation and translation function analyses. The results of these analyses are presented in the second half of this chapter.

5.2 Phasing by Isomorphous Replacement

The method of isomorphous replacement is the most commonly utilised technique in macromolecular crystallography for phase determination. It is enormously important in cases where there is little or no *a priori* structural information available for the macromolecule under study. The theory of the phasing of X-ray diffraction data by isomorphous replacement is discussed in detail in a variety of texts (for example Blundell and Johnson (1976) and Giacovazzo *et al.*, (1992)). The following section presents an outline of the theory and its practical application.

Assuming that a heavy-atom compound has been introduced isomorphously into the crystal of a macromolecule, then its contribution to the diffraction pattern is given by the vector addition in equation 5.1.

$$\mathbf{F}_{\text{PH}} = \mathbf{F}_{\text{P}} + \mathbf{F}_{\text{H}} \quad (5.1)$$

In equation 5.1 \mathbf{F}_{P} represents the structure factor of the native macromolecule,

F_{PH} that of the derivative and F_H that of the heavy-atom contribution.

If the position of the heavy-atom within the unit cell is known then the phase for the macromolecule, α_P , can be determined from the simultaneous equations presented in equation 5.2.

$$F_{PH}^2 = F_P^2 + F_H^2 + 2F_P F_H \cos(\alpha_P - \alpha_H) \quad (5.2)$$

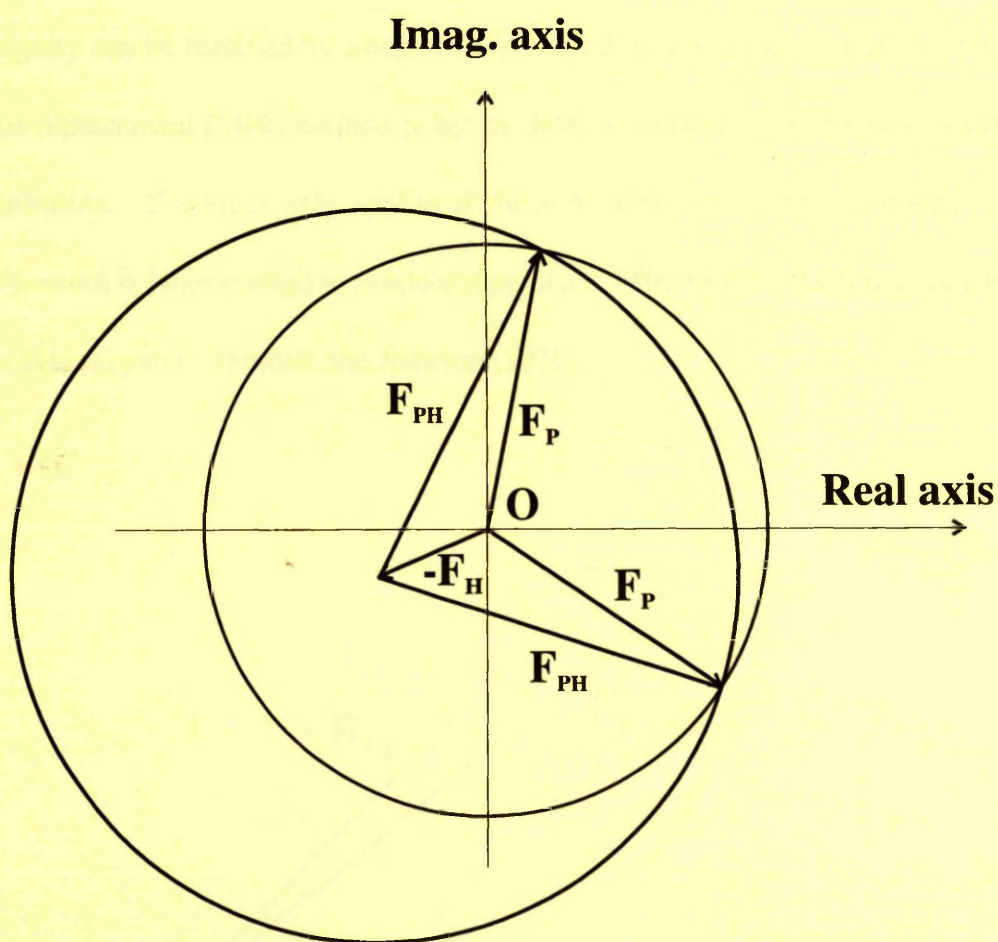


Figure 5.1: The Ideal Situation for the Single Isomorphous Replacement Method

In equation 5.2 the structural amplitude terms F_P and F_{PH} can be measured experimentally and F_H and α_H can be calculated.

This situation is represented graphically in Figure 5.1, adapted from Blundell and Johnson(1976), which illustrates the ideal situation for the single isomorphous replacement (SIR) method. The structure factor of the native macromolecule is represented by a vector lying on a circle of radius F_P and centre O . Another circle is drawn with its centre at the end of the vector $-F_H$ with a radius F_{PH} . The two circles intersect at two points. These two points of intersection indicate that there are two possible values for the phase angle of the native structure factor F_P from a single derivative. This ambiguity can be resolved by using either a second derivative in the multiple isomorphous replacement (MIR) method or by the determination of an anomalous scattering contribution. The latter only applies if the anomalous scattering component of the heavy-atom is large enough to produce significant differences in the reflections which form Friedel pairs (Blundell and Johnson (1976)).

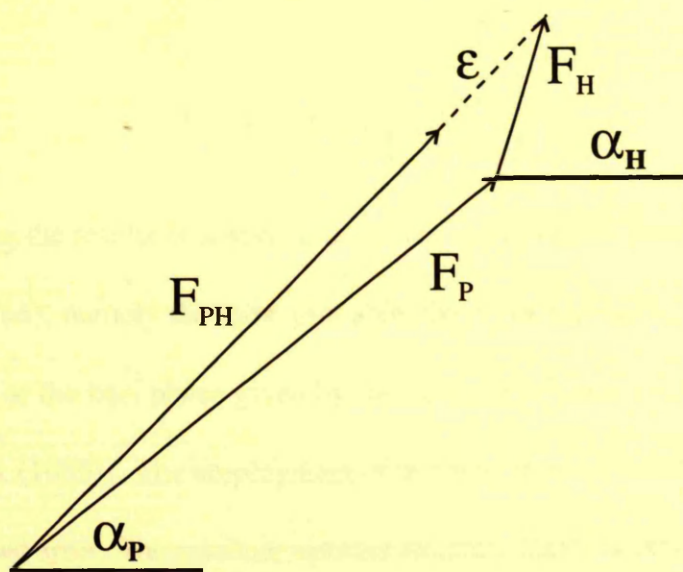


Figure 5.2: Illustration of Lack of Closure in the SIR Method

Empirically, the phasing triangles illustrated in Figure 5.1 rarely close as shown

due to errors largely incurred in the measurement and scaling of F_{PH} and F_P , and in F_H due to lack of isomorphism. In their treatment of the error incurred, Blow and Crick (1959) demonstrated that a reasonable approximation was to assume that the lack of closure resulted from the measurement of F_{PH} . This situation is illustrated in Figure 5.2, adapted from Blow and Crick (1959). The lack of closure error, ε , is defined in equation 5.3.

$$\varepsilon = (F_{PH}^{obs} - F_{PH}^{calc}) \quad (5.3)$$

Assuming a Gaussian distribution of the errors, then the probability for the j^{th} derivative that α_P is correct is given by equation 5.4, where N is a normalization factor; $\varepsilon(\alpha)$ is the lack of closure of the phase triangle along F_{PH} at native phase angle α ; and E_j is the standard deviation of the distribution.

$$P_j(\alpha) = N \exp \left[\frac{-\varepsilon_j(\alpha)^2}{2E_j^2} \right] \quad (5.4)$$

In combining the results of several derivatives there are two possible phase angles derived from $P(\alpha)$, namely the most probable phase corresponding to the maximum value for $P(\alpha)$ or the best phase given by the centroid of the probability distribution (Blow and Crick (1959)). The employment of this best phase gives an electron density map with the least error. The resultant optimal structure factor is expressed in equation 5.5, where m is a figure of merit, defined in equation 5.6, dependent on the sharpness of the phase probability, and which is equivalent to a weighting function.

$$\mathbf{F}_{\text{best}} = m|\mathbf{F}_{\mathbf{P}}|e^{i\alpha_{\text{best}}} \quad (5.5)$$

$$m = \frac{\int_{\alpha=0}^{2\pi} e^{i\alpha} P(\alpha) d\alpha}{\int_{\alpha=0}^{2\pi} P(\alpha) d\alpha} \quad (5.6)$$

5.3 Determination of Heavy-Atom Positions

From the theory outlined in the previous section, it follows that in order to determine the phases it is necessary to obtain a model for the vector $\mathbf{F}_{\mathbf{H}}$ which is dependent upon the knowledge of the positions of the heavy-atoms in the cell. There are a number of possible approaches to resolving this problem. Certain methodologies require some previous knowledge of the number of heavy-atoms present in the cell, or the positions of heavy-atoms of at least one already determined derivative crystal. The most general approach, which requires no previous knowledge of the heavy-atom positions within the derivative crystal, is the Patterson method. This method was employed to evaluate the derivative data-sets of the d(TG₄T) oligonucleotide crystals, and the results of this analysis are presented in the next section. The method itself is discussed below.

5.3.1 The Patterson Function

The Patterson function (Patterson (1934)) is a convolution function of the electron density which can always be calculated from a set of X-ray diffraction data since it is the Fourier transform of $F^2(hkl)$ rather than $\mathbf{F}(hkl)$. The function itself is defined in equation 5.7.

$$P(uvw) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F^2(hkl) \cos 2\pi(hu + kv + lw) \quad (5.7)$$

The Patterson function represents a map, not of the individual atomic positions, but of the interatomic vectors between all the atoms in the cell. Generally speaking, if an atom i contains Z_i electrons and an atom j contains Z_j electrons, then the peak in the Patterson summations which represent the vector between the atoms i and j will have a size proportional to the product $Z_i Z_j$. Therefore, the heavier atoms will give rise to larger peaks in the Patterson map. The function summation always contains a large peak at the origin as a result of the superposition of all the self-self vectors (see equation 5.8), and contains a centre of symmetry at the origin even although the structure itself may not be centrosymmetric.

$$P(000) \propto \sum_{j=1}^N Z_j^2 \quad (5.8)$$

5.3.2 The Patterson Method

The 'true' difference Patterson synthesis would be calculated using the coefficients $(F_{PH}^2 - F_P^2)$ (Zanotti (1992)). This map represents the difference between the Patterson of the derivative minus the Patterson of the native structure. However, this would result in a vector map containing peaks corresponding to the interatomic heavy-atom \longleftrightarrow heavy-atom vectors, as well as heavy-atom \longleftrightarrow macromolecule vectors, and would be extremely difficult to interpret. A better method, and one commonly used in macromolecular crystallography, is to use the coefficients $(F_{PH} - F_P)^2$ in what is called an isomorphous difference-Patterson. Using these coefficients, the expression for the

Patterson summation expands to give the terms shown in the array below.

$$(F_{PH} - F_P)^2 = 4F_P^2 \sin^4 \left(\frac{\alpha_P - \alpha_{PH}}{2} \right)$$

(a)

$$+F_H^2 \cos^2(\alpha_{PH} - \alpha_H)$$

(b)

$$-4F_P F_H \sin^2 \left(\frac{\alpha_P - \alpha_{PH}}{2} \right) \cos(\alpha_{PH} - \alpha_H)$$

(c)

If $F_H \ll F_P$ the terms (a) and (c) will generate background noise since, for small values of F_H , $(\alpha_P - \alpha_{PH})$ tends to be small (Blundell and Johnson (1976)). In addition, the term (b) will result in a map containing the expected vectors at about half the expected weight for the situation where $F_H \ll F_P$, since $(\alpha_{PH} - \alpha_H)$ will be effectively random (Blundell and Johnson (1976)). In general, the magnitude of the differences between the native and derivative amplitudes is not equal to the heavy-atom structure factors due to errors in scaling and measurement, lack of isomorphism and the fact that for acentric hkl , F_{PH} and F_P are not usually colinear. All such effects contribute noise to the Patterson summation. Other difficulties can occur when there are several sites with low occupancies, where there is high symmetry or where the macromolecule is large, all of which result in a Patterson map which can be difficult to interpret.

The symmetry operators present in the crystal leave a trace in the Patterson function, consisting of particular clusterings of vector maxima on specific lines or planes of the

map, called Harker lines and sections (Blundell and Johnson (1976)). These are produced by vectors, with one or two constant components, between equivalent atoms related by symmetry elements other than the inversion centre. Harker sections are useful in isolating the heavy-atom \longleftrightarrow heavy-atom vectors since the vector between equivalent heavy-atoms gives rise to a peak on the Harker section. The space group P1, however, contains no symmetry operators apart from the identity and, therefore, contains no Harker sections that might be useful in determining heavy-atom positions.

A difference Patterson may also be calculated using the coefficients ($F_{PH}(+) - F_{PH}(-)$)² using anomalous data (Rossman (1961)). The resultant Patterson summation produces a map which ideally contains positive peaks of half weight at the ends of interatomic vectors relating anomalous scatterers. The diffraction observations used are measured from the same crystal so that problems with systematic errors are minimised, no scaling is necessary and isomorphism is not a factor. Consequently, these maps are potentially very useful. However, the difficulty in accurately measuring very small differences results in the anomalous difference Pattersons generally being much more noisy than the respective isomorphous difference Pattersons. In addition, the Patterson summation can be dominated by the largest terms. Should these terms have large errors associated with them then the maps will be impossible to interpret. Extreme caution must be taken to exclude any spuriously large differences from the Patterson calculation, regardless of whether the calculation is of an isomorphous or anomalous difference synthesis. When sufficient care is observed anomalous difference Pattersons are extremely useful, especially when used in conjunction with the isomorphous map for checking solutions.

5.4 Scaling of Data and Patterson Analysis

All of the derivative data-sets were anisotropically scaled to the native data using the CCP4 program *SCALEIT*. This program calculates a derivative to native scaling function using F^2 which is then applied to the derivative data F_{PH} . The scaling factor used is of the form given in equation 5.9.

$$K = C \exp(-(h^2 B_{11} + k^2 B_{22} + l^2 B_{33} + 2hk B_{12} + 2hl B_{13} + 2kl B_{23})) \quad (5.9)$$

In equation 5.9, C is an overall scaling factor and the B terms are anisotropic temperature factors. Anisotropic modelling of the scaling of the native and derivative data involves a more general assumption that the variation in intensities between data-sets is not necessarily the same for each individual index h , k and l . This consideration is necessary since the data-sets originate from different crystals. The six parameters B_{11} , B_{22} , B_{33} , B_{12} , B_{13} and B_{23} describe the anisotropy of the reflection intensities in terms of the individual indices h , k and l . An initial scaling factor, K_{init} is calculated using relative Wilson scaling as detailed in equation 5.10. The scale and anisotropic temperature factors are then refined using a modification of the Fox and Holmes (1966) method.

$$K_{init} = \left(\frac{\sum F_P^2}{\sum F_{PH}^2} \right)^{\frac{1}{2}} \quad (5.10)$$

The difference between the measured intensity and the mean intensity is minimized

for all measurements of all reflections. With reference to the equation array immediately below, the expression (a) is minimized with respect to all parameters represented by expressions (b) and (c). This amounts to two scale factors and six B values in the anisotropic case.

$$(a) \sum_h \sum_i w(h_i) [I(h_i) - G_i I_h]^2$$

$$(b) \quad G1 = 1.0$$

$$(c) \quad G2 = \left(\frac{1}{C} \right) \exp(2\bar{h}B_h)$$

A qualitative test for the degree of heavy-atom substitution is the calculation of an agreement index R_{iso} between the native and derivative structure factors, as shown in equation 5.11. Experience indicates that a value from 0.15 to 0.25 can be considered as a reasonable agreement (Zanotti (1992)). A smaller value is an index of little substitution, whilst too high a value can be taken as a clue to non-isomorphism.

$$R_{iso} = \frac{\sum |F_P - F_{PH}|}{\sum |F_P|} \quad (5.11)$$

Assuming a perfect isomorphous case, with no experimental errors, then, evaluating R_{iso} for equivalent resolution bins, the agreement should be consistent with increasing resolution. Empirically, R_{iso} will increase with resolution as a result of random errors

and lack of isomorphism arising from heavy-atoms perturbing the native structure. However, a dramatic increase in R_{iso} at higher resolution assists in the derivative evaluation as it is indicative of non-isomorphism. The overall value of R_{iso} can still be misleading. Large values do not always indicate a high degree of substitution and can instead be due to non-isomorphism or problems in scaling. Similarly, a derivative with small values of R_{iso} may be isomorphous and still contain useful phase information.

Derivative	Res (Å)	No. Com. Refins.	Scale Fac.	Temp. Fac.	R_{iso}	N_{anom}	K_{emp}
K_2PtCl_4	1.6	22821	0.987	-0.816	0.110	22623	3.70
$Pb(CH_3CO_2)_2$	1.8	17923	1.020	-0.495	0.075	12452	1.89
Me-I	2.0	13518	0.991	-1.552	0.107	13467	4.60
I_2	2.0	13349	0.986	-0.576	0.086	8485	2.32
$UO_2(CH_3CO_2)_2$	1.8	6203	1.026	-0.962	0.179	—	—

$$R_{iso} = \frac{\sum |F_P - F_{PH}|}{\sum F_P}$$

$$K_{emp} = 2 \left(\frac{\overline{(F_{PH} - F_P)^2}}{(\overline{F_{PH}(+) - F_{PH}(-)})^2} \right)^{\frac{1}{2}}$$

Table 5.1: Scaling Statistics for Native to Derivative Datasets

Certain heavy-metals will absorb X-rays and this results in anomalous diffraction effects which manifest as a non-equivalence in reflections related by Friedel symmetry.

This signal is generally small but can be measured if intensity measurements of Friedel partners are collected carefully. This anomalous data can also be used to determine the degree of substitution. These measurements usually have larger experimental errors but have the advantage that they are independent of the degree of isomorphism.

Data from derivative and native crystals were collected using the R-Axis IIC Imaging Plate detector at York University, as detailed in Chapter 4. The derivative data-sets were processed with *DENZO* and anisotropically scaled with the 1.2 Å *DENZO* processed native data-set. The statistics for this anisotropic scaling are given in Table 5.1.

All of the data-sets contain data to high-resolution. In the case of the K_2PtCl_4 , $\text{Pb}(\text{CH}_3\text{CO}_2)_2$, Me-I and I_2 putative derivatives, the data are nearly complete to 1.6 Å, 1.8 Å, 2.0 Å and 2.0 Å, respectively. All of these forementioned putative derivatives have R_{iso} values in the region of 10% to the limit of their resolution range. This would appear to be indicative of very little in the way of heavy-atom substitution, especially considering the high-resolution range of the anisotropic scaling. The $\text{UO}_2(\text{CH}_3\text{CO}_2)_2$ derivative has a value of R_{iso} of 17.9%. Whilst this might be indicative of some significant heavy-atom substitution, it is noteworthy that the crystal soaked with this compound suffered severe radiation damage, producing a poor quality data-set with a reflection measured multiplicity of only one. It is also instructive to compare the R_{iso} values of the different derivative data-sets with the R_{merge} values for the data-sets described in Chapter 4. Whilst the 1.2 Å native data has a R_{merge} of 0.061, the K_2PtCl_4 , $\text{Pb}(\text{CH}_3\text{CO}_2)_2$, Me-I, I_2 and $\text{UO}_2(\text{CH}_3\text{CO}_2)_2$ data-sets have R_{merge} values of 0.047, 0.046, 0.033, 0.072 and 0.064 respectively. Thus, with the noted exception of

the $\text{UO}_2(\text{CH}_3\text{CO}_2)_2$ data, all of the derivative data-sets possess R_{iso} values not much greater than might be expected were two native data-sets being scaled together.

Patterson syntheses of isomorphous and anomalous differences were calculated for all of the above derivatives using the program *FFT* in the *CCP4* suite. Data where F is less than $3\sigma(F)$ were excluded from the calculation along with unacceptably large differences as indicated by the program *SCALEIT*. None of the Patterson functions calculated displayed any peaks representative of heavy-atom \longleftrightarrow heavy-atom cross vectors.

In $P1$, where there are no symmetry elements, a single heavy-atom substitution within the unit cell, or asymmetric unit, would only contribute to the accumulative self-self vector noise at the origin since there are no restrictions on the origin position within this spacegroup. In order to observe a heavy-atom cross vector peak for a macromolecule in the $P1$ cell at least two distinct heavy-atom sites would be required. Supposing that one or more of the above derivatives might contain a single heavy-atom substitution, such a substitution would be undetectable without an accurate knowledge of the native unit cell contents.

With the failure of the collected derivative data-sets to yield any heavy-atom position solutions for subsequent phasing, it was therefore necessary to re-evaluate the phasing strategy. As indicated in Chapter 3, the vast majority of time spent in the search for heavy-atom derivatives was invested in the attempted crystallization of the $\text{d}(\text{U}_{\text{Br}}\text{G}_4\text{T})$ and $\text{d}(\text{U}_{\text{I}}\text{G}_4\text{T})$ oligonucleotides. Whilst crystals were obtained they were not suitable for crystallographic analysis. Further investigations of heavy-atom compounds would involve attempts to improve crystals of the bromo and iodo oligonu-

cleotides, as well as a considerably more extensive screening of heavy-atom compound soaks.

An alternative method of phasing is the molecular replacement approach. This experimental method is particularly useful when a model is available for the macromolecule under study. During the data collection and evaluation of the putative derivatives discussed above, a potential search model became available for the parallel-stranded tetraplex formed by d(TG₄T), based on the NMR characterisation of the tetraplex (Aboul-ela *et al.*, (1992)). The molecular replacement method, and its application to the structural study of the d(TG₄T) tetraplex, is discussed in the remainder of this chapter.

5.5 The Molecular Replacement Method

The molecular replacement method is often useful in the determination of macromolecular structures in cases where a similar or homologous structure is already known. In the study of proteins the method often constitutes the most economical approach under a variety of circumstances. Examples of such suitable cases might be where the molecule crystallizes in a different unit cell from a previous structural determination, where it is the product of site-direct mutagenesis experiments leading to a different crystal form, or where the protein belongs to a structural family containing certain distinctive folding patterns. As a result of the greater structural organization and regularity of the nucleic acids, molecular replacement has become the most successful means of determining the structures of oligonucleotides (Kennard and Hunter (1989, 1991)). The models

used for the nucleic acids are derived from extant crystal structures, fibre diffraction co-ordinates or NMR structural characterisations.

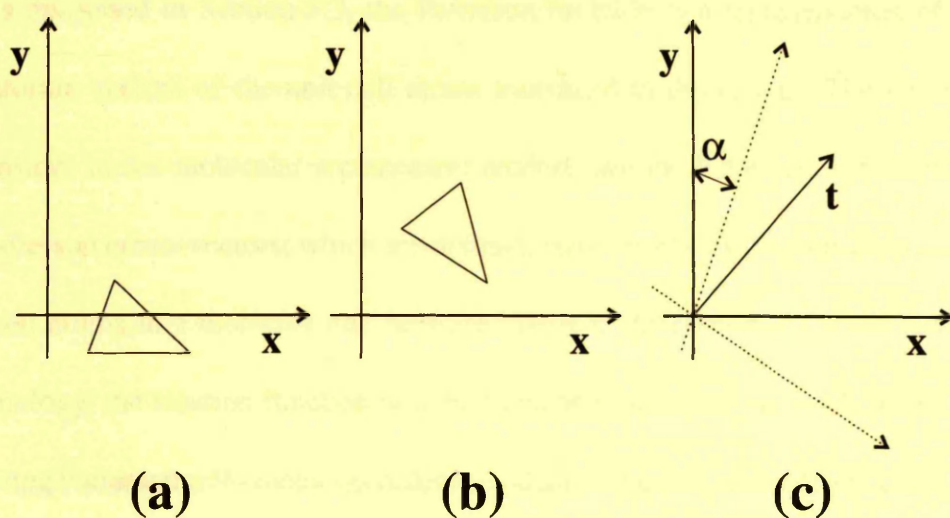


Figure 5.3: 2-D Depiction of the Rotation And Translation Operations

The rotational and translational analyses involved in molecular replacement studies involve the comparison of Patterson functions generated from the native X-ray crystallographic data and the experimental search model. The nature of the rotational and translational operations are depicted in two dimensions in Figure 5.3, which is adapted from Giacovazzo (1992). In Figure 5.3(a) a triangle is depicted, representing a general molecular object. Figure 5.3(b) illustrates this object after being subjected to a rotation (α) and a translation (t). The rotation, α , is shown defined as the angle to the y-axis, and the translation, t , as a vector from the origin in Figure 5.3(c). In the study of the crystal structures of biological macromolecules the same principles apply as in Figure 5.3, but are extended to three dimensional space. Determination of the corresponding three dimensional rotational and translational parameters can be achieved by analysis of the Patterson function. The structure factors thus obtained, and the resultant electron

density map, contain a strong bias towards the starting model. However, they are often sufficiently close to the correct values for a successful refinement to be attempted.

As discussed in Section 5.3, the Patterson function is a representation of all the interatomic vectors of the unit cell atoms translated to the origin. The key vectors to consider in the molecular replacement analysis are the Patterson self-vectors and the Patterson cross-vectors, which are defined, respectively, as the interatomic vectors between atoms in a molecule and between atoms in different molecules. Using this terminology, the rotation function may be thought of as being primarily a calculation involving Patterson self-vectors in order to evaluate the orientation of molecules, whilst the translation function may be considered to be a Patterson cross-vector calculation in order to evaluate the relative positions of oriented molecules.

In the first instance, before an experimental search model is available, molecular replacement can be utilized in order to search for the orientation of non-crystallographic symmetry elements within the asymmetric unit. This information can be obtained using a self-rotation function, and can be useful in deriving an outline of the cell contents and general features of their packing.

When in possession of an experimental search structure, molecular replacement may then be used to obtain a preliminary phasing model for the macromolecule under study. The search model ought to be the best that is available, and preferably refined. Atoms which are known to differ from the target structure ought to be removed from the search model, although care must be taken on this regard. The inclusion of erroneous structural atoms generally results in an increase in the overall noise level, but the removal of correct atoms will reduce the strength of the solution signal.

The rotational and translational searches described above constitute a six dimensional problem involving three rotation angles and three components of a translation vector for each search model thus positioned. A rotation function is used to determine the three rotation angles and a translation function, or some other positioning technique, is used to correctly position the properly orientated molecule. Finally, a procedure may be employed to optimize the derived orientations and positions before crystallographic refinement is attempted.

The theory and applications of the molecular replacement method, and the various formulations and implementations of the rotation and translation functions, are discussed in detail in a variety of texts (Rossmann (1972); Blundell and Johnson (1976); Giacovazzo *et al.*, (1992)). The following section presents an outline of this theory and its practical applications.

5.5.1 The Rotation Function

The term molecular replacement was coined by Michael Rossmann and the theory of the rotation function was first developed by Rossmann and Blow (1962). In their exposition of the rotation function theory Rossmann and Blow describe the function shown in equation 5.12 below.

$$R(C) = \int_{-\infty}^{\infty} P_2(x)U(x)P_1(Cx)dV \quad (5.12)$$

In the function shown in equation 5.12, $P_2(x)$ is the Patterson function of crystal 2 at real space point x ; $P_1(x)$ is the Patterson function of crystal 1 at real space point x ;

C is a rotation matrix; and U is a shape function, most commonly a sphere, that has a value of 1 inside some volume and 0 elsewhere. When R is large a value for C has been found that maximizes the overlap of the two Patterson functions.

The function shown in equation 5.12 can be evaluated in either real or reciprocal space. In their treatment of the rotation function Rossmann and Blow (1962) implement the function in reciprocal space. The expression of the R function in reciprocal space is given in equation 5.13 below.

$$R(C) = \sum_p \sum_h F_1^2(p) F_2^2(h) G(h + C^t p) \quad (5.13)$$

In the expression for the R function shown in equation 5.13, $F_1^2(p)$ is the intensity for crystal 1 at reciprocal lattice point p ; $F_2^2(h)$ is the intensity for crystal 2 at lattice point h ; C^t is the transpose of the rotation matrix C ; and G is an interference function which is the Fourier transform of the shape function U given in equation 5.12. The general formulation given in equation 5.13 can be calculated using any Patterson function. Thus the calculation of the R function of the native Patterson with respect to itself, the self-rotation function, facilitates the determination of the order and orientation of non-crystallographic rotational symmetry elements in the asymmetric unit of the molecule under study. Evaluation of this function using the Patterson of an experimental search model with respect to the Patterson of the native data, the cross-rotation function, can lead to the determination of the orientation of the search model in the asymmetric unit.

Another formulation of the rotation function has been proposed by Lattman and Love (1970). This formulation is given in equation 5.14.

$$R(C) = \sum_p F_1^2(p) C^t F_m^2(p) \quad (5.14)$$

In the formulation of Lattman and Love, $F_1^2(p)$ is the intensity for crystal 1 at reciprocal lattice point p ; $F_m^2(p)$ is the intensity for crystal 2 at reciprocal lattice point p ; and C^t is the transpose of rotational matrix C . The F_2 term in equation 5.13 has been replaced by the F_m term which is defined as the intensities for a crystal consisting of an isolated molecule in a large cell. This assumption that F_m is an isolated molecule in a large cell has the consequence that the Lattman and Love formulation is not general but is computationally faster than the formulation of Rossmann and Blow.

The general slowness of the calculation of the R function previously required that the data used in the calculation be limited. This problem was largely overcome by the use of the Crowther fast rotation function (Crowther (1972)). The Crowther fast rotation function involves the expansion of the two Pattersons being compared in the R function in terms of spherical Bessel functions. This reduces the computational problem to a series of two-dimensional Fourier series summations that can be quickly evaluated using a fast Fourier transform algorithm, and hence allows the inclusion of all of the data.

Certain crystallographic problems can lead to difficulties in the evaluation of the rotation function. For example, should the search model deviate greatly from sphericalness or the unit cell deviate greatly from being equidimensional, vectors in the search model Patterson may overlap the origin peak of the native Patterson function of an adjacent unit cell, leading to false peaks. Also, the Crowther fast rotation function places restrictions on the radii of integration used. These restrictions are dependent

on the resolution of the data used. This can present a problem for large structures. Fundamentally, the success of the rotation function analysis relies in the first instance in the implementation of an accurate search model of the unknown macromolecule under study.

5.5.2 The Translation Function

Application of a translation function involves the optimization of the position of at least two correctly orientated molecules so that their cross-Patterson vectors correlate with the Patterson of the structure under study. Fitzgerald (1991) draws a distinction between translation functions, which she defines as the reciprocal space implementation of this positional search method by means of a Fourier transform, and correlation searches, which she takes to be the real space implementation of the positional search procedure which advances by the evaluation of the function at each sample point in real space.

Adhering to Fitzgerald's definition, the most widely used formulation of the translation function in the study of macromolecules is that proposed by Crowther and Blow (1967). This formulation is given in the expression in equation 5.15 below.

$$T(t) = \sum_h F_1^2(h) F_{2,j}(h) F_{2,k}^*(h) \exp^{-2\pi i h t} \quad (5.15)$$

In the formulation of the translation function given in equation 5.15, $F_1^2(h)$ is the intensity for crystal 1 at reciprocal lattice point h ; $F_{2,j}(h)$ is the structure factor amplitude for molecule j of crystal 2 at reciprocal lattice point h ; $F_{2,k}^*(h)$ is the complex conjugate of the structure factor amplitude of molecule k of crystal 2 at reciprocal lattice

point h ; and t is the translation vector. In a modification of equation 5.15, $T_1(t)$ replaces the term $F_1^2(h)$, where $T_1(t)$ is equivalent to $F_1^2(h) - \sum_{i=1,n} F_{1,i}^2(h)$, thus removing the self-vectors from the Patterson function of crystal 1. This modification usually has the effect of reducing the noise encountered in the search. Other translation functions and their formulation are described elsewhere (for a review see Fitzgerald (1991)).

The singular and most substantial limitation of the translation function analysis is the very high sensitivity of the function to the accuracy of the rotational angles of the search model rotation solutions. In addition, the positional algorithms are very sensitive to systematically missing data. This may result from a faulty collection strategy, as well as particular situations in which there is local symmetry parallel to crystallographic symmetry or strong vector patterns parallel to crystallographic symmetry axes.

5.5.3 Correlation Searches

Correlation searches, like translation functions, operate from the assumption that the orientation of the search models within the cell has already been determined. The correlation function then proceeds by the real space systematic translation of the molecule through the cell, accompanied by the calculation of a measure of correctness at each sample point. This may take the form of an agreement index, or R-value, or, as an alternative, a correlation coefficient.

An example of a correlation coefficient is found in the positioning technique of Harada *et al.*, (1981). This correlation coefficient is given in equation 5.16.

$$C(t) = \left(\frac{\sum_h F_1^2(h) F_2^2(h, t)}{[\sum_h F_1^4(h) \sum_h F_2^4(h, t)]^{\frac{1}{2}}} \right) \quad (5.16)$$

The method of Harada *et al.*, (1981) uses this correlation coefficient as a measure of correctness. In equation 5.16, $F_1^2(h)$ is the intensity for crystal 1 at reciprocal lattice point h and $F_2^2(h, t)$ is the intensity for crystal 2 at reciprocal lattice point h given the translation vector t . The correlation coefficient of Harada *et al.*, (1981) has been deployed by Fujinaga and Read (1987), with modification of the terms $F_1^2(h)$ and $F_2^2(h)$ to give $(F_1^2(h) - \langle |F_1(h)| \rangle^2)$ and $(F_2^2(h) - \langle |F_2(h)| \rangle^2)$, to facilitate the removal of the origin from the Patterson function.

Correlation coefficient functions are generally flat over the majority of the translational space, with a sharp feature appearing at the correct position of the search structure. This results in the necessary fine sampling of the function to ensure that the solution is not missed. As a consequence, correlation coefficient calculations are disadvantaged by being very computationally intensive and particularly lengthy when there is more than one search model in the asymmetric unit, since the dimensions of the search problem are increased accordingly. However, using the correlation coefficient as the measure of correctness, the search computation may be treated as a series of simpler, independent three dimensional searches.

5.5.4 Packing Function Analysis

In the same manner as translation function and correlation searches, packing function analyses assume that the orientations of the search model in the cell are already known. The packing function analysis proceeds by utilising some index of physical reasonable-

ness, such as the minimization of bad contacts or the maximization of the occupation of space, in order to determine the allowed positions of the search model structure.

In the case of packing analysis by the minimization of bad contacts, the search model is systematically translated through the cell of the target structure and all intermolecular distances are calculated at each sample position. Permitted positions of the search structure correspond to positions with few or no short intermolecular distances. The formulation and implementation of such an approach is discussed by Fitzgerald (1991).

Alternatively, packing analysis may be investigated using the criterion of the maximization of the occupation of space. Hendrickson and Ward (1976) describe a shape function $M(x)$ which approximates the search structure, where $M(x)$ is 1 if the positional coordinate x is intramolecular and 0 if x is elsewhere. A grid is defined in translation space with all points initialized with a value of 0. The shape function is then translated across the grid. The value at a grid point is set to 1 if $M(x)$, or a symmetry mate of $M(x)$, is 1. At each sample translation a summation is performed over all grid points to give a function $T(t)$. Large values of $T(t)$ represent possible translation function solutions.

As with translation functions and correlation searches, the accuracy of the search model orientations is crucial in the packing function analysis. However, packing analysis is ineffective where only a fragment of the unknown structure can be modelled, since the unit cell does not effectively constrain the packing of a partial model. As a result, packing function analysis is not appropriate in cases where the asymmetric unit contains more than one copy of the search structure. Usually, in addition to a packing analysis, a translation function or correlation search is required in order to determine

the precise parameters of the translation.

5.6 Native Patterson Function Analysis

The native Patterson function, as described in Section 5.3, is calculated from the experimentally derived value of $F^2(hkl)$ for the native crystal and is composed of N^2 interatomic vectors for a unit cell containing N atoms. Of these peaks, N lie at the origin and, since the function is centrosymmetric, the calculated Patterson map contains $(N^2 - N)/2$ independent vectors. The vector peaks have approximately the same width as atomic peaks and occupy a volume equivalent to that occupied by the N atoms of the structure. Consequently these peaks overlap as N become large. Thus, for a macromolecular structure such as a protein or multimeric oligonucleotide, the peak overlap is irresolvable with the exception of peaks resulting from the interatomic vectors of atoms of high atomic number. However, in conjunction with certain assumptions regarding the organization of the unknown structure, features of the native Patterson can lead to a qualitative description of the packing of the unit cell, particularly in the case of helical nucleic acid structures.

The native Patterson function for the d(TG₄T) oligonucleotide crystal data was computed employing data in the resolution range 20 Å to 1.6 Å, using the program *FFT* in the *CCP4* suite. Data where F is less than $3\sigma(F)$ were excluded from the calculation. Figure 5.4 illustrates a section of the Patterson map calculated along the w -axis near the origin of the function, and contoured in 1σ intervals.

The uniformly spaced vector streaks approximately perpendicular to the v -axis are

characteristic of a helical nucleic acid structure. Within the nucleic acid helix bases are organized, to a reasonable approximation, in a planar arrangement with parallel base planes related by a uniform translation and rotation along the helical axis. Consequently, in the Patterson map, the planar base arrangement gives rise to an approximately planar set of interatomic vectors perpendicular to the helical axis.

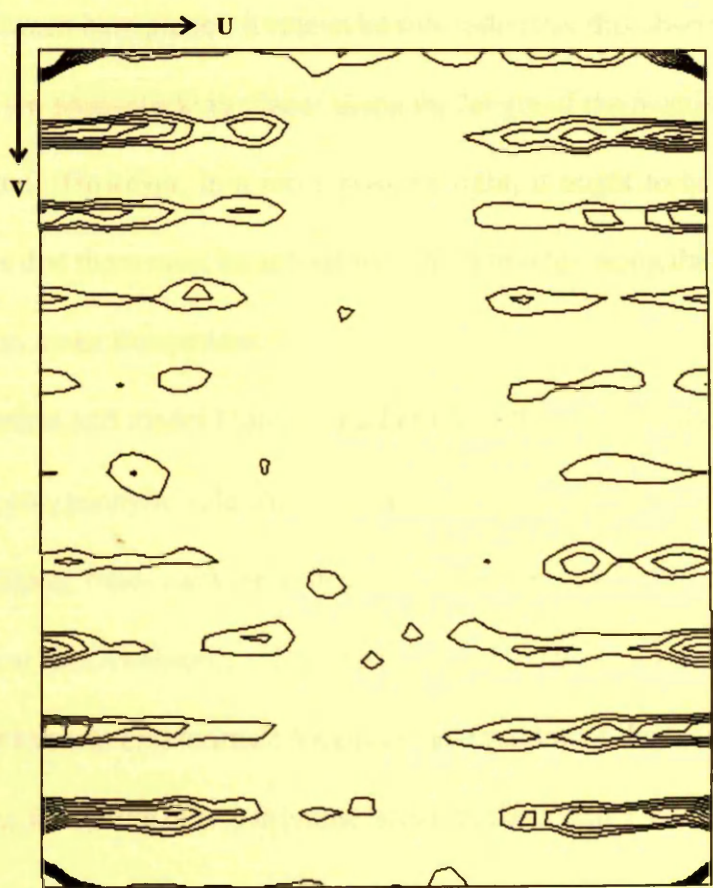


Figure 5.4: Native Patterson Function uv Section Calculated At the Cell Origin

Further interatomic vectors between planes uniformly translated and rotated along the direction of the helical axis build up a characteristic set of streaks or interatomic

vector planes, all of which lie approximately perpendicular to the helical axis direction. Thus it can be concluded from inspection of Figure 5.4 that the structure adopted by the oligonucleotide d(TG₄T) packs with its helical axis approximately coincident with the *b*-axis of the crystal unit cell in real space.

Inspection of Figure 5.4 reveals a pattern of ten vector streaks along the direction of the *v*-axis. Due to the extensive overlap of the interatomic vector peaks, probable deviations from planarity in the base arrangements and the possible non-uniformity of the axial rise between base planes, it cannot be concluded that this observation indicates the presence of ten base-stacking planes along the length of the *b*-axis within the unit cell of the crystal. However, in a more positive light, it ought to be noted that this pattern indicates that there must be at least five uniform steps along the direction of the *b*-axis in order to make this pattern.

Fibre diffraction and model building studies (Arnott *et al.*, (1974); Zimmerman *et al.*, (1975)) on polyguanylic acid indicated an axial rise per residue for the polymer of 3.4Å, and strong base-stacking reflections are observed in the d(TG₄T) X-ray diffraction data at this resolution. Thus, assuming an axial rise of 3.4Å, the d(TG₄T) tetraplex would have an approximate length in the direction of the helical axis of 17Å. This assumes the formation of two thymine tetrads at the 5' and 3' ends of the tetraplex, whereas thymine bases have been observed to be conformationally disordered in NMR studies of parallel-stranded tetraplexes (Aboul-ela *et al.*, (1992); Wang and Patel (1993a)).

Whilst no information is available to indicate the organization of the thymine residues within the unit cell, it is reasonable to propose, from the information presented

in Figure 5.4 and geometric considerations, that d(TG₄T) tetraplexes pack in the cell with two tetraplexes stacked end-to-end and approximately aligned with the *b*-axis of the unit cell which has a length of 35.5Å. This length is equivalent to ten helical steps, thereby suggesting that there must be some intermeshing of the tetraplex structures in order to achieve the repeat. Such a packing is consistent with the features of the Patterson synthesis. Two tetraplexes containing thymine tetrads, stacked end-to-end and with a further axial rise of 3.4Å separating the two T₄ tetrads would, with a total length of approximately 37.4Å, be too large to pack reasonably within the cell in this orientation.

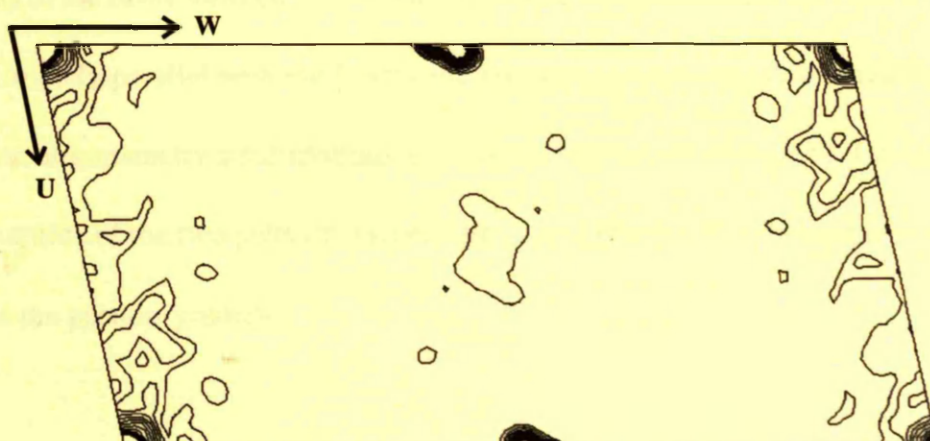


Figure 5.5: Native Patterson Function *uw* Section Calculated At the Cell Origin

The native Patterson function also revealed a strong peak at approximately (0, 0, 0.5), and this is illustrated in Figure 5.5. The peak is also slightly less than half the height of the origin peak. The presence of this peak led to initial concerns that the cell had perhaps been misindexed, as discussed in Chapter 2 Section 2.7, since such a peak is indicative of an additional translational symmetry whilst the space group P1 contains

no symmetry elements of any kind, except for the three principal crystallographic translations. Once the unit cell had been verified, the peak in the native Patterson at (0, 0, 0.5) was interpreted to arise from pseudotranslational symmetry, indicating that a significant proportion of the unit cell contents are related by a translation of approximately one half the length of the c -axis.

Assuming an approximate width of the d(TG₄T) tetraplex of 23 Å from fibre diffraction and model building studies of polyguanylic acid (Arnott *et al.*, (1974); Zimmerman *et al.*, (1975)), the observed peak at (0, 0, 0.5), taken in conjunction with the observations derived from Figure 5.4, facilitates the proposal of a model for the packing contents of the entire unit cell. In this packing model two stacked tetraplexes are aligned approximately parallel with the b -axis and are related to another two tetraplexes with the same alignment by a pseudotranslation of approximately (0, 0, 0.5). Certain structural features of the two pairs of stacked tetraplexes must render them non-equivalent to yield the pseudosymmetry.

5.7 Self-Rotation Function Analysis

With the exception of evidence that the unit cell contains a base-stacking arrangement, the Patterson analysis described in the previous section gives no indication of the structural form adopted by the oligonucleotide d(TG₄T) in the crystal. In particular, the Patterson map gives no indication that the oligonucleotide d(TG₄T) is in fact forming a tetraplex structure. The formation of a tetraplex composed of guanine tetrads would result in the presence of non-crystallographic symmetry within the cell. Such non-

crystallographic symmetry can be detected by the computation of the self-rotation function.

The calculation of the self-rotation function allows the determination of the relative orientation of the molecules in the cell by the comparison of the the native Patterson function with itself. The interatomic vectors relevant to the non-crystallographic symmetry resulting from the arrangement of the molecules in the cell may be considered to be contained within a sphere of radius r . Care should be taken to ensure that r is appreciably shorter than the shortest cell dimension in order to avoid spurious overlaps with the origin peak of the next unit cell in the Patterson. In addition, vectors within at least 2Å of the origin of the Patterson should be omitted so as to exclude the strong origin peak and the considerable noise in its vicinity (Blundell and Johnson (1976)).

Data of a resolution lower than approximately 10Å contain information that is also related to solvent, as well as macromolecular, structure and should be omitted from the self-rotation computation. Terms of too high a resolution may make the self-rotation function difficult to interpret due to the fine structure of the Patterson maps. Generally speaking however, resolution and integration limits have to be varied empirically in order to find the optimal solution with due attention being given to the most persistent features in the self-rotation map.

The self-rotation function for the native d(TG₄T) data was calculated using the program *ALMN* in the *CCP4* suite. The program *ALMN* utilizes Crowther's FFT rotation function (Crowther (1972)) and computes the values of rotation function overlaps using FFT techniques. The position of peaks in the rotation function map is given for all symmetry related positions in both Eulerian and polar angles. The

program deals with all space groups and provides a number of orthogonalisation axes conventions.

Peak No.	Omega (ω)	Phi (ϕ)	Kappa (κ)	Peak Height	RMS Deviation
1	92.6	347.7	179.8	25.7	13.1
2	87.9	166.0	178.6	24.9	12.7
3	87.9	167.7	90.2	21.2	10.8
4	92.1	347.7	90.2	21.2	10.8

Table 5.2: Self-Rotation Peaks For 10.0–4.00Å Data

Self-rotation functions were determined for a variety of resolution and integration limits. Orthogonal axes (X, Y, Z) were defined with the crystal c axis along x , c^*a along y and c^* along z . The self-rotation function was sampled at 5° intervals. The most persistent features of the self-rotation map were two four-fold peaks and two two-fold peaks with orientations coincident to the four-folds. These four peaks were best resolved using data within the resolution range 10.0–4.00Å, with an integration sphere of inner radius 5Å and outer radius 23Å. The four-folds and their corresponding two-folds are summarised in Table 5.2, with the co-ordinates in the spherical polar form.

In the self-rotation map, corresponding to the resolution and integration limits specified above, the *rms* deviation from mean density was 1.96. The two highest peaks, not shown in Table 5.2, corresponded to the origin peaks and were the largest

peaks in the map with a height of 50 – 25.5 times the *rms* deviation. The four peaks shown in Table 5.2 represent the next four highest peaks in the self-rotation map.

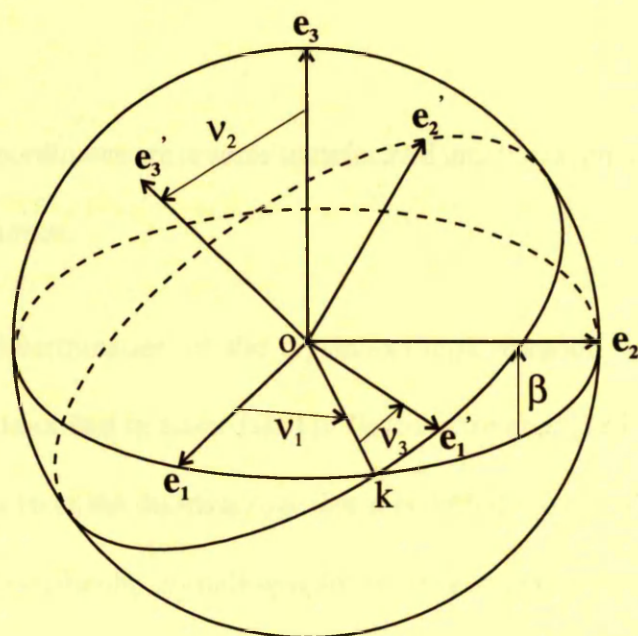
The conventional definition for the Eulerian and spherical polar coordinate systems is presented in Figure 5.6, adapted from Giacovazzo *et al.*, (1992). Figure 5.6(a) illustrates the Eulerian angle convention. Two orthonormal frameworks $\mathbf{E}[\mathbf{O}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$ and $\mathbf{E}'[\mathbf{O}, \mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3]$ are shown. The axis \mathbf{ok} is the intersection of the $(\mathbf{e}_1, \mathbf{e}_2)$ and $(\mathbf{e}'_1, \mathbf{e}'_2)$ planes and is perpendicular both to \mathbf{e}_3 and \mathbf{e}'_3 . The Eulerian system \mathbf{E} may be superimposed to the Eulerian system \mathbf{E}' by three anti-clockwise rotations in the following order:

1. rotate about \mathbf{e}_3 by the angle ν_1 so that \mathbf{ok} and \mathbf{e}_1 are now identical.
2. rotate through ν_2 about \mathbf{ok} which will bring \mathbf{e}_3 into coincidence with \mathbf{e}'_3 .
3. rotate about \mathbf{e}'_3 by ν_3 which then brings \mathbf{e}_1 to \mathbf{e}'_1 and \mathbf{e}_2 to \mathbf{e}'_2 .

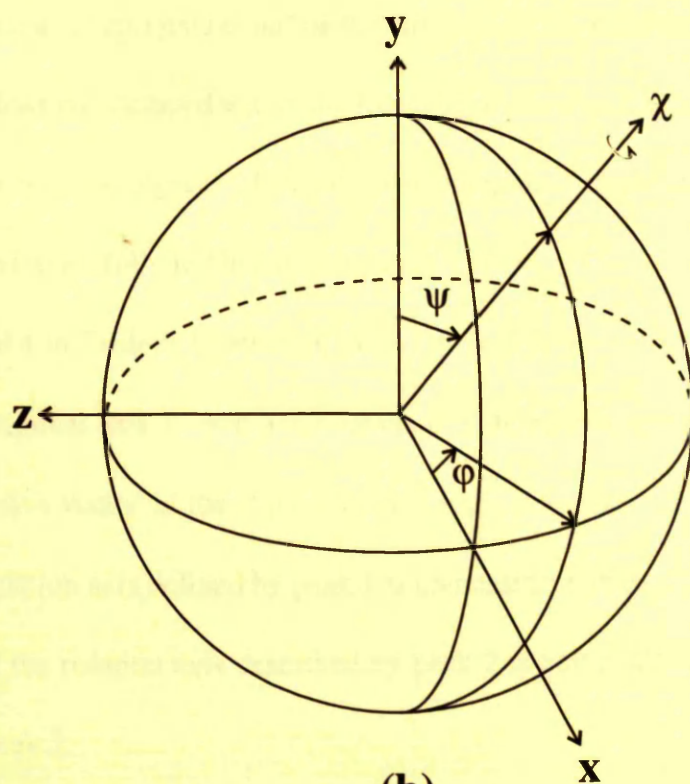
The Eulerian angles can be used in order to calculate, in any crystallographic system represented by the unit cell axes a, b and c , the rotation function corresponding to any desired rotation. This procedure involves three basic steps:

1. Transformation of the coordinates within the crystallographic system described above into coordinates in the Eulerian system \mathbf{E} .
2. The resulting Cartesian coordinates are then transformed into a rotated set of axes. The values of ν_1, ν_2 and ν_3 define this rotation within the Eulerian system and are often referred to as α, β and γ when used to describe the crystallographic

Figure 5.6.1. Illustration of the Eulerian angle convention.



(a)



(b)

Figure 5.6: Definitions of the Eulerian and Polar Coordinate Systems

rotation function. The full range of rotation operations is $0 \leq \nu_1 < \pi$, $0 \leq \nu_2 < 2\pi$ and $0 \leq \nu_3 < 2\pi$.

3. Finally, the coordinates are reverse transformed into the original crystallographic frame of reference.

The algebraic determination of the crystallographic rotation function using the Eulerian system is described in more detail in Giacovazzo *et al.*, (1992). The Eulerian angle system suffers from the disadvantage that it is difficult to visualize. Therefore, it is conventional to describe the crystallographic rotation function in terms of spherical polar coordinates. The spherical polar coordinate system is illustrated in Figure 5.6(b). In this system the rotation about a given axis is defined by the angle kappa (κ). The polar coordinates omega (ω) and phi (ϕ) define the direction of the rotation axis. All possible rotation operations are included within the limits $0 \leq \kappa < 2\pi$, $0 \leq \omega < \pi$ and $0 \leq \phi < \pi$. As for the Eulerian system, the algebra of the rotation function in spherical polar coordinate space is detailed more fully in Giacovazzo *et al.*, (1992).

Peaks 3 and 4 in Table 5.2 define an axis of four-fold symmetry at approximately 12° to the orthogonal axis Y , and are orientated with regards to each other such that one is the negative vector of the other. Peaks 1 and 2 describe two-fold rotation axes such that the rotation axis defined by peak 1 is coincident with the four-fold described by peak 4, and the rotation axis described by peak 2 is coincident with the four-fold described by peak 3.

The detection of four-fold non-crystallographic symmetry in the self-rotation map permits the conclusion that the oligonucleotide d(TG₄T) most probably occupies the

unit cell of the crystal as a four-stranded tetraplex structure containing cyclic guanine tetrads. However, an anti-parallel tetraplex would approximate local four-fold symmetry at the resolution limits employed in the calculation, and hence it cannot be concluded from the self-rotation function alone that the tetraplex is parallel-stranded. The fact that the parallel-stranded form is thermodynamically favoured by the oligonucleotide in solution (Aboul-ela *et al.*, (1992)) does suggest that d(TG₄T) will crystallize in the preferred parallel-stranded conformation.

The reciprocal cell vector c^* is defined as being perpendicular to the unit cell vectors a and b . The unit cell for d(TG₄T) was determined such that the angle between a and b , γ , is very nearly 90° (see Chapter 2 Section 2.7). Therefore, the vector cross product $c^* \times a$ defines an axis which is approximately coincident with the direction of the b -axis in the unit cell. Thus, the orthogonal axis Y is approximately equivalent to the unit cell axis b , and it can be concluded that the four-fold helical axis of the tetraplex is incident to the b -axis of the unit cell at an angle of approximately 12°. This observation is in agreement with the analysis of the native Patterson, detailed in the previous section, which indicated that the direction of the helical axis was approximately coincident with the b -axis.

Two peaks with $\kappa=180^\circ$ can be seen in the self-rotation function which are orientated at approximately 90° to the four-folds defined in Table 5.2. These peaks, which are the next highest peaks in the map after the peaks defined in Table 5.2, have a height of 18.1 (9.2 times the rms deviation) and are orientated at approximately 90° to the four-folds. As such, these peaks may relate the two members of a stacked pair of tetraplexes aligned such that the helical axis is at approximately 12° to the b -axis of the

unit cell. The remaining peaks in the self-rotation function have peak heights of less than 16 (8.1 times the rms standard deviation) and diminish in size with no interpretable features.

The discussed features of the self-rotation function analysis can be clearly illustrated by inspection of the $\kappa=180^\circ$ and $\kappa=90^\circ$ sections of the self-rotation function map. Sections of the self-rotation function map were generated with the *CCP4* program *POLARRFN* and are represented as stereographic projections along the κ -axis in Figures 5.7 and 5.8.

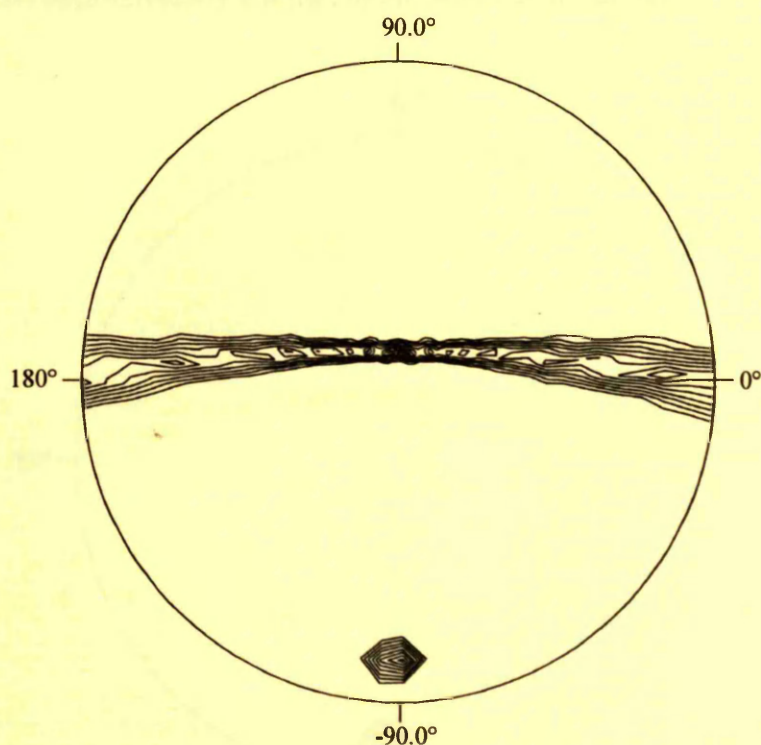


Figure 5.7: Stereographic Projection of the Self-Rotation $\kappa=180^\circ$ Section

Figure 5.7 illustrates the stereographic projection of the $\kappa=180^\circ$ section of the self-rotation map calculated using the X-ray data in the resolution range 10.0–2.0Å. A single peak is clearly visible at a position inclined at approximately 11° to the b -axis of

the crystal unit cell, as determined from the assignment of the orthogonal axes X , Y , Z above. A large 'streak' in the map approximately perpendicular to the crystal cell b -axis shows a succession of two-fold symmetry peaks perpendicular to the b -axis. As previously discussed, the detail of the self-rotation function is generally insufficient to allow the unequivocal determination of molecular features such as nucleic acid chain polarity. Thus, non-crystallographic symmetries observed in the self-rotation function may not be strictly observed at the atomic level. Therefore, the observed 'streak' can at least be said to be consistent with a model for the cell of tetraplexes stacking along a common axis approximately parallel to the b -axis in the crystal.

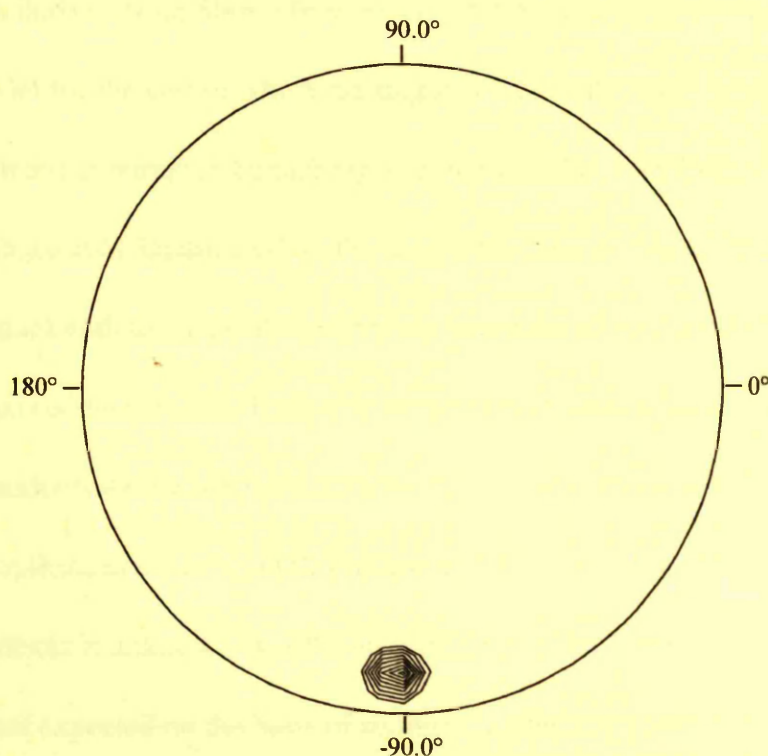


Figure 5.8: Stereographic Projection of the Self-Rotation $\kappa=90^\circ$ Section

The second stereographic projection of the self-rotation function, Figure 5.8, illustrates the $\kappa=90^\circ$ section. This section was also calculated using the X-ray data within

the resolution limits 10.0–2.0Å and the *CCP4* program *POLARRFN*. In this case no ‘streak’ is observed as in Figure 5.7, indicating that the multiple two–folds which produced this feature in Figure 5.7 have no coincident four–fold symmetry. However, a four–fold symmetry peak is clearly observed in the map in Figure 5.8, coincident with the position for the two–fold peak observed in Figure 5.7, at approximately 11° to the crystal *b*–axis. This singular four–fold peak strongly suggests the presence of tetrameric DNA within the crystal and also designates a direction for this tetraplex helical axis, be the tetraplex parallel or anti–parallel stranded.

In conclusion, the native Patterson and self–rotation analyses, taken in conjunction with models derived from fibre diffraction and NMR studies, allow the proposal of a packing model for the cell in which the oligonucleotide d(TG₄T) is assumed to form a parallel–stranded tetraplex containing guanine tetrads. However, this assumption is not unambiguously legitimated by the results of these analyses. In this model, two tetraplexes stack end–to–end with their helical four–fold axes approximately coincident with the *b*–axis of the unit cell. These two tetraplexes are then related to another stacked pair by a pseudocrystallographic translation of approximately (0, 0, 0.5), such that there are four tetraplexes in the irreducible repeating unit of the crystal. The relative polarity of the tetraplexes is unknown as is the arrangement of the thymine residues. Thymine tetrads are not expected on the basis of solution studies and geometric considerations.

5.8 Cross-Rotation Function Analysis

5.8.1 Generation of Search Models for Molecular Replacement

The NMR characterisation of the parallel-stranded tetraplex formed by the oligonucleotide d(TG₄T) (Aboul-ela *et al.*, (1992)) facilitated the generation of a family of tetraplex search models for cross-rotation and translation function analyses. As discussed in Chapter 2 Section 2.1, NMR studies by Aboul-ela *et al.*, (1992) determined that the oligonucleotide d(TG₄T) formed a parallel-stranded tetraplex containing guanine tetrads which exhibited differential stabilization by Na⁺ and K⁺ monovalent cations. Furthermore, it was demonstrated that the individual nucleotides exhibited conformational features similar to the canonical B-DNA topology of double helices (Saenger (1984)), with the β -glycosyl linkage between sugar and base exclusively *anti* and deoxyribose sugars in the C_{3'}-*exo* envelope conformation.

Tetraplex models for molecular replacement were generated by Dr. David G. Norman of the Cancer Research Campaign Nucleic Acid Structure Research Group, Department of Biochemistry, University of Dundee. An initial model was generated based on the conformational details determined in solution by NMR, with the conformational parameters for the individual nucleotides similar to those for canonical B-DNA. At this stage the thymine nucleotides were incorporated into the model as cyclic tetrads at the 3' and 5' ends of the tetraplex, although NMR analyses indicate that these nucleotides are in fact conformationally disordered in solution (Aboul-ela *et al.*, (1992); Wang and Patel (1993a)). The original model was then incrementally over- and underwound by an integral number of degrees per tetrad step, or unit axial rise, to

generate a series of experimental structures for molecular replacement. The extent of the variation in the unit twist per tetrad step was -20° , for the most underwound, to $+20^\circ$, for the most overwound model. A total of 41 models were generated, including the original model derived from the NMR data. The nomenclature adopted for the individual models was $Tpm\varphi$, for the underwound, and $Tp\varphi$ for the overwound models, where φ designates the integral number of degrees of over- or underwind for each individual model. The original model $Tp0$ is depicted in Figure 5.9. In Figure 5.9(a) the tetraplex is depicted side-on to the four-fold helical axis and the four d(TG₄T) strands constituting the tetraplex are coloured red, green, blue and yellow. Figure 5.9(b) presents an axial view of the model down the helical four-fold axis.

5.8.2 The AMoRe Package

Cross-rotation and subsequent translation function computations were carried out using the *AMoRe* package (Navaza (1992)) which combines cross-rotation, translation and rigid-body calculations in an automated environment. The *AMoRe* package is composed of a number of functional program units, namely *SORTING*, *TABLING*, *ROTING*, *TRAINING* and *FITING*. The X-ray diffraction data are formatted for subsequent calculations using the program *SORTING*, whilst the program *TABLING* manipulates the search model co-ordinates and generates a table of continuous Fourier coefficients from the structure. The program *ROTING* calculates the cross-rotation function, whilst *TRAINING* and *FITING* calculate the translation function and rigid-body refinement of the translation function solutions, respectively. The programs *TRAINING* and *FITING* are discussed in the next section.

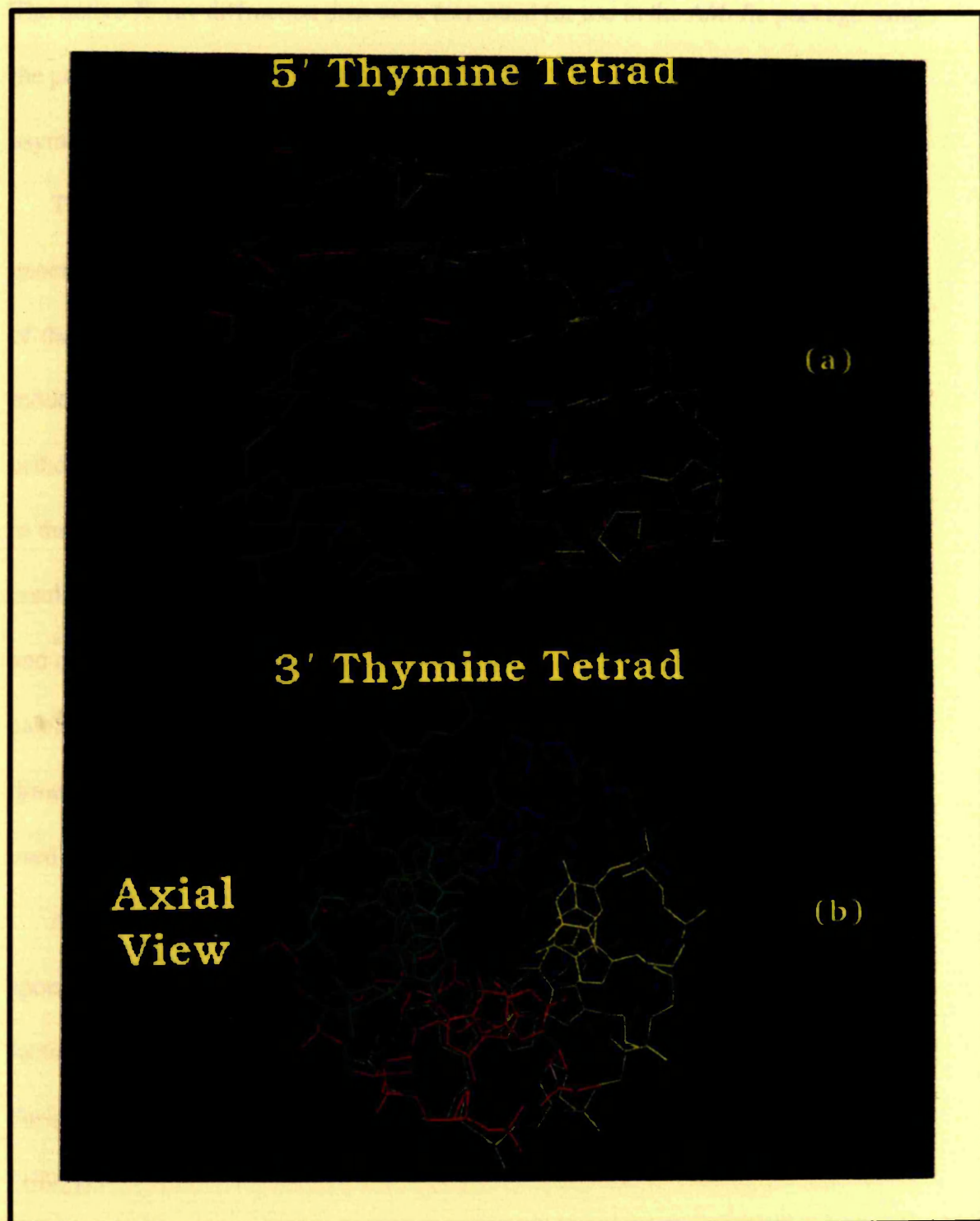


Figure 5.9: The Tp0 Model as used in Molecular Replacement Studies

5.8.3 Manipulation of Search Model Coordinates

The native X-ray diffraction data were formatted for use in the *AMoRe* package using the program *SORTING* which produces a packed file containing h , k , l , and F in the P1 asymmetric unit.

The program *TABLING* was then applied to each search model in turn in order to generate a table of continuous Fourier coefficients. In this procedure, the coordinates of the search model are translated so that their centre of gravity is at the origin. The model is then rotated so that the principle axes of inertia of the model are parallel to the orthogonal axes given as $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$. By fitting all search models to the rotated coordinates output from the first model it is then possible to compare the results from the various models employed. A minimal box with dimensions a_{\min} , b_{\min} and c_{\min} is then found which just covers the coordinates. The structure factors are then calculated from the modified coordinates for a cell in which all angles are 90° , and with dimensions Ka_{\min} , Kb_{\min} and Kc_{\min} , where K is a scale factor which defines the grid used for the structure factor calculations and is selected by the user.

Atomic coordinates are used only once to compute the Fourier coefficients corresponding to the model electron densities of the search molecules. Subsequent structure factor calculations for each orientation of the search model, and for each iteration during rigid-body refinement, are calculated using an interpolation procedure (Navaza (1992)).

Structure factors were prepared for each search model of the d(TG₄T) tetraplex in turn using the P1 cell dimensions of the native crystal. The orthogonalisation

convention was set with the crystal c axis along X , c^*a along Y and c^* along Z , which is the Brookhaven default. The temperature factors of all the non-hydrogen atoms in the model files were set to 10\AA^2 and the finest sampling of reciprocal space was employed. The resolution for the interpolation of structure factors was set to a maximum of 2.00\AA .

5.8.4 Cross-Rotation Function Calculation using ROTING

Cross-rotation functions were calculated for each tetraplex search model using the program *ROTING*. The formulation of the *AMoRe* rotation function is based upon that of Crowther (1972). However, the *AMoRe* fast rotation function utilizes a numerical integration procedure in place of Fourier-Bessel expansions (Navaza (1987)).

The fast rotation function was calculated using all of the X-ray diffraction data between 10.0 and 3.00\AA . Structure factors for each model were generated in a large cell with dimensions $a = b = c = 100\text{\AA}$, with all angles defaulting to 90° . An integration sphere of radius 14.0\AA was employed for the cross-rotation search, with 14.00\AA corresponding to the maximal distance from the centre of mass of the model, as determined from *TABLING*. The whole of the asymmetric unit of rotation space was explored at 2.5° on the Eulerian angles alpha (α), beta (β) and gamma (γ). The orthogonalization convention employed throughout was the Brookhaven default, coded as number one.

ROTING produces a table of the cross-rotation function solutions in terms of their Eulerian angles and provides a correlation coefficient for each individual cross-rotation solution. The Eulerian angles can be modified in the program in order that they can be

directly applied to the search model for subsequent translation function calculations in *TRAINING*.

Model	Peak No.	Alpha (α)	Beta (β)	Gamma (γ)	Correlation Coefficient
Tp4	1	266.3	79.5	240.8	0.425
Tp4	2	89.0	105.8	252.1	0.384
Tp4	3	89.6	105.5	73.5	0.353
Tp4	4	267.8	81.3	328.9	0.349
Tp5	1	88.1	103.2	75.5	0.375
Tp5	2	88.6	104.1	250.6	0.363
Tp5	3	265.2	82.1	241.0	0.351
Tp5	4	265.4	81.7	326.6	0.345
Tpm5	1	88.4	106.9	249.7	0.306
Tpm5	2	270.3	80.3	241.6	0.295
Tpm5	3	90.4	106.2	76.3	0.268
Tpm5	4	270.3	81.5	327.5	0.260

Table 5.3: *AMoRe* Cross-Rotation Peaks, 10.0–3.00Å Data

The cross-rotation solutions for the various search models of the d(TG₄T) tetraplex exhibited a number of striking features. In nearly all cases the top four or five cross-rotation solutions, on the basis of the correlation coefficients, clustered together with a substantial drop-off in the correlation coefficient between the lowest ranking solution of the cluster and the next solution in the list. This result inspired initial confidence that the four tetraplex model proposed for the cell was possibly correct, and that the search models were accurate enough for molecular replacement investigation of the packing

model. Of all the models analysed in this manner, the highest correlation coefficients were obtained for the model Tp4. The four top ranking solutions for Tp4, along with those of Tp5 and Tpm5 for comparison, are shown in Table 5.3.

Inspection of Table 5.3 reveals another interesting aspect of the cross-rotation solutions found. For all three of the models shown in Table 5.3, and for most of the models for which cross-rotation functions were calculated, the top four cross-rotation solutions separate into two sets. In each case, the two members of a set have corresponding values of α and β , with the γ angles being related by a difference that is an approximate multiple of 90° . It was suspected at this stage that the relationship between the γ angles might correspond to the four-fold rotation axis of the tetraplex. This can be seen for model Tp4 in which the peaks 1 and 4, as well as 2 and 3, correlate in the described manner. Inspection of the cross-rotation solutions using computer graphics confirmed this suspicion. In each case the direction of the helical four-fold axis was approximately the same for each tetraplex. Tetraplexes with equivalent values of α and β possessed the same polarity but had opposite polarity to the other pair. Furthermore, tetraplexes with the same polarity superimposed without any need for any further rotational operations.

It was decided on the basis of these results to proceed with the computation of the translation function, rather than making any attempts to improve the model for further cross-rotation analyses. On the basis of the correlation coefficients observed for the various search structures, the model Tp4 was chosen as the prime candidate for this experiment. The calculation of the translation function for Tp4 and the results obtained are described in the next section.

5.9 Translation Function Analysis

5.9.1 The Programs TRAINING and FITING

Translation functions are calculated in the *AMoRe* package using the program *TRAINING*. The *CCP4* implementation of the package utilizes the method of Crowther and Blow (1967), although other translation function methods are being incorporated into the system (Navaza (1992)). The output from *TRAINING* includes the correlation coefficient and the R-factor corresponding to the highest peaks of the translation function employed. Good solutions should give high correlation coefficients and low R-factors. The *TRAINING* program offers the option of fitting and fixing a certain number of positioned fragments in cases where the translation function is computed for the entire unit cell of the crystal.

The program *FITING* performs fast rigid-body refinement upon the orientated and translated contents of the unit cell after analysis with *ROTING* and *TRAINING*. The Fourier coefficient for the individual search models in the cell are placed on a common scale and the function given in equation 5.17 is minimized

$$Q = \sum_{\mathbf{H}} \left\{ |F^{obs}(\mathbf{H})| e^{-B|\mathbf{H}|^2} - \lambda \left| \sum_{n=1}^N F_n[\alpha_n, \beta_n, \gamma_n, x_n, y_n, z_n](\mathbf{H}) \right| \right\}^2 \quad (5.17)$$

The minimization is alternately performed with respect to the positional parameters of each search model, whilst the positional parameters of the other models are kept fixed. During each iteration of the refinement procedure, the unique scale factor λ and

the overall temperature factor B are chosen so as to minimize the function Q . (Navaza (1992)).

5.9.2 Translation Function Solutions for d(TGGGGT)

Translation functions for the oligonucleotide d(TG₄T) were calculated in the native crystal P1 cell using the program *TRAINING* and the top four orientation solutions for the model Tp4 as determined by *ROTING*. All data in the resolution range 10.0–3.00 Å were employed and the calculation was performed using the function of Crowther and Blow (1967). This takes each orientation solution in turn and searches for the best translational X_i , Y_i and Z_i for this orientation. Translation function solutions were evaluated using the output correlation coefficients and R-factors for each solution.

The proposed packing model of the P1 cell described in Sections 5.6 and 5.7, in which four tetraplexes occupy the irreducible asymmetric unit, results in a total of twenty four distinct packing arrangements for the four orientation solutions although, because there is no unique origin in the cell, there are many redundant solutions among the twenty four possibilities. The translation search procedure therefore took the form of a branching structure. In this approach, each orientation was fixed at the origin of the P1 native cell, and the translation function for the other orientations was computed. For each resulting branch, the structure at the origin and the other orientation translated relative to it were positionally fixed. The translation function was then computed for the other orientations. Finally, this procedure was repeated in the situation where three orientations were positionally fixed and the translation function was computed for the final orientation. This procedure resulted in twenty four solutions for four tetraplexes

in the cell, which could then be compared on the basis of their correlation coefficients and R-factors.

The initial calculation of the case where only one orientation was fixed at the origin appeared to be completely dominated by the strong vectors at approximately (0, 0, 0.5), as described in Section 5.6, in relation to Figure 5.5, which it was proposed resulted from pseudosymmetry relating axial tetraplex pairs packed along the *b*-axis. This provoked initial concerns that the translation function calculation might prove to be highly sensitive to this particular crystal characteristic. However, the various two model cases did display differing correlation coefficients and R-factors, and the translations involving two and three fixed models yielded best solutions with translations of (0, 0.5, 0) and (0, 0.5, 0.5) as anticipated for the proposed packing arrangement. The various packing models were then refined using the individual tetraplexes as rigid bodies using the program *FITING*.

Alpha (α)	Beta (β)	Gamma (γ)	Fractional Translation (X) (Y) (Z)			Correlation Coefficient	R-Factor
266.3	79.5	240.8	-0.0016	-0.0071	-0.0007	0.775	—
267.8	81.3	328.9	0.0175	0.0011	0.5132	0.537	0.519
89.6	105.5	73.5	0.0713	0.5348	0.4385	0.540	0.514
86.0	105.8	252.1	0.0602	0.5109	0.9293	0.631	0.487

Table 5.4: *AMoRe* Translation Peaks, 10.0–3.00Å Data

The best translation function solutions for the four orientations of the Tp4 model

are shown in Table 5.4. This solution produced the highest correlation coefficient and lowest R-factor of any of the solutions obtained by the branching procedure described. The Eulerian angles are given for each solution along with the fractional translations applied to each solution to give the final packing arrangement in the cell. The correlation coefficients provided are such that 0.775 is the correlation coefficient for the single model, 0.537 is the two model result, 0.540 the three model result and 0.631 the four model result. The R-factor for the four model result was 0.487 for data in the resolution range 10.0–3.00Å.

Figure 5.10 illustrates the packing arrangement of the four tetraplexes in the unit cell. The individual tetraplexes are lettered and coloured for descriptive purposes. The direction of the *b*- and *c*-axes are indicated by the yellow arrows at the base of the figure. The packing arrangement derived from the best cross-rotation and translation function solutions shows four parallel-stranded tetraplexes occupying the asymmetric unit as anticipated. These are organized as two pairs, (A, B) and (C, D), with the four-fold rotation axis of each tetraplex approximately coincident with the *b*-axis of the cell. Tetraplex A(red) is the (-0.0016, -0.0071, -0.0007) translation solution, B(green) the (0.0602, 0.5109, 0.9293) solution, C(yellow) the (0.0175, 0.0011, 0.5132) solution and D(blue) the (0.0713, 0.5348, 0.4385) solution. The axial pairs (A, B) and (C, D) are orientated such that they stack with opposite polarity and there is stacking between tetraplexes within the cell at the 5' ends. Tetraplex A translates onto tetraplex C such that no additional reorientation of either tetraplex is required. The same holds for tetraplexes C and D. As such this packing arrangement satisfies the observations derived from the native Patterson and the self-rotation function, at least with regards

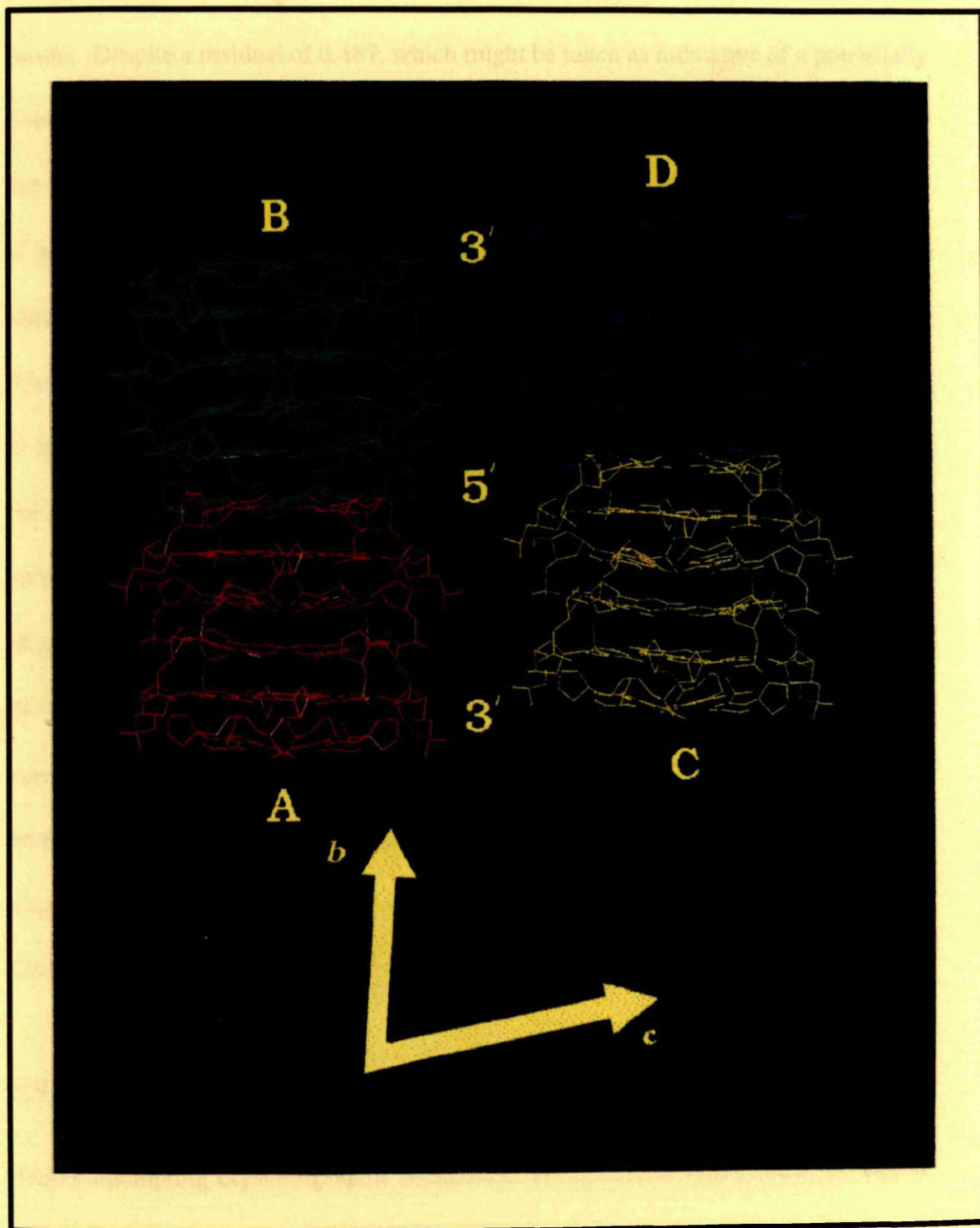


Figure 5.10: Packing Arrangement of the Tp4 Tetraplexes in the Unit Cell

to the guanine residues.

However, the packing model also illustrates the limitations inherent in the search model. Despite a residual of 0.487, which might be taken as indicative of a potentially correct solution (Fitzgerald (1991)), the molecules interpenetrate at the 5' ends within the cell, and at the 3' ends, not shown in the figure, by overlap of the corresponding 5' and 3' thymine tetrads. The demonstration by NMR that thymine residues were disordered in solution for parallel-stranded DNA tetraplexes (Aboul-ela *et al.*, (1992); Wang and Patel (1993a)) would suggest that such an arrangement would not be adopted in the crystal. This can also be deduced from the molecular replacement solution for the Tp4 model. Approximate geometrical considerations, outlined in Section 5.6, anticipated that a packing problem such as that outlined above might occur i.e. ten steps would fit in the cell, whilst two tetraplexes, including T₄ tetrads, would generate twelve steps. This model was used, however, because initially it was thought that the thymines might intermesh comfortably, although it was now clear that they do not. Molecular replacement calculations using the latest version of the *AMoRe* suite, which also includes a packing function, produced an identical result (J. Navaza, Personal Communication).

5.9.3 Progression from the Molecular Replacement Solution

Before attempting crystallographic refinement, the molecular replacement procedure was repeated for an improved molecular model. Such a model was provided by the removal of the thymine residues from the original Tp4 model. Thus, the search model contained no assumptions about the arrangement of the thymine residues. This model

is shown in Figure 5.11, where the four DNA strands are coloured blue, red, yellow and green. Figure 5.11(a) shows the model side-on to the four-fold helical axis, whilst Figure 5.11(b) presents an axial view.

The cross-rotation and translation functions were calculated using this model and the *AMoRe* package as previously described. Similar cross-rotation and translation function solutions were found for the model as those calculated for the model Tp4. The final best solution is illustrated graphically in Figure 5.12.

The labelling A(red), B(green), C(yellow) and D(blue) for the individual tetraplexes corresponds to those used in Figure 5.10 for the model containing the thymine tetrads. The direction of the *b*- and *c*-axes are indicated by the yellow arrows at the base of the figure. The tetraplexes labelled A and C in Figure 5.12 superimpose on the tetraplexes A and C in Figure 5.10 with respect to their common atoms. The guanine tetrads of tetraplexes B and D in Figure 5.12 are related to those of B and D in Figure 5.10 by an axial translation of approximately one tetrad repeat towards A and C along their respective, shared axes, and a further rotation about these four-fold axes so that the phosphates, the strongest scatterers in the tetraplex, coincide.

Both of the packing models in Figure 5.10 and 5.12 are indistinguishable on the basis of the crystallographic residual. Furthermore, rigid-body refinement of the model in Figure 5.12 using *XPLOR* (Brünger (1993)), which is described in detail in the next chapter, demonstrated that the guanine tetrads of tetraplexes B and D could translate and rotate such that they superimposed with the tetrads in tetraplexes B and D in Figure 5.10, but with no change in the crystallographic residual.

As such, the lack of any knowledge about the arrangement of the thymines in the

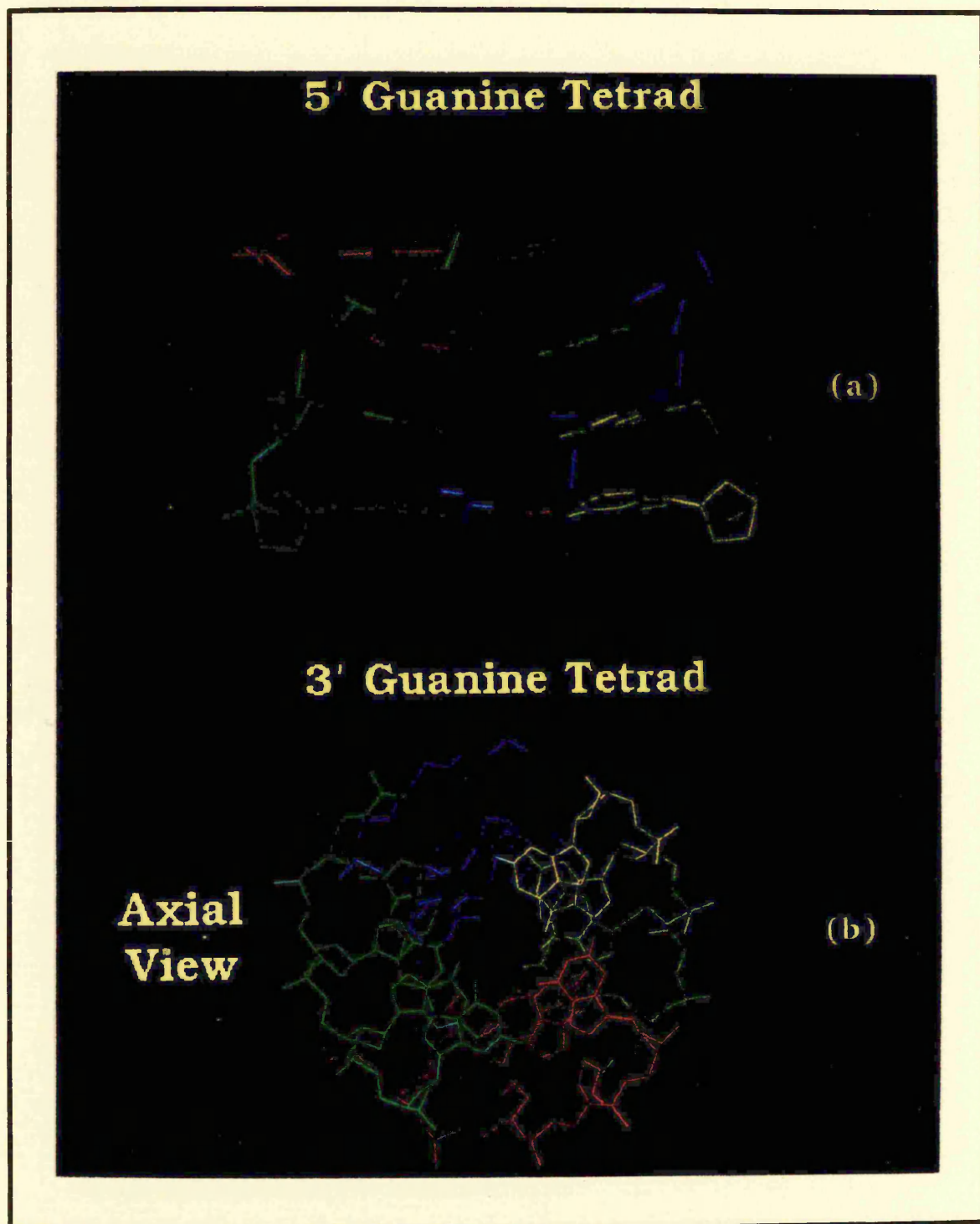


Figure 5.11: The Model Tp4-T as used in Molecular Replacement Studies

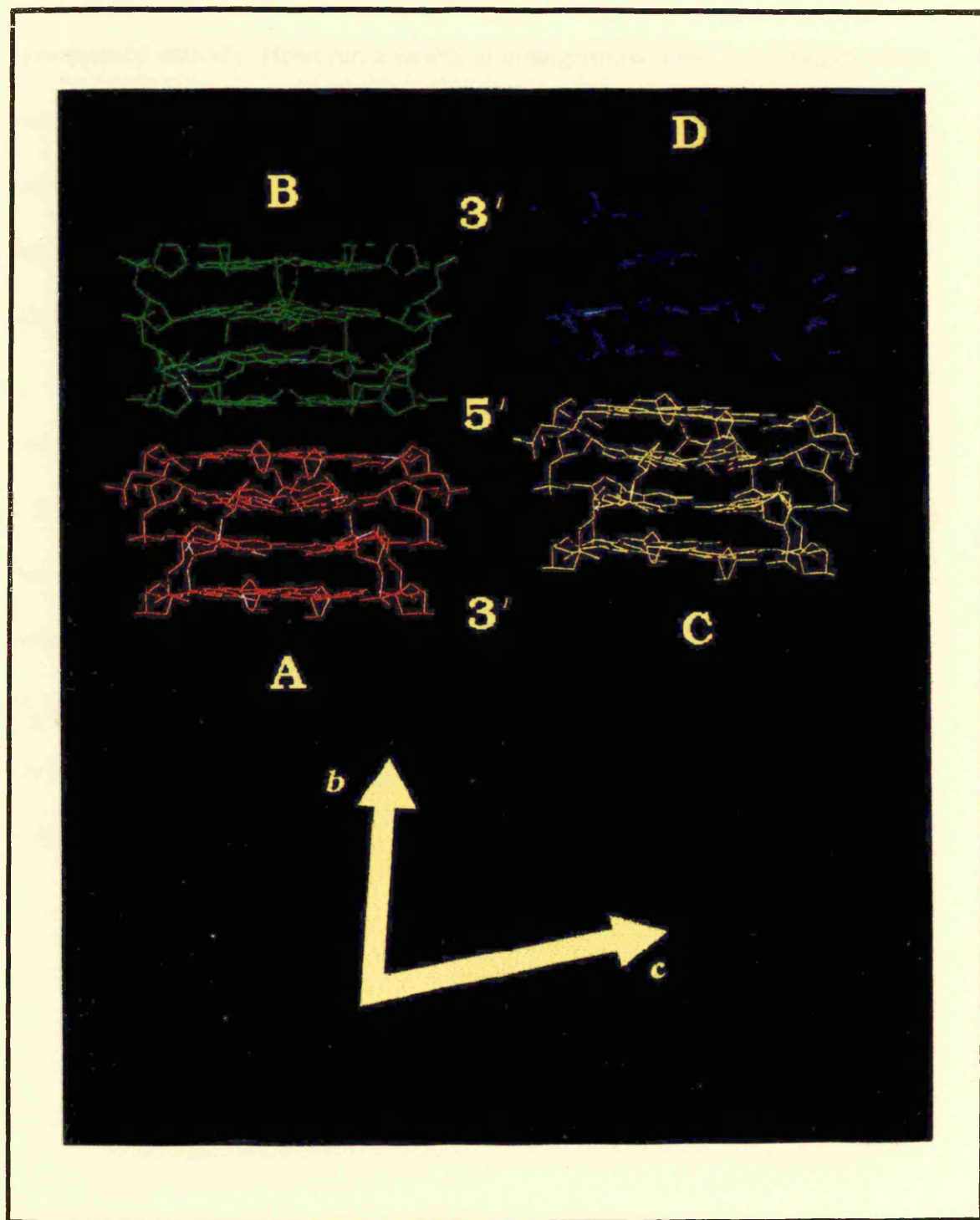


Figure 5.12: Packing Arrangement of the Tp4-T Tetraplexes in the Unit Cell

cell led to an indeterminacy regarding the relative positioning and orientation of the guanine tetrads. Clearly, thymine tetrads at both the 3' and 5' end of the tetraplexes now seemed unlikely. However, a variety of arrangements could be envisaged which would most probably involve the displacement of some or all of the thymine residues into the surrounding solvent channels. Thymine residues might interdigitate at either the 3' or 5' interface between the tetraplexes, or possibly form T₄ tetrads at certain selective positions. None of these arrangement could be effectively modeled.

The correct relative orientation and positioning of the tetraplexes was ascertained by the refinement procedure. It later transpired that the packing model shown in Figure 5.12 was correct. However, the arrangement of the thymine residues being unknown, the model in Figure 5.10 was employed in the refinement procedure since some planar arrangement of thymines was anticipated. Despite the fact this model was later seen to be in error, it still yielded the correct solution in a refinement procedure that exploited the high-resolution of the X-ray diffraction data collected. The derivation of this result is described in Chapter 6.

Abstraction today is no longer that of the map, the double, the mirror or the concept. Simulation is no longer that of a territory, a referential being or a substance. It is the generation by models of a real without origin or reality: a hyperreal. The territory no longer precedes the map, nor survives it. Henceforth, it is the map that precedes the territory – **PRECESSION OF SIMULACRA** – it is the map that engenders the territory and if we were to revive [Borges'] fable today, it would be the territory whose shreds are slowly rotting across the map. It is the real, and not the map, whose vestiges subsist here and there, in the deserts which are no longer those of the Empire, but our own. *The desert of the real itself.*

Jean Baudrillard, *The Precession of Simulacra*

Our taverns and our metropolitan streets, our offices and furnished rooms, our railroad stations and our factories appeared to have us locked up hopelessly. Then came the film and burst this prison-world asunder by the dynamite of a tenth of a second, so that now, in the midst of its far-flung ruins and debris, we calmly and adventurously go travelling. With the close-up, space expands, with slow motion, movement is extended . . . Evidently a different nature opens itself to the camera than opens to the naked eye – if only because an unconsciously penetrated space is substituted for a space consciously explored.

Walter Benjamin, *Illuminations*

In the essential movement of the Spectacle, which consists of taking up all that existed in human activity in a fluid state so as to possess it in a congealed state as things which have become the exclusive value by their formulation in the negative of lived value, we recognise our old enemy, the commodity, who knows so well how to seem at first glance something trivial and obvious, while on the contrary it is so full of metaphysical subtleties. This is the principle of commodity fetishism, the domination of society by 'intangible as well as tangible things', which reaches its absolute fulfillment in the spectacle, where the tangible world is replaced by a selection of images which exist above it, and which simultaneously impose themselves as the tangible par excellence.

Guy Debord, *The Society of the Spectacle*

We have all had enough of hearing about the death of God, of man, of art and so on since the nineteenth century. What in fact happened was simply the progressive disintegration of a faith in perception founded in the Middle Ages, after animism, on the basis of the unicity of divine creation, the absolute intimacy between the universe and the God-man of Augustinian Christianity, a material world which loved itself and contemplated itself in its one God. In the West, the death of God and the death of art are indissociable; *the zero degree of representation* merely fulfilled the prophecy voiced a thousand years earlier by Nicephorus, Patriarch of Constantinople, during the quarrel with the iconoclasts: 'If we remove the image, not only Christ but the whole universe disappears.'

Paul Virilio, *The Vision Machine*

Chapter 6

Structure Refinement

6.1 Refinement in Macromolecular Crystallography

In the previous chapter, the problem of obtaining a three dimensional model of the contents of the crystallographic unit cell was discussed with regards to the proposed parallel-stranded tetraplex formed by the oligonucleotide d(TG₄T). In this particular case, as a consequence of the availability of a search model for the tetraplex derived from NMR studies (Aboul-ela *et al.*, (1992)), the molecular replacement method proved to be the more instructive approach.

In general, whether the initial model of the packing of the macromolecule under study in the unit cell is derived by molecular replacement techniques, or by other approaches such as multiple isomorphous replacement (MIR) or anomalous scattering contributions, the coordinates of the atomic positions obtained are fairly inaccurate. Typically, the accuracy of these atomic coordinates is likely to be no better than 0.5Å, and may well be as large as 2.0Å. This accuracy is dependent upon the quality of the

packing solution and hence the validity of the search model, in the case of molecular replacement, or on the electron density maps calculated using phases derived from other methods such as those listed above.

Despite the appreciable degree of inaccuracy encountered in the initial three dimensional models of macromolecules, such models can be useful in the determination of the global features of the structure – for example, chain folding and the interrelation of distinct domains within the structure. However, in order to obtain a detailed and meaningful interpretation of the chemistry of the macromolecular assembly, it is essential to refine the atomic parameters against the X-ray diffraction data available.

6.1.1 Definition of Refinement

The definition of refinement, as understood in X-ray crystallography in general, refers to the procedure whereby one adjusts the parameters of the molecular model with the objective of improving the agreement between the structure factor amplitudes (F_o) or intensities of the experimentally derived X-ray diffraction data and the corresponding values (F_c) calculated from the parameters of the model. Understood in this way, the model parameters which are manipulated are the atomic positions (x_j, y_j, z_j) and the thermal parameters of the j atoms within the unit cell of the structure. A simplified schematic illustration of this process, with reference to the least-squares method of crystallographic refinement, is depicted in Figure 6.1.

Inspection of Figure 6.1 reveals that this process is an iterative one. The individual atomic parameters are altered from one cycle to the next and the corresponding calculated phases are derived from the modified model. The refinement procedure is

understood to have converged when the shifts applied to the parameters are deemed to be insignificant and oscillate about a minimum. Clearly, the refinement procedure can serve to test the validity of a solution of a macromolecular structure since an incorrect solution will not necessarily improve significantly with refinement. Likewise, if the atomic parameters of the starting model differ too greatly from their true values, the refinement procedure may converge to a false local energy minimum and not the overall global minimum of the structure.

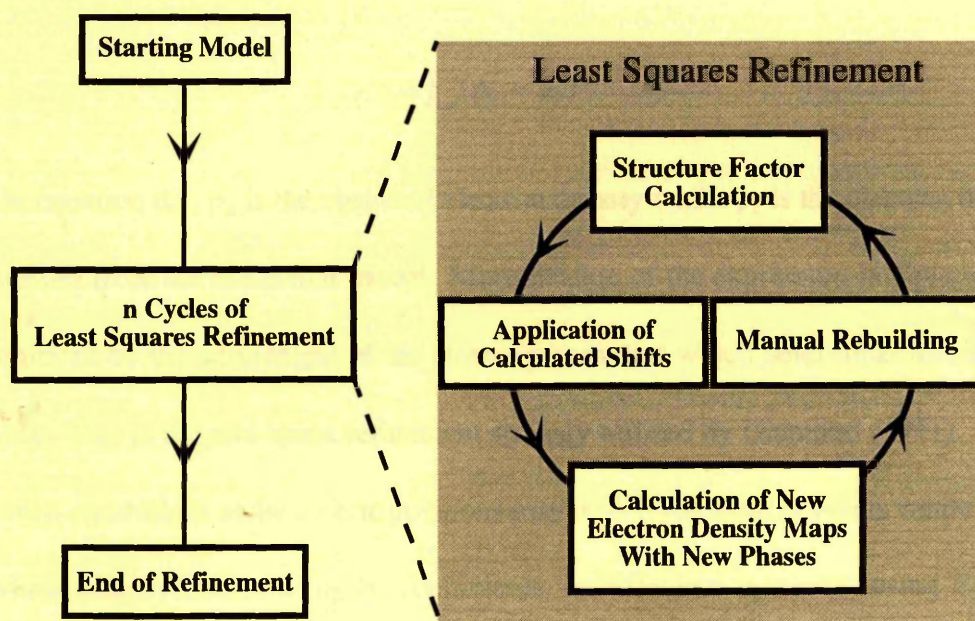


Figure 6.1: Illustrative Overview of the Refinement Procedure

Within macromolecular crystallography the term refinement may serve to refer to a variety of different methodological approaches to the key problem of obtaining optimal parameters for the atomic model (reviewed in Jensen (1985) and Sussman (1991)). Amongst these methods, the principle of least-squares minimization constitutes the basis of the most widely used approach to macromolecular refinement employed at the

current time.

6.1.2 The Method of Least-Squares Refinement

Approaches to the problem of refinement are divided into two categories, real space refinement and reciprocal space refinement. In the case of real space refinement, the model of the macromolecule under study is iteratively improved by the minimisation of the quantity expressed in equation 6.1.

$$\sum (\rho_o - \rho_c)^2 \quad (6.1)$$

In equation 6.1, ρ_o is the observed electron density whilst ρ_c is the electron density calculated from the molecular model. Minimization of the expression in equation 6.1 is achieved by the adjustment of the atomic parameters which determine the electron density. This is the real space refinement strategy utilised by Diamond (1971). Alternatively, corrections to the atomic positions may be derived from slope/curvature ratios in whole map syntheses using F_o coefficients, or difference syntheses using $F_o - F_c$ coefficients (Freer (1985)).

In the case of refinement in reciprocal space, the molecular model is improved by the iterative minimisation of the differences between the observed (F_o) and calculated (F_c) structure factor amplitudes. This method proceeds by the adjustment of the atomic parameters which are used in the calculation of the structure factors of the model. The quantity minimised in the reciprocal space refinement is given in equation 6.2.

$$\phi = \sum_{hkl} w_{hkl} (F_o - F_c)^2 \quad (6.2)$$

In equation 6.2, F_o is the observed structure factor amplitude derived from the experimental X-ray data; F_c is the structure factor amplitude calculated from the molecular model; and w_{hkl} is a weighting parameter. The summation is over all individual observed reflection with Miller Indices hkl . The calculated structure factor amplitude, F_c , is a function of the n independent atomic parameters p . These parameters can be the scale factor relating F_o and F_c , the positional parameters (x_j, y_j, z_j) and the thermal parameters of the j individual atoms of the structure. In accordance with the principle of least-squares, the optimal values for the refinement parameters are those which minimise the quantity ϕ in equation 6.2.

Minimisation of the quantity ϕ is achieved by taking the partial derivative of the function with respect to each individual refinement parameter p in turn and equating each derivative to zero. This results in the derivation of a set of n normal equations, the form of which is given in equation 6.3.

$$\frac{\delta \phi}{\delta p_i} = \sum_{hkl} w_{hkl} (F_o - F_c) \frac{\delta F_c}{\delta p_i} = 0 \quad (6.3)$$

The expression for F_c , given later in this section, is a non-linear function and hence the normal equations represented by equation 6.3 cannot be solved directly. Consequently, these normal expressions are linearised by approximating the value for F_c to a Taylor series, neglecting the contribution of second and higher derivatives. The resultant expansion for F_c is given in equation 6.4.

$$F_c = F(p_o) + \sum_j \frac{\delta F_c}{\delta p_j} (p_j - p_o) \quad (6.4)$$

In equation 6.4, p_o is an approximate value for the parameter p_j ; $(p_o - p_j)$ is the parameter shift Δp_j ; and $\delta F_c / \delta p_i$ is evaluated at p_o . Incorporating the Taylor expansion for F_c , the normal expression given in equation 6.3 assumes the form shown in equation 6.5, or alternatively, equation 6.6.

$$\frac{\delta \phi}{\delta p_i} = \sum_{hkl} w_{hkl} \left(F_o - F(p_o) - \sum_j \frac{\delta F_c}{\delta p_j} \Delta p_j \right) \frac{\delta F_c}{\delta p_i} = 0 \quad (6.5)$$

$$\sum_{hkl} \sum_j w_{hkl} \frac{\delta F_c}{\delta p_i} \frac{\delta F_c}{\delta p_j} \Delta p_j = \sum_{hkl} w_{hkl} (F_o - F(p_o)) \frac{\delta F_c}{\delta p_i} \quad (6.6)$$

The values of the parameter shifts which are derived from the normal equations outlined above will result in new, improved approximations for the atomic parameters. Since the function for F_c is non-linear and an approximation, the least-squares method does not converge in a single cycle. The procedure is therefore an iterative one and, as a result, refinement cycles are repeated until convergence is obtained. Since the radius of convergence for the least-squares calculations is no more than about 1\AA , the refinement may converge to a local energy minimum. This stresses the importance of the accuracy of the initial atomic parameters of the macromolecular model.

Alternate inspection of the electron density of the model during the refinement procedure facilitates manual adjustment of the molecular model using computer graphics. The manual readjustment of the molecular model to give a better fit with the electron density map may lift the model out of a local minimum, hence allowing the refinement

to continue. A combination of alternate rounds of automatic refinement and manual remodelling is often an effective strategy in approaching a global minimum for the refined model.

6.1.3 Calculation of Electron Density

The X-ray diffraction pattern obtained from a single crystal of a molecular structure is the basis of the experimental determination of the molecular structure itself. This structure occupies the regular, repeating unit cell which makes up the crystal. The relationship between the diffraction pattern and the molecular structure, the latter understood to be the electron density of the structure, is expressed as a Fourier transform. The diffraction pattern is the Fourier transform of the crystal structure. Conversely, the electron density of the structure is the inverse Fourier transform of the diffraction pattern. The Fourier transformation relating the electron density of the cell to the X-ray diffraction data is given in equation 6.7

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl) \exp(i\alpha(hkl)) \exp(-2\pi i(hx + ky + lz)) \quad (6.7)$$

Equation 6.7 shows the Fourier transform expression for the electron density ρ , given as a function of the coordinates of the crystallographic unit cell, x , y and z . The Miller indices h , k and l identify the individual crystallographic reflections which are obtained experimentally in the X-ray diffraction experiment. $F(hkl)$ is the structure factor amplitude of each individual reflection and is obtained experimentally from

the diffracted intensities, as described in Chapter 4. In the experiment, only the intensities of the individual reflections can be measured. The phase angle information for the reflections, $\alpha(hkl)$ in equation 6.7, is inaccessible from the native X-ray diffraction data alone for a macromolecular structure. Without this phase information it is impossible to determine the structure of the molecule under study. This is known as the phase problem and it constitutes the central problem in the determination of any crystal structure (Blundell and Johnson (1976)). Attempts to tackle this problem with regards to the crystal structure of the oligonucleotide d(TG₄T) are discussed in Chapter 5.

The structure factors of the model structure can be calculated from the atomic parameters of the model itself. Hendrickson (1985) has expressed this relationship in the form given in equation 6.8.

$$F_c(hkl) = K \sum_j f_j(s_{hkl}) \exp(-B_j s_{hkl}^2) \exp[2\pi i(hx_j + ky_j + lz_j)] \quad (6.8)$$

In equation 6.8, $F_c(hkl)$ is the model derived structure factor of the X-ray reflection with Miller indices hkl ; K is a scale factor; f_j is the atomic scattering factor of atom j ; B_j is an isotropic temperature parameter related by $B = 8\pi\overline{u^2}$ to the mean square displacement, $\overline{u^2}$, of the atom from its average position; x_j , y_j , and z_j are the fractional coordinates of atom j ; and $s_{hkl} = \sin \theta / \lambda$. The calculated structure factor amplitudes F_c can be calculated from $\sqrt{A_c^2 + B_c^2}$, where $A_c = \sum f_j \cos 2\pi(hx_j + ky_j + lz_j)$, and $B_c = \sum f_j \sin 2\pi(hx_j + ky_j + lz_j)$.

During the refinement procedure, graphical inspection of electron density maps and the corresponding refined model facilitates the manual readjustment of the model to give a better fit to the observed electron density. This assists in lifting the model out of a local energy minimum and hence allowing the refinement to continue. It is common practice to calculate two types of electron density maps using the observed structure factor amplitudes, F_o , and the calculated structure factor amplitudes, F_c . In the first type of map, $F(hkl)$ in equation 6.7 is replaced by the coefficients $|F_o - F_c|$, which corresponds to a Fourier-difference synthesis. The second type of map may have coefficients $(2F_o - F_c)$ or $(3F_o - 2F_c)$ which correspond to a combination of a Fourier electron density map and a difference-Fourier synthesis. Both types of map are useful in detecting movements in atomic positions necessary to improve the fit of the model with the electron density.

When a structure solution has been obtained using only the molecular replacement method, as is the case for the d(TG₄T) crystal structure, the phases used in the electron density syntheses are derived solely from the molecular packing model. In this situation, the value of the Fourier syntheses described above lies in their ability to allow correction for any erroneous bias which may be inherent in the model at any time during the refinement procedure. Rectification of the model in this way permits the refinement to proceed towards the global minimum and, ultimately, towards a more reliable molecular structure, based upon the experimental X-ray diffraction observations.

6.1.4 Resolution of Macromolecular Structures

For macromolecular structures, as opposed to small molecules, significant problems are encountered due to the disorder which is often inherent in atomic positions in the crystalline state, and which may be static or dynamic in nature. In particular, such disorder can often be found in surface groups which protrude into the surrounding solvent channels which also occupy the crystal volume, and which are characteristic of macromolecular crystals in general. As a consequence, the resolution, and hence number of observed diffraction intensities, will be strictly limited since atomic disorder leads to a diminution of observed diffraction intensities with increase in the angle of diffraction. Indeed, whilst small molecule crystals can commonly be expected to diffract to a resolution of 0.7\AA or better, crystals of macromolecules typically diffract to a limit in the range $3.0\text{--}2.5\text{\AA}$, and sometimes 2.0\AA . Higher resolutions are generally considered to be very good, whilst diffraction as high as 1.0\AA is extremely rare (Zanotti (1992)).

$$d_{\min} = \frac{\lambda}{2 \sin \theta_{\max}} \quad (6.9)$$

Equation 6.9 illustrates a formulation of Bragg's Law for crystal diffraction (Blundell and Johnson (1976)). The expression above demonstrates that the resolution limit of the data-set, d_{\min} , is determined by the maximum diffraction angle, θ_{\max} . The concept of resolution employed in X-ray diffraction is the same as that generally applied in other areas of image formation in the optical sense, such as photography, microscopy or television. Understood in this way the resolution limit determines the degree of

detail visible in the electron density map. Thus, in the case of the refined structure of the d(TG₄T) oligonucleotide structure, objects separated by more than 1.2 Å should be clearly and separately resolved.

6.1.5 Resolution and Determinacy in Refinement

The fall off of intensity of diffraction observations for macromolecules with increasing resolution is associated with the key point regarding the refinement, and thus determination, of a macromolecular structure. In general, the number of experimental observation utilised in a refinement procedure must at least equal, and preferably exceed, the number of parameters being refined. The greater the ratio of observables to parameters, the more reliable the final refined parameter values will be. Should there be less than one observation per parameter, or degree of freedom, then the refinement problem is said to be underdetermined.

$$N \cong \frac{(4.2)V}{n d_{\min}^3} \quad (6.10)$$

Equation 6.10 gives an approximate value for the number of unique reflections accessible for any given structure at a given resolution for CuK α radiation (Stout and Jensen (1968)). N is the approximate number of reflections accessible; V is the volume of the unit cell in Å³; d_{\min} is the resolution limit; and n has a value dependent upon the symmetry of the unit cell.

In the case of the crystals of the oligonucleotide d(TG₄T), V is 54549 Å³ as derived from the cell dimensions determined in Chapter 2; and n is 2 since the unit cell of the

crystals belongs to the space group P1, so that one half of the reflections in reciprocal space are unique, as discussed in Chapter 4. Hence, at a resolution limit of 1.2Å, one would expect 66,290 unique reflections. This corresponds well to the observed number of unique reflections for the *DENZO* processed native data, discussed in Chapter 4, which is 65,264 to a limit of 1.2Å.

Assuming that the crystal unit cell of the d(TG₄T) oligonucleotide crystals contains four copies of the d(TG₄T) tetraplex, then the cell would contain at least 2,000 non-hydrogen atoms. In addition to these, assuming that the ordered water molecules and the water in the solvent continuum is about 20% of the non-hydrogen atoms in the structure (Lamzin and Wilson (1993)), the approximate total number of non-hydrogen atoms contained in the cell would be approximately 2,500. Taking into consideration the three positional parameters (x, y, z) and an isotropic temperature factor $B = 8\pi u^2$ for each atom, at least 10,000 reflections would be required in order to attain a 1:1 ratio of observables and parameters for the refinement problem, thus making the refinement experimentally feasible. According to equation 6.10, this would require a resolution of at least 2.25Å.

With the X-ray diffraction data available to 1.2Å, the ratio of observables to the four parameters considered above is approximately 6.5:1, so that the refinement problem is significantly overdetermined and refinement against the data is possible. Furthermore, adopting an anisotropic model for the temperature factor results in nine parameters for each atom and a observable-parameter ratio of 2.9:1 to 1.2Å for the proposed crystal structure. Therefore, at the resolution limit available, anisotropic refinement is also feasible for the d(TG₄T) structure. Anisotropic temperature factors are discussed later

in the chapter in relation to the *SHELXL-93* package (Sheldrick (1993)).

6.1.6 Restraints and Constraints

In the study of small molecules which diffract to very high-resolution, the conventional least-squares approach involves the independent refinement of the individual atomic positions. This requires that the ratio of observables and parameters must be of the order of 10 or more (Zanotti (1992)). In the case of macromolecules, even when the refinement problem is significantly overdetermined, this approach is not feasible since the ratio of observables to parameters is usually not sufficient. This most commonly occurs because the resolution of the experimental data is too low. As a result, meaningless shifts of atomic positions will result from such a refinement procedure.

In order to improve the ratio of observations to parameters in the least-squares refinement of macromolecules restraints or constraints can be applied in the form of bond lengths, bond angles and torsion angles to the structure under study. Such quantities are derived from high-resolution crystal studies of small molecule species. For DNA and RNA relevant data are obtained from the small molecules crystal structures of nucleotides and nucleosides species (Saenger (1984)).

Sussman (1991) describes restraints and constraints in terms of a six atom ring with 18 positional degrees of freedom, three for each atom. This diagram is reproduced in Figure 6.2. In the case of constraints, data from small molecule species are included in the refinement as rigid and inflexible restrictions upon the values of bond lengths, bond angles and torsion angles. This can be seen in Figure 6.2(a) where the six-membered ring, with the imposition of hard constraints, now has only six degrees of freedom,

namely the rigid body parameters describing the position and orientation of the ring. When included in the refinement procedure in this way, constraints have the effect of reducing the number of independent parameters and hence improving the ratio of observables to parameters in the refinement.

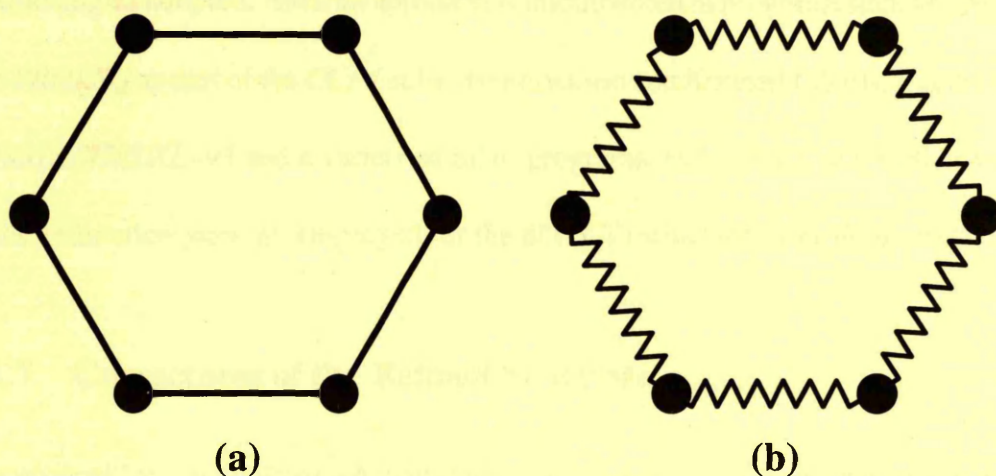


Figure 6.2: Illustration of the Imposition of Restraints and Constraints

Figure 6.2(b) illustrates the same ring but with restraints imposed in place of constraints. As with constraints, the same data derived from small molecule studies are used. However, restraints act as complementary information by introducing ideal stereochemical values which are not rigidly fixed, so that the corresponding stereochemistry of the refined macromolecule can deviate from the introduced ideal values. Since these values are well known from high-resolution studies, deviations are not expected to be very significant. The restraints imposed on the six-membered ring in Figure 6.2(b) are therefore depicted as springs pulling the atoms towards ideal stereochemistry, without reducing the number of degrees of freedom. The inclusion of restraints in the refinement procedure effectively increases the number of observations and thus improves the

observation–parameter ratio also.

The use of stereochemical restraints in macromolecular refinement is arguably the more realistic approach, since the chemical environment of the various parts of the structure within the cell may subtly effect the values of bond lengths, bond angles and torsion angles adopted. Such an approach is implimented in programs such as *PROTIN* and *PROLSQ* as part of the *CCP4* suite (Hendrickson and Konnert (1980); Hendrickson (1985)), *SHELXL-93* and a variety of other programs, and is discussed further as part of the refinement strategy employed for the d(TG₄T) structure, later in this chapter.

6.1.7 Correctness of the Refined Structure

The progress of a crystallographic refinement is normally monitored using the conventional R–factor index. This index employs the observed structure factor amplitudes, F_o , and the corresponding calculated structure factor amplitudes, F_c , derived from the refined model, and is defined in equation 6.11.

$$R = \frac{\sum_{hkl} ||F_o(hkl)| - |F_c(hkl)||}{\sum_{hkl} |F_o(hkl)|} \quad (6.11)$$

As defined in in the equation above, the R–index operates as an indication of relative error, the summation being taken over all observed reflections in the experimental X–ray diffraction data–set. As such, the better the corresponding agreement between the observed and calculated structure factor amplitudes, the lower the R–index will be. The presumption here is that the lower the R–index the better the refined model, and this is generally accepted as an approximate guide–line to the progress of the refinement

procedure. Care is necessary since reduction in the R-index can always be observed but is not always meaningful, for example, if data have been poorly recorded or if an incorrect structure is refined. All of this taken into account, a refinement procedure is considered to be complete when the R-index has attained a suitably low value. It is worth noting that the R-index of a macromolecular crystal structure refined to about 2Å resolution using restrained least-squares refinement is typically in the range 0.15–0.20, whilst studies of extensively refined structures with data greater than 2Å have recorded R-indices approaching 0.12 (Jensen (1985)), depending on the quality of the data. Comparatively, small molecule structures commonly have R-indices in the range 0.03–0.05.

It has been demonstrated for a random structure occupying the unit cell that the corresponding R-index for the structure will be 0.83 if the structure is centric and 0.59 if the structure is acentric (Wilson (1949); Wilson (1950); Phillips *et al.*, (1950)). On this matter Jensen (1985) has commented that, since biological macromolecules are chiral and their crystals must be noncentrosymmetric, it is then reasonable to conclude that R-index values in the area of about 0.6 indicate a model structure that does not fit the observed data any better than a randomly generated structure. This is significant when one considers that initial molecular models of macromolecules, derived from heavy-atom maps or molecular replacement solutions, commonly have R-indices in the range 0.4–0.5, only a little below that expected for a random structure.

The final refined structure solution can be analysed in a number of other ways in order to evaluate its correctness. A key evaluation of the structures correctness is the accuracy and feasibility of its geometry and its chemistry. The structural principles of

oligonucleotide organization and chemistry are extensively detailed in Saenger (1984). The final structure can be checked using $3F_o - 2F_c$, $2F_o - F_c$, and $F_o - F_c$ electron density syntheses, which should give good indication of misplaced macromolecular or solvent molecules, or inappropriate atomic B values. The spread of B values in the structure should correspond generally to those residues which are buried and those which are exposed. Typically, atoms with smaller B values tend to be well localised and are commonly tightly packed within the structure or participate in well coordinated hydrogen-bonding. Residues with higher B values are commonly on the outer surface of the molecule, exposed to the solvent, and are generally less well localised and may be involved in a dynamic equilibrium about a mean position.

Variations in the B factors of different atoms and residues ought also to be reflected in the estimated standard deviation (σ) of their positional parameters. Atoms which are well determined ought to have low B values and smaller standard deviations in their positional parameters. As such, the σ values of the atoms are an indication of the accuracy of the structure as a whole or of individual parts of the structure in particular. An upper-limit value for the errors in atomic positions, based on the differences in F_o and F_c for the final structure, can be derived from a Luzzati plot (Luzzati (1952)). The Luzzati plot involves a plot of the R-index for the structure against $(\sin \theta / \lambda)$. For an error in coordinates of about 0.05 \AA the plot is approximately linear over all ranges of resolution. For larger errors the plot becomes progressively more nonlinear, and for an error of 0.5 \AA the value of the R-index of 0.59 corresponds to a random acentric structure (see Jensen (1985)). The values of the standard deviations in the positional parameters of the structure should be taken into account when interpreting

its chemistry.

6.2 Refinement Strategy for the d(TGGGGT) Tetraplex

6.2.1 Overview of the Refinement Protocol

An overall view of the refinement procedure employed in the elucidation of the crystal structure of the parallel-stranded tetraplex formed by the oligonucleotide d(TG₄T) is presented in Figure 6.3.

The first step in the strategy devised utilised the rigid body minimization procedure provided by the program *XPLOR* (Brünger (1993)). Molecular replacement solutions obtained using tetraplex models derived from NMR studies (Aboul-ela *et al.*, (1992)), discussed in Chapter 5, were investigated. Two similar models for the unit cell packing were derived using a search structure which contained the 3' and 5' thymine residues in an unlikely planar formation, and another which omitted the thymine residues completely. The two resultant packing models are shown in Figures 5.10 and 5.12 respectively. Consequently, rigid body minimization was employed initially to ascertain whether the two models could be distinguished, on the basis of their R index, at this early stage. This might help facilitate a clearer path for the following refinement steps and perhaps give some indication of the correct packing model, at least in relation to the guanine tetrads.

The next step in the refinement procedure was the application of the technique of simulated annealing, as applied using molecular dynamics to the problem of macromolecular refinement (Brünger (1991)). This procedure is provided by the program

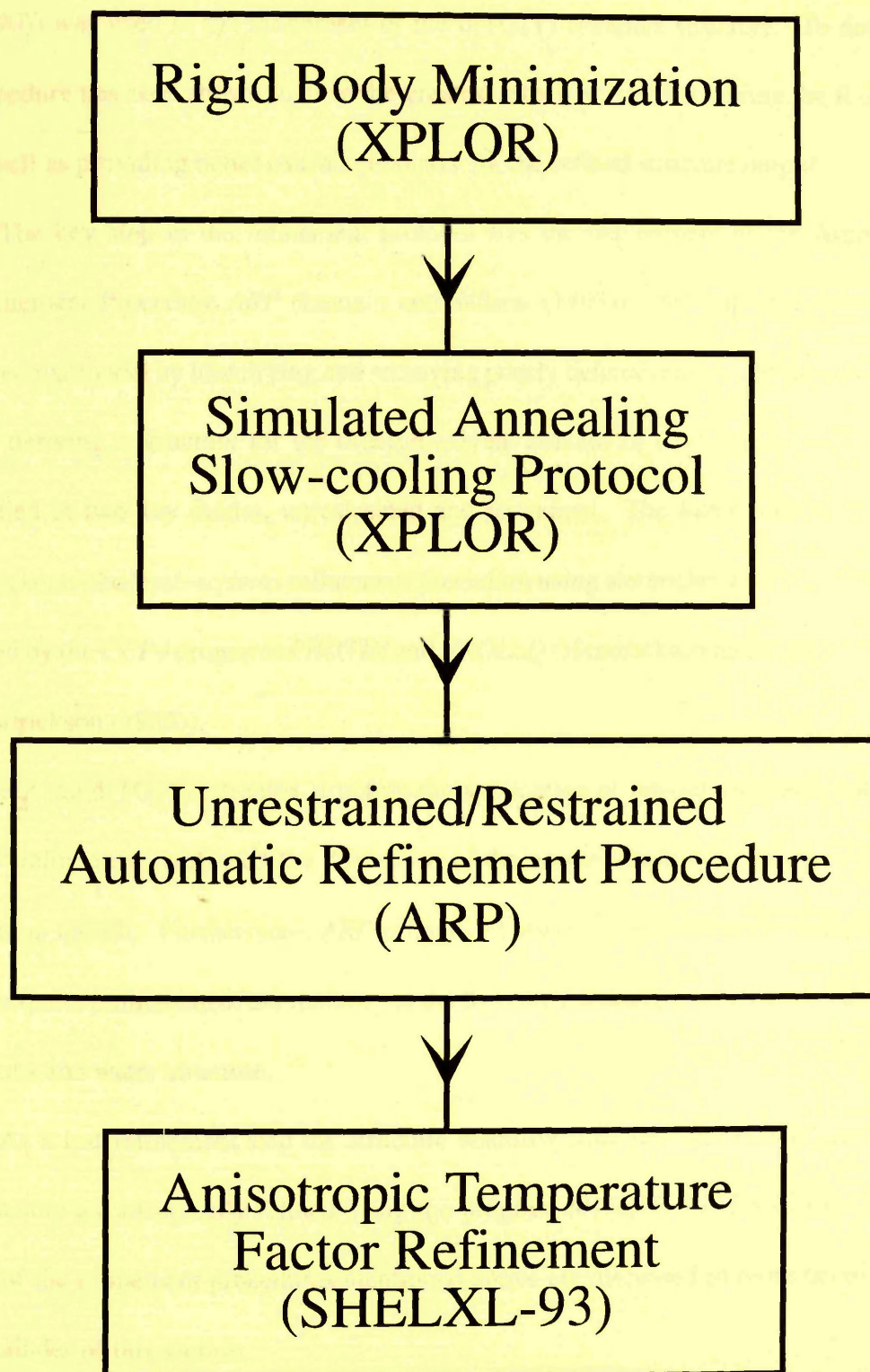


Figure 6.3: Overview of the d(TGGGGT) Structure Refinement Procedure

XPLOR (Brünger (1993)). In particular, the slow-cooling protocol (Brünger *et al.*, (1990)) was used in the refinement of the d(TG₄T) tetraplex structure. To date this procedure has been shown to have the greatest effectiveness in lowering the R-index, as well as providing better overall geometry for the refined structure output.

The key step in the refinement protocol was the deployment of the Automated Refinement Procedure *ARP* (Lamzin and Wilson (1993)). *ARP* updates the macromolecular model by identifying and removing poorly defined atoms, adding new atoms and deriving a structure for the ordered solvent content of the crystal. *ARP* can be applied in two key modes, unrestrained and restrained. The latter couples the *ARP* program to the least-squares refinement procedure using stereochemical restraints provided by the *CCP4* programs *PROTIN* and *PROLSQ* (Hendrickson and Konnert (1980); Hendrickson (1985)).

For the d(TG₄T) tetraplex structure the application of unrestrained and restrained *ARP* refinement facilitated the derivation of the correct packing arrangement of the guanine tetrads. Furthermore, *ARP* refinement also allowed the model building and subsequent refinement of the majority of the thymine residues, as well as relevant metal cations and water structure.

As a last refinement step the structure resulting from the *ARP* procedure had its *B*-factors anisotropically refined using the program *SHELXL-93* (Sheldrick (1993)). All of the refinement procedures mentioned above are discussed in more detail in the remainder of this section.

6.2.2 Rigid Body Minimization

Rigid body minimization can be understood as a special case of constrained refinement. In this instance a part or the whole of a macromolecule or macromolecules occupying the unit cell can be treated as a rigid body. In effect this means that all of its atomic parameters are constrained to the values inherited from the molecular model. In this way the refinement problem reduces to the determination of the optimal values of the six parameters describing the position and orientation of the molecule or fragments treated as a rigid body (see Figure 6.2(a) and relevant discussion). The objective is to find the values of the positional parameters which produce the best possible value of the R-index. The magnitude of this improvement is necessarily very limited.

Equation 5.17 gives an example of the type of function minimized, in this case as part of the suite of programs *AMoRe* (Navaza (1992)). User defined parts of the macromolecular assemblage may be rigid body refined whilst the rest of the model structure is held fixed. This procedure is often necessary following the derivation of a molecular replacement solution of the unit cell contents since the sensitivity of the rotation function is about $3\text{--}5^\circ$, and the sensitivity of translation function and correlation function searches is typically $0.2\text{--}0.5\text{\AA}$ (Fitzgerald (1991)). With n molecules in the asymmetric unit, the generalised rigid body refinement problem therefore becomes a $6n$ dimensional problem. Clearly the accuracy, appropriateness and completeness of the molecular model is of prime significance.

6.2.3 Simulated Annealing

The term annealing denotes a general physical process which begins when a solid is placed in a heat bath and heated due to the increase in the heat bath temperature. At a certain temperature all of the particles of the solid arrange themselves randomly in the liquid phase. Thereafter, the substance is slowly cooled by lowering the temperature of the bath such that the particles of the substance arrange themselves in the lowest energy state of the solid. Ensuring that the maximum temperature is sufficiently high, and that the cooling procedure is carried out at a slow enough pace, the substance is said to have been annealed.

In general terms, by formally identifying the target of an optimization problem with the energy of a particular system, this annealing procedure can be simulated for the problem in hand. This approach has been applied to the specific problem of macromolecular refinement in the form of simulated annealing using molecular dynamics (reviewed in Brünger (1991)). The simulated annealing approach to macromolecular refinement presents an advance over more conventional restrained least-squares refinement procedures. This results from the fact that, whereas gradient descent methods such as restrained least-squares are restricted to shifts in the parameters which minimise the target function, and require interactive model building to lift the target function out of a local energy minimum, simulated annealing allows movements uphill in the energy target function. This allows the method to effectively jump out of local energy minima without time consuming interactive rebuilding of the model. The likelihood of a movement uphill in the target energy function is determined by a control

parameter, referred to as “temperature”, which has no physical meaning but rather determines the likelihood of overcoming local minima in the target function. The higher the temperature, the more likely the optimization can overcome local minima energy barriers.

In the simulated annealing algorithm a target function $E(q_1, \dots, q_i)$ is defined as a function of the parameters q_i of the system. A generation mechanism is then deployed which creates the Boltzmann distribution at a given temperature T , as shown in equation 6.12, where k_b is the Boltzmann constant. An annealing schedule is then devised which comprises of a series of temperatures $T_1 > T_2 > \dots > T_l$ at which the Boltzmann distribution is calculated.

$$B(q_1, \dots, q_i) = \exp \left(\frac{-E(q_1, \dots, q_i)}{k_b T} \right) \quad (6.12)$$

The generation mechanism can be implemented in a number of different ways, and these implementations differ in the manner in which they generate a movement from one set of parameters to another, consistent with the Boltzmann distribution at a given temperature. The Monte Carlo method is one of the most popular approaches used in this regard. In this approach, given the current state of the systems in terms of the parameters q_i , a move is applied by a shift of a randomly chosen parameter q_i . If the difference in the target energy function before and after the shift is negative, the process continues with the newfound state. Should the energy difference be greater than or equal to zero, then the move is still accepted with a probability P as given in equation 6.13. This permits the simulated annealing approach to jump over local minima energy

barriers, and effectively gives it a greater radius of convergence than gradient descent methods such as restrained least-squares refinement. The special case of $T = 0$ results in the acceptance of moves which only lower the target function, and is equivalent to the restrained least-squares refinement procedure.

$$P = \exp\left(\frac{\Delta E}{k_b T}\right) \quad (6.13)$$

Molecular Dynamics

In the application of the simulated annealing procedure to macromolecular refinement, the Monte Carlo approach becomes more complicated since it is necessary to carry out the shifts in an especially chosen coordinate frame, such as torsion angles, in order that the shifts be sensible. Instead of this it was found to be more advantageous to utilise a molecular dynamics implementation of the simulated annealing technique (Brünger (1991)), and this approach was first shown to be successful using crambin, a small protein containing 46 amino acids (Brünger *et al.*, (1987)).

In the molecular dynamics scenario, a chosen set of continuous parameters q_i can be viewed as generalised coordinates, propagated in time by the Hamilton equations given in equation 6.14 and 6.15. In these expressions, $H(p, q)$ is the Hamiltonian of the system and p_i are the generalized momenta conjugate to q_i .

$$\frac{\delta H(p, q)}{\delta p_i} = \frac{dq_i}{dt} \quad (6.14)$$

$$\frac{\delta H(p, q)}{\delta q_i} = \frac{dp_i}{dt} \quad (6.15)$$

If the generalized coordinates represent the degrees of freedom of a molecular system, the approach is referred as molecular dynamics. A sequence of molecular dynamics simulations of decreasing values of the temperature T represents another implementation of the simulated annealing approach. In the special case where the generalized coordinates q_i represent the cartesian coordinates of n point masses, and momenta can be separated from coordinates in the Hamiltonian H , the Hamilton equations of motion reduce to the Newtonian equations of motion given in equation 6.16.

$$m_i \frac{\delta^2 r_i}{\delta t^2} = -\Delta_i E \quad (6.16)$$

In equation 6.16, the quantities r_i and m_i represent the coordinates and masses of atom i , respectively, and E is the potential energy. In the context of simulated annealing, E denotes the target function of the optimization problem which may contain molecular energies such as torsion angles and covalent bonding energy terms, as well as energies defined with respect to the correlation of the experiment X-ray diffraction data and the calculated diffraction data from the model.

Temperature Control

The implementation of simulated annealing requires a means to control the temperature during molecular dynamics. The three most commonly used methods are velocity

scaling, Langevin dynamics and temperature coupling.

Velocity scaling consists of periodic uniform scaling of the velocities v_i . This is shown in equation 6.17 for all atoms i , where T is the target temperature and T_{curr} is the current temperature.

$$v_i^{\text{new}} = v_i^{\text{old}} \sqrt{\frac{T}{T_{\text{curr}}}} \quad (6.17)$$

Langevin dynamics incorporates the influence of a heat bath into the Newtonian equations of motion. The resultant expression is shown in equation 6.18, where γ_i specifies a friction coefficient for atom i and $R(t)$ is a random force. $R(t)$ is assumed to be uncorrelated with the positions and velocities of the atoms.

$$m_i \frac{\delta^2 r_i}{\delta t^2} = -\Delta_i E_{\text{total}} - m_i \gamma_i \frac{\delta r_i}{\delta t} + R(t) \quad (6.18)$$

Finally, there is the temperature coupling method. This is related to Langevin dynamics. However, random forces are not used and a scale factor is applied to the friction coefficient. This is shown in equation 6.19.

$$m_i \frac{\delta^2 r_i}{\delta t^2} = -\Delta_i E^{\text{total}} - m_i \gamma_i v_i \left(1 - \frac{T}{T_{\text{curr}}}\right) \quad (6.19)$$

The second term on the right-hand side of equation 6.19 represents positive friction if $T_{\text{curr}} > T$, thus lowering the temperature, and negative friction if $T_{\text{curr}} < T$, hence raising the temperature.

Crystallographic Refinement Using Simulated Annealing

In applying the simulated annealing method to crystallographic refinement, a target function is required. The general form of such a target function is shown in equation 6.20 below.

$$E = E_{\text{chem}} + wE_{\text{exp}} \quad (6.20)$$

The crystallographic term E_{exp} is the effective crystallographic potential energy, consisting of the weighted differences between the observed and the calculated structure factor amplitudes. The most common form of E_{exp} is the crystallographic residual, as defined in equation 6.2. The term E_{chem} represents a chemical term relating to the properties of the structure under refinement. Commonly, this is expressed in the form of an empirical potential energy, which is a function of all the atomic positions of the system that describe covalent and nonbonded interactions. Covalent interactions are typically bond lengths, bond angles, dihedral torsion angles, chiral centres and the planarity of aromatic rings. Nonbonded interactions include Van der Waals, hydrogen-bonding and electrostatic interactions. The array below illustrates the formulation of such a function for E_{chem} .

$$\begin{aligned} E_{\text{chem}} = & \sum_{\text{bonds}} k_b(r - r_o)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_o)^2 \\ & + \sum_{\text{dihedrals}} k_\phi \cos(n\phi + d) + \sum_{\text{chiral, planar}} k_\omega(\omega - \omega_o)^2 \\ & + \sum_{\text{atom-pairs}} (ar^{-12} + br^{-6} + cr^{-1}) \end{aligned}$$

The parameters of E_{chem} are derived from experimental and theoretical investigations. In particular, in the case of nucleic acids, these parameters result from the high-resolution crystal structures of nucleosides and nucleotides. In order to prevent distortion of aromatic rings during the simulated annealing procedure, certain energy constants such as k_{ϕ} and k_{ω} in the above array may have to be increased. Since bulk solvent is generally not included in the refinement of crystal structures, highly charged parts of the structure need to be screened, and this is accomplished by setting the charges to zero. This observation applies particularly to the densely charged phosphate-sugar backbone found in nucleic acid structures. Since crystallographic refinement is not very sensitive to the accuracy of the empirical energy function, electrostatic and hydrogen-bonding terms can be omitted altogether without incurring large uncertainties, and their omission may in fact prevent bias towards the empirical energy function (Brünger (1991)).

The choice of the weight w in equation 6.20 is crucial. This term relates the effective crystallographic potential E_{exp} to the physical energy term E_{chem} . If w is too big the refined structure exhibits large deviations from ideal geometry, and if w is too small the refined structure does not satisfy the X-ray diffraction data. Brünger *et al.*, (1989) have suggested running a preliminary molecular dynamics simulation with w set to zero, using the initial structure. The value of w is then adjusted in order to make the magnitude of the gradients of E_{chem} and E_{exp} of the molecular dynamics structure equal. This procedure is implemented as part of the *XPLOR* package (Brünger (1993)).

The Slow-Cooling Protocol

The choice of an annealing schedule is of prime importance with regards to the success of the simulated annealing molecular dynamics procedure. Factors such as the weighting of the energy terms, the cooling rate and the duration of the heating stage have been studied as to their influence on the success of the chosen simulated annealing refinement protocol (Brünger *et al.*, (1990)).

In their study, Brünger and colleagues found that the use of temperature coupling to control the temperature change during molecular dynamics was more effective than either velocity scaling or Langevin dynamics. Furthermore, a slow-cooling protocol produced lower R-index values than those obtained using faster-cooling approaches. If the cooling is sufficiently slow an initial heating stage is not required and iterations of the slow-cooling approach can lead to an improvement in the overall result. On the basis of these results, the following slow-cooling protocol was recommended;

1. Determinations of weights; highly charged residues set to neutral
2. Minimization, 40 conjugate gradient steps with soft repulsive potential for non-bonded contacts followed by 120 conjugate gradient steps with Van der Waals non-bonded potential
3. Slow-cooling molecular dynamics, 3-7ps, $T = 4000-300\text{K}$ (or lower initial temperature should instabilities occur), time step = 0.5fs, temperature coupling, temperature decrement 25K every 25fs
4. Minimization, 80 conjugate gradient steps

The slow-cooling procedure above is implemented in the *XPLOR* package and is detailed in the *XPLOR* manual (Brünger (1993)). Following on from the rigid body refinement analysis, this slow-cooling simulated annealing protocol was the basis of the next refinement step in the determination of the crystal structure of the d(TG₄T) tetraplex.

6.2.4 ARP Refinement

The Automated Refinement Procedure *ARP* (Lamzin and Wilson (1993)) is a program package designed for the refinement of protein structures, but which is applicable to macromolecules in general. The central principle underlying refinement with *ARP* is that the molecular model is iteratively updated in two ways. Atoms poorly defined in the electron density are identified and removed, whilst new atoms are added on the basis of difference density calculations and distance constraints. Atom rejection is carried out on the basis of the density at the atomic centre, deviation of density shape from sphericity and certain distance evaluations.

The *ARP* protocol is coupled to an iterative least-squares minimization procedure, utilising code from other programs in the *CCP4* suite. Atom rejection is based upon the analysis of a $3F_o - 2F_c$ electron density synthesis using calculated phases, α_c , derived from the model. Assuming that the above synthesis ought to be considerably positive at an atomic centre, and that the atomic density distribution is spherical, *ARP* can reject atoms if the density at an atomic centre is questionably low, or if the density is highly non-spherical. The latter assumption applies for resolutions greater than 2.0 Å for macromolecular atoms, and greater than 2.5 Å for water molecules, and is truly

valid at resolutions greater than 1.2\AA . Atom addition follows from analysis of the $F_o - F_c$ synthesis, again using α_c phases, where distance constraints are applied and grid points in the difference synthesis above a certain threshold are analysed in turn. New atoms can then be more accurately positioned in the next round of least-squares minimization.

Since rejection and addition are executed on the basis of electron density maps calculated with model phases, certain standards are stipulated for the quality of the original model. Experimental trials (Lamzin and Wilson (1993)) have shown that a model can be successfully refined if it contains no waters and has, in the case of proteins, $C\alpha$ atoms to within 0.5\AA of their correct position. Furthermore, if the model is about 75% complete, with $C\alpha$ atoms between 1.0\AA and 1.5\AA of their correct positions, significant improvements in the electron density are possible. Experimental X-ray data should generally be greater than 2.0\AA . If, however, the model is nearly right and only requires minor corrections and the adjustment of water structure, a limit of 2.5\AA in the data is feasible. This restriction follows from the fact that *ARP* operates on the basis of the interpretation of atomic electron density. Also, the data should be complete in the range $4\text{--}10\text{\AA}$ since, should low-resolution strong data be missing, the resultant electron density map is typically discontinuous and inconsistent with the model. This can lead to erroneous interpretation of the electron density and hinderance of convergence in the *ARP* procedure. In addition, the low-resolution data contains information on the solvent structure and the *ARP* procedure operates on the basis of the definition and updating of solvent and molecular structure.

When a preliminary model of the unit cell contents is available, *ARP* can be

implemented in two separate modes – unrestrained and restrained *ARP* refinement. Both of these modes were used in the successful determination of the d(TG₄T) tetraplex structure, and are described in this section.

Unrestrained *ARP* Refinement

Unrestrained *ARP* is the application of the *ARP* protocol where the program is coupled with an unrestrained least-squares minimization procedure. In this application all of the atoms of the model are updated. Such an approach should be considered when the *R*-index of the model is 0.3 or worse. Experimental implementation of the *ARP* program has shown, if the resolution is high enough and the initial model essentially correct, that the *ARP* atoms are found at approximately the true atomic positions of the macromolecule under study. This facilitates the rebuilding of the model as well as the modelling of previously omitted features of the structure. The unrestrained procedure is represented schematically in Figure 6.4.

All of the programs utilised in conjunction with *ARP* are incorporated as part of the *CCP4* suite of programs. *SFALL* is used in two capacities: first of all, to perform the unrestrained least-squares minimization; secondly, to calculate the structure factors from the model after this minimization is complete. The program *RSTATS* scales the observed structure factor amplitudes to the calculated structure factor amplitudes and applies an overall temperature factor to the calculated structure factors. *FFT* is employed to calculate the two electron density syntheses with coefficients $3F_o - 2F_c$ and $F_o - F_c$, using the calculated phases from the model, α_c . *EXTEND* then enlarges the maps to cover the true asymmetric unit of the unit cell.

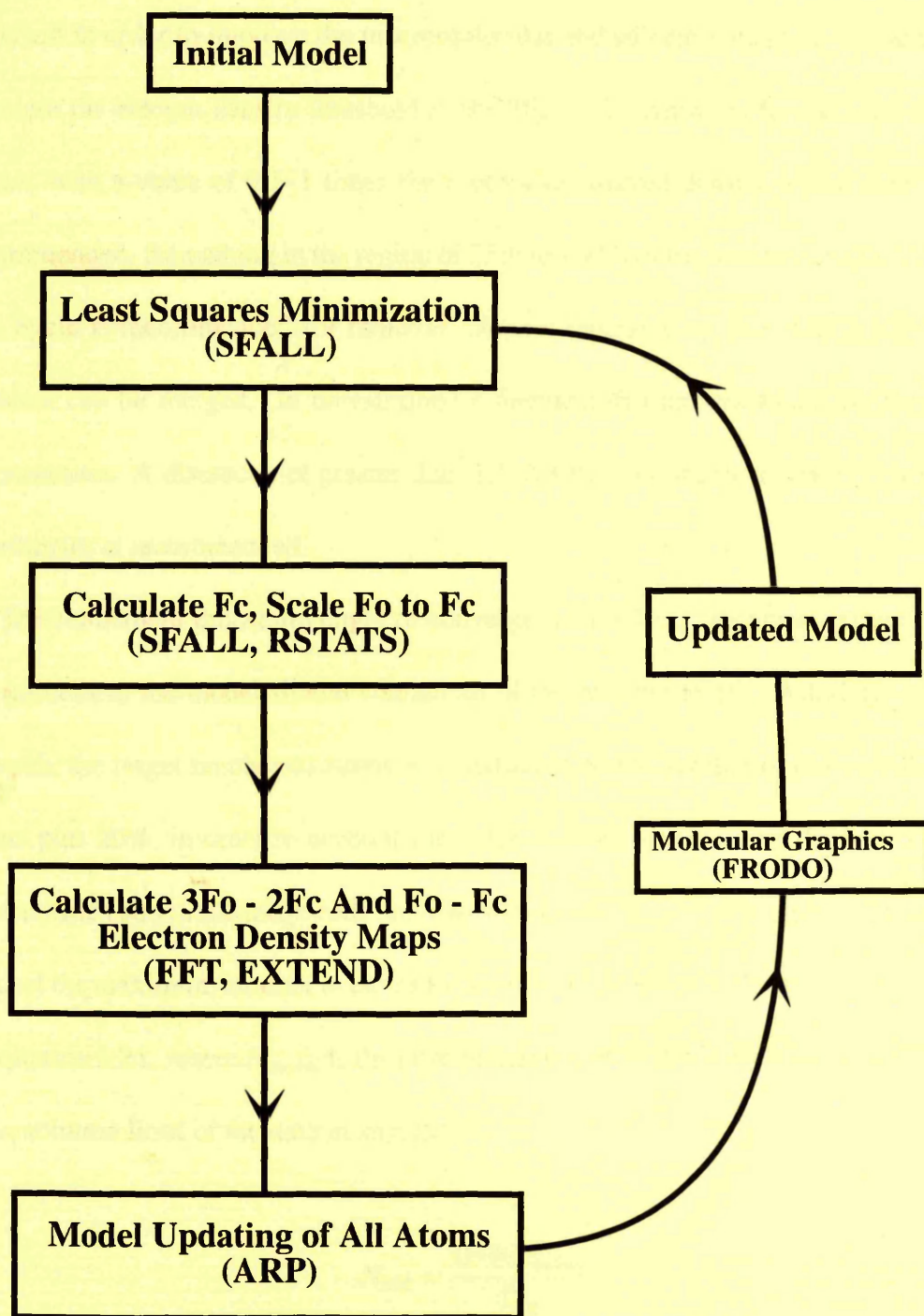


Figure 6.4: The Unrestrained ARP Refinement Protocol

ARP itself then analyses the computed $3F_o - 2F_c$ and $F_o - F_c$ syntheses in order to perform the model updating. In the case of unrestrained refinement, all atoms are analysed in order to improve the macromolecular and solvent structure. The user can stipulate the electron density threshold in the $3F_o - 2F_c$ synthesis for the rejection of atoms, with a value of 0.5–1 times the root mean squared density of the map being recommended. Something in the region of 25% to 100% of the number of atoms added in a cycle is recommended for removal. Atoms separated by less than a stipulated distance can be merged. In unrestrained refinement this applies to all the atoms of the structure. A distance not greater than 1.3–2.3 the root mean squared error in the coordinates is recommended.

The refinement procedure ought to converge within 20 to 50 cycles. At the end of the refinement the model should contain all of the relevant atoms, including solvent. As such, the target number of atoms is proposed to be the number of macromolecular atoms plus 20%, in order to account for ordered water structure and the less ordered waters which constitute the pseudo solvent continuum. As a working guide the authors suggest the maximum number of atoms to be added may be derived from the expression in equation 6.21, where N_{model} is the current number of atoms in the model, and d_{min} is the resolution limit of the data in angstroms.

$$N_{\text{add}} = \frac{0.08 N_{\text{model}}}{d_{\text{min}}^3} \quad (6.21)$$

Hence, for the model of the d(TG₄T) tetraplex unit cell, containing approximately 2,000 non-hydrogen atoms, the maximum number to add would be 92. Isotropic

temperature factors assigned to new atoms should be equal to the average temperature factor of the macromolecule in the early cycles. Near refinement completion, temperature factors should be set to the average temperature factors of the existing solvent molecules. In the same manner as for the removal of atoms, the user can define the threshold, this time in the $F_o - F_c$ synthesis, for the location of new atoms. A value of 3–4 times the root mean squared density of the map is suggested, although the program can assign this automatically.

The location of new atoms requires the imposition of distance constraints. The minimum and maximum distances between new atoms and existing atoms in the structure is user defined. In the case of unrestrained refinement, the minimum distance is usually set at 0.7 times the maximum resolution of the X-ray diffraction data. For the maximum distance a value of 3.3 Å, equivalent to the maximum water–protein distance, is recommended since features further than this for any current model are likely to be picked up in later cycles. The minimum distance between new atoms is also defined in order to prevent a number of atoms being placed in the one peak, and is commonly set at 2.2 Å, the minimum, fully occupied water–water distance. At any stage in the iterative ARP refinement procedure the model may be manually adjusted using a molecular graphics package such as, for example, *FRODO* (Jones (1978, 1985)) or *O* (Jones *et al.*, (1991)).

Restrained ARP Refinement

Once a model has been obtained that is more or less correct and requires little more than the updating of the solvent structure and the application of minor shifts, the restrained

mode of *ARP* refinement may then be used. This procedure is schematically illustrated in Figure 6.5.

The unrestrained *ARP* protocol couples the *ARP* program with the restrained, reciprocal space, least-squares refinement procedure of Hendrickson and Konnert (Hendrickson and Konnert (1980); Hendrickson (1985)), implemented in the *CCP4* programs *PROLSQ* and *PROTIN*. In this restrained *ARP* procedure, *SFALL* performs the least-squares minimization and output the gradients for input to *PROLSQ*. The program *PROTIN* then prepares the geometrical restraints for *PROLSQ* and the program *PROLSQ* then applies the calculated shifts to the model. Structure factors and electron density maps $3F_o - 2F_c$ and $F_o - F_c$ are then calculated as for unrestrained *ARP*, using the programs *SFALL*, *RSTATS*, *FFT* and *EXTEND*.

The *ARP* restrained refinement is then engaged in much the same manner as for unrestrained *ARP*, but with the exception that only water molecules are analysed for addition and rejection. It is recommended that the density threshold for finding new atoms be defaulted to the automatic mode, in order to prevent too many extra atoms being added. A real space refinement procedure, based on the assumption of the sphericity of the atomic density, is also provided and is recommended for the restrained refinement mode of water analysis.

The problem of the degree of determinacy in the crystallographic refinement of macromolecules, discussed earlier in this chapter, leads to the need for some form of compensation for the less than ideal ratio of observations to parameters. In the real space refinement procedure of Diamond (1971) this was achieved by constraining bond lengths and some bond angles. Konnert and Hendrickson (1980) have devised

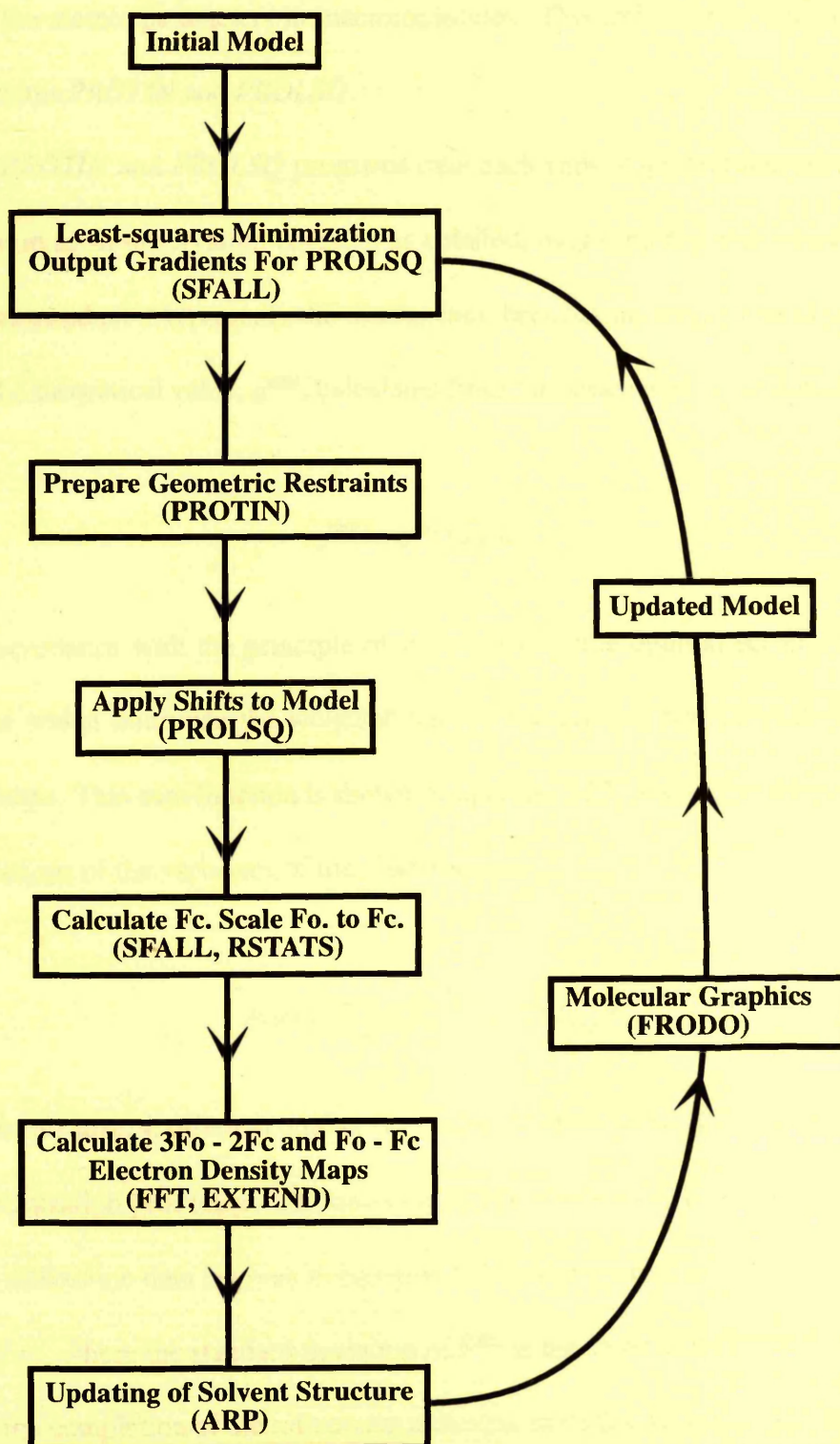


Figure 6.5: The Restrained ARP Refinement Protocol

a reciprocal space refinement technique which utilizes predetermined knowledge regarding the atomic parameters in macromolecules. This technique is implemented in the programs *PROTIN* and *PROLSQ*.

The *PROTIN* and *PROLSQ* programs treat each item of generalised experimental information as an observation equation as detailed, in general form, in equation 6.22, where the residual ε represents the discrepancy between an observational parameter g^{obs} , and a theoretical value, g^{calc} , calculated from the parameters, \mathbf{x} , of a model.

$$g^{\text{obs}} = g^{\text{calc}}(\mathbf{x}) + \varepsilon \quad (6.22)$$

In accordance with the principle of least-squares, the optimal set of parameters are those which minimize the weighted sum of the squared residual, taken over all observations. This sum function is shown in equation 6.23, where the weights, w_h , are the inversions of the variances of the observations.

$$\phi(\mathbf{x}) = \sum_h w_h [g_h^{\text{obs}} - g_h^{\text{calc}}(\mathbf{x})]^2 \quad (6.23)$$

There are several different qualitative classes of observation that can be handled in this minimization procedure. The observation function which relates to the crystallographic diffraction data is given in equation 6.24. In this case the variance is given in terms of σ_F , where the standard deviation of F^{obs} is the value that should be attributed to σ_F at the completion of the refinement although, in earlier stages, σ_F is related to the current average residuals.

$$\phi_{\text{diffraction}} = \sum_{hkl} \frac{1}{\sigma_F^2(hkl)} (|F^{\text{obs}}(hkl)| - |F^{\text{calc}}(hkl)|)^2 \quad (6.24)$$

Another class of observation function is that which describes the stereochemistry of the atomic positions. One example is the interatomic distances relating to chemical bonding. These may include bond distances, next-nearest-neighbour distances, in sets of three atoms describing bond angles, and the distances which describe dihedral and torsion angles. The derived observation function is given in equation 6.25, where σ_D is the standard deviation for the distribution of values expected for distances of a particular type.

$$\phi_{\text{bonding}} = \sum_j \frac{1}{\sigma_D^2(j)} (d_j^{\text{ideal}} - d_j^{\text{model}})^2 \quad (6.25)$$

Another example of restraints relating to the stereochemistry of the atomic positions are the conformational torsion angles. These possess a certain degree of inherent flexibility and are consequently one of the least restrained of stereochemical parameters. Restrictions relating to non-bonded contacts can be applied and the observation function deployed by Hendrickson and Konnert is shown in equation 6.26. Here, χ^{ideal} is the value of the torsion angle minimum to which χ^{model} is closest to, whilst σ_T is the standard deviation of the expected distribution. Examples of cases pertinent to DNA structures would be the β -glycosyl linkage connecting base and sugar, the endocyclic torsion angles describing the conformationally restricted β -D-2'-deoxyribose sugar form, staggered potentials for the $O_{5'}-C_{5'}-C_{4'}-C_{3'}$ torsion, as well as target sugar-phosphate backbone conformation angles, the restrained values of which are

dependent on the morphic form of the structure.

$$\phi_{\text{torsion}} = \sum_t \frac{1}{\sigma_T^2(t)} (\chi_t^{\text{ideal}} - \chi_t^{\text{model}})^2 \quad (6.26)$$

Restraints may also be modelled in order to incorporate the effects of atomic motions. Whilst variation in bond and bond angle distances is slight, rms displacements from mean atomic positions are typically in the range of several tenths of an angstrom for macromolecules. Consequently, atomic movements need to be highly correlated in order to allow stereochemically related atoms to move in conjunction with an atom that experiences large displacements. This may be modelled in terms of the isotropic thermal parameters $B = 8\pi\overline{u^2}$, resulting in the observation function in equation 6.27, where B^{origin} and B^{target} represent the thermal parameters of an atom pair related by a particular bonding distance.

$$\phi_{\text{isotropic}} = \sum_j \frac{1}{\sigma_B^2(j)} (B_j^{\text{origin}} - B_j^{\text{target}})^2 \quad (6.27)$$

Ultimately, all of the introduced macromolecular restraints are treated as additional observations which are included in the normal-matrix derived from the normal equation described in the least-squares treatment outlined in Section 6.1.2. Hence, the total quantity to be minimized, Φ , is as shown in equation 6.28.

$$\Phi = \phi_{\text{diffraction}} + \phi_{\text{bonding}} + \phi_{\text{torsion}} + \phi_{\text{isotropic}} + \dots \quad (6.28)$$

Other observation functions may include those for describing planar groups, chiral

volumes, nonbonded contacts, anisotropic thermal parameters, noncrystallographic symmetry, resistance to excessive shifts and occupancy factors. All of the above are described in Hendrickson and Konnert (1980) and Hendrickson (1985).

In the normal-matrix defined by the expressions in equations 6.5 and 6.6, the structure factor derivatives contribute only to the diagonal elements of the least-squares matrix. All of the additional observational equations, or restraints, contribute in full to diagonal and off-diagonal terms. In this conjugate-gradient least-squares (CGLS) refinement procedure, only the diagonal terms of the normal-matrix are computed in order to facilitate the calculation of shifts to the structure under study. However, this imposes the significant disadvantage that the conjugate-gradient least-squares procedure does not provide estimated standard deviations on measurements of interest e.g. individual bond lengths and bond angles of the refined structure.

6.2.5 SHELXL-93 Refinement

SHELXL-93 (Sheldrick (1993)) is a program designed for the refinement of crystal structures and is primarily designed with small molecule crystal structures in mind. It can, however, be applied to macromolecular structures in cases where the resolution of the X-ray diffraction data is sufficiently high. The refinement method used in this case is again the restrained, least-squares, reciprocal space refinement procedure.

SHELXL-93 was deployed as the final stage in the refinement of the d(TG₄T) tetraplex structure, after the restrained *ARP* refinement of the positional parameters (x , y , z) and isotropic thermal parameters $B = 8\pi\overline{u^2}$ for the macromolecular and solvent atoms. This final *ARP* refined structure was then subjected to a further round

of restrained, least-squares refinement in order to anisotropically refine its temperature factor parameters.

$$\exp[-(\beta_{11}h^2 + \beta_{22}k^2 + \beta_{33}l^2 + 2\beta_{12}hk + 2\beta_{23}kl + 2\beta_{32}lh)] \quad (6.29)$$

Equation 6.29 describes the standard expression for the anisotropic temperature factor. This expression describes the more general case where atoms within the structure are not free to vibrate equally in all directions. The six parameters β_{11} , β_{22} , β_{33} , β_{12} , β_{23} and β_{32} define an ellipsoid which describes the anisotropy of the individual atomic thermal vibrations. The refinement of anisotropic temperature factors for macromolecules is unusual since, when taken in conjunction with the three positional parameters, a total of nine independent parameters are required to be refined for each atom. The resolution limit of macromolecular X-ray diffraction data does not typically warrant this type of refinement. However, as discussed in Section 6.1.5, assuming a model of four tetraplexes in the unit cell, the ratio of observables to parameters for nine parameters per atom is approximately 2.9:1 to the limit of the 1.2Å resolution data-set.

Unlike most crystallographic refinement packages, *SHELXL-93* refines, as an option, against F_o^2 rather than F_o and takes as input the untruncated data containing the diffraction intensities and their standard deviations for all data measured. For a well-behaved structure the geometrical parameters and their estimated standard deviations are almost identical for refinement based on all F_o^2 values and for a more traditional refinement against F_o ignoring data with F_o less than approximately 3σ (*SHELXL-93* documentation, Sheldrick (1993)). For weakly diffracting crystals, and in particular for

pseudosymmetry problems, the refinement against all data is superior. The estimated standard deviations are reduced because more information is used and the chance of getting stuck in a local minimum is reduced. In addition, the use of a threshold admits a systematic error which introduces bias into the displacement parameters β_{ij} above. Also, it is impossible to refine against F_o using all the data since this would involve taking the square root of a negative number for reflections with negative F_o^2 , due to a higher background than the peak height as a result of statistical fluctuations. Hence, in the refinement procedure, all data should be used except for reflections known to suffer from systematic errors. *SHELXL-93* output its own statistics based on F_o^2 and but also produces a conventional R-index based on F_o values larger than 4σ , as well as a R-index against all the data.

SHELXL-93 also facilitates the calculation of the free R-index for the refinement procedure (Brünger (1992)). In the computation of the free R-value a unique set of the observed reflections is partitioned into a randomly selected test set T and a working set A . Typically, T comprises 10% of the total data and in general is a data-set substantially smaller than the working set A . Sets T and A have no reflections in common and when added constitute the entire set of observed X-ray diffraction data. The free R-index is defined in equation 6.30.

$$R_T^{\text{free}} = \frac{\sum_{(hkl) \in T} ||F_{\text{obs}}(hkl)| - k|F_{\text{calc}}(hkl)||}{\sum_{(hkl) \in T} |F_{\text{obs}}(hkl)|} \quad (6.30)$$

The reflection set T is omitted during any generalized modelling procedure such as, as in the case of the d(TG₄T) tetraplex structure presented here, restrained, least-

squares refinement of anisotropic temperature factors. The resultant crystallographic residual to be minimised is given in equation 6.31.

$$R_A = \sum_{(hkl) \in A} (|F_{\text{obs}}(hkl)| - k|F_{\text{calc}}(hkl)|)^2 \quad (6.31)$$

The free R-index indicates the information content of the molecular model. The conventional R-index, given in equation 6.11, can be made arbitrarily small due to the overfitting of noise in the electron density during refinement. Since noise is independent amongst individual reflections, the overfitting of the refinement using the data in set *A* will not bias the free R-index. Consequently, the calculation of the free R-index can reflect the validity of the structural detail of a model and also evaluate the validity of applying a particular refinement procedure. The value of the free R-index is greater than the value of the conventional R-index by definition. Furthermore, the difference between the free R-index and the conventional R-index is uniformly distributed as a function of resolution.

6.3 Refinement of the d(TGGGGT) Tetraplex Structure

The molecular replacement studies described in Chapter 5 resulted in two distinct initial packing models for the parallel-stranded d(TG₄T) tetraplex. These models are depicted in Figures 5.10 and 5.12, and are discussed in Sections 5.9.2 and 5.9.3 respectively.

In both cases the cell is occupied by four parallel-stranded tetraplex structures which are arranged as two axial pairs of opposite polarity, (A, B) and (C, D), with A and B related to C and D by a translation of approximately (0, 0, 0.5). This translation

would appear to be pseudosymmetric on the basis of the interpretation of the native Patterson function, described in Section 5.6, and also on the evidence of the self-rotation function, detailed in Section 5.7. The structural origin of this pseudosymmetry is not evident in the models at this stage.

The model in Figure 5.10 was obtained from a search structure which included the thymine nucleotides as T₄ quartets, analogous to the U₄ quartets detected in the NMR study of the ribonucleotide r(UG₄U) (Cheong and Moore (1992)). However, NMR studies of parallel-stranded tetraplexes have indicated that the thymine nucleotides are in fact disordered in solution (Aboul-ela *et al.*, (1992); Wang and Patel (1993a)). The resultant packing model derived from molecular replacement calculations contains sterically unfeasible overlap of the thymine quartets at the cell interface and within the cell itself between the coaxial pairs. As a result of this clash, tetraplexes B and D are slightly misaligned with respect to the helical axis defined by tetraplexes A and C. The model depicted in Figure 5.12 is more stereochemically reasonable, with the coaxial pairs stacking along a common helical axis, giving the impression of a continuous helix. In relation to the model in Figure 5.10, tetraplexes B and D are translated towards tetraplexes A and C respectively by about 3.5 Å, or one helical step, and rotated by 30–40° such that the phosphates, the largest scattering masses in the structure, are approximately coincident.

Thus it is clearly apparent that an ambiguity exists with regards to the relative packing and orientation of the tetraplex structures, arising from the lack of any information concerning the structural organization of the thymine nucleotides. Therefore, the first step taken in the refinement analysis was the calculation of rigid body minimisations

for the four tetraplex structures. The aim of this procedure was twofold: correct for the errors in the calculation of the molecular replacement solutions (Fitzgerald (1991)) and attempt to resolve the forementioned packing ambiguity.

6.3.1 Rigid Body Minimization

Rigid body minimisations were calculated using the package *XPLOR* (Brünger (1993)). The models depicted in Figures 5.10 and 5.12 were analysed. In this procedure each tetraplex was treated as a rigid body in turn with the other three held fixed. In the final step, all four tetraplexes were rigid body refined simultaneously. In each case the packing solutions were refined successively using data in the resolution ranges 17.0 to 9.0Å, 17.0 to 8.0Å, 15.0 to 7.0Å, 12.0 to 6.0Å, 10.0 to 5.0Å and, finally, 8.0 to 3.0Å. All of the isotropic thermal parameters were set to 15Å² and the atomic occupancies to 1.0.

In the case of the model shown in Figure 5.12, it was observed that this model could refine to the packing arrangement illustrated in Figure 6.6. In Figure 6.6 the tetraplexes A(red) and C(yellow) remain unchanged from the original model in Figure 5.12, with the exception of some minor adjustments. The tetraplexes B(green) and D(blue) have however changed position considerably, having been displaced by approximately 3.5Å along the helical axis with an accompanying rotation about this axis which practically superimposes the guanine tetrads with the corresponding tetrads of tetraplexes B and D in Figure 5.10. In Figure 6.6, however, the axial pairs (A, B) and (C, D) are not misaligned but share a common four-fold axis. The R-index for both of the models in Figures 5.12 and 6.6 is 0.514 for all the 3955 reflections in the final resolution range of

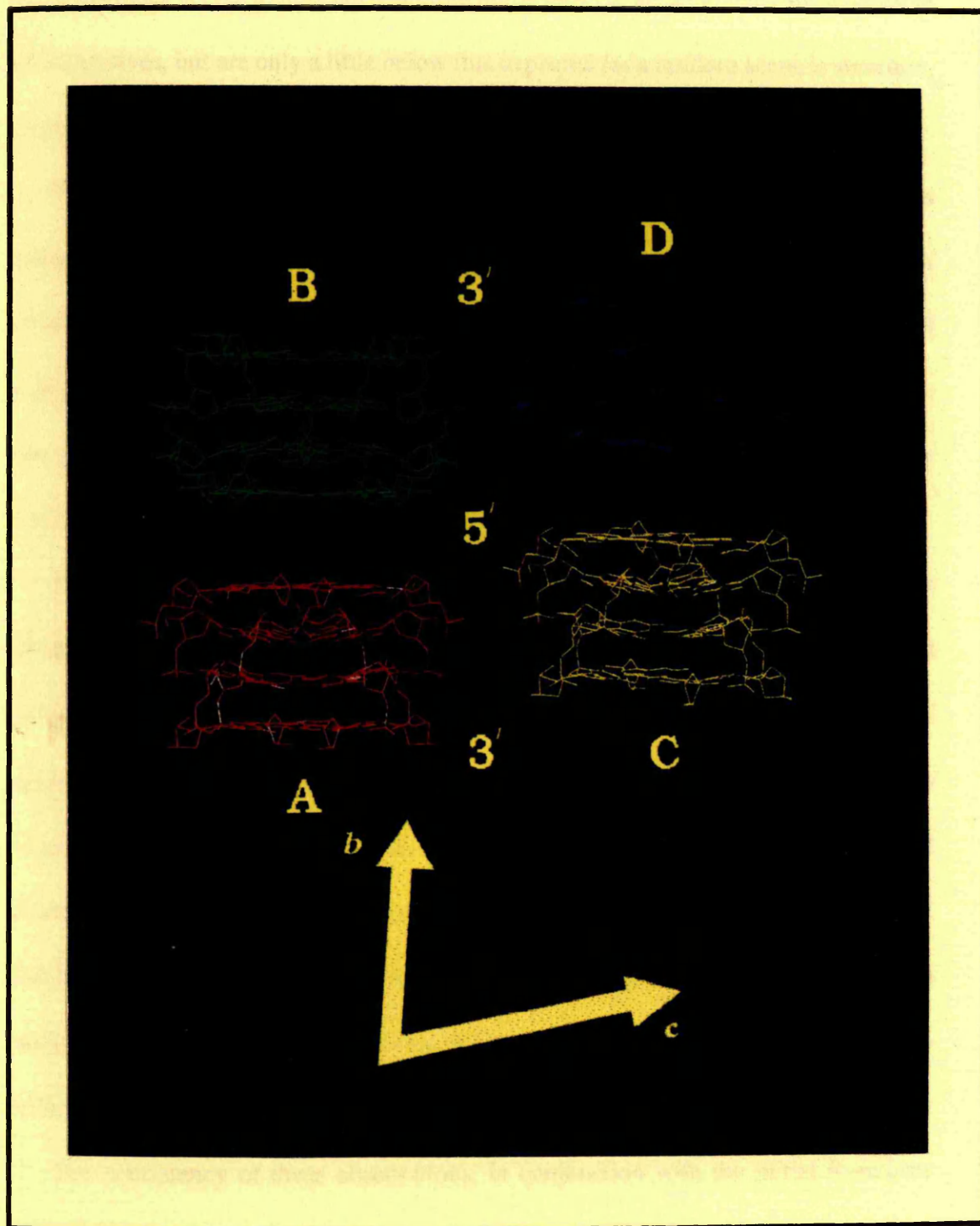


Figure 6.6: Alternative Cell-Packing Arrangement of the Tp4-T Tetraplexes

8.0–3.0Å, making them indistinguishable on the basis of the R-index alone. This value of the R-index serves to indicate that these initial packing arrangements are acceptable in themselves, but are only a little below that expected for a random acentric structure, as discussed in Section 6.1.7.

The model contained in Figure 5.10 did not change under rigid body refinement as a result of the considerable steric restrictions imposed by the clashing thymine tetrads, as witnessed at the cell interface and within the cell itself. The R-index for this model is 0.520 for all the 3955 reflections in the resolution range 8.0–3.0Å, and as such it is not significantly distinct from the values obtained from the models shown in Figures 5.12 and 6.6.

Further calculations were carried out using the model in Figure 5.10 by stripping out all of the thymine nucleotides and then subjecting the model to the rigid body refinement procedure detailed above. In this procedure, the guanine tetrads of tetraplexes A and C again remained relatively unchanged in their position and orientation. The G₄ tetrads of tetraplexes B and D, initially misaligned to tetraplexes A and C respectively, aligned themselves such that two coaxial pairs of tetraplexes were formed as in Figures 5.12 and 6.6. Indeed, iterations of the rigid body refinement procedure illustrated that the model did in fact refine to the packing arrangements shown in Figures 5.12 and 6.6 with identical R-index value in the resolution range 8.0–3.0Å.

The consistency of these observations, in conjunction with the initial R-indices from the models, inspired confidence that the overall packing arrangement was sensible. Electron density maps were calculated for each of the models using $3F_o - 2F_c$ coefficients and the phases α_c , derived from the models. Maps were calculated in the

range 8.0–3.0Å, utilizing all 3955 reflections with $F \geq 3\sigma(F)$. The maps were displayed on a Silicon Graphics Indigo graphics system using the program *O* (Jones *et al.*, (1991)). Inspection of these maps indicated that the model packing solutions were correct in terms of the coaxial arrangement of the two pairs of tetraplex structures. However, the quality of the maps did not clarify whether the tetraplexes were stacking in the manner of a pseudocontinuous helix of G₄ tetrads, as in Figure 5.12, or with an additional stepwise spacing separating the tetraplexes, as in Figure 6.6. Indeed, a combination of these two arrangements within the cell was conceivable at this stage.

Thus, whilst the rigid body analysis served to confirm the general features of the packing arrangement with regards to the G₄ tetrads, it could not resolve the ambiguity concerning the relative positioning of the coaxial tetraplex pairs. This ambiguity arises from the incompleteness of the search model employed, since no information is available concerning the possible arrangement of the thymine nucleotides within the crystal cell. It is unlikely on steric grounds that all the thymine nucleotides form T₄ tetrads, although it is impossible to rule out the formation of such an arrangement altogether. Whilst it appeared likely that some or possibly all of the thymine residues might be extended into the solvent continuum, such an arrangement cannot be conceivably modelled, nor is there any evidence for it in the electron density at this stage. Clearly, clarification of this problem was dependent on further refinement of the packing solutions.

6.3.2 Simulated Annealing

The next stage in the refinement procedure utilised a simulated annealing protocol as implemented in the package *XPLOR* (Brünger (1993)). In particular, the slow cooling protocol (Brünger *et al.*, (1990)), described in Section 6.2.3, was employed. The only alteration from the protocol described is that the initial temperature was set to 3000K, since test simulations at 4000K had proved to be unstable. All of the charges on the phosphate backbone were turned off. Planarity restraints were deployed which correspond to the energy constant k_w in the expression for E_{chem} , detailed in Section 6.2.3. A restraint of 300.0Kcal/molÅ² was imposed to maintain the planarity of the individual heterocyclic bases, whilst a milder restraint of 10.0Kcal/molÅ² was added in order to maintain the planarity of the G₄ tetrad planes. In the same manner as for rigid body refinement, all of the atomic thermal parameters were set to 15.0Å² and the occupancy to 1.0.

Simulations were carried out for the three models depicted in Figures 5.10, 5.12 and 6.6. Refinement was initially carried out in the resolution range 8.0–3.0Å using all 3955 X-ray data measurements. The resolution range was then extended to encompass the limits 8.0–2.5Å. The latter resolution range contained 7007 unique reflections with $F \geq 3\sigma(F)$. Typically it was observed that each refinement simulation should be repeated before a minimum for the procedure was reached.

The model depicted in Figure 5.10 had a final R-index value of 0.325 for all the data in the resolution range 8.0–3.0Å after refinement against this data, and a R-index of 0.384 after refinement against the data in the resolution range 8.0–2.5Å, again for all

the data included within the limits stated. The model in Figure 5.12 had R-indices of 0.310 and 0.366 for all the data in the ranges 8.0–3.0 Å and 8.0–2.5 Å respectively after each refinement simulation. Lastly, the model in Figure 6.6 had R-indices of 0.315 and 0.371 for all the data in the corresponding ranges 8.0–3.0 Å and 8.0–2.5 Å.

Electron density maps were calculated employing $3F_o - 2F_c$ and $F_o - F_c$ coefficients and using the calculated phases, α_c . These maps were inspected using the program *O* (Jones *et al.*, (1991)) on a Silicon Graphics Indigo workstation. For all three models the electron density for the G₄ tetrads in tetraplexes A and C was well-defined and continuous. Correspondingly, the overall structural details of the A and C tetraplex guanine nucleotides appeared well-defined with no suspect alterations or deformations in the stereochemistry from the original model.

However, the quality of the density for the tetraplexes B and D varied between the models. In the model refined from the arrangement in Figure 5.12, where the tetraplexes form a pseudocontinuous helix, the density was clearer and better defined than for the other models and again there appeared to be no aberrations from reasonable stereochemistry or from the assumed four-fold G₄ tetraplex structure. In the case of the models refined from Figures 5.10 and 6.6 the density for the G₄ tetrads in tetraplexes B and D was poorer and noticeably more discontinuous. This density was particularly poor for the G₄ tetrad planes at the 3' end of the B and D tetraplexes, where guanine nucleotides were observed to lift significantly out of their planar arrangement and reach towards the neighbouring cell, despite the imposition of planar restraints. The sugar conformations for these residues were also noticeably deformed.

The model in Figure 5.10 was subjected to the simulated annealing refinement

procedure since there was an interest in observing what might happen to the thymine nucleotides. At this stage a planar arrangement of some kind was considered to be a reasonable possibility. The overlapping tetrads at the cell interface had acquired poor stereochemistry, with no intelligible features, and the density in this area was extremely poor. However, a surprising result was observed for the T₄ tetrads overlapping at the interface between the coaxial pairs (A, B) and (C, D). At this interface, four of the thymine nucleotides had been expelled into the surrounding solvent continuum, with the four remaining nucleotides adopting a planar arrangement with two bases donated from each tetraplex. The electron density was poorer than that for the guanine tetrads in tetraplexes A and C, but appeared to indicate a planar arrangement of bases, although it was not clear if the bases were in fact thymines or guanines.

Taking into consideration the above observations regarding the electron density, the deformation of the 3' guanine tetrads in tetraplexes B and D, and the spontaneous adoption of a planar arrangement by some of the thymine bases situated between the coaxial pairs in the cell, a number of explanations appeared likely. The model refined from Figure 5.12 would appear to be the better of the three, with sharper density and more consistent stereochemistry evident for all four tetraplexes, indicating a packing arrangement involving the formation of a pseudocontinuous helix with all of the thymine at the interface of the coaxial pairs somehow expelled into the surrounding solvent continuum. If this were the case then the thymine planes observed in the model refined from Figure 5.10 would appear to be mimicking the guanine tetrad plane which in fact ought to be present at this position in the cell. Furthermore, the observed significant deviations of guanine nucleotides from the planar arrangement at the 3'

end of tetraplexes B and D in the models refined from Figures 5.10 and 6.6 would suggest that there are in fact no guanine tetrads present at this position, although the remaining three tetrads planes might be assumed to occupy reasonable positions even if the tetraplexes B and D have been displaced. This evidence further supported the hypothesis that the model derived from Figure 5.12 is the correct one with regards to the relative positioning of the guanine tetrads.

6.3.3 ARP Refinement

At this stage in the refinement procedure it appeared as if an impasse had been reached. The central problem stemmed from the fact that whilst the results of the simulated annealing refinements provided some evidence for the relative packing of the G₄ tetrads, it did not resolve the question of the location of the thymine nucleotides. Inspection of the electron density syntheses above gave no indication of the formation of a T₄ tetrad at the cell interface, whilst the electron density in the solvent continuum proved too difficult to interpret with regards to locating individual thymine nucleotide sugar, base and phosphate, most probably present in a variety of conformations.

With no information available on the positioning of these thymine residues, in effect approximately 30% of the scattering mass of the macromolecular unit cell contents alone remained completely unaccounted for. The absence of such a significant scattering mass would have a significantly disadvantageous effect on the quality of the phases α_c used in the electron density calculations. The only apparent way forward was the iterative manual modelling of water structure coupled with a least-squares or simulated annealing refinement procedure. Such a procedure would have been difficult

and extremely time consuming to undertake.

It was at this point that the decision was made to undertake refinement of the model using the *ARP* refinement procedure of Lamzin and Wilson (1993), discussed previously in Section 6.2.4. This procedure seemed potentially advantageous for two principle reasons; first of all, *ARP* has been shown to be particularly effective using high-resolution data, and the native d(TG₄T) data are of exceptional high-resolution and quality; secondly, solvent molecules are often important in stabilising helix structure (Saenger (1984, 1987); Saenger *et al.*, (1986); Westhof (1988); Berman (1991, 1994)), and *ARP* detects and updates ordered solvent structure in the crystal. In particular, for such a large unit cell or asymmetric unit as that observed for the d(TG₄T) tetraplex structure, putatively containing 96 individual nucleotides, ordered hydration might reasonably be expected to be quite extensive. Furthermore, the B-DNA form which experimental evidence suggests the d(TG₄T) tetraplex structure adopts (Aboul-ela *et al.*, (1992)) is favoured by a high humidity relative to the A-DNA form and has been observed to be extensively hydrated (Saenger (1984, 1987); Westhof (1988)).

As such, ordered water structure might be expected to contribute significantly to the scattering mass within the d(TG₄T) asymmetric unit cell. It might then be reasonable suspected that the overall structure solution ought to be greatly enhanced following the effective modelling of such water structure.

Unrestrained ARP

The unrestrained *ARP* protocol was implemented as described in Section 6.2.4. The refinement procedure was carried out using all 65,295 unique experimental X-ray

diffraction measurements in the resolution range 10.0–1.2 Å. Of this data 80.2% of the reflections possessed $F \geq 3\sigma(F)$. Of the 19.8% of the data with $F < 3\sigma(F)$, 19.7% of the reflections had $F \geq 2\sigma(F)$ and 0.02% $1\sigma(F) < F < 2\sigma(F)$. A high-resolution limit of 1.2 Å was chosen since the data become very weak beyond this limit. In particular, the value of $\bar{I}/\sigma(\bar{I})$ for the higher resolution bins beyond 1.2 Å falls below 3.0, as discussed in Chapter 4.

Since at this stage in the model refinement there was still some expectation that a planar T_4 tetrad arrangement might be contained within the unit cell arrangement of tetraplexes, it was decided to employ the simulated annealing refined model derived from Figure 5.10 in the initial test refinement runs. All of the atoms of the model were assigned atomic B values of 15 Å^2 and the model contained no solvent molecules initially. An R-index calculation for this model, before *ARP* refinement, yielded a value of 0.497 for all the 65,295 reflections in the resolution range 10.0–1.2 Å. Therefore, whilst unsuitable for restrained *ARP* refinement, the model was considered to be suitable for the unrestrained procedure.

The unrestrained refinement procedure typically sought to locate 10 new atoms and delete up to 5 poorly defined atoms in each refinement cycle, as illustrated in Figure 6.4. Executed in this way, the unrestrained *ARP* refinement procedure took typically 50 to 100 iterative cycles before there were no more observed improvements in the value of the R-index. For each implementation, starting from a model containing no solvent structure, the unrestrained refinement protocol added something in the region of 600–700 new atoms. Whilst this is greater than the 500 or so water molecules that might be expected to be located in the unit cell when the structure solution is determined,

the primary interest at this stage lay in determining ways in which the model might be augmented and the phases improved. Other *ARP* parameters were defined as described for unrestrained *ARP* refinement in Section 6.2.4.

Following the initial application of the unrestrained *ARP* refinement procedure to the model derived from Figure 5.10, the R-index dropped after convergence to 0.220 for the 65,295 reflections in the resolution range 10.0–1.2 Å. Electron density syntheses were calculated using all data in the forementioned range and with coefficients $3F_o - 2F_c$ and $F_o - F_c$. The resultant maps were then inspected using a Silicon Graphics Indigo Workstation and the program *O* (Jones *et al.*, (1991)).

The electron density syntheses were superimposed upon the atomic coordinates of the model resulting from the unrestrained refinement procedure, as well as the coordinates of the initial model structure. The model resulting from the unrestrained refinement proved to be significantly dismantled, particularly in those areas which did not match up well with the calculated electron density maps. However, the definition and connectivity of the electron density was of sufficiently high quality to reveal that the molecular replacement solution depicted in Figure 5.10, and hence in Figure 6.6, was false.

Inspection of the density syntheses superimposed upon the initial model revealed that the guanine tetrads defined by tetraplexes A and C were more or less correctly situated, as had been suspected from previous refinement experiments. Furthermore, it was clearly apparent that tetraplexes B and D were spatially out of register by one base step along the helical axis. The interface between tetraplexes in the initial model, previously occupied by a T₄ arrangement, was clearly the site of a guanine G₄ tetrad

for both of the coaxial pairs. The connectivity of the density was unambiguous in asserting that these tetrads belong to the tetraplexes stacked upon tetraplexes A and C. It had been unequivocally determined, and from the incorrect molecular replacement solution, that the packing solution of the guanine tetrad tetraplexes, illustrated in Figure 5.12, was correct. Consequently, the initial model was manually rectified in preparation for further refinement steps.

The unrestrained *ARP* refinement procedure was then further deployed in order to augment the now clarified packing model. This procedure required the execution of iterative blocks of the unrestrained *ARP* procedure. This involved the repetitive application of the refinement protocol combined with the graphics analysis described above. After each block of unrestrained refinement cycles, $3F_o - 2F_c$ and $F_o - F_c$ electron density syntheses were calculated from the final output model and inspected, superimposed upon the initial model for that particular block of refinement cycles. Previously undetected features of the unit cell contents were then manually built in and existing parts of the model were adjusted to better fit the observed density. This operation was carried out using the program *FRODO* (Jones (1978, 1985)) on an Evans and Sutherland graphics workstation with extensive help from Dr. Ben Luisi. Once the model could be improved no further the augmented model, containing no solvent structure, was then utilised as the starting model for the next block of unrestrained *ARP* refinement cycles.

A total of five blocks of unrestrained *ARP* cycles were carried out. The R-index for the converged unrestrained refinement procedure was observed to further reduce for each subsequent procedure involving an improved initial model. After the final

block of unrestrained refinement, the R-index was 0.163 for all 65,295 reflections in the resolution range 10.0–1.2Å. Clear density was observed for the majority of the thymine nucleotides which did not participate in the formation of a T₄ tetrad arrangement. Rather, these residues were observed to participate in an extensive array of lattice contacts, to which the high crystalline order of the structure can be attributed. In total, 27 of the 32 possible thymine nucleotides were located and incorporated into the unit cell model. There was no apparent evidence of any density to describe the location of the remaining 5 thymine residues. A string of spherical, atomic-sized densities were observed along the helical axes, running through the core of the coaxial tetraplex pairs. These spherical densities were modelled as sodium cations. Furthermore, calcium cations were included in the crystallisation solution and nine ions were detected as well ordered with clearly defined hydration shells. Since the calciums ions are the highest scattering masses in the cell, they were still visible in the $3F_o - 2F_c$ electron density syntheses at a contouring of 9σ , where σ is the root mean squared standard deviation from mean density. When the model had been augmented as much as was possible using the unrestrained *ARP* procedure, the refinement progressed to the use of the restrained *ARP* refinement protocol.

Restrained ARP

The final model derived from the unrestrained *ARP* refinement procedure contained all 64 guanine nucleotides as well as 27 of the 32 thymine nucleotides, to give a total of 1,916 non-hydrogen macromolecular atoms. In addition, the model contained 14 sodium and 9 calcium cations. At this stage the model had a R-index value of 0.291

for all 65,295 reflections in the resolution range 10.0–1.2Å. The inclusion of 100 well-defined solvent molecules from the unrestrained refinement procedure lowered the R-index to 0.236 for the forementioned resolution range. Solvent molecules were modelled as oxygen atoms and were only accepted if they possessed reasonable hydrogen-bonding geometry or, as in the case of the calcium ions, the coordination sphere was well-defined. With the R-index now less than 0.300, the model was deemed appropriate for the restrained *ARP* refinement procedure.

The restrained *ARP* refinement procedure, as discussed in Section 6.2.4, utilises the conjugate gradient restrained, reciprocal space, least-squares refinement procedure of Hendrickson and Konnert (1980), as implemented in the *CCP4* programs *PROTIN* and *PROLSQ*. Whilst written with proteins in mind, *PROTIN* and *PROLSQ* were deployed in the refinement of the d(TG₄T) tetraplex structure by utilising a modified stereochemical restraints dictionary for nucleic acids, derived from the orthogonal coordinates of Taylor and Kennard (1982). This dictionary was kindly supplied by Dr. Madeleine Moore of the University of York.

The starting model was prepared by setting the atomic *B*-values of all the atoms in the unit cell to 15.0Å², with atomic occupancies of 1.0. Atomic positions and individual isotropic temperature factors were refined with restraints. Up until this point atomic scattering factors for all atoms had been expressed as 2-Gaussian approximations. These were now increased to 5-Gaussian approximations for all C, N, O and P atoms, as well as for Na⁺ and Ca²⁺ cations. The restrained *ARP* procedure was set to locate 6 new water atoms per cycle and to delete a maximum of 3 poorly defined water atoms per cycle. Additionally, the real space refinement options was applied to the water

structure only. All other *ARP* parameters were defined as described in Section 6.2.4. In this manner, the macromolecular structure was refined by restrained, reciprocal space, least-squares calculation, with simultaneous updating and refinement of the surrounding ordered solvent continuum structure.

Cycles of refinement were interspaced with calculations of $3F_o - 2F_c$ and $F_o - F_c$ electron density maps, using all the data in the resolution range 10.0–1.2 Å. The maps were displayed on an Evans and Sutherland graphics workstation using the program *FRODO* (Jones (1978, 1985)). Every few refinement cycles the model was inspected and manually adjusted to better fit the observed electron density. Analysis of the electron density syntheses up until the point of refinement convergence revealed no details that might be interpreted as the 5 missing thymine nucleotides. Consequently, these were excluded from the final model, although it was possible to model in the phosphate groups belonging to the 5' thymine nucleotides. The remaining 64 guanine nucleotides, 27 thymine nucleotides, 14 sodium cations and 9 calcium cations were all well-defined.

Refinement was continued until the $F_o - F_c$ difference density maps were featureless at a contour level of 3σ , the deoxyribonucleic acid geometry was as good as or better than that specified by the target root mean square difference values and the *B*-value distribution was sensible. The model output from restrained *ARP* refinement contained 640 solvent molecules, modelled as oxygen atoms. However, inspection of the electron density syntheses resulted in the manual removal of 126 which were not justified on the basis of the electron density at or near the atomic centre.

The final model contains a total of 2,453 non-hydrogen atoms, consisting of 1,916

Parameter	Restraint	RMS Δ
Bond Lengths (1–2) (\AA)	0.020	0.014
Angle Related Distances (1–3) (\AA)	0.040	0.035
Planar Groups (\AA)	0.020	0.010
Chiral Volumes (\AA^3)	0.100	0.064
Van Der Waals' Distances:		
Multiple Torsion Contacts	0.300	0.302
Possible Hydrogen–Bonds	0.300	0.212
Thermal Factors:		
Mainchain Bond (1–2) (\AA^2)	1.500	0.951
Mainchain Angle (1–3) (\AA^2)	3.00	2.691
Sidechain Bond (\AA^2)	4.500	1.721
Sidechain Angle (\AA^2)	9.000	2.445
Final R–Factor (all data)		
	0.1874	
Total Number of Atoms:		
Tetraplex	1916	
Sodium	14	
Calcium	9	
Water Molecules	514	
Average Temperature Factors:		
Sugar–Phosphate Backbone	12.50	
Heterocyclic Bases	8.71	
Sodium Ions	6.34	
Calcium Ions	6.56	
Water Molecules	28.62	

Table 6.1: Refinement Statistics for the Restrained ARP Refinement Procedure

macromolecular atoms, 14 sodium cations, 9 calcium cations and 514 water molecules. The final *R*-index value for this model, calculated for all the data in the resolution range 10.0–1.2 Å is 0.1874. The results of this refinement procedure are presented in Table 6.1, where RMS Δ is the final root mean square deviation of the parameter restrained. A simple breakdown of the distribution of the thermal *B*-values of the various components of the structure is also given. A more detailed breakdown of the thermal parameters is described following the results of the *SHELXL-93* refinement procedure. However, it is worth noting that the sodium and calcium cations are well-defined, and the heterocyclic bases, which are predominantly buried within the guanine tetraplexes, have lower overall average *B*-values than the solvent exposed sugar-phosphate backbone, as would be expected.

6.3.4 Refinement Using SHELXL-93

The program *SHELXL-93* (Sheldrick (1993)) was employed in the final stage of the refinement protocol. The final model obtained from the restrained *ARP* procedure was used as the initial model for this refinement procedure. As discussed in Section 6.1.5, taking into consideration the three positional and six anisotropic thermal parameters, the refinement procedure for the d(TG₄T) tetraplex unit cell model has an approximate ratio of observables to parameters of 2.9:1 for the 65,264 experimental X-ray measurements in the resolution range 10.0–1.2 Å. Thus, anisotropic refinement of the structure is feasible with the application of suitable restraints.

Anisotropic temperature factor refinement was carried out using the data in the resolution range 10.0–1.2 Å. The *SHELXL-93* refinement procedure utilises the method

Atoms Restrained	Target Distance	Sigma(σ)	Restraint Description
N ₁ -C ₂	1.379	0.03	1-2 Bond Distances
N ₁ -C ₆	1.380	0.03	
C ₂ -N ₃	1.373	0.03	
C ₂ -O ₂	1.218	0.03	
N ₃ -C ₄	1.383	0.03	
C ₄ -C ₅	1.440	0.03	
C ₄ -O ₄	1.227	0.03	
C ₅ -C ₆	1.338	0.03	
C ₅ -C _{5α}	1.520	0.03	
N ₁ -N ₃	2.319	0.05	1-3 Bond Distances
N ₁ -C ₅	2.386	0.05	
N ₁ -O ₂	2.286	0.05	
C ₂ -C ₄	2.467	0.05	
C ₂ -C ₆	2.405	0.05	
N ₃ -O ₂	2.267	0.05	
N ₃ -O ₄	2.260	0.05	
N ₃ -C ₅	2.378	0.05	
C ₄ -C ₆	2.397	0.05	
O ₄ -C ₅	2.372	0.05	
C ₄ -C _{5α}	2.569	0.05	
C ₆ -C _{5α}	2.482	0.05	

Table 6.2: *SHELXL-93* Distance Restraints for Thymine Heterocycles

Atoms Restrained	Target Distance	Sigma(σ)	Restraint Description
N ₁ –C ₂	1.375	0.03	1–2 Bond Distances
N ₁ –C ₆	1.392	0.03	
C ₂ –N ₂	1.341	0.03	
C ₂ –N ₃	1.327	0.03	
N ₃ –C ₄	1.355	0.03	
C ₄ –C ₅	1.377	0.03	
C ₄ –N ₉	1.377	0.03	
C ₅ –C ₆	1.425	0.03	
C ₅ –N ₇	1.389	0.03	
C ₆ –O ₆	1.239	0.03	
N ₇ –C ₈	1.304	0.03	
C ₈ –N ₉	1.374	0.03	
N ₁ –N ₃	2.386	0.05	1–3 Bond Distances
N ₁ –C ₅	2.324	0.05	
N ₁ –N ₂	2.307	0.05	
N ₁ –O ₆	2.280	0.05	
C ₂ –C ₄	2.221	0.05	
C ₂ –C ₆	2.453	0.05	
N ₂ –N ₃	2.307	0.05	
N ₃ –C ₅	2.461	0.05	
N ₃ –N ₉	2.434	0.05	
C ₄ –C ₆	2.408	0.05	
C ₄ –N ₇	2.277	0.05	
C ₄ –C ₈	2.197	0.05	
C ₅ –C ₈	2.215	0.05	
C ₅ –N ₉	2.193	0.05	
C ₅ –O ₆	2.390	0.05	
C ₆ –N ₇	2.542	0.05	
N ₇ –N ₉	2.240	0.05	

Table 6.3: *SHELXL-93* Distance Restraints for Guanine Heterocycles

Atoms Restrained	Sigma(σ)	Restraint Description
C _{1'} -C _{2'}	0.03	1-2 Bond Distances
C _{2'} -C _{3'}	0.03	
C _{3'} -C _{4'}	0.03	
C _{4'} -C _{5'}	0.03	
C _{1'} -O _{4'}	0.03	
C _{3'} -O _{3'}	0.03	
C _{4'} -O _{4'}	0.03	
C _{5'} -O _{5'}	0.03	
C _{1'} -C _{3'}	0.05	1-3 Bond Distances
C _{1'} -C _{4'}	0.05	
C _{2'} -C _{4'}	0.05	
C _{3'} -C _{5'}	0.05	
C _{2'} -O _{3'}	0.05	
C _{2'} -O _{4'}	0.05	
C _{3'} -O _{4'}	0.05	
C _{4'} -O _{3'}	0.05	
C _{4'} -O _{5'}	0.05	
C _{5'} -O _{4'}	0.05	

Table 6.4: *SHELXL-93* Similarity Restraints for Deoxyriboses

Atoms Restrained	Sigma(σ)	Restraint Description
O _{3'} -P	0.03	1-2 Bond Distances
O _{5'} -P	0.03	
O _{1P} -P	0.03	
O _{2P} -P	0.03	
O _{1P} -O _{3'}	0.05	1-3 Bond Distances
O _{2P} -O _{3'}	0.05	
O _{1P} -O _{5'}	0.05	
O _{2P} -O _{5'}	0.05	
O _{3'} -O _{5'}	0.05	
C _{5'} -P	0.15	
C _{3'} -P	0.15	

Table 6.5: *SHELXL-93* Similarity Restraints for Phosphates

Atoms Restrained	Sigma(σ)	Restraint Description
N ₁ -C _{1'}	0.03	Thymine β-Glycosyl
N ₁ -C _{2'}	0.05	
N ₁ -O _{4'}	0.05	
C _{1'} -C ₂	0.05	
C _{1'} -C ₆	0.05	
N ₉ -C _{1'}	0.03	Guanine β-Glycosyl
N ₉ -C _{2'}	0.05	
N ₉ -O _{4'}	0.05	
C _{1'} -C ₄	0.05	
C _{1'} -C ₈	0.05	

Table 6.6: *SHELXL-93* Similarity Restraints for Base-Deoxyribose Linkages

of structure refinement against F_o^2 so that all 65,264 reflections in the resolution range 10.0–1.2 Å were used, as discussed in Section 6.2.5. Restraints for standard bases were calculated from the orthogonal coordinates contained in Taylor and Kennard (1982). Tables 6.2 and 6.3 show the applied 1–2 and 1–3 bond distances along with their allowed standard deviations, σ . The 1–2 bond distances represent explicit restraints of chemically bonded distances. The 1–3 bond distances, on the other hand, represent restraints on the distances between non-bonded atoms, and hence an implied restraint on the bond angle involving the atom chemically bonded to both e.g. the C₄–C₆ distance restraints implies a restraint on the bond angle C₄–C₅–C₆.

SHELXL-93 also provides the option of applying similarity restraints. When these restraints are applied, specified atomic distances, which occur more than once in the entire structure, are restrained to be similar within an effective standard deviation. This is particularly powerful in the case of macromolecular structures where nucleotides and/or amino acids, which occur many times within the asymmetric unit, contain chemically identical constituents. Consequently, similarity restraints were applied in the *SHELXL-93* refinement of the d(TG₄T) structure in the form of restraints on the deoxyribose sugar dimensions, the phosphate groups and the base–glycosyl linkage. The applied similarity restraints, along with their effective standard deviations are given in Tables 6.4, 6.5 and 6.6, respectively. Weak restraints were applied to the anisotropic thermal parameters such that anisotropic parameters along bond lengths were restrained to be the same. Neighbouring atoms were also restrained to have the similar anisotropic thermal parameters. Since these restraints are weak, the values of the parameters for neighbouring atoms can diverge as the refinement progresses. Water

molecules were restrained to be approximately isotropic in behaviour. Combining distance and similarity restraints, a total of 162,106 restraints were generated for the refinement procedure.

Hydrogen atoms were included in the model in order to better model the density around atoms in the model which form hydrogen-bonds. *SHELXL-93* automatically includes the hydrogen atoms in the model. This requires the specification by the user of the nature of the orbital configurations of the non-hydrogen atoms to which the hydrogens are to be attached. The position of the hydrogen atoms themselves are not refined. Rather, a riding model is adopted whereby the positional shifts applied to the non-hydrogen atoms are also applied to the hydrogen atoms riding upon them, and the geometries of the covalent bonds formed by the hydrogen atoms to the non-hydrogen atoms are idealised after each refinement cycle.

The refinement procedure was broken down into two stages. In the first stage the structure was refined using the conjugate-gradient least-squares approach of Hendrickson and Konnert (1980). This procedure was monitored by calculation of the free R-index (Brünger (1992)), defined in equation 6.30. A randomly selected tenth of all the reflections were excluded from the CGLS refinement procedure. The resultant free R-index for the CGLS anisotropic refinement was 0.1758. This served to indicate that the refinement procedure was resulting in meaningful modification of the molecular model, since overfitting of the model using the remaining 90% of the data would not have resulted in the observed significant lowering in the free R-index value. It was therefore concluded that anisotropic refinement was justified for the d(TG₄T) cell model.

The CGLS refinement procedure converged after 15 cycles and the second stage of the *SHELXL-93* refinement procedure was commenced. This involved the refinement of overlapping full-matrix blocks in order to obtain values of the estimated standard deviations in bond lengths and bond angles from the normal-matrix. A total of 9 cycles of least-squares block-matrix refinement were executed. Each of the first 8 cycles defined matrix blocks describing two of the sixteen d(TG₄T) strands contained within the model. The final cycle defined a block describing the Na⁺ and Ca²⁺ cations, and the solvent structure.

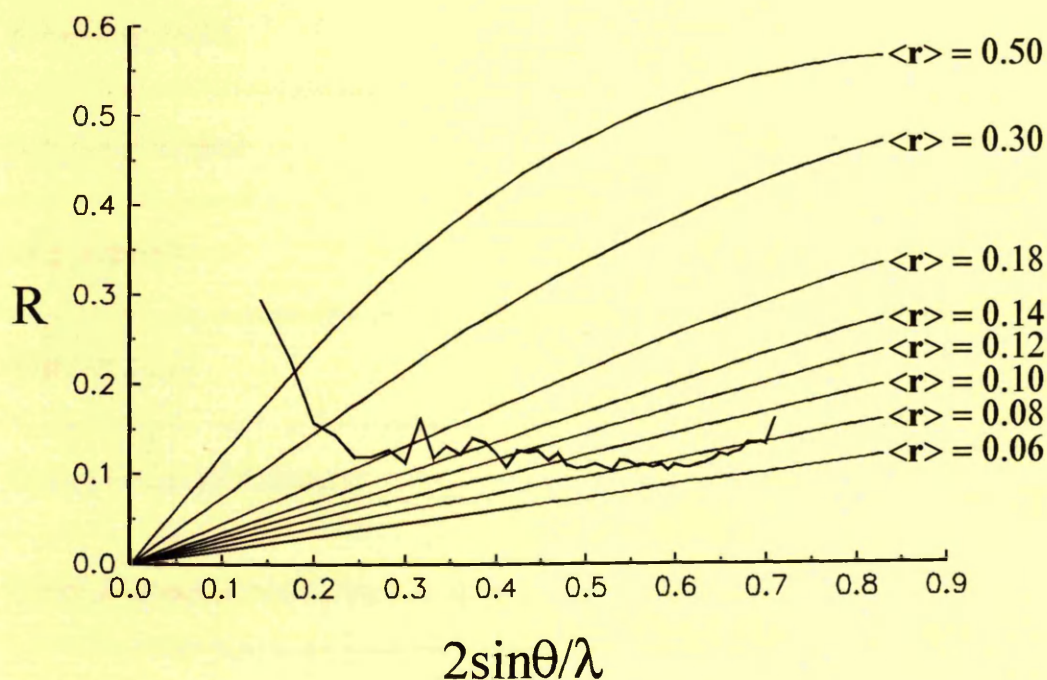


Figure 6.7: Luzzati Plot for the Refined d(TGGGGT) Tetraplex Structure

The final model has an overall deviation from ideality of 0.016 Å on bond lengths and 1.61° in bond angles. A Luzzati plot, shown in Figure 6.7, illustrates the upper bound of the error in atomic positions for the final model. A number of standard Luzzati curves are shown superimposed for comparison. In the *SHELXL-93* refinement a bulk

Empirical formula	$\text{C}_{910}\text{H}_{846}\text{Ca}_9\text{N}_{374}\text{Na}_{14}\text{O}_{1113}\text{P}_{80}$
Unit Cell Lengths	$a = 28.76(3)\text{\AA} \quad b = 35.47(4) \text{\AA} \quad c = 56.77(6)\text{\AA}$
Unit Cell Angles	$\alpha = 74.39(5)^\circ \quad \beta = 77.64(5)^\circ \quad \gamma = 89.73(5)^\circ$
Data(<i>n</i>)/Restraints(<i>r</i>)/Parameters(<i>p</i>)	64357/162106/22115
Goodness-of-fit(GooF)	1.229
R-Index [<i>I</i>>2σ(<i>I</i>)]	0.1156
R-Index [All Data]	0.1655
wR2 [<i>I</i>>2σ(<i>I</i>)]	0.3113
wR2 [All Data]	0.4074
Free R-Index [10 % of Data]	0.1758
Residual Electron Density(Peak & Hole)	$0.675\text{e}\text{\AA}^{-3} \quad -0.446\text{e}\text{\AA}^{-3}$

$$\text{GooF} = \left[\frac{\sum_{hkl} w(F_{\text{obs}}^2 - F_{\text{calc}}^2)^2 + \sum_i [w_r(r_i^{\text{ideal}} - r_i^{\text{final}})^2]}{r_n + (n - p)} \right]^{\frac{1}{2}}$$

$$\text{wR2} = \left[\frac{\sum_{hkl} w(F_{\text{obs}}^2 - F_{\text{calc}}^2)^2}{\sum_{hkl} w(F_{\text{obs}}^2)^2} \right]^{\frac{1}{2}}$$

Table 6.7: Summary of Final Refinement Statistics from *SHELXL-93*

solvent correction, to compensate for the scattering contribution of the unordered solvent within the crystal, was applied to the data to the upper limit of refinement of 10.0 Å resolution. It can be seen from Figure 6.7 that the R-index rises dramatically at very low-resolution, probably due to the incompleteness of the modelling of this bulk solvent contribution to the crystal scattering. This rise may also result from extinction effects.

Refinement statistics from the *SHELXL-93* refinement procedure, based upon the intensities, I , are detailed in Table 6.7. The conventional R-index has a final value of 0.1190 for all reflections with $I \geq 2\sigma(I)$, and 0.1843 for all the intensity data. The R-index for all the structure amplitude data in the range 10.0–1.2 Å is 12.4%. The total data recorded represents 97.7% of the data theoretically possible in this resolution range. The statistical quantities GooF and wR2 are defined as used in *SHELXL-93*. For the restrained Goodness-of-fit (GooF), r_n is the total number of restraints, r^{ideal} is the target value, r^{final} is the final value of the restrained distance from the molecular model and w_r is the weighting. Typically, the value of GooF ought to approach unity for a well refined structure and is normally slightly augmented when calculated using F_o^2 .

6.4 Quality of the Final Structure Solution

The resolution of the data collected acts as an approximate guide to the quality of the structure solution. The nominal resolution is fairly uninformative in itself without knowledge of the completeness of the data that could be experimentally measured. In

the case of the d(TG₄T) tetraplex structure, 97.7% of the theoretically possible data were collected within the ranges 10.0–1.2 Å. However, in macromolecular structure analysis, a 2 or 3 σ cut-off is normally applied. In this respect it is worth noting that 97.6% of the data have $F \geq 2\sigma(F)$ in the forementioned resolution range.

The most commonly accepted measure of the accuracy of a crystal structure is the crystallographic R-factor, or R-index. However, as noted in Section 6.2.5, the R-index can be made meaninglessly low by the modelling of structurally irrelevant noise as crystal solvent content. In this regard the free R-index of Brünger (1992) constitutes a more reliable measurement of the structural validity of the refinement procedure.

Whilst a certain degree of solvent structure most definitely forms part of an oligonucleotide structure, it is an interesting indicator of the accuracy of the macromolecular model itself to know how much the R-index rises when the solvent structure is removed. The model output from the restrained *ARP* refinement procedure had an R-index value of 0.1874 for all 65,264 reflections in the resolution range 10.0–1.2 Å. The R-index value of this model with the water structure removed is 0.281 for all the data in the same resolution range.

Dickerson (1992) suggests a number of other simple quantities which can be considered in evaluating the quality of an oligonucleotide structure solution. With reference to double helical structures, the volume per base-pair acts as a measure of how tightly an oligonucleotide structure is packed. The d(TG₄T) tetraplex structure packs with a volume of 568 Å³ per base and, for comparison with double helical structures, 1136 Å³ per 'imaginary' base-pair. Thus, on the whole, the structure is extremely tightly packed.

Helices in the A- and B-DNA form typically pack with a volume per base-pair

of 1400\AA^3 . Indeed, the d(TG₄T) tetraplex structure is comparable to Z-DNA which typically has a volume per base-pair of 1100\AA^3 (see Dickerson (1992) and references within). As the refinement procedure has shown, a densely packed crystal is potentially advantageous since the contents are held in a more ordered array, leading to a higher resolution diffraction pattern. However, dense packing can lead to localised steric deformations. Such deformations are observed in two of the four d(TG₄T) tetraplexes and shall be discussed in the next chapter.

Another empirical indication of the level of knowledge attained for the oligonucleotide structure is the number of observed data per unique base-pair. The d(TG₄T) tetraplex unit cell has a total of 680 observed reflections per base, and hence 1,360 per 'imaginary' base-pair. Again, for the sake of comparison with most other oligonucleotide structure determinations, the number of observed reflections with $F \geq 2\sigma(F)$ is a more informative indicator.

In this regard, the four tetraplex unit cell has a total of 679 reflections per base and 1359 reflections per base-pair observed. On the basis of this criterion, the best determined B-DNA double helical structure is the decamer C-C-A-A-C-G-T-T-G-G determined to 1.4\AA resolution with 880 reflections observed per base-pair. This however assumes a 1σ cut-off in the experimental data. The best determined structure by this standard is the Z-DNA helical sequence C-G-C-G-C-G determined with spermine in the crystallization liquor. Determined to a resolution of 0.9\AA with a 2σ cut off in the experimental data, this structure has a total of 2,500 reflections per base-pair (see Dickerson (1992) and references within). The d(TG₄T) tetraplex structure has been observed to diffract to a maximum resolution of 0.88\AA using synchrotron radiation

at 4° in experiments conducted by Dr. Ben Luisi. Successful data collection to this limit would result in an approximate total of 168,096 unique reflections as determined by equation 6.10. This would result in a total of 1,751 observations per base and 3,502 per base-pair. Flash freezing experiments might further extend this resolution limit.

Whilst water structure is undoubtedly an important aspect of the structural stabilisation of DNA helical structure, it is nevertheless crucial to have a degree of confidence that one hasn't been overcautious or overzealous in modelling water structure. In this respect, the number of water molecules located and accepted per base-pair acts as a useful index of the likelihood of this water structure being convincing, as does the number of observed data per water molecule found.

The final d(TG₄T) tetraplex structure, containing 514 water molecules, has a total of 5.35 solvent molecules per base and 10.7 per base-pair. For all the data observed in the resolution range 10.0–1.2 Å this corresponds to 127 observed reflections per solvent molecule. At the 2 σ cut-off the number of reflections per water molecule is also approximately 127. This compares favourably with other B-DNA helical studies as reviewed by Dickerson (1992). In particular, the determination of the structure of C-C-A-A-C-G-T-T-G-G to 1.4 Å resolution contained a total of 14.2 waters per base-pair and 62 observed reflections per water molecule at a cut-off of 1 σ .

Finally, an important consideration in any macromolecular structure determination is the values and distributions of the thermal *B*-parameters. A summary of the *B*-values for the d(TG₄T) unit cell model is given in Table 6.8. The averages displayed were calculated from the final model output from the *SHELXL-93* refinement where the atoms are given anisotropic and equivalent isotropic thermal parameters. The averages

are taken over the sum of isotropic values.

Chemical Description	No. of Atoms	Average B-Factor (\AA^2)
Four Individual Tetraplexes	467/483/483/483	11.04/13.11/13.79/11.30
All Four Tetraplexes	1916	12.32
Guanine Nucleotides	1408	10.72
Thymine Nucleotides	508	16.79
Phosphate Groups(PO_4)	423	16.45
Guanine Phosphate Groups	320	14.90
Thymine Phosphate Groups	103	21.26
Deoxyribose Sugars	729	13.71
Guanine Deoxyribose Sugars	512	12.17
Thymine Deoxyribose Sugars	217	17.34
Sugar-Phosphate Backbone	969	14.54
Heterocyclic Bases	947	10.05
Guanine Heterocyclic Bases	704	8.31
Thymine Heterocyclic Bases	243	15.08
Sodium Ions	14	8.14
Calcium Ions	9	8.68
Solvent Structure(Oxygens)	514	33.65

Table 6.8: Summary of Average Isotropic Temperature Factors

The values for the various components of the structure are, overall, noticeably low as might be anticipated for a highly ordered, well diffracting structure. The Na^+ and Ca^{2+} cations are well-defined with average values of 8.14\AA^2 and 8.68\AA^2 for the 14 sodium and 9 calcium cations respectively. The water structure has an average isotropic B -value of 33.65\AA^2 for the 514 waters modelled as oxygen atoms in the unit

cell. It is worth noting that some of the water molecules have high temperature factors, on occasion above 50\AA^2 . Waters with high temperature factors were accepted in the final model if the final electron density syntheses provided evidence supporting their validity.

The distribution of temperature factors between the thymine and guanine nucleotides is noticeable. For all of the nucleotide structural components – base, sugar and phosphate – the chemical groupings belonging to the guanine nucleotides have significantly lower *B*-factors than those observed for the thymine nucleotides. This correlates with the structural organisation of the different nucleotides within the tetraplex. The guanine bases are all packed within the four stranded tetraplex helical structure, whilst the thymine nucleotides are exclusively involved in the formation of lattice contacts in the surrounding solvent continuum. The sugar–phosphate backbone attached to the guanine bases has higher average temperature factor values, on the outside of the helix and in contact with the solvent, than the values observed for the guanine bases packed tightly within the helix.

The values of the atomic *B*-parameters ought to be correlated to the estimated standard deviations on the atomic positions, expressed in such quantities as bond length and bond angle. Tables 6.9–6.14 detail bond lengths and estimated standard deviations for a single d(TG₄T) strand, whilst Tables 6.15–6.20 detail bond angles and estimated standard deviations for the same oligonucleotide strand. These values were determined from the least-squares block-matrix refinement undertaken using *SHELXL-93*. Atoms highlighted in bold belong to the previous 5' orientated residue and are involved in bond lengths and bond angles across nucleotide units as the latter

are defined.

A noticeable correlation can be seen between the thermal parameters of the thymine (5'T(1) and 3'T(6)) and guanine (G(2)–G(5)) nucleotides. The estimated standard deviations from the normal–matrix are substantially higher for the thymine nucleotides (Tables 6.9, 6.14, 6.15, 6.20) than for the guanine nucleotides (Tables 6.10–6.13 and 6.16–6.19). This corresponds to the observed differences in the isotropic thermal parameters for the two residues. The exact nature of this correlation, as determined from plots of isotropic B -values against estimated standard deviations, for example, has not been determined.

Table 6.8 demonstrates that for the four tetraplexes occupying the unit cell there is a discrepancy in the overall B -values for the individual tetraplexes themselves. In particular, two of the tetraplexes have lower average B -values of 11.04\AA^2 and 11.30\AA^2 , whilst the other two tetraplexes have average B -values of 13.11\AA^2 and 13.79\AA^2 . This difference in temperature factors for two of the tetraplexes appears to arise from the nature of the packing of the four tetraplexes. One tetraplex in each coaxial pair is deformed at the interface of the pairs. This results in a slightly deformed geometry, and it is these deformed tetraplexes which have higher average B -factors than the other two tetraplexes in the cell. Thus, whilst the tetraplexes with lower B -value averages correspond well to the B–DNA architecture expected for the structure (Aboul–ela *et al.*, (1992)), the other two tetraplexes exhibit characteristics of A–DNA at the packing interface. The strand used to generate Tables 6.9–6.20 comes from the tetraplex with average B -value 11.04\AA^2 and is representative in its estimate standard deviation of all the oligonucleotide strands constituting the two tetraplex structures with lower average

Residue	Bonded Atom Pair	Bond Length & ESD
5'T(1)	O _{5'} -C _{5'}	1.443(14)
	C _{5'} -C _{4'}	1.535(14)
	C _{4'} -O _{4'}	1.466(13)
	C _{4'} -C _{3'}	1.498(13)
	O _{4'} -C _{1'}	1.437(12)
	C _{1'} -N ₁	1.485(13)
	C _{1'} -C _{2'}	1.525(13)
	N ₁ -C ₂	1.37(2)
	N ₁ -C ₆	1.41(2)
	C ₆ -C ₅	1.33(2)
	C ₂ -O ₂	1.21(2)
	C ₂ -N ₃	1.37(2)
	N ₃ -C ₄	1.38(2)
	C ₄ -O ₄	1.23(2)
	C ₄ -C ₅	1.44(2)
	C ₅ -C _{5α}	1.54(2)
	C _{2'} -C _{3'}	1.543(14)
	C _{3'} -O _{3'}	1.449(13)

Table 6.9: Bond Lengths and Estimated Standard Deviations – Residue 5'T(1)

Residue	Bonded Atom Pair	Bond Length & ESD
G(2)	O _{3'} -P	1.612(9)
	P-O _{2P}	1.465(10)
	P-O _{1P}	1.487(9)
	P-O _{5'}	1.595(9)
	O _{5'} -C _{5'}	1.465(12)
	C _{5'} -C _{4'}	1.526(13)
	C _{4'} -O _{4'}	1.468(12)
	C _{4'} -C _{3'}	1.518(13)
	O _{4'} -C _{1'}	1.382(10)
	C _{1'} -N ₉	1.482(11)
	C _{1'} -C _{2'}	1.600(12)
	N ₉ -C ₈	1.367(13)
	N ₉ -C ₄	1.386(12)
	C ₄ -C ₅	1.371(14)
	C ₄ -N ₃	1.379(13)
	N ₃ -C ₂	1.352(13)
	C ₂ -N ₂	1.324(13)
	C ₂ -N ₁	1.383(14)
	N ₁ -C ₆	1.382(13)
	C ₆ -O ₆	1.255(13)
	C ₆ -C ₅	1.427(14)
	C ₅ -N ₇	1.389(13)
	N ₇ -C ₈	1.311(13)
	C _{2'} -C _{3'}	1.524(13)
	C _{3'} -O _{3'}	1.467(12)

Table 6.10: Bond Lengths and Estimated Standard Deviations – Residue G(2)

Residue	Bonded Atom Pair	Bond Length & ESD
G(3)	O_{3'}-P	1.596(9)
	P-O_{2P}	1.474(10)
	P-O_{1P}	1.477(10)
	P-O_{5'}	1.585(10)
	O_{5'}-C_{5'}	1.448(14)
	C_{5'}-C_{4'}	1.523(14)
	C_{4'}-O_{4'}	1.429(13)
	C_{4'}-C_{3'}	1.512(14)
	O_{4'}-C_{1'}	1.435(11)
	C_{1'}-N₉	1.474(12)
	C_{1'}-C_{2'}	1.519(14)
	N₉-C₄	1.367(13)
	N₉-C₈	1.382(14)
	C₄-N₃	1.340(14)
	C₄-C₅	1.39(2)
	N₃-C₂	1.336(14)
	C₂-N₂	1.335(14)
	C₂-N₁	1.397(14)
	N₁-C₆	1.41(2)
	C₆-O₆	1.241(13)
	C₆-C₅	1.44(2)
	C₅-N₇	1.382(14)
	N₇-C₈	1.330(14)
	C_{2'}-C_{3'}	1.530(14)
	C_{3'}-O_{3'}	1.435(14)

Table 6.11: Bond Lengths and Estimated Standard Deviations – Residue G(3)

Residue	Bonded Atom Pair	Bond Length & ESD
G(4)	O _{3'} -P	1.592(11)
	P-O _{1P}	1.426(12)
	P-O _{2P}	1.464(13)
	P-O _{5'}	1.621(11)
	O _{5'} -C _{5'}	1.45(2)
	C _{5'} -C _{4'}	1.520(14)
	C _{4'} -O _{4'}	1.473(14)
	C _{4'} -C _{3'}	1.52(2)
	O _{4'} -C _{1'}	1.485(12)
	C _{1'} -N ₉	1.478(13)
	C _{1'} -C _{2'}	1.496(14)
	N ₉ -C ₈	1.34(2)
	N ₉ -C ₄	1.38(2)
	C ₄ -N ₃	1.36(2)
	C ₄ -C ₅	1.37(2)
	N ₃ -C ₂	1.32(2)
	C ₂ -N ₂	1.34(2)
	C ₂ -N ₁	1.40(2)
	N ₁ -C ₆	1.39(2)
	C ₆ -O ₆	1.246(14)
	C ₆ -C ₅	1.42(2)
	C ₅ -N ₇	1.39(2)
	N ₇ -C ₈	1.30(2)
	C _{2'} -C _{3'}	1.53(2)
	C _{3'} -O _{3'}	1.45(2)

Table 6.12: Bond Lengths and Estimated Standard Deviations – Residue G(4)

Residue	Bonded Atom Pair	Bond Length & ESD
G(5)	O _{3'} -P	1.585(11)
	P-O _{1P}	1.451(13)
	P-O _{2P}	1.487(13)
	P-O _{5'}	1.599(12)
	O _{5'} -C _{5'}	1.46(2)
	C _{5'} -C _{4'}	1.52(2)
	C _{4'} -O _{4'}	1.462(14)
	C _{4'} -C _{3'}	1.511(14)
	O _{4'} -C _{1'}	1.481(12)
	C _{1'} -N ₉	1.463(13)
	C _{1'} -C _{2'}	1.505(14)
	N ₉ -C ₄	1.37(2)
	N ₉ -C ₈	1.39(2)
	C ₄ -N ₃	1.36(2)
	C ₄ -C ₅	1.37(2)
	N ₃ -C ₂	1.33(2)
	C ₂ -N ₂	1.36(2)
	C ₂ -N ₁	1.36(2)
	N ₁ -C ₆	1.39(2)
	C ₆ -O ₆	1.24(2)
	C ₆ -C ₅	1.44(2)
	C ₅ -N ₇	1.38(2)
	N ₇ -C ₈	1.31(2)
	C _{2'} -C _{3'}	1.54(2)
	C _{3'} -O _{3'}	1.44(2)

Table 6.13: Bond Lengths and Estimated Standard Deviations – Residue G(5)

Residue	Bonded Atom Pair	Bond Length & ESD
3'T(6)	O _{3'} -P	1.606(12)
	P-O _{2P}	1.448(13)
	P-O _{1P}	1.458(13)
	P-O _{5'}	1.581(12)
	O _{5'} -C _{5'}	1.47(2)
	C _{5'} -C _{4'}	1.53(2)
	C _{4'} -O _{4'}	1.475(14)
	C _{4'} -C _{3'}	1.51(2)
	O _{4'} -C _{1'}	1.478(13)
	C _{1'} -N ₁	1.491(14)
	C _{1'} -C _{2'}	1.532(14)
	N ₁ -C ₆	1.37(2)
	N ₁ -C ₂	1.40(2)
	C ₆ -C ₅	1.37(2)
	C ₂ -O ₂	1.20(2)
	C ₂ -N ₃	1.39(2)
	N ₃ -C ₄	1.39(2)
	C ₄ -O ₄	1.24(2)
	C ₄ -C ₅	1.43(2)
	C ₅ -C _{5α}	1.50(2)
	C _{2'} -C _{3'}	1.53(2)
	C _{3'} -O _{3'}	1.45(2)

Table 6.14: Bond Lengths and Estimated Standard Deviations – Residue 3'T(6)

Residue	Bonded Atom Triplet	Bond Angle & ESD
5'T(1)	O _{5'} -C _{5'} -C _{4'}	110.20(12)
	O _{4'} -C _{4'} -C _{3'}	106.2(9)
	O _{4'} -C _{4'} -C _{5'}	107.50(11)
	C _{3'} -C _{4'} -C _{5'}	115.00(12)
	C _{1'} -O _{4'} -C _{4'}	109.0(8)
	O _{4'} -C _{1'} -N ₁	106.1(9)
	O _{4'} -C _{1'} -C _{2'}	106.50(10)
	N ₁ -C _{1'} -C _{2'}	111.50(10)
	C ₂ -N ₁ -C ₆	120.90(13)
	C ₂ -N ₁ -C _{1'}	118.70(13)
	C ₆ -N ₁ -C _{1'}	120.30(12)
	C ₅ -C ₆ -N ₁	121.60(14)
	O ₂ -C ₂ -N ₁	123(2)
	O ₂ -C ₂ -N ₃	120(2)
	N ₁ -C ₂ -N ₃	117(2)
	C ₂ -N ₃ -C ₄	123.00(14)
	O ₄ -C ₄ -N ₃	120(2)
	O ₄ -C ₄ -C ₅	121(2)
	N ₃ -C ₄ -C ₅	118.50(14)
	C ₆ -C ₅ -C ₄	118.30(14)
	C ₆ -C ₅ -C _{5α}	123(2)
	C ₄ -C ₅ -C _{5α}	118(2)
	C _{1'} -C _{2'} -C _{3'}	100.1(9)
	O _{3'} -C _{3'} -C _{4'}	108.80(11)
	O _{3'} -C _{3'} -C _{2'}	112.80(11)
	C _{4'} -C _{3'} -C _{2'}	105.8(9)

Table 6.15: Bond Angles and Estimated Standard Deviations – Residue 5'T(1)

Residue	Bonded Atom Triplet	Bond Angle & ESD
G(2)	C _{3'} -O _{3'} -P	117.7(8)
	O _{2P} -P-O _{1P}	120.1(6)
	O _{2P} -P-O _{5'}	111.0(6)
	O _{1P} -P-O _{5'}	104.9(6)
	O _{2P} -P-O _{3'}	105.0(6)
	O _{1P} -P-O _{3'}	110.8(6)
	O _{5'} -P-O _{3'}	104.1(5)
	C _{5'} -O _{5'} -P	120.4(7)
	O _{5'} -C _{5'} -C _{4'}	106.6(9)
	O _{4'} -C _{4'} -C _{3'}	106.3(8)
	O _{4'} -C _{4'} -C _{5'}	110.00(10)
	C _{3'} -C _{4'} -C _{5'}	115.20(10)
	C _{1'} -O _{4'} -C _{4'}	111.5(8)
	O _{4'} -C _{1'} -N ₉	114.4(8)
	O _{4'} -C _{1'} -C _{2'}	106.4(9)
	N ₉ -C _{1'} -C _{2'}	106.7(7)
	C ₈ -N ₉ -C ₄	106.4(9)
	C ₈ -N ₉ -C _{1'}	127.40(10)
	C ₄ -N ₉ -C _{1'}	126.20(10)
	C ₅ -C ₄ -N ₃	129.40(11)
	C ₅ -C ₄ -N ₉	104.60(10)
	N ₃ -C ₄ -N ₉	126.00(11)
	C ₂ -N ₃ -C ₄	111.20(10)
	N ₂ -C ₂ -N ₃	118.80(12)
	N ₂ -C ₂ -N ₁	118.10(12)
	N ₃ -C ₂ -N ₁	123.10(11)
	C ₆ -N ₁ -C ₂	125.30(11)
	O ₆ -C ₆ -N ₁	119.20(12)
	O ₆ -C ₆ -C ₅	127.80(11)
	N ₁ -C ₆ -C ₅	113.00(11)
	C ₄ -C ₅ -N ₇	112.00(10)
	C ₄ -C ₅ -C ₆	117.90(11)
	N ₇ -C ₅ -C ₆	130.00(11)
	C ₈ -N ₇ -C ₅	103.20(10)
	N ₇ -C ₈ -N ₉	113.70(11)
	C _{3'} -C _{2'} -C _{1'}	100.9(8)
	O _{3'} -C _{3'} -C _{4'}	106.60(10)
	O _{3'} -C _{3'} -C _{2'}	111.90(10)
	C _{4'} -C _{3'} -C _{2'}	104.4(9)

Table 6.16: Bond Angles and Estimated Standard Deviations – Residue G(2)

Residue	Bonded Atom Triplet	Bond Angle & ESD
G(3)	C _{3'} -O _{3'} -P	123.1(8)
	O _{2P} -P-O _{1P}	118.8(7)
	O _{2P} -P-O _{5'}	106.0(7)
	O _{1P} -P-O _{5'}	110.6(7)
	O _{2P} -P-O _{3'}	111.5(6)
	O _{1P} -P-O _{3'}	105.0(6)
	O _{5'} -P-O _{3'}	104.0(6)
	C _{5'} -O _{5'} -P	120.6(9)
	O _{5'} -C _{5'} -C _{4'}	108.20(11)
	O _{4'} -C _{4'} -C _{3'}	105.30(10)
	O _{4'} -C _{4'} -C _{5'}	110.50(11)
	C _{3'} -C _{4'} -C _{5'}	113.90(12)
	C _{4'} -O _{4'} -C _{1'}	112.6(9)
	O _{4'} -C _{1'} -N ₉	106.4(8)
	O _{4'} -C _{1'} -C _{2'}	105.20(10)
	N ₉ -C _{1'} -C _{2'}	115.10(10)
	C ₄ -N ₉ -C ₈	108.50(10)
	C ₄ -N ₉ -C _{1'}	126.40(11)
	C ₈ -N ₉ -C _{1'}	125.10(11)
	N ₃ -C ₄ -N ₉	127.60(13)
	N ₃ -C ₄ -C ₅	127.80(12)
	N ₉ -C ₄ -C ₅	104.60(11)
	C ₂ -N ₃ -C ₄	115.00(11)
	N ₂ -C ₂ -N ₃	119.80(12)
	N ₂ -C ₂ -N ₁	118.70(12)
	N ₃ -C ₂ -N ₁	121.50(12)
	C ₂ -N ₁ -C ₆	124.90(12)
	O ₆ -C ₆ -N ₁	120.20(12)
	O ₆ -C ₆ -C ₅	127.70(13)
	N ₁ -C ₆ -C ₅	112.10(12)
	N ₇ -C ₅ -C ₄	111.00(11)
	N ₇ -C ₅ -C ₆	130.30(13)
	C ₄ -C ₅ -C ₆	118.70(12)
	C ₈ -N ₇ -C ₅	105.30(11)
	N ₇ -C ₈ -N ₉	110.50(12)
	C _{1'} -C _{2'} -C _{3'}	103.00(10)
	O _{3'} -C _{3'} -C _{4'}	107.20(12)
	O _{3'} -C _{3'} -C _{2'}	112.60(13)
	C _{4'} -C _{3'} -C _{2'}	104.70(10)

Table 6.17: Bond Angles and Estimated Standard Deviations – Residue G(3)

Residue	Bonded Atom Triplet	Bond Angle & ESD
G(4)	C _{3'} -O _{3'} -P	126.20(11)
	O _{1P} -P-O _{2P}	119.20(12)
	O _{1P} -P-O _{3'}	110.80(10)
	O _{2P} -P-O _{3'}	110.9(9)
	O _{1P} -P-O _{5'}	108.2(9)
	O _{2P} -P-O _{5'}	105.10(10)
	O _{3'} -P-O _{5'}	100.9(7)
	C _{5'} -O _{5'} -P	117.30(12)
	O _{5'} -C _{5'} -C _{4'}	108.20(13)
	O _{4'} -C _{4'} -C _{3'}	104.80(11)
	O _{4'} -C _{4'} -C _{5'}	109.60(12)
	C _{3'} -C _{4'} -C _{5'}	113.90(13)
	C _{4'} -O _{4'} -C _{1'}	107.40(10)
	N ₉ -C _{1'} -O _{4'}	99.1(9)
	N ₉ -C _{1'} -C _{2'}	119.70(11)
	O _{4'} -C _{1'} -C _{2'}	104.70(11)
	C ₈ -N ₉ -C ₄	106.30(12)
	C ₈ -N ₉ -C _{1'}	128.40(13)
	C ₄ -N ₉ -C _{1'}	125.30(13)
	N ₃ -C ₄ -C ₅	127.70(13)
	N ₃ -C ₄ -N ₉	126.40(14)
	C ₅ -C ₄ -N ₉	105.80(12)
	C ₂ -N ₃ -C ₄	113.50(13)
	N ₃ -C ₂ -N ₂	121(2)
	N ₃ -C ₂ -N ₁	123.00(14)
	N ₂ -C ₂ -N ₁	115.70(14)
	C ₆ -N ₁ -C ₂	124.10(13)
	O ₆ -C ₆ -N ₁	120.10(14)
	O ₆ -C ₆ -C ₅	127(2)
	N ₁ -C ₆ -C ₅	112.60(13)
	C ₄ -C ₅ -N ₇	109.50(12)
	C ₄ -C ₅ -C ₆	119.00(13)
	N ₇ -C ₅ -C ₆	131.50(14)
	C ₈ -N ₇ -C ₅	104.70(13)
	N ₇ -C ₈ -N ₉	113.70(14)
	C _{1'} -C _{2'} -C _{3'}	101.80(11)
	O _{3'} -C _{3'} -C _{4'}	105.10(13)
	O _{3'} -C _{3'} -C _{2'}	113.10(14)
	C _{4'} -C _{3'} -C _{2'}	107.60(11)

Table 6.18: Bond Angles and Estimated Standard Deviations – Residue G(4)

Residue	Bonded Atom Triplet	Bond Angle & ESD
G(5)	C _{3'} -O _{3'} -P	124.90(12)
	O _{1P} -P-O _{2P}	121.60(13)
	O _{1P} -P-O _{3'}	109.40(11)
	O _{2P} -P-O _{3'}	108.7(9)
	O _{1P} -P-O _{5'}	109.00(10)
	O _{2P} -P-O _{5'}	103.50(10)
	O _{3'} -P-O _{5'}	102.9(8)
	C _{5'} -O _{5'} -P	120.20(12)
	O _{5'} -C _{5'} -C _{4'}	104.60(12)
	O _{4'} -C _{4'} -C _{3'}	105.50(10)
	O _{4'} -C _{4'} -C _{5'}	108.30(13)
	C _{3'} -C _{4'} -C _{5'}	115.60(14)
	C _{4'} -O _{4'} -C _{1'}	107.0(9)
	N ₉ -C _{1'} -O _{4'}	103.8(9)
	N ₉ -C _{1'} -C _{2'}	116.00(12)
	O _{4'} -C _{1'} -C _{2'}	105.00(11)
	C ₄ -N ₉ -C ₈	105.40(12)
	C ₄ -N ₉ -C _{1'}	127.60(14)
	C ₈ -N ₉ -C _{1'}	127.10(14)
	N ₃ -C ₄ -C ₅	127.70(14)
	N ₃ -C ₄ -N ₉	126(2)
	C ₅ -C ₄ -N ₉	106.40(13)
	C ₂ -N ₃ -C ₄	112.60(14)
	N ₃ -C ₂ -N ₂	118(2)
	N ₃ -C ₂ -N ₁	125(2)
	N ₂ -C ₂ -N ₁	117(2)
	C ₂ -N ₁ -C ₆	123.60(14)
	O ₆ -C ₆ -N ₁	122(2)
	O ₆ -C ₆ -C ₅	125(2)
	N ₁ -C ₆ -C ₅	113.10(14)
	C ₄ -C ₅ -N ₇	111.00(13)
	C ₄ -C ₅ -C ₆	118.10(14)
	N ₇ -C ₅ -C ₆	131(2)
	C ₈ -N ₇ -C ₅	104.10(13)
	N ₇ -C ₈ -N ₉	113.00(14)
	C _{1'} -C _{2'} -C _{3'}	99.00(10)
	O _{3'} -C _{3'} -C _{4'}	106.10(13)
	O _{3'} -C _{3'} -C _{2'}	109.60(13)
	C _{4'} -C _{3'} -C _{2'}	106.60(11)

Table 6.19: Bond Angles and Estimated Standard Deviations – Residue G(5)

Residue	Bonded Atom Triplet	Bond Angle & ESD
3'T(6)	C _{3'} -O _{3'} -P	122.80(12)
	O _{2P} -P-O _{1P}	116.00(13)
	O _{2P} -P-O _{5'}	114.40(11)
	O _{1P} -P-O _{5'}	106.60(11)
	O _{2P} -P-O _{3'}	107.30(10)
	O _{1P} -P-O _{3'}	110.00(10)
	O _{5'} -P-O _{3'}	101.5(8)
	C _{5'} -O _{5'} -P	121.10(13)
	O _{5'} -C _{5'} -C _{4'}	104.40(13)
	O _{4'} -C _{4'} -C _{3'}	100.50(11)
	O _{4'} -C _{4'} -C _{5'}	108.80(14)
	C _{3'} -C _{4'} -C _{5'}	114(2)
	C _{1'} -O _{4'} -C _{4'}	104.90(10)
	O _{4'} -C _{1'} -N ₁	104.20(10)
	O _{4'} -C _{1'} -C _{2'}	106.20(11)
	N ₁ -C _{1'} -C _{2'}	114.60(13)
	C ₆ -N ₁ -C ₂	122.00(14)
	C ₆ -N ₁ -C _{1'}	122.00(14)
	C ₂ -N ₁ -C _{1'}	116.00(14)
	N ₁ -C ₆ -C ₅	122(2)
	O ₂ -C ₂ -N ₃	121(2)
	O ₂ -C ₂ -N ₁	124(2)
	N ₃ -C ₂ -N ₁	115(2)
	C ₂ -N ₃ -C ₄	126(2)
	O ₄ -C ₄ -N ₃	118(2)
	O ₄ -C ₄ -C ₅	126(2)
	N ₃ -C ₄ -C ₅	115(2)
	C ₆ -C ₅ -C ₄	120(2)
	C ₆ -C ₅ -C _{5α}	120(2)
	C ₄ -C ₅ -C _{5α}	120(2)
	C _{1'} -C _{2'} -C _{3'}	103.00(11)
	O _{3'} -C _{3'} -C _{4'}	108(2)
	O _{3'} -C _{3'} -C _{2'}	113(2)
	C _{4'} -C _{3'} -C _{2'}	105.30(13)

Table 6.20: Bond Angles and Estimated Standard Deviations – Residue 3'T(6)

The very definition of the real becomes: *that of which it is possible to give an equivalent reproduction*. This is contemporaneous with a science that postulates that a process can be perfectly reproduced in a set of given conditions, and also with the industrial rationality that postulates a universal system of equivalency (classical representation is not equivalence, it is transcription, interpretation, commentary). At the limit of this process of reproducibility, the real is not only what can be reproduced, but *that which is always already reproduced*. The hyperreal.

Jean Baudrillard, *The Orders of Simulacra*

Soon the signs started appearing. THE MOST PHOTOGRAPHED BARN IN AMERICA. We counted five signs before we reached the site . . . All the people had cameras; some had tripods, telephoto lenses, filter kits. A man in a booth sold postcards and slides – pictures of the barn taken from the elevated spot . . .

“No one sees the barn,” he said finally . . . “Once you’ve seen the signs about the barn, it becomes impossible to see the barn.” . . . “We’re not here to capture an image, we’re here to maintain one. Every photograph reinforces the aura. Can you feel it, Jack? An accumulation of nameless energies.” . . . “Being here is a kind of spiritual surrender. We see only what the others see. The thousands who were here in the past, those who will come in the future. We’ve agreed to be part of a collective perception. This literally colors our vision. A religious experience in a way, like all tourism.” . . . “They are taking pictures of taking pictures.” . . . “What was the barn like before it was photographed?” . . . “What did it look like, how was it different from other barns, how was it similar to other barns? We can’t answer these questions because we’ve read the signs, seen the people snapping the pictures. We can’t get outside the aura. We’re part of the aura. We’re here. We’re now.”

He seemed immensely pleased by this.

Don De Lillo, *White Noise*

In the theatre one is well aware of the place from which the play cannot immediately be detected as illusionary. There is no such place for the movie scene that is being shot. Its illusionary nature is that of the second degree, the result of cutting. That is to say, in the studio the mechanical equipment has penetrated so deeply into reality that its pure aspect freed from the foreign substance of equipment is the result of a special procedure, namely, the shooting by the specially adjusted camera and the mounting of the shot together with other similar ones. The equipment-free aspect of reality here has become the height of artifice; the sight of immediate reality has become an orchid in the land of technology.

Walter Benjamin, *Illuminations*

Chapter 7

The Structure of the d(TGGGGT)

Parallel–Stranded Tetraplex

7.1 Introduction

The previous chapters in this thesis have described the method of crystallographic determination of the high-resolution crystal structure of the parallel-stranded guanine tetraplex formed by the oligonucleotide d(TG₄T), covering the topics of crystallization, heavy-atom derivative trials, data collection, phase determination and structure refinement. In this chapter the structure of the tetraplex itself is described in detail. Initially, an overview of the packing arrangement of the unit cell contents is presented, alongside a description of the character of the guanine tetrads which form the basis of the tetraplex structure. Following on from this, the morphology of the tetraplex is discussed in depth, encompassing topics such as the deoxyribose sugar-puckers adopted throughout the structure, the characteristic P···P atom separations relating the various

nucleotide units and the conformation of the sugar-phosphate backbone. The four independently refined tetraplexes which occupy the crystal unit cell are then compared by means of a root-mean-square (RMS) superposition of the atomic coordinates.

Once the overall morphology of the tetraplex has been established, the other salient features of the crystal structure are introduced. Of primary importance in this regard is the location and environment of the monovalent sodium cations which are essential for the stability of the tetraplex structure. A total of 14 sodium ions were located, and their position, environment and coordination are fully detailed. The tetraplex is extensively hydrated and the key features of this hydration are reviewed and illustrated. Base-stacking interactions are another crucial factor in stabilising nucleic acid structures in general. For the tetraplexes reported here, distinct intra- and inter-molecular stacking patterns are observed, and their character and relevance discussed. The terminal thymine nucleotides, whilst not assembling as T₄ tetrads, participate in an extensive and intricate variety of intermolecular contacts. These fortuitous contacts account for the remarkable order of the crystal lattice and are thoroughly characterised within this chapter. Finally, the role of divalent calcium cations in consolidating the packing of the crystal contents is described. Calcium ions were included in the crystallization liquor and were found to be essential for crystal growth; nine calcium ions were located in the final structure presented here.

All of the colour figures presented in this chapter, and in Chapters 5 and 6, were generated using the molecular graphics package *O* (Jones *et al.*, (1991)), and the digital-image manipulation programs *XV* and *CLEARPAINT*. In many of these figures the nucleic acid substituents are depicted using a contrived colour scheme, which is

explained in the accompanying discussion. However, in some cases, the conventional default atom colours are used. These defaults paint carbon atoms yellow, nitrogens blue, oxygens red and phosphorus white. Water molecules are represented generally as red, and sodium and calcium cations as yellow, asterisks. Electron density maps are represented in two forms – the conventional, transparent chickenwire and a semi-transparent solid-surface rendering. The electron density maps were calculated using $(3F_o - 2F_c)$ coefficients and, unless otherwise indicated, the contouring of the electron density is 1.5σ . The calculated structure factors were taken from the output of *SHELXL-93* (Sheldrick (1993)), and the electron density maps were computed as for the *ARP* procedures (Lamzin and Wilson (1993)), described in Chapter 6.

Throughout this discussion, in situations where multiple measurements of a particular parameter are reported as an average value, a standard deviation, σ , for this value is also given. The calculation of this standard deviation is described by equation 7.1, where X_i and \bar{X} are the individual and mean values of the parameter, respectively, and N is the number of observations.

$$\sigma = \left[\frac{\sum_i (\bar{X} - X_i)^2}{(N - 1)} \right]^{\frac{1}{2}} \quad (7.1)$$

The key structural findings of this work have already been reported elsewhere (Laughlan *et al.*, (1994); Surridge (1994)), and are also to be presented in the context of a general review of parallel-stranded guanine tetraplex structures (Murchie *et al.*, (1995)). The solution structure of the parallel-stranded tetraplex formed by d(TG₄T) has been published recently, based on the refinement of the basic tetraplex model

(Aboul-ela *et al.*, (1992)) against the NMR data (Aboul-ela *et al.*, (1994)). A detailed comparison of the crystal and solution structures of the d(TG₄T) tetraplex is planned for future publication.

7.2 Packing of Tetraplexes in the Cell

As detailed in the previous chapters the hexanucleotide d(TG₄T) crystallised as the anticipated four-stranded tetraplex in the presence of sodium ions. A total of four tetraplexes, which were individually refined without the use of any non crystallographic symmetry restraints, occupy the irreducible repeating unit of the crystal. The strands of the tetraplexes form a right-handed helix and are aligned parallel to each other.

Figure 7.1 presents a schematic representation of the strand orientation of the DNA tetraplex and their relative packing within the unit cell. Four parallel-stranded tetraplexes occupy the asymmetric unit. These four tetraplexes are organised into two sets of stacking tetraplex pairs. The tetraplex constituting a pair stack with opposite polarity in each case. The arrangement of the stacking tetraplex pairs is such that the 3' end of each tetraplex in the pair is positioned at the interface between cells, coincident with the *ac* face of the crystal. The overlapping interface of the tetraplexes is formed by the 5' ends in each case. A second stacked pair, not shown in Figure 7.1, is related to the first by a non-crystallographic translation. The numbering of the guanine tetrad planes in Figure 7.1 is derived from the oligonucleotide sequence 5'-T(1)G(2)G(3)G(4)G(5)T(6)-3', and the bases of the lower tetrads have been given negative integers for clarity. The numbering sequence G5 to G-5 shall be referred to

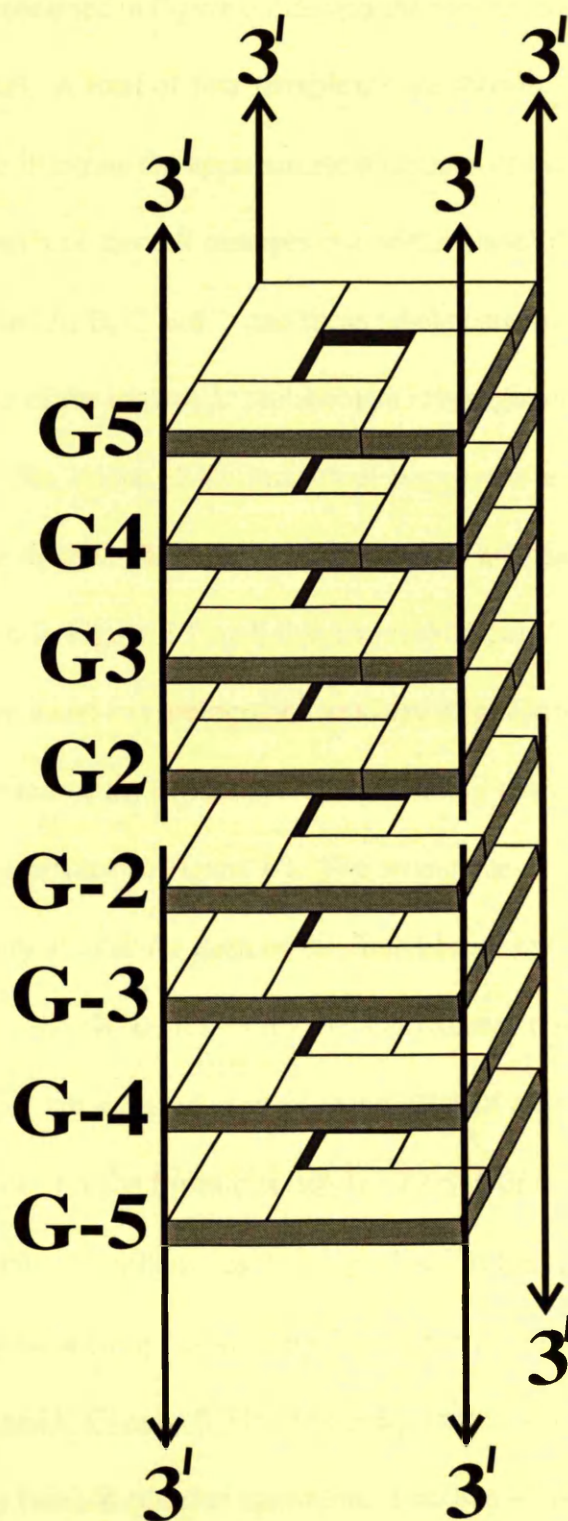


Figure 7.1: Schematic of Strand Orientation and Packing in the Cell

throughout the discussion in this chapter.

The illustration contained in Figure 7.2 depicts the entire contents of the irreducible unit cell of the crystal. A total of four tetraplexes are shown. The yellow arrows at the base of the figure illustrate the approximate directions of the *b*- and *c*-axes of the crystal cell. The *a*-axis of the cell emerges out of the plane of the figure. The four tetraplexes are labelled A, B, C and D and these labels correspond to the labels used to describe the results of the molecular replacement investigations depicted in Figures 5.10, 5.12 and 6.6. The colour of the individual tetraplexes also corresponds to the scheme used in these figures. The coaxial tetraplex pairs are stacked with the relative orientations described in Figure 7.1 such that the coaxial pair (A, B) is related to the coaxial pair (C, D) by a non-crystallographic translation of approximately (0, 0, 0.5).

The apparent source of the pseudosymmetry relating the coaxial tetraplex pairs can be seen to be in evidence in Figure 7.2. The arrangement of thymine nucleotides can be seen to be very similar for each of the four tetraplexes and, in particular, for the tetraplexes (A, C) and (B, D) related by the approximate translation of (0, 0, 0.5). These thymine residues are involved in an extensive array of inter- and intra-molecular contacts which account for the fortuitous degree of crystalline order observed in the crystals obtained. A number of these residues are related by the approximate translation of (0, 0, 0.5), as can be seen by inspection of the thymine residues situated at the 3' ends of the tetraplexes (A, C) and (B, D). However, other thymine residues are related by an approximately twofold rotation operation. This can be observed in Figure 7.2 for some of the residues situated at the 5' ends of the tetraplexes where they stack in the cell. As a result the thymine residues of the coaxial pairs are related by two

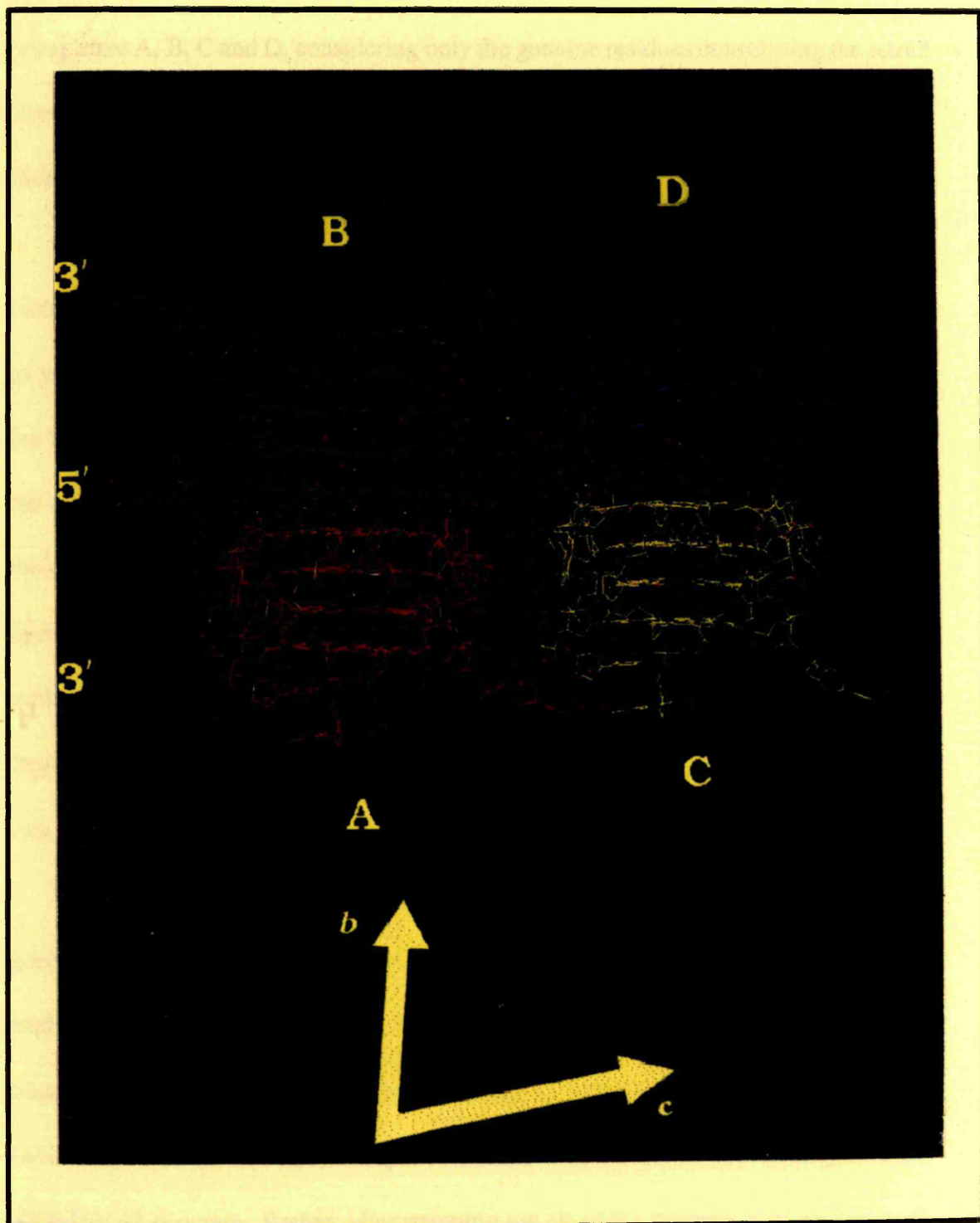


Figure 7.2: The Unit Cell Contents

main symmetry operations and the arrangements do not commute. This results in the translation relating the coaxial pairs being pseudosymmetric. In actual fact the tetraplexes A, B, C and D, considering only the guanine residues constituting the tetrad structure, are not identical as evidenced from root-mean-square comparison of the individual tetraplexes, discussed later in the chapter.

A string of sodium ions is visible in Figure 7.2, in yellow, running down the central cavity of the coaxially stacked tetraplex pairs. Calcium ions are also in evidence, again in yellow. One such ion can be seen in the left of tetraplex A, in red, midway down the length of the tetraplex. The remaining eight of the nine calcium cations located in the refinement procedure are situated in an approximate planar arrangement between the coaxially stacked tetraplex pairs. This plane also includes the sodium ions situated between the stacked pairs. The four tetraplexes are extensively hydrated and the water molecules, obtained from the *ARP* refinement procedure discussed in Chapter 6, are depicted as oxygen atoms and coloured red. All of the forementioned features of the unit cell contents shall be described more fully in the following sections of the chapter.

The helical parameters of the individual tetraplexes were analysed using the program *NEWHEL93* freely distributed by R. E. Dickerson. The *NEWHEL93* program analyses the input coordinates of nucleic acid helices and outputs a variety of helical parameters, such as those described in Dickerson (1992), for analysis. Due to the exotic nature of the d(TG₄T) tetraplex structure it could not be fed directly into the *NEWHEL93* program. Rather, after stripping out all of the thymine nucleotides, each tetraplex was broken down into two helix models consisting of two diagonally opposite parallel strands. Each of these model helices was then analysed for each of the four

Tetraplex	G₄ Tetrad Step	Tetraplex Twist Angle(°)	G₄ Tetrads Per Turn	Rise Per Tetrad (Å)
A	G2–G3	30.51±.12	11.79±.04	3.351±.044
	G3–G4	26.44±.98	13.62±.50	3.546±.028
	G4–G5	30.99±.12	11.61±.04	3.195±.033
B	G2–G3	33.53±.12	10.74±.04	3.346±.054
	G3–G4	26.85±.06	13.41±.03	3.716±.016
	G4–G5	33.12±.05	10.87±.02	3.375±.028
C	G2–G3	31.87±.38	11.29±.13	3.179±.057
	G3–G4	26.34±.41	13.67±.21	3.532±.019
	G4–G5	30.18±.29	11.92±.11	3.330±.081
D	G2–G3	33.75±.39	10.67±.12	3.313±.057
	G3–G4	26.68±.01	13.49±.01	3.728±.001
	G4–G5	33.31±.24	10.81±.08	3.384±.089

Table 7.1: Local Helical Parameters for Tetraplexes A, B, C, and D

tetraplexes in turn.

Inspection of Figure 7.2 illustrates that all of the tetrad planes exhibit a high degree of planarity. This is particularly evident for the G2 planes of tetraplexes B and D, where the tetrad plane is slightly deformed, presumably due to packing effects, as discussed later. Closer inspection reveals that the individual tetrad planes are slightly concave with respect to the 3' end of each individual tetraplex, and that this concave proclivity appears slightly more pronounced the nearer the tetrad plane is to the 3' end of the tetraplex, being most pronounced for the G5 tetrad planes. Table 7.1 presents values for the helical parameters defined as the tetraplex twist angle, the number of G₄ tetrads per turn and the rise per tetrad. These values, derived from *NEWHEL93*, were averaged in each case for the values obtained from each of the two-stranded model helices analysed, and the spread for each value is given.

The third column of Table 7.1 describes the helical twist observed for each individual tetrad step. It can be seen by inspection that the helical twist, which involves measuring the angle between the C_{1'}–C_{1'} vectors of stacking base-pairs, is approximately 31° for the tetrad steps G2–G3 and G4–G5. As has been indicated, the tetraplexes are not identical and it can be seen that, in tetraplexes B and D, the G2–G3 and G4–G5 tetrad steps are slightly overwound in comparison to tetraplexes A and C by a couple of degrees. The observed rise per tetrad, measured between C_{1'} atoms at the ends of the base-pairs, for all of the G2–G3 and G4–G5 steps is approximately 3.35 Å, the only noticeable deviation being in tetraplexes A and C. All of the G3–G4 tetrad steps, in each of the four individual cases, are marginally underwound in comparison to the G2–G3 and G4–G5 steps by about 5°. Since twist and rise are closely correlated

helical parameters (Dickerson (1992)) this variation can also be observed in the rise per tetrad for the G3–G4 steps, where there is an increment of approximately 0.2Å in each case. Cheong and Moore (1992) also reported a slight increment in distance between the central two tetrads within their model of the parallel-stranded RNA tetraplex of r(UG₄U) based on the NMR data, and speculated that a potassium ion could be present within the resultant pocket. A similar conclusion was reached for the parallel-stranded tetraplex formed by d(T₂G₄T) in the presence of potassium (Wang and Patel (1993a)). It is possible that in the tetraplex structure formed by d(TG₄T) in the presence of sodium ions this variation results from the positioning of the sodium ions at this tetrad interval, discussed later in the chapter.

Table 7.1 also presents figures for the number of G₄ tetrads per turn for each tetrad step, calculated from the corresponding tetraplex twist angle. The four independent tetraplexes are not entirely equivalent structurally, and this difference occurs principally at the interface between coaxial tetraplexes, where one member of the coaxial tetraplex pair switches conformation, as discussed later. Elsewhere in the molecule, where the morphology is more uniform, the helical structure has a repeat corresponding to 10.4 tetrads per turn.

Finally in this section, Figure 7.3 presents a view of the unit cell contents looking down the common fourfold axes of the coaxial tetraplex pairs, approximately coincident with the unit cell *b*-axis. At the top of the figure the approximate directions of the unit cell *a*- and *c*-axes are shown in yellow. The unit cell contents at the centre of the figure are shown using the default colourings and the string of sodium ions can be observed running down the channel of the coaxially stacked tetraplex pairs. A number

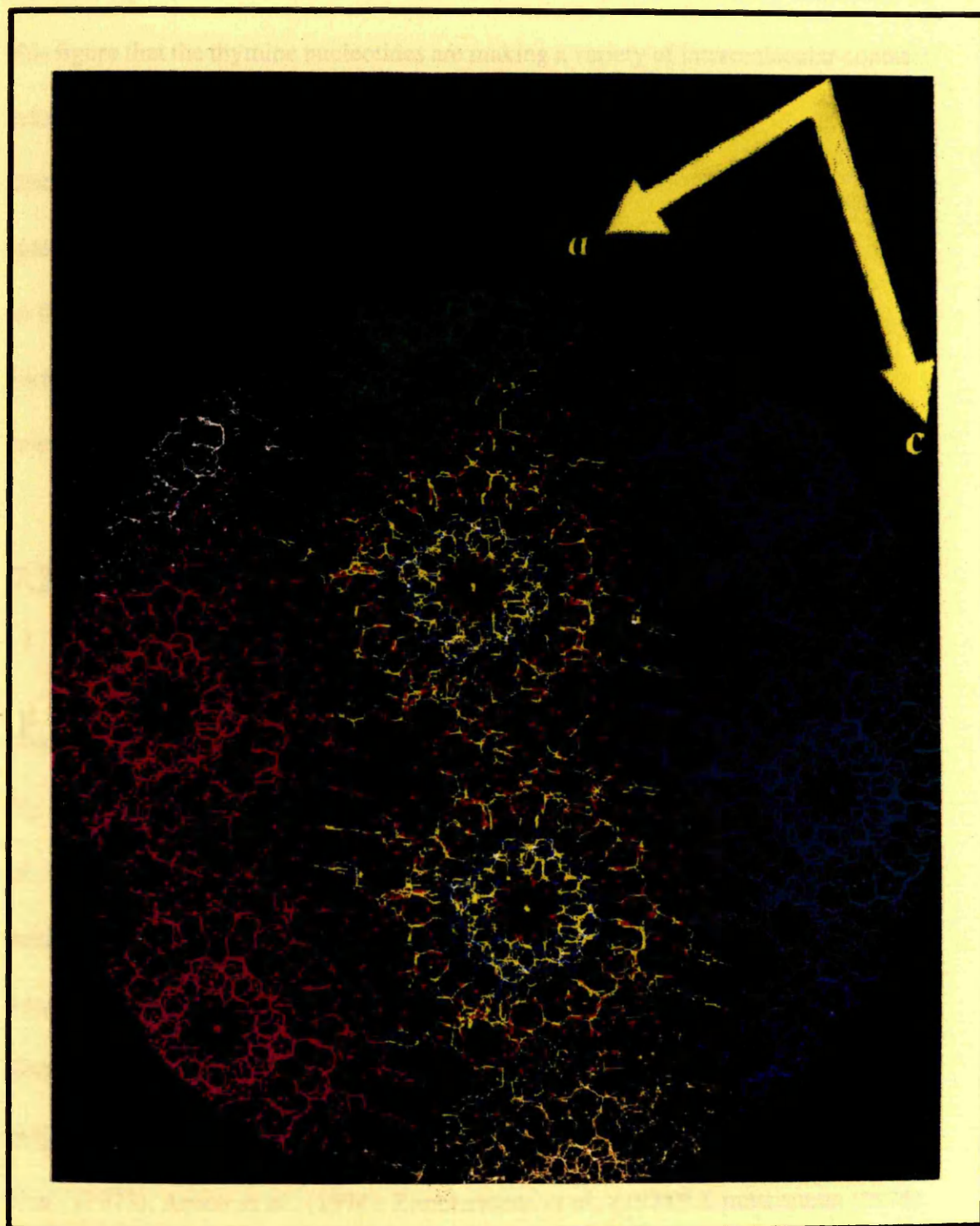


Figure 7.3: View of the Unit Cell Contents Along the Helical Axis

of symmetry related cells, generated by translations along the *a*- and *c*-axes of the cell, are depicted in a number of monochromatic colours. It is evident by inspection of this figure that the thymine nucleotides are making a variety of intermolecular contacts with symmetry related molecules at the *ab* and *bc* faces of the unit cell. In particular, a distinctive arrangement of stacked thymine bases can be seen at the *bc* face. Intricate intermolecular contacts also occur at the *ac* face of the unit cell but are not visible in this figure. The nature and variety of these intermolecular contacts, which are one fortuitous consequence of the dense packing of the unit cell contents, are described towards the end of the chapter.

7.3 The Guanine Tetrad

Within the four individual tetraplexes, the four strands are found to associate by means of the expected hydrogen-bonding between N_{1H} and O₆, and between N_{2H} and N₇. This cyclic arrangement of guanines in a tetrad was first proposed by Gellert *et al.*, (1962) in order to explain the self-ordered helical structures formed by guanylic acid, where infinite stacking of guanylic acid tetrads was demonstrated to generate a quadruple helical structure which was consistent with the experimental fibre-diffraction data. Subsequent experiments demonstrated that most other guanosine derivatives and polymers could associate into similar guanine tetrad-based quadruple helices (Tougaard *et al.*, (1973); Arnott *et al.*, (1974); Zimmermann *et al.*, (1975); Zimmermann (1976); Mariani *et al.*, (1989)). The guanine tetrad has been detected in numerous NMR studies, discussed in Chapter 1, and in the only other crystallographic determination of

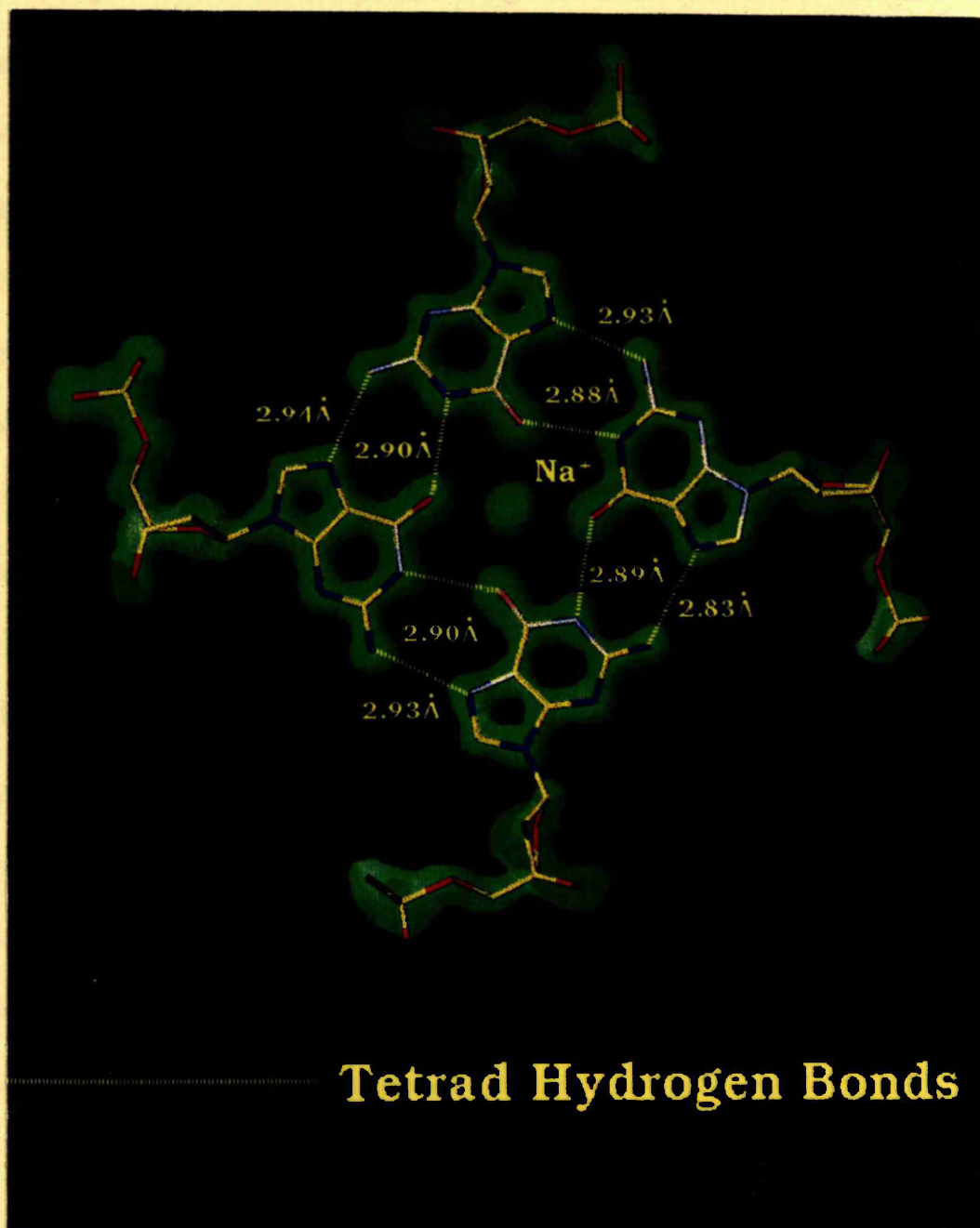


Figure 7.4: The Guanine Tetrad

a guanine quadruplex to date (Kang *et al.*, (1992)).

Figure 7.4 presents an illustration of a representative guanine tetrad from the structure. The guanine residues are painted using the default scheme and the electron density, in green, is rendered in the semi-transparent form. The map was calculated with $(3F_o - 2F_c)$ coefficients and is contoured at a level of 1.5σ . At the 1.2\AA resolution limit of this structure determination individual atoms of the guanine nucleotides are resolvable in the electron density, and the holes in the aromatic six-membered pyrimidine and five-membered imidazole rings are clearly visible. The tetrad depicted in Figure 7.4 is taken from the 3' end of the tetraplex, corresponding to G5 in Figure 7.1, and a central sodium ion can be seen in the tetrad plane. The environment of this and the other sodium cations is discussed later in the chapter. This figure can be compared with the proposed chemical structure for the tetrad in Figure 1.9.

The hydrogen-bonds stabilising the tetrad between the N_{1H} and O_6 atoms, and between the N_{2H} and N_7 atoms, are depicted as broken yellow lines relating the hydrogen-bonded atoms. The lengths of the hydrogen-bonds are also presented, measured between N_1 and O_6 , and N_2 and N_7 atoms, in each case, whilst the hydrogen atoms themselves are not shown in the figure. For a hydrogen-bonding system of the $N-H \cdots N$ type, the $N \cdots N$ distance has a minimum value of 2.75\AA , a maximum value of 3.15\AA and a mean value of 2.90\AA (Saenger (1984)). Likewise, for a hydrogen-bonding system of the $N-H \cdots O$ type, the $N \cdots O$ distance has a minimum value of 2.74\AA , a maximum value of 3.07\AA and a mean distance of 2.95\AA . It can be derived from inspection of Figure 7.4 that the observed values of the $N_1 \cdots O_6$ and $N_2 \cdots N_7$ hydrogen-bonding arrangements within the tetrad fall comfortably within the respective limits.

In Chapter 1, section 1.2.2, it was noted that four distinct G•G base-pairs could be constructed by the major tautomeric form of guanine at neutral pH (Donohue (1956)), and these four arrangements were depicted in Figure 1.8. The guanine tetrad from the tetraplex, Figure 7.4, clearly involves a hydrogen-bonding interaction of the type G:G 3, as shown in Figure 1.8. This G•G base-pairing arrangement has not been observed before in a crystal structure, with the exception of the 2.5Å resolution study of d(G₄T₄G₄) (Kang *et al.*, (1992)), and is now definitively characterised crystallographically in this 1.2Å study. Therefore, it can be clearly seen in Figure 7.4 that the G:G 3 base-pair is the structural basis of the guanine tetrad stabilising the d(TG₄T) parallel-stranded tetraplex.

The guanine tetrad depicted in Figure 7.4 is four-fold symmetric and is representative in its dimensions of all of the guanine tetrads to be found in each of the four tetraplexes occupying the unit cell. In their 2.5Å crystallographic study of the intramolecular hairpin quadruplex formed by the *Oxytricha* telomere sequence d(G₄T₄G₄) in the presence of potassium ions, Kang *et al.*, (1992) report that the first and fourth cyclically bonded guanine tetrads in the quadruplex do not form a completely square array, but rather are slightly distorted. This effect is less dramatic for the second and third planes, and Kang and colleagues propose that this may be due to the observation that these planes have a potassium ion located between them.

As a result of the distortions from a square array in some planes, several hydrogen-bonds in the d(G₄T₄G₄) structure tetrads are apparently three-centred or bifurcated. In the structure of the d(TG₄T) parallel-stranded tetraplex reported here such distortions are not in evidence. With only minor local variations, all of the tetrad planes approxi-

mate a perfect square planar array, and the hydrogen-bonds between N_{1H} and O_6 , and N_{2H} and N_7 , appear to be conventionally two-centred. Thus it may be the case that the distortions witnessed in the $d(G_4T_4G_4)$ structure arise from the steric constraints experienced in the dimerization of intramolecular hairpins, constraints which are not encountered in the tetramerization of four strands in a parallel-stranded tetraplex, such as that formed by $d(TG_4T)$.

7.4 Sugar-Puckering Modes of the Tetraplex

The preference of the five-membered furanose ring, constituting the sugar of the DNA backbone, for non-planar conformations has been introduced in Chapter 1. The ring can be puckered in an envelope (E) or twist (T) conformation as illustrated in Figure 1.3. Furanose rings pucker for a number of reasons. Most obviously, the planar furanose arrangement is energetically unfavourable, since in this configuration all of the torsion angles are set to 0° and the substituents attached to the carbon atoms are fully eclipsed. The ring system reduces its energy by puckering and, since the barriers to rotation about C–O bonds are lower than those about C–C bonds, the furanose ring adopts conformations with C–O torsion angles ν_0 and ν_4 nearly eclipsed and C–C torsion angles ν_1 , ν_2 and ν_3 maximally staggered (Saenger (1984)). This results in the $C_{2'}$ -*endo* or -*exo* puckering modes for $\nu_4 = 0^\circ$ and the $C_{3'}$ -*endo* or -*exo* puckering modes for $\nu_0 = 0^\circ$. In general, the $C_{2'}$ -*endo* and $C_{3'}$ -*endo* puckering modes are preferred because nonbonding interactions between furanose ring substituents are at a minimum.

The description of sugar-puckering given in Chapter 1 is only an approximate one

and can fall short, especially if intermediate twist modes are being described. The puckering of five-membered rings has been treated analytically in a form represented by the concept of pseudorotation (Saenger (1984)). This concept was initially introduced for cyclopentane rings and later extended to include substituted furanoses such as the deoxyribose sugar of the DNA structure. In cyclopentane, conformational changes do not proceed via a planar intermediate but rather the maximum pucker rotates virtually without potential energy barriers, giving rise to an infinite number of conformations which can be described in terms of the maximum torsion angle, ν_{\max} , and the pseudorotation phase angle P . When the five-membered ring is unsymmetrically substituted, as in nucleotide structures, potential energy thresholds are created which limit the pseudorotation and lead to preferred puckering modes.

In nucleotides, the pseudorotation phase angle P is calculated from the endocyclic sugar torsion angles $\nu_0, \nu_1, \nu_2, \nu_3$ and ν_4 , as defined in Figure 1.2. The expression for the calculation is expressed in equation 7.2 below.

$$\tan(P) = \frac{(\nu_4 + \nu_1) - (\nu_3 + \nu_0)}{2\nu_2[\sin(36^\circ) + \sin(72^\circ)]} \quad (7.2)$$

The phase angle $P = 0^\circ$ is defined such that the torsion angle ν_2 is maximally positive, corresponding to a symmetrical $C_{2'}\text{-exo}-C_{3'}\text{-endo}$ twist form. The mirror image $C_{2'}\text{-endo}-C_{3'}\text{-exo}$ configuration is represented by $P = 180^\circ$. Given the phase angle P , the five endocyclic torsion angles are related by equation 7.3, where $j = 0$ to 4 and $\varphi = 144^\circ$.

$$\nu_j = \nu_{\max} \cos(P + j\varphi) \quad (7.3)$$

The maximum torsion angle, ν_{\max} , is derived by setting $j = 0$, as shown in equation 7.4 below.

$$\nu_{\max} = \frac{\nu_0}{\cos(P)} \quad (7.4)$$

Equation 7.3 emphasises the fact that a change in P of 180° reverses the sign of all the endocyclic torsion angles, ν_j , resulting in the mirror image of the sugar-pucker configuration at phase angle P . Also, at every phase angle P , the sum of the positive furanose torsion angles equals that of the negative torsion angles such that the total sum is zero. In the pseudorotation cycle, envelope (E) and twist (T) conformations alternate every 18° . Twist (T) configurations are observed at even multiples of 18° , whilst envelope (E) configurations are found at odd multiples. In nucleotide structure analysis in general, two ranges of pseudorotation phase angles are found to be preferred. These are $C_{3'}$ -endo at $0^\circ \leq P \leq 36^\circ$, in the north (N) of the pseudorotation cycle, and $C_{2'}$ -endo at $144^\circ \leq P \leq 190^\circ$, in the south (S) of the cycle.

Tables 7.2–7.5 summarise the analysis of the endocyclic sugar torsion angles, and the adopted sugar-puckers, for the deoxyribose sugars of the guanine residues constituting the four d(TG₄T) tetraplexes A, B, C and D in the unit cell. The tables list the values for the five endocyclic torsion angles ν_0 , ν_1 , ν_2 , ν_3 and ν_4 , the pseudorotation phase angle P , as obtained from equation 7.2, and the resultant sugar-pucker configuration in each case. Numerical values were derived from analysis carried out

Strand	Residue	ν_0	ν_1	ν_2	ν_3	ν_4	P	Pucker
1	G2	-19.6	30.6	-29.5	19.2	0.0	161.1	C _{2'} -endo-C _{1'} -exo
	G3	-17.1	28.9	-30.8	21.1	-2.5	166.7	C _{2'} -endo-C _{3'} -exo
	G4	-34.8	33.4	-20.6	-0.1	21.8	125.1	C _{1'} -exo-O _{4'} -endo
	G5	-35.8	40.7	-31.6	10.9	15.3	140.2	C _{1'} -exo-C _{2'} -endo
2	G2	-18.0	31.0	-32.8	23.5	-3.4	167.6	C _{2'} -endo-C _{3'} -exo
	G3	-20.6	32.6	-33.9	22.1	-1.1	163.9	C _{2'} -endo-C _{3'} -exo
	G4	-40.9	41.5	-28.9	6.0	21.8	132.1	C _{1'} -exo-C _{2'} -endo
	G5	-34.2	39.0	-29.3	10.4	15.0	139.1	C _{1'} -exo-C _{2'} -endo
3	G2	-20.2	32.9	-33.1	22.0	-0.9	163.5	C _{2'} -endo-C _{3'} -exo
	G3	-17.2	30.1	-32.3	22.3	-3.6	167.9	C _{2'} -endo-C _{3'} -exo
	G4	-39.2	37.4	-23.7	0.7	23.8	126.2	C _{1'} -exo-C _{2'} -endo
	G5	-31.0	33.2	-23.9	6.4	15.4	135.1	C _{1'} -exo-C _{2'} -endo
4	G2	-19.6	32.2	-31.6	22.1	-1.3	163.7	C _{2'} -endo-C _{3'} -exo
	G3	-17.0	30.2	-32.3	22.9	-3.9	168.4	C _{2'} -endo-C _{3'} -exo
	G4	-31.4	33.5	-24.3	5.7	15.7	134.9	C _{1'} -exo-C _{2'} -endo
	G5	-33.2	38.5	-31.3	12.1	12.9	143.1	C _{1'} -exo-C _{2'} -endo
	Average B-DNA	—	—	—	—	—	141	C _{2'} -endo

Table 7.2: Sugar Torsion Angles(°) and Furanose Conformations for Tetraplex A

Strand	Residue	ν_0	ν_1	ν_2	ν_3	ν_4	P	Pucker
1	G2	-19.5	-6.4	28.6	-40.2	36.6	45.6	C _{4'} -exo-C _{3'} -endo
	G3	-21.0	31.6	-31.1	19.2	1.0	160.2	C _{2'} -endo-C _{1'} -exo
	G4	-32.7	35.9	-27.1	8.3	15.2	137.8	C _{1'} -exo-C _{2'} -endo
	G5	-33.9	38.6	-30.6	11.9	13.6	141.8	C _{1'} -exo-C _{2'} -endo
2	G2	-25.3	0.5	23.2	-38.6	39.8	55.5	C _{4'} -exo-O _{4'} -endo
	G3	-16.6	27.6	-28.5	19.2	-1.7	165.2	C _{2'} -endo-C _{3'} -exo
	G4	-35.6	38.4	-27.9	6.9	17.6	135.4	C _{1'} -exo-C _{2'} -endo
	G5	-32.5	40.1	-34.1	15.2	10.4	147.2	C _{2'} -endo-C _{1'} -exo
3	G2	-13.3	-11.2	30.6	-39.6	33.2	38.5	C _{4'} -exo-C _{3'} -endo
	G3	-22.0	32.2	-30.3	18.1	2.4	157.6	C _{2'} -endo-C _{1'} -exo
	G4	-25.3	34.4	-30.4	16.1	5.5	152.2	C _{2'} -endo-C _{1'} -exo
	G5	-31.2	38.1	-33.2	15.8	9.6	148.2	C _{2'} -endo-C _{1'} -exo
4	G2	-13.6	-12.3	32.9	-40.8	34.0	37.0	C _{4'} -exo-C _{3'} -endo
	G3	-21.3	30.6	-29.2	17.5	2.5	157.7	C _{2'} -endo-C _{1'} -exo
	G4	-28.8	35.9	-30.5	13.7	9.2	147.4	C _{2'} -endo-C _{1'} -exo
	G5	-36.7	33.9	-20.8	-0.4	23.3	124.1	C _{1'} -exo-O _{4'} -endo
	Average B-DNA	—	—	—	—	—	141	C _{2'} -endo

Table 7.3: Sugar Torsion Angles(°) and Furanose Conformations for Tetraplex B

Strand	Residue	ν_0	ν_1	ν_2	ν_3	ν_4	P	Pucker
1	G2	-14.3	28.7	-31.4	24.8	-6.2	172.9	C _{2'} -endo-C _{3'} -exo
	G3	-13.3	26.5	-30.0	22.6	-6.0	173.1	C _{2'} -endo-C _{3'} -exo
	G4	-35.6	36.0	-24.5	3.8	19.2	131.0	C _{1'} -exo-C _{2'} -endo
	G5	-34.6	40.6	-33.2	13.7	12.3	144.2	C _{2'} -endo-C _{3'} -exo
2	G2	-23.6	34.2	-33.1	20.7	2.5	158.8	C _{2'} -endo-C _{1'} -exo
	G3	-18.6	31.4	-32.9	22.2	-2.3	165.9	C _{2'} -endo-C _{3'} -exo
	G4	-31.6	35.7	-27.2	8.5	13.9	139.0	C _{1'} -exo-C _{2'} -endo
	G5	-31.2	39.3	-33.6	16.0	9.4	148.2	C _{2'} -endo-C _{1'} -exo
3	G2	-18.7	30.8	-30.9	21.9	-1.6	164.7	C _{2'} -endo-C _{3'} -exo
	G3	-32.6	35.7	-26.6	7.8	15.2	137.3	C _{1'} -exo-C _{2'} -endo
	G4	-18.0	30.2	-30.8	20.9	-2.2	165.2	C _{2'} -endo-C _{3'} -exo
	G5	-31.0	36.3	-29.4	12.3	11.6	143.6	C _{1'} -exo-C _{2'} -endo
4	G2	-19.8	30.9	-30.6	19.5	0.3	161.6	C _{2'} -endo-C _{1'} -exo
	G3	-16.8	31.4	-35.5	25.2	-5.5	170.9	C _{2'} -endo-C _{3'} -exo
	G4	-42.1	37.2	-20.6	-4.5	-28.5	119.5	C _{1'} -exo-O _{4'} -endo
	G5	-34.5	37.0	-27.9	8.2	16.8	137.0	C _{1'} -exo-C _{2'} -endo
	Average B-DNA	-	-	-	-	-	141	C _{2'} -endo

Table 7.4: Sugar Torsion Angles(°) and Furanose Conformations for Tetraplex C

Strand	Residue	ν_0	ν_1	ν_2	ν_3	ν_4	P	Pucker
1	G2	-20.1	-4.3	25.2	-37.2	36.1	49.0	C _{4'} -exo-C _{3'} -endo
	G3	-17.1	29.1	-30.9	21.5	-2.6	166.9	C _{2'} -endo-C _{3'} -exo
	G4	-24.1	31.6	-28.2	14.9	5.6	151.8	C _{2'} -endo-C _{1'} -exo
	G5	-32.2	39.1	-32.4	14.2	11.1	145.5	C _{2'} -endo-C _{1'} -exo
2	G2	-25.6	-0.6	25.6	-41.1	41.9	53.9	C _{4'} -exo-C _{3'} -endo
	G3	-21.3	31.3	-30.3	18.1	1.9	158.7	C _{2'} -endo-C _{1'} -exo
	G4	-32.8	37.3	-29.5	10.3	14.1	140.9	C _{1'} -exo-C _{2'} -endo
	G5	-34.5	38.3	-28.5	8.2	16.0	137.5	C _{1'} -exo-C _{2'} -endo
3	G2	-5.7	-17.0	32.2	-36.2	26.2	27.3	C _{3'} -endo-C _{4'} -exo
	G3	-22.3	30.6	-28.5	16.1	3.7	155.2	C _{2'} -endo-C _{1'} -exo
	G4	-27.4	33.2	-27.5	12.2	9.3	145.6	C _{2'} -endo-C _{1'} -exo
	G5	-32.6	39.8	-33.9	14.9	11.0	146.8	C _{2'} -endo-C _{1'} -exo
4	G2	-15.1	-10.0	30.4	-39.5	34.3	40.1	C _{4'} -exo-C _{3'} -endo
	G3	-19.3	28.4	-27.8	16.6	1.6	159.1	C _{2'} -endo-C _{1'} -exo
	G4	-27.7	33.8	-29.3	12.4	9.4	147.1	C _{2'} -endo-C _{1'} -exo
	G5	-36.5	43.6	-35.8	14.1	13.8	144.1	C _{2'} -endo-C _{1'} -exo
	Average B-DNA	-	-	-	-	-	141	C _{2'} -endo

Table 7.5: Sugar Torsion Angles(°) and Furanose Conformations for Tetraplex D

using the *NEWHEL93* program. The labels A, B, C and D for the tetraplexes are as defined in Figure 7.2. For each of the individual guanine residues the sugar-pucker is represented with the major-puckering in bold type, and the minor-puckering in plain type. At the foot of each table the average value for the B-helical form of double helical DNA for the pseudorotation phase angle, P , is presented, as derived from the structures d(CGCGAATTCGCG) (Dickerson and Drew (1981)) and d(CGTAGATCTACG) (Leonard and Hunter (1993)), and provided by Dr. Madeleine Moore of the University of York. The major-pucker of these double helical B-DNA structures is the expected C_2' -endo conformation, with an average pseudorotation phase angle of 141° .

Inspection of Tables 7.2–7.5 reveals that most of the guanine sugars exhibit pseudorotation phase angles in the south of the pseudorotation cycle, with values varying in the range 119.5° (tetraplex C) to 173.1° (tetraplex C). Therefore it can be concluded that most of the guanine residue sugars exhibit a B-DNA character. On average this is the case for fifty-six of the sixty-four guanine sugars and, for these fifty-six deoxyribose sugars, the average pseudorotation phase angle is $158.5 \pm 11.3^\circ$. However, at the tetraplex interface all of the G2 sugars of tetraplexes B and D (Tables 7.3 and 7.5) switch to configurations that exhibits pseudorotation phase angles in the north of the pseudorotation cycle and, therefore, which resemble those of A-DNA. For the eight G2 sugars of tetraplexes B and D the average pseudorotation phase angle is $45.6 \pm 8.0^\circ$. This switch in the sugar conformations for the G2 residues of tetraplexes B and D corresponds to the observed variations in the phosphate separation between the residues G2 and G3 in section 7.5, and the striking discrepancies noted in the backbone torsion angles δ and ϵ , discussed in section 7.6. It would appear that the

switch in the sugar-puckering from south to north for the G2 residues of tetraplexes B and D is required in order to avoid steric clash at the tightly packed interface between coaxially stacked tetraplexes, and constitutes the only apparent major anomaly in the conformations observed for the four tetraplex structures of d(TG₄T) in the unit cell.

The sugar-puckering modes of the d(TG₄T) structure, described above, are illustrated in Figures 7.5 and 7.6. In each case a guanine nucleotide is shown with most of the atoms painted yellow, with selected atoms coloured red and labelled in yellow. Each nucleotide is depicted surrounded by the electron density for the nucleotide, represented in the semi-transparent form and coloured in blue. Figure 7.5 depicts a guanine nucleotide in a puckering conformation located in the south of the pseudorotation phase cycle, and is therefore representative of the majority of guanine nucleotide sugars. The C_{2'} atom, painted red, is puckered towards the face of the sugar containing the O_{5'} atom, also coloured red. The C_{1'} atom, on the other hand, is puckered to a lesser degree away from the O_{5'} atom. Thus, with C_{2'}-*endo* major-puckering and C_{1'}-*exo* minor-puckering, the sugar conformation is C_{2'}-*endo*-C_{1'}-*exo*.

In Figure 7.6 the observed sugar-pucker is located in the north of the pseudorotation cycle and is, therefore, representative of the furanoses of the G2 residues of the tetraplexes B and D. The C_{3'} atom, painted red, is situated *endo* to the red O_{5'} atom, whilst the C_{4'} atom can be located in the *exo* position. The puckering of the C_{4'} atom is major whilst that of the C_{3'} atom is minor. Thus the sugar conformation is C_{4'}-*exo*-C_{3'}-*endo*.

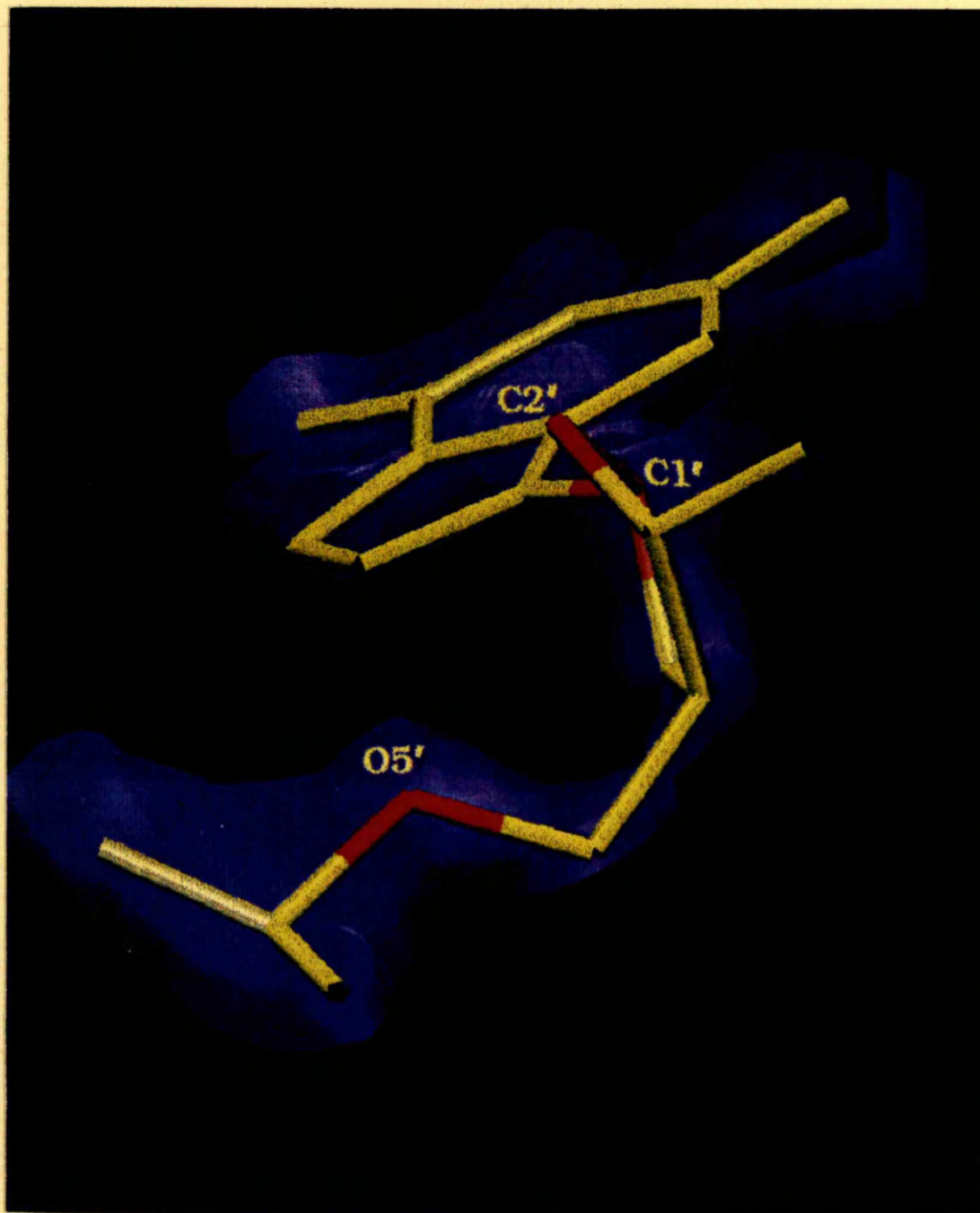


Figure 7.5: Deoxyribose Sugar in a Southern Conformation

7.3 Phosphate-Phosphate Repulsions

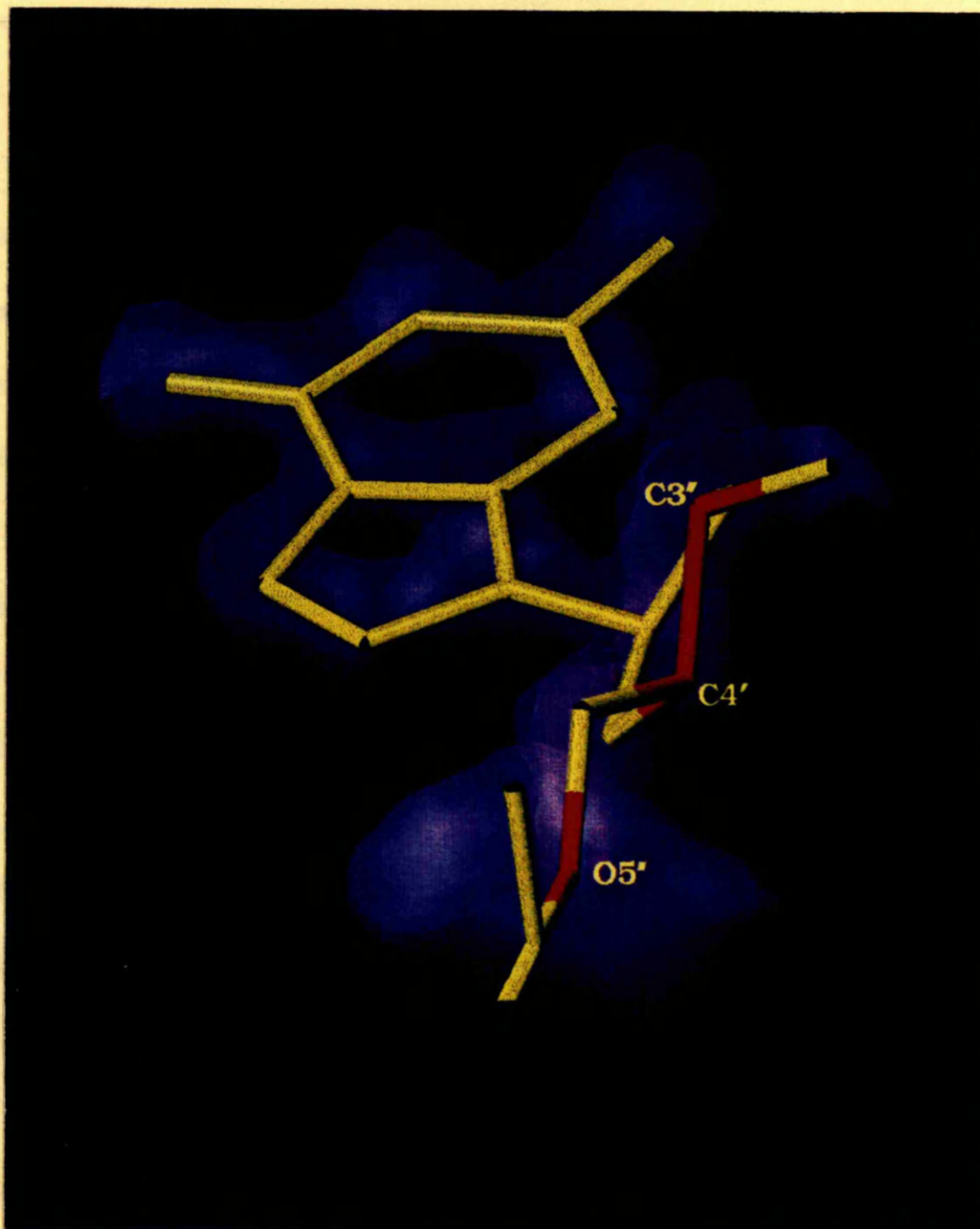


Figure 7.6: Deoxyribose Sugar in a Northern Conformation

7.5 Phosphate–Phosphate Separations

In comparison to RNA, DNA exhibits a considerable deal of structural polymorphism. Double–helical RNA, of natural or synthetic origin, is confined to the closely related A– and A'–forms. The A–type polynucleotide helix family is essentially characterised by the adoption of the $C_{3'}$ –*endo* sugar conformation by the furanose sugar of the structure backbone. The adoption of this sugar–pucker leads to a characteristic spacing between adjacent phosphorous atoms in the same polynucleotide chain of about 5.9Å. Structural features of the A–form helix were observed, for example, in the NMR study of the parallel–stranded tetraplex formed by the RNA analogue of d(TG₄T), r(UG₄U) (Cheong and Moore (1992)). In contrast to RNA, DNA displays a wide range of structures dependent upon factors such as sequence, base composition, cation environment and temperature (Saenger (1984)). The majority of DNA polynucleotide helices fall into either the A– or B–form families. The essential distinction between A– and B–type polynucleotide helices lies in the sugar–puckering modes. B–form helices generally adopt the $C_{2'}$ –*endo* sugar–pucker, and this results in a larger adjacent phosphate separation on the same strand in comparison to the A–form, of approximately 7.0Å. This generic feature of A– and B–type helices is illustrated in Figure 1.5.

Table 7.6 summarises the direct distance separation observed between adjacent phosphorous atoms for all sixteen strands of the four tetraplexes occupying the unit cell of the crystal. The labels A, B, C and D for the tetraplexes correspond to the labels depicted in Figure 7.2. A total of four P···P distances are given for the five phosphates in the d(TG₄T) oligonucleotide chain. There is no P···P distance for the

Tetraplex	Residues	Adjacent P-Atom Separation (Å)			
		Strand 1	Strand 2	Strand 3	Strand 4
A	G2-G3	7.07	7.09	7.03	6.93
	G3-G4	6.54	6.64	6.58	6.64
	G4-G5	6.65	6.51	6.79	6.77
	G5-3'T6	6.79	6.67	6.36	6.44
B	G2-G3	6.66	7.01	6.88	6.77
	G3-G4	6.64	6.67	6.56	6.72
	G4-G5	6.68	6.75	6.70	6.59
	G5-3'T6	6.27	6.54	6.91	5.60
C	G2-G3	7.03	7.01	7.04	7.01
	G3-G4	6.39	6.55	6.62	6.63
	G4-G5	6.72	6.73	6.65	6.76
	G5-3'T6	6.44	6.78	6.49	6.12
D	G2-G3	6.97	6.87	6.59	6.90
	G3-G4	6.64	6.68	6.60	6.70
	G4-G5	6.67	6.68	6.64	6.76
	G5-3'T6	6.97	6.34	6.47	6.49

Table 7.6: Adjacent P...P Distances for Tetraplexes A, B, C, and D

5'T1 residue since this residue, which terminates the 5' end of the sequence, does not contain a phosphate group. An overall inspection of the contents of Table 7.6 reveals that nearly all of the observed P...P distances fall in the range 6.6–7.0Å, hence indicating that the P...P separation of the adjacent phosphates on the d(TG₄T) strands is characteristic of B-DNA. Interesting variations can be observed for the P...P separations at the 3' and 5' ends. Whilst distances measured between the G3–G4 and G4–G5 residues are consistently in the vicinity of 6.60Å, values at the 5' end of the tetraplex, measured between the residues G2–G3, are somewhat larger, generally falling in the region of 7.0Å. This variation would appear to occur as a result of the previously mentioned deformation of the 5' guanine nucleotides, where the 5'T1 thymine nucleotides adopt extended conformations in the surrounding solvent. At the 3' end of the tetraplex, P...P distances measured between residues G5–3'T6 demonstrate a tendency, in some cases, to adopt values lower than that expected for B-form DNA, stretching as low as 5.60Å, as witnessed for tetraplex B. Again this variation is as a result of the variety of conformations adopted by the terminal thymine nucleotides in the surrounding solvent. For the G2–G3 P...P distances at the 5' end of the four tetraplexes, the average phosphate separation is $6.93 \pm 0.15 \text{Å}$, for the sixteen individual measurements recorded in Table 7.6. For the remaining forty-eight P...P separation distances, where the character exhibited is more consistent with B-DNA, the average phosphate separation is $6.63 \pm 0.14 \text{Å}$.

The findings of Table 7.6 are illustrated in Figure 7.7. Figure 7.7 presents two examples of adjacent phosphate distances measured from tetraplexes A and B, the coaxial stacked pair defined in Figures 7.1 and 7.2. In Figure 7.7(a) the two adjacent

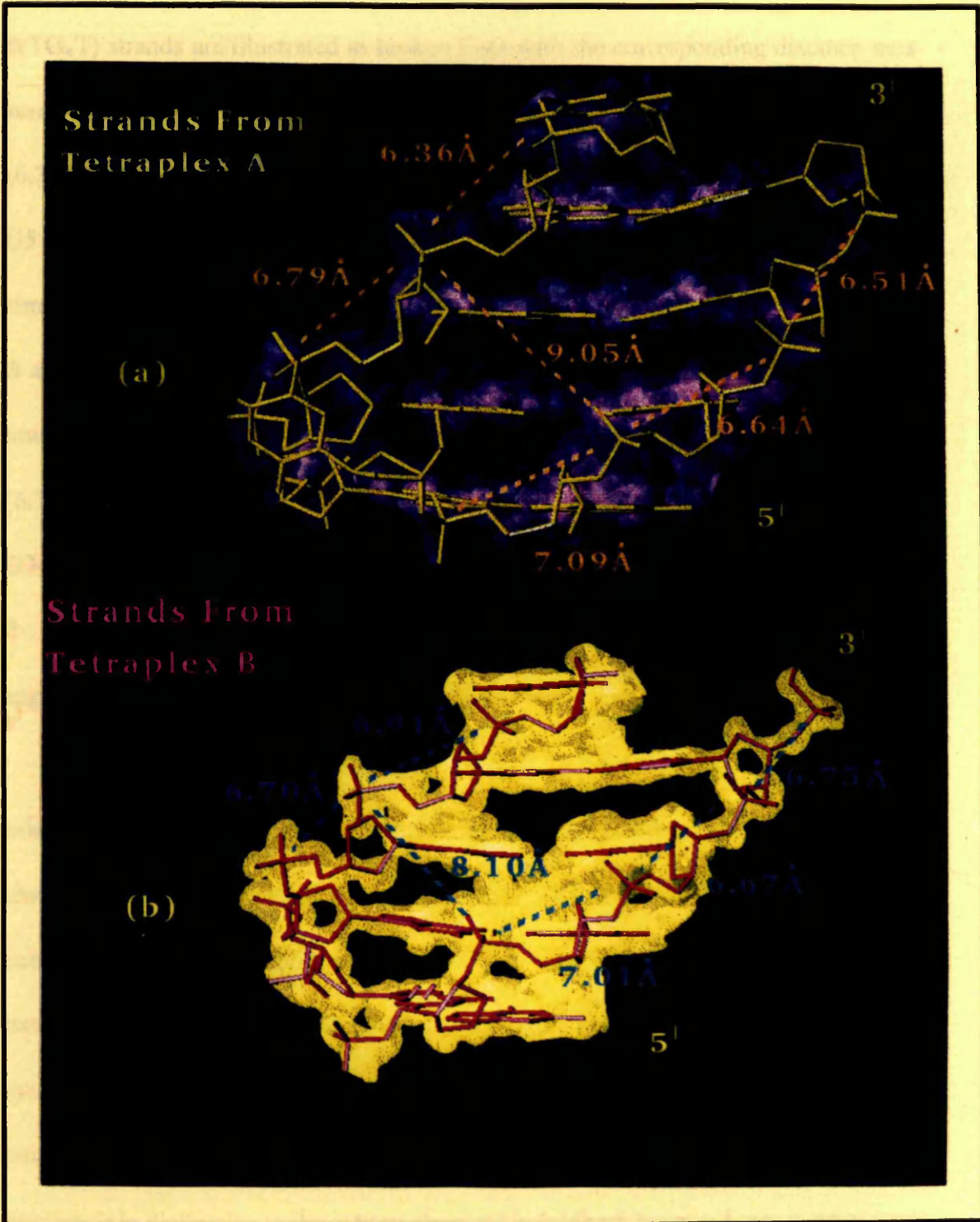


Figure 7.7: The P...P Separation Distances of the Tetraplex

nucleotide chains are shown in yellow, with the surrounding electron density depicted in the semi-transparent form in blue. The adjacent phosphate separations of the two d(TG₄T) strands are illustrated as broken lines with the corresponding distance measurements in orange. The d(TG₄T) strand on the left demonstrates the G5–3'T6 distance (6.36Å) and the G4–G5 distance (6.79Å), whilst the strand on the right depicts the G4–G5 (6.51Å), G3–G4 (6.64Å) and G2–G3 (7.09Å) distances for the oligonucleotide. A similar set-up is illustrated in Figure 7.7(b) where the adjacent strands from tetraplex B are shown in purple, with the surrounding semi-transparent electron density in yellow. In Figure 7.7(b) the strands on the left depicts the G5–3'T6 (6.91Å) and G4–G5 (6.70Å) distances in blue, while the strand on the right illustrates the G4–G5 (6.76Å), G3–G4 (6.67Å) and G2–G3 (7.01Å) distances for the adjacent strand. The observed P...P separation of adjacent phosphates is correlated with the character of the adopted sugar-puckering mode, as discussed in section 7.4.

Figure 7.7 also presents examples of the width of the tetraplex groove. The parallel orientation of strands generates four equivalent grooves of essentially minor-groove character. The term minor-groove designates the fact that the sugars to which a base-pair are attached lie closer to one side of the base-pair than the other, with the edge that lies closer to an imaginary line drawn between the two sugars designated the minor-groove side (see Calladine and Drew (1992)). The four equivalent grooves of the tetraplex contain the atoms N₂, N₃ and C₈ of the guanine bases, and this arrangement of grooves is in distinction to the pattern observed in foldback hairpin dimer quadruplexes of d(G₄T₄G₄) in which two wide and two narrow (Kang *et al.*, (1992)), and two medium, one wide and one narrow (Smith and Feigon (1992)) grooves are observed. In the

d(TG₄T) tetraplex the edge of the guanine bases that would define the major-groove in duplex DNA lies on the interior of the tetraplex.

In general the characterisation of the shapes of the grooves is crucial since it can influence access to the functional groups of the bases by some sequence-specific reagent such as a drug or a protein. However, groove shape will also influence the pattern of hydration. The groove width can be estimated as the shortest phosphorous-phosphorous distance across a groove less 5.8Å to compensate for the van der Waals radii of two phosphate groups. As Figure 7.7 demonstrates, the direct cross-groove P...P distances vary in the range 8.10Å to 9.05Å, and this corresponds to a groove width of about 2.30Å to 3.25Å. This groove width cannot be measured at all base steps but does tend to be at its widest at the ends of the tetraplex, and this may be influenced by the multiple conformations adopted by the thymine nucleotides at both ends of the tetraplex. As shall be covered later in the chapter, the relatively narrow grooves of the tetraplex are favourable binding sites for water molecules.

The minor-groove depth in duplex DNA is determined by the distance between phosphorous and the N₂ atom for a guanine nucleotide, less the sum of the appropriate van der Waals radii. Taking values for the van der Waals radii of phosphorous and nitrogen of 1.9Å and 1.5Å, respectively, the observed measurement corresponding to minor-groove depth for the d(TG₄T) tetraplex tends to occur in the region 2.6Å to 3.6Å along the length of the tetraplex.

7.6 Conformation of the Sugar–Phosphate Backbone

The conformation of the sugar–phosphate backbone in DNA is described, following the sequence of atoms $P \rightarrow O_{5'} \rightarrow C_{5'} \rightarrow C_{4'}$, by the angles (and atoms) $\alpha(O_{n-1}O_{3'}-P-O_{5'}-C_{5'})$, $\beta(P-O_{5'}-C_{5'}-C_{4'})$, $\gamma(O_{5'}-C_{5'}-C_{4'}-C_{3'})$, $\delta(C_{5'}-C_{4'}-C_{3'}-O_{3'})$, $\epsilon(C_{4'}-C_{3'}-O_{3'}-P)$ and $\zeta(C_{3'}-O_{3'}-P-O_{5'(n+1)})$. The definitions of α , β , γ , δ , ϵ and ζ are depicted in Figure 1.2 in Chapter 1 for a representative nucleotide unit.

Values for the backbone torsion angles for the d(TG₄T) tetraplex structure were determined using the program *NEWHEL93*, distributed by R. E. Dickerson. As detailed in section 7.2, this analysis was carried out by stripping out the non-helical thymine nucleotides. All four tetraplexes were analysed in this fashion. The results for the backbone torsion angles of tetraplexes A, B, C and D, where the letters A, B, C and D correspond to the usage adopted in Figure 7.2, are organised in Tables 7.7, 7.8, 7.9 and 7.10, respectively. Due to the stripping out of the 5' and 3' thymine nucleotides, the torsion angles α and β for each G2 residue, and the torsion angles ϵ and ζ for all the G5 residues are not reported. Tables 7.7 to 7.10 also contain averaged values for a DNA double helix of the B-form for comparison. These values were calculated from the sequences d(CGCGAATTCGCG) (Dickerson and Drew (1981)) and d(CGTAGATCTACG) (Leonard and Hunter (1993)), and were provided by Dr. Madeleine Moore of the University of York.

The torsion angle δ describes the orientation about the $C_{4'}-C_{3'}$ bond described by the atoms $C_{5'}-C_{4'}-C_{3'}-O_{3'}$. This angle is strongly correlated with the endocyclic sugar torsion angle ν_3 , described by the atoms $C_{2'}-C_{3'}-C_{4'}-O_{4'}$, and hence with the

Strand	Residue	α	β	γ	δ	ϵ	ζ	χ
1	G2	–	–	167.2	50.2	143.6	177.3	237.3
	G3	261.5	286.5	203.2	44.4	141.2	180.8	248.8
	G4	261.5	294.1	184.6	49.2	123.2	180.2	235.1
	G5	266.1	290.3	186.1	46.3	–	–	238.5
2	G2	–	–	169.7	49.1	149.5	184.0	238.1
	G3	261.4	285.5	201.9	42.2	145.2	190.0	250.5
	G4	254.6	287.5	181.8	42.2	134.8	209.3	249.8
	G5	223.6	294.3	166.0	42.5	–	–	244.7
3	G2	–	–	164.5	45.0	143.0	176.1	236.6
	G3	263.3	291.8	200.7	47.9	136.6	189.8	248.7
	G4	267.1	285.8	176.8	46.2	123.2	185.4	240.9
	G5	253.5	280.8	190.8	44.9	–	–	254.1
4	G2	–	–	177.5	49.1	146.4	178.1	238.0
	G3	258.3	290.4	202.0	46.4	143.4	184.9	247.0
	G4	256.0	290.0	184.6	52.6	128.0	178.0	243.0
	G5	258.7	294.1	189.4	43.5	–	–	250.6
	Average B-DNA	295	167	50	126	194	234	249

Table 7.7: Backbone Torsion Angles(°) for Tetraplex A

Strand	Residue	α	β	γ	δ	ϵ	ζ	χ
1	G2	–	–	111.6	174.0	87.6	199.1	180.9
	G3	290.6	291.8	198.2	46.8	141.2	182.8	240.5
	G4	262.3	287.0	188.6	51.3	130.6	185.2	240.9
	G5	256.4	290.9	188.1	43.4	–	–	247.1
2	G2	–	–	190.2	185.1	85.0	194.1	183.3
	G3	284.1	290.2	197.4	54.9	139.9	183.9	242.3
	G4	265.2	286.5	186.7	50.0	128.3	181.5	241.3
	G5	258.1	291.7	190.8	43.8	–	–	251.2
3	G2	–	–	180.4	187.6	91.5	196.9	182.7
	G3	286.4	291.9	200.3	47.3	142.0	184.2	241.6
	G4	260.5	288.7	189.6	46.8	139.5	185.1	248.9
	G5	245.0	296.8	180.4	45.5	–	–	239.2
4	G2	–	–	146.9	182.4	88.5	197.8	180.2
	G3	292.9	288.0	201.5	50.5	140.9	181.3	237.5
	G4	266.2	288.3	195.9	48.0	136.9	184.7	243.6
	G5	254.3	297.6	188.1	32.8	–	–	246.8
	Average B-DNA	295	167	50	126	194	234	249

Table 7.8: Backbone Torsion Angles(°) for Tetraplex B

Strand	Residue	α	β	γ	δ	ϵ	ζ	χ
1	G2	–	–	174.2	52.3	144.1	170.4	238.9
	G3	265.3	293.2	211.7	41.8	141.7	180.4	251.4
	G4	262.3	293.3	181.9	50.4	123.3	184.4	239.1
	G5	261.2	295.5	185.7	39.3	–	–	254.0
2	G2	–	–	164.6	48.9	143.4	178.8	237.5
	G3	263.3	288.7	205.8	42.2	141.2	179.2	247.9
	G4	264.0	291.1	188.3	48.5	126.9	179.6	240.3
	G5	260.5	293.2	186.7	47.3	–	–	242.0
3	G2	–	–	170.8	46.8	146.2	178.5	239.2
	G3	259.4	288.3	201.8	45.2	143.5	179.4	248.0
	G4	258.8	290.0	188.9	44.9	130.0	187.4	246.4
	G5	254.9	292.6	188.4	38.8	–	–	249.1
4	G2	–	–	161.3	46.2	142.1	176.2	236.8
	G3	260.1	291.0	203.0	47.0	142.1	188.2	249.2
	G4	263.4	285.4	179.4	47.6	120.9	184.0	236.6
	G5	262.8	277.2	194.5	45.7	–	–	256.7
	Average B-DNA	295	167	50	126	194	234	249

Table 7.9: Backbone Torsion Angles($^{\circ}$) for Tetraplex C

Strand	Residue	α	β	γ	δ	ϵ	ζ	χ
1	G2	–	–	182.3	188.3	87.4	195.2	183.0
	G3	287.1	285.8	203.5	50.4	145.5	184.8	246.3
	G4	256.9	288.6	191.9	45.4	138.8	183.5	248.7
	G5	246.0	292.2	179.6	50.5	–	–	237.2
2	G2	–	–	152.1	180.6	93.8	203.2	181.8
	G3	288.1	287.5	200.6	50.0	142.5	185.8	238.4
	G4	257.4	289.9	188.9	48.2	133.2	184.3	244.3
	G5	254.3	292.3	191.6	43.5	–	–	249.6
3	G2	–	–	171.1	190.2	91.1	195.3	180.5
	G3	285.3	292.4	200.6	46.9	142.9	183.2	243.1
	G4	264.8	281.3	189.3	50.0	133.5	184.2	245.1
	G5	251.8	295.3	184.8	43.5	–	–	245.4
4	G2	–	–	175.1	189.2	85.4	198.6	183.2
	G3	286.4	293.3	198.3	54.1	140.3	184.8	237.1
	G4	265.1	286.3	191.3	53.0	130.4	187.3	242.8
	G5	259.3	287.3	184.4	44.6	–	–	250.5
	Average B-DNA	295	167	50	126	194	234	249

Table 7.10: Backbone Torsion Angles(°) for Tetraplex D

pucker conformation adopted by the nucleic acid sugar. Inspection of Tables 7.7–7.10 reveals that values for the δ torsion angle for the tetraplex fall consistently, for the majority of residues, within the range 40–50°, or *+synclinal* to use the Klyne–Prelog notation (Saenger (1984)). However, for the tetraplexes labelled B and D, the value of δ converts dramatically to values within the range 170–190° (*+antiperiplanar* and *–antiperiplanar* in the Klyne–Prelog notation) for all the G2 residues in tetraplexes B and D. This shift in the value of δ correlates with similar dramatic shifts in ν_3 for the same residue, and hence in the sugar–pucker observed for these residues, reported in section 7.4 and in Tables 7.3 and 7.5. This conformational change is apparently required at the interface where the coaxially stacked tetraplexes overlap in order to avoid steric clash at the tightly packed tetraplex interface. A considerable effect of this packing deformity for the G2 tetrad of the B and D tetraplexes is also observed in the β –glycosyl orientation describing the relative positioning of sugar and base, and described later in this section.

Rotation about the $C_{4'}-C_{5'}$ bond of the DNA backbone allows the $O_{5'}$ atom to assume different positions relative to the deoxyribose sugar and is described by the torsion angle $\gamma(O_{5'}-C_{5'}-C_{4'}-C_{3'})$. Three main conformations with all substituents in staggered conformations are possible. In the *+synclinal*(30–60°) configuration the $O_{5'}$ atom is positioned over the deoxyribose, whilst in the *–synclinal*(270–330°) and *+antiperiplanar*(150–180°) orientations the $O_{5'}$ atom is directed away from the sugar (Saenger (1984)). In purine nucleoside studies the *+synclinal* and *+antiperiplanar* ranges for the torsion angle γ show a similar frequency for $C_{2'}$ –*endo* and $C_{3'}$ –*endo* sugar–puckers, whilst the *–synclinal* configuration is rarely observed. Inspection of

Tables 7.7–7.10 reveals that this propensity is conserved for the guanine tetraplexes of d(TG₄T) in the crystal. Most of the values for the γ torsion angle fall in the *+antiperiplanar* range (150–180°) with some values extending slightly beyond into the *–antiperiplanar* (180–210°) range, thus indicating that the O_{5'} atom is directed away from the sugar. For the G2 nucleotides the value of γ is generally lower than for the other nucleotides, with values approaching the boundary of 150° between *+antiperiplanar* and *+anticlinal* (90–150°). In fact, for residue G2 of strand 1 of tetraplex B in Table 7.8, γ falls as low as 111.6°, in the low end of the *+anticlinal* region. This aberration results from the noted deformation of the G2 residues of the B tetraplex at the tightly packed stacking interface, and in particular from the extended conformation of the 5' thymine residue which makes a broad intermolecular contact.

The torsion angles β and ϵ denote the rotations about the the C_{5'}–O_{5'} bond, defined by the atoms P–O_{5'}–C_{5'}–C_{4'}, and the bond C_{3'}–O_{3'}, defined by the atoms C_{4'}–C_{3'}–O_{3'}–P, respectively. Theoretical considerations and experimental data, largely derived from NMR spectroscopy, indicate that severe steric hinderance between the phosphate group and the sugar moiety restricts the β and ϵ torsion angles to fairly well-defined ranges (Saenger (1984)). Generally speaking, β is largely limited to the *+antiperiplanar* (150–180°) range, with only few deviations into *+anticlinal* (90–120°) and *–anticlinal* (210–270°), whilst ϵ is restricted to the *+antiperiplanar* (150–180°) and *–anticlinal* (210–270°) ranges. In the tetraplexes formed in the crystals by the oligonucleotide d(TG₄T), the values of the β torsion angle are confined to the upper reaches of the rarer, but still theoretically and experimentally valid, *–anticlinal* (210–270°) region, extending into the *–synclinal* (270–330°) domain. Likewise, whilst the torsion angle

ϵ commonly greatly populates the *-antiperiplanar*(180–210°) and *-anticlinal*(210–270°) conformations in determined crystal structures (Saenger (1984)), in the d(TG₄T) tetraplex, ϵ falls in the lower reaches of the *+antiperiplanar*(150–180°) region, and extends into the *+anticlinal*(90–150°) domain. Again, for the tetraplexes B and D, the values of ϵ are observed to dip into the unusual *+synclinal*(30–90°) conformation as a result of the steric stress experienced by the G2 nucleotides in these two tetraplexes.

Whilst the rotations about the C–O ester bonds are fairly restricted, rotations about P–O bonds are less so, and therefore P–O bonds act as the major pivots affecting polynucleotide structure. A total of seven basic sets of conformations for the torsion angles α and ζ , which describe the torsion angles defined by the atoms $_{n-1}\text{O}_3'-\text{P}-\text{O}_{5'}-\text{C}_{5'}$ and $\text{C}_{3'}-\text{O}_3'-\text{P}-\text{O}_{5'(n+1)}$, respectively, generate the fundamental properties of single-stranded polynucleotides (Saenger (1984)). Only one conformation, where α and ζ are both in the range of 270°, leads to the Watson–Crick type double helices with parallel bases orientated towards the helix axis. The oligonucleotide d(TG₄T) does not, however, form a Watson–Crick double helix but rather forms a parallel-stranded tetraplex. Tables 7.7–7.10 reveal that the α torsion angle has values in the *-anticlinal*(210–270°) and *-synclinal*(270–330°) ranges, whilst ζ falls in the *-antiperiplanar*(180–210°) and *+antiperiplanar*(150–180°) ranges. Thus the tetraplex corresponds approximately to the basic conformation where α is roughly 270° and ζ is 170°. This ensures that, like the Watson–Crick model, all of the sugars of the polynucleotide are in parallel alignment, base parallelism is feasible and the conformation is repeatable. However, unlike the Watson–Crick arrangement where the base is positioned towards the propagation axis, in the tetraplex conformation the relative position of the base is now away, as can be

seen in Figure 7.4.

Tables 7.7–7.10 also contain values for the torsion angle χ , which describes the conformation of the β -glycosyl linkage, and which is described by the atoms $O_{4'}$ – $C_{1'}$ – N_9 – C_4 in guanines. This rotational positioning of the base relative to the sugar is sterically restricted and two conformational states, *syn* and *anti*, are preferred. In the *anti* conformation, the heterocycle is orientated away from the sugar, whilst in the *syn* conformation it is orientated towards or above the sugar. The allowed ranges for the *anti* conformation for purine nucleosides typically extend from 180° , the classical *anti* domain, to about 300° , the *high anti* range (Saenger (1984)). All of the guanines within the planar G_4 tetrads of the tetraplex are connected to the sugar–phosphate backbones by glycosyl bonds that are exclusively *anti*. This is in contrast to the regular alternation of *syn* and *anti* glycosyl bonds in anti–parallel guanine tetraplexes (Wang *et al.*, (1991a, 1991b); Kang *et al.*, (1992); Wang and Patel (1992, 1993b); Smith and Feigon (1993)), but is in agreement with models originally proposed for poly(G) (Arnott *et al.*, (1974); Zimmerman (1976)) and NMR studies of parallel–stranded tetraplexes (Aboul–ela *et al.*, (1992); Cheong and Moore (1992); Wang and Patel (1992, 1993a)).

Most of the values for the torsion angle for the glycosyl bonds fall in the centre of the allowed range for purine residues, between the classical *anti* and *high anti* conformations, with values for χ between the limits of 230° and 250° . A representative guanine nucleotide is shown in Figure 7.8. All of the atoms are coloured yellow, with the exception of C_4 , N_9 , $C_{1'}$ and $O_{4'}$ which define the glycosyl linkage and are coloured red. The electron density surrounding the nucleotide is coloured blue and is depicted in the semi–transparent mode. It can be clearly seen that the nucleotide is in

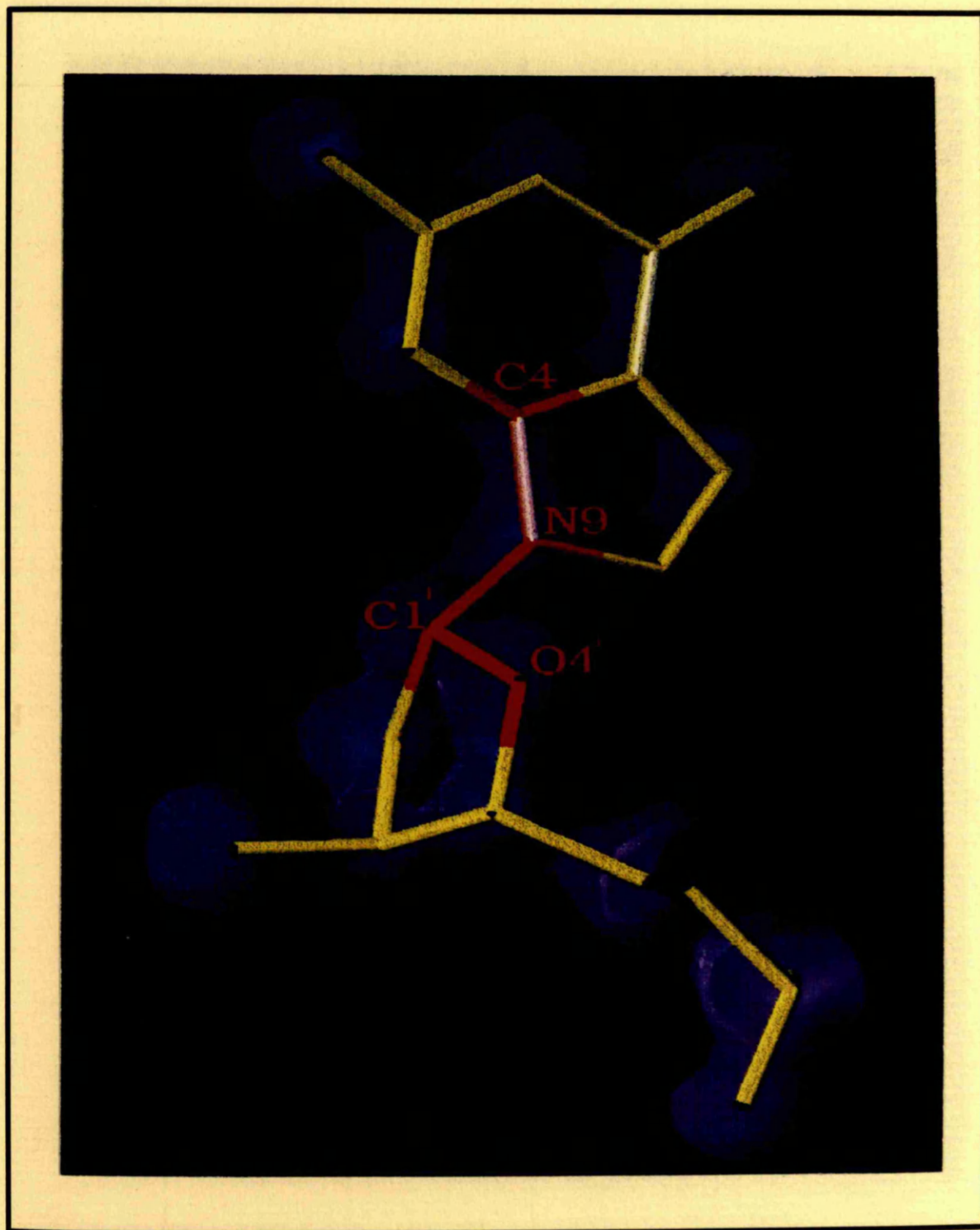


Figure 7.8: Guanine Nucleotide in the *anti* Configuration

a conformation. Interconversion of the two isomers is rapid at room temperature.

This can be compared with Figure 7.8, which shows the *syn* conformation of the base.

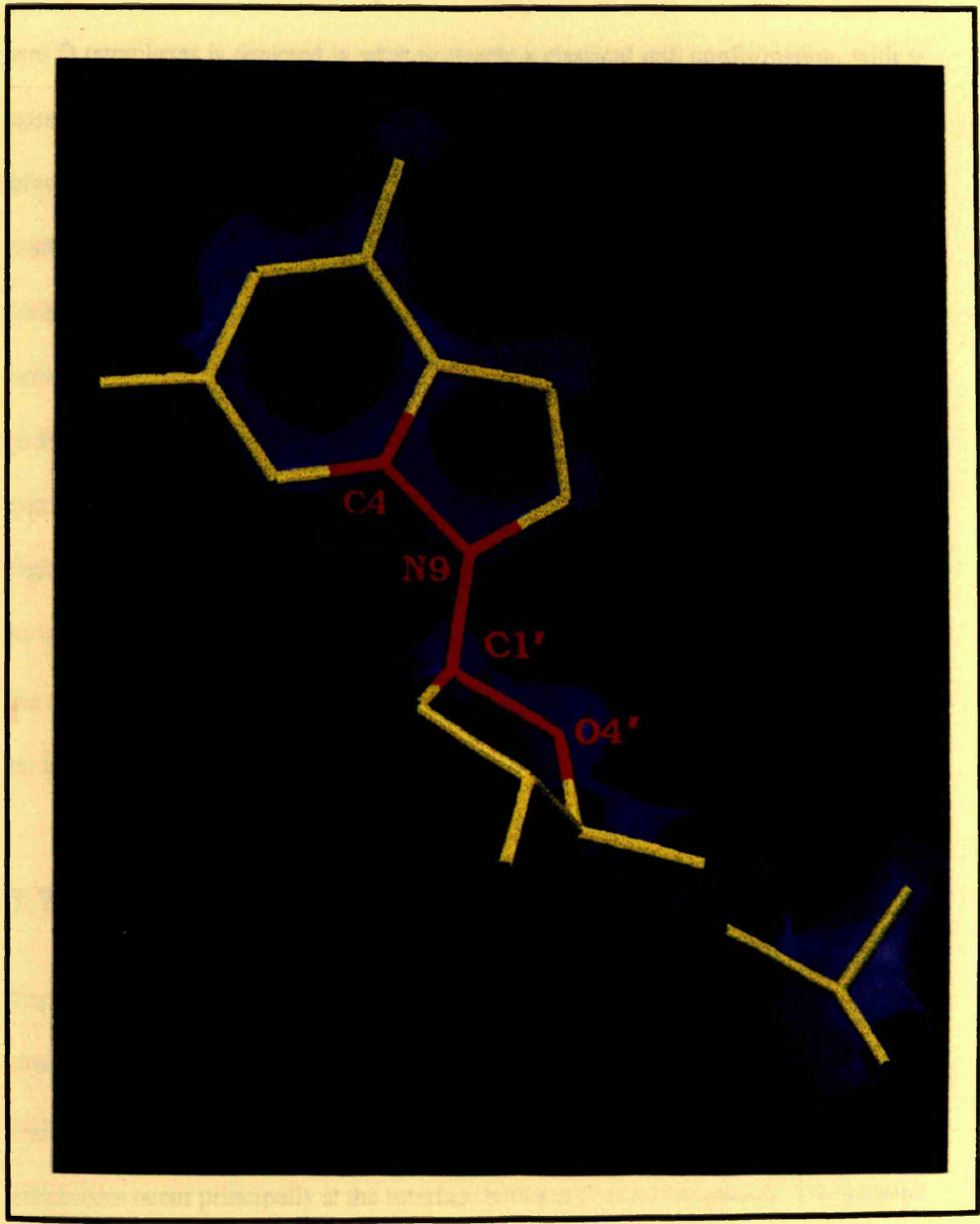


Figure 7.9: Guanine Nucleotide in the Classical *anti* Configuration

a conformation intermediate to the classical *anti* and *high anti* positions.

This can be contrasted with Figure 7.9 where a G2 guanine nucleotide from the B and D tetraplexes is depicted in what is clearly a classical *anti* conformation, with χ approximately 180° . Thus, from Tables 7.7–7.10 and Figure 7.9, it can be seen that glycosyl bonds for the G2 residues of tetraplexes B and D adopt the classical *anti* conformation, in distinction to all the other guanine nucleotides in the tetraplex. This switch appears to be necessary to avoid steric clash at the tightly packed interface between coaxially stacked tetraplex pairs. This can be further illustrated by reference to Figure 7.2. Whilst all of the tetrad planes exhibit a mild concavity towards the 3' end, as discussed in section 7.2, the G2 tetrads of tetraplexes B and D can be seen to be rigidly planar in comparison, as a result of the induced classically *anti* conformation of the guanine tetrads. It is only with regards to the β -glycosyl bond orientation that the tetraplex torsion angles resemble those of the B–DNA double helix, as represented by the values at the foot of Tables 7.7–7.10.

7.7 RMS Comparison of the Tetraplexes

The previous sections of this chapter have established the morphology of the parallel-stranded tetraplex formed by the oligonucleotide d(TG₄T) in the crystal unit cell. The four independently refined tetraplexes are not entirely equivalent structurally, and the differences occur principally at the interface between coaxial tetraplexes. The guanine tetrads exhibit a high degree of planarity and are connected to the backbones by glycosyl bonds that are exclusively *anti*. Most of the guanine sugars exhibit a B–DNA character,

however, at the tetraplex interface all of the G2 sugars of the B and D tetraplexes switch to a conformation resembling that of A-DNA. This switch is apparently required to avoid steric clash at the tightly packed interface, with distortions also evident in the manner of the *anti* glycosyl linkage and the torsion angles δ and ϵ . Elsewhere in the molecule, where the sugars are uniformly like B-DNA, the helical structure has a repeat corresponding to 10.4 tetrads per turn. The separation of adjacent phosphorous atoms on the same strand is also characteristic of B-DNA with variations observed most noticeably at the 5' and 3' ends of the tetraplexes where the thymine residues adopt multifarious conformations in the surrounding solvent. The conformation of the crystal structure tetraplex is in reasonable agreement with that determined for the same molecule in solution by NMR (Aboul-ela *et al.*, (1992, 1994)), except at the 5' end: this discrepancy appears to arise from the interaction of tetraplexes in the crystal.

The four tetraplexes occupying the crystal unit cell were compared by means of root-mean-square (RMS) fitting of the tetraplex to itself. This analysis was conducted using the CCP4 program *LSQKAB* with the assistance of Eleanor Dodson of the University of York. Due to the large variety of conformations adopted by the terminal thymine residues, the fitting of the tetraplexes was performed using only the guanine bases. The overall RMS displacement for tetraplexes A and B, A and C and A and D is 0.297Å, 0.267Å and 0.288Å, respectively, whilst the RMS displacement for tetraplexes B to C and B to D is 0.238Å and 0.122Å, respectively.

These results indicate that the four tetraplex structures, considered on the basis of the guanine residues only, are all fairly similar, but that there is a consistent variation in the conformation of all four tetraplexes in the unit cell. This result came as something of

a surprise since it was anticipated, on the basis of the main features of the packing of the four tetraplexes, that the tetraplexes would fall into two main groups, where tetraplexes A and C would be virtually identical, as would tetraplexes B and D. The results above suggest that whilst the tetraplexes adopt a consistent morphology, this morphology is, to a degree, flexible. The variations observed for all the tetraplex structures, where the match of the coaxial pair (A, B) is comparable to the match for tetraplexes A and C, would appear to result from local discreet variations in the conformation of the tetraplexes required to accommodate the packing of all four tetraplexes in the dense unit cell. It is interesting to note that the lowest observed RMS displacement occurs for the tetraplexes B and D (0.122\AA), where the G2 residues switch conformation at the tightly packed tetraplex interface. It is possible that the better match of these two tetraplexes, in comparison to all the others, may be as a result of the enforced conformational change witnessed at the tetraplex 5' interface.

Figure 7.10 illustrates the superposition of the coaxially stacked tetraplexes A and B. The tetraplexes are fitted on the basis of the guanine nucleotides only. However, the thymine residues, which adopt a wide variety of conformations, are also depicted in the figure. Tetraplex A is painted in red and tetraplex B in green. Both tetraplexes are orientated such that the 3' end of each tetraplex is positioned at the top of the figure, and the 5' end is at the bottom. The RMS displacement for the two tetraplexes is 0.297\AA . The two central guanine tetrads and the guanine tetrad at the 3' end of the tetraplexes appear to match fairly well, but there are clearly evident discrepancies in the conformation of the sugar-phosphate backbone of the individual strands. The most obvious discrepancy between the two tetraplex structures can be observed for the G2



Figure 7.10: RMS Superposition of the Coaxial Pair (A, B)

residues which constitute the 5' tetrads. The G2 tetrad of tetraplex B is considerably more planar than that of tetraplex A, where the β -glycosyl linkage of the G2 guanines of the B tetraplex adopt the classical *anti* conformation with a torsion angle value for χ of approximately 180° . The switch in the sugar-puckers of the G2 deoxyriboses, where the sugars of the B tetraplex adopt conformations in the north of the pseudorotation phase cycle, unlike all of the other sugars which have southern conformations, is clearly apparent.

A superposition of tetraplexes A and C, not shown here, is very similar to the imposition of tetraplexes A and B in Figure 7.10, with the exception of the terminal thymine residues. As such, these two comparisons illustrate the manner in which the concept of quasi-equivalence (Casper and Klug (1962)) has some application in considering the crystal packing of the d(TG₄T) tetraplex. This point is discussed further in section 7.11. The structural difference of tetraplexes A and C, and B and D, yield the imperfection of the non-crystallographic symmetry.

7.8 Location of the Sodium Cations

As discussed in Chapter 1, studies of the self-association of guanine derivatives and polymers have revealed that size-specific, monovalent cation binding is integrally involved in the association of guanine base-tetrads into quadruple helical structures. The specific cation binding site(s) of guanine quadruple helices appear to be of optimal size for potassium binding since the relative stability of quadruplexes decreases markedly as the size of the monovalent cation is decreased ($K > Na > Li$) or increased ($K > Rb > Cs$)

(Miles and Frazier (1978); Pinnavaia *et al.*, (1978); Howard and Miles (1982)). It has been postulated that this cation-specific stability arises from the ability of potassium to bind in eight coordinate chelation cages between guanine tetrads (Miles and Frazier (1978)), as illustrated in Figure 1.10, which have optimal size and ligand stoichiometry for potassium chelation.

Two different proposals for the binding of sodium cations have been put forward. Firstly, since the axial cages between guanine tetrads appear to be too large for the optimal binding of sodium, it has been suggested that sodium binds instead within tetrad planes (Howard and Miles (1982); Williamson *et al.*, (1989)), although this requires a square planar four coordinate geometry not typically observed for sodium. However, since ligand cages which bind potassium also bind sodium, albeit with reduced affinity (Gandour *et al.*, (1986); Cram *et al.*, (1986)), Sundquist and Klug (1989) have proposed that both sodium and potassium bind in eight coordinate axial cages.

As discussed in Chapter 6, the electron density map of the tetraplex structure was of sufficient quality at the early stages of the refinement procedure to identify sodium cations lying along the axis of the tetraplex. Seven sodium ions were observed per stacked pair of tetraplexes, including one located at the tetraplex interface. Most of the ions are located between consecutive tetrads and are coordinated by the tetrads buried O₆ atoms.

Figure 7.11 illustrates the location of the inter-tetrad sodium ions. Shown in the figure is a $(3F_o - 2F_c)$ electron density map contoured at 1.5σ in the semi-transparent form, depicting the string of sodium cations running along the central cavity of the

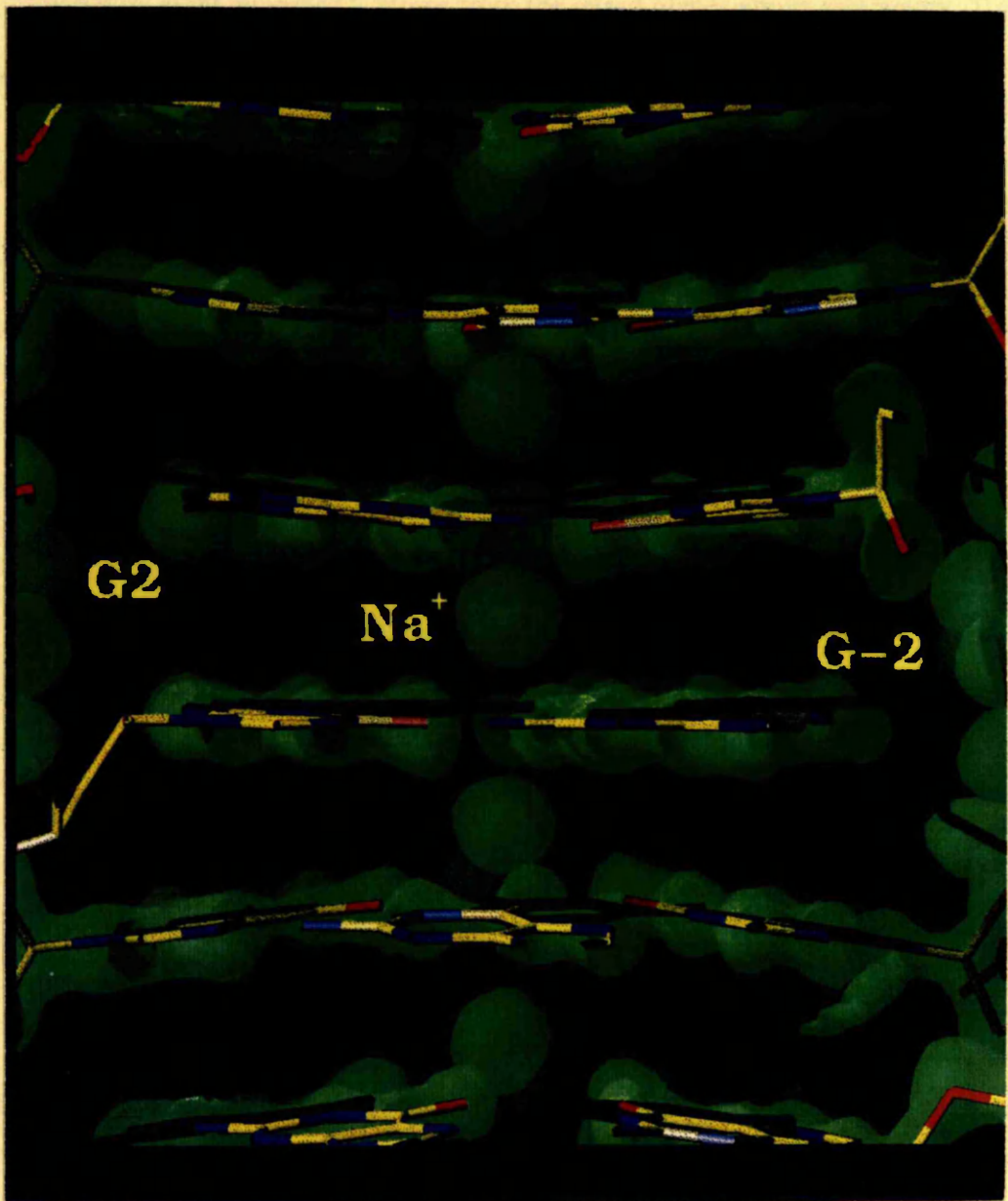


Figure 7.11: The Inter-Tetrad Sodium Cations

tetraplex. The molecule is rendered in the default colours and the G2 tetrad, belonging to tetraplex A, and the G-2 tetrad, belonging to tetraplex B, are labelled as is the central sodium ion located between tetrads G2 and G-2. Each sodium ion in the figure is coordinated by eight guanine O₆ atoms from the tetrad planes above and below. The G5 tetrads, and the ions coordinating them, are not shown in this figure.

Plane of Ion Coordination	Tetraplex Na ⁺ ...O ₆ Distances(Å)			
	Tetraplex A	Tetraplex B	Tetraplex C	Tetraplex D
G2	2.69	—	2.70	—
	2.71	—	2.71	—
	2.73	—	2.71	—
	2.73	—	2.71	—
G-2	—	2.75	—	2.74
	—	2.75	—	2.77
	—	2.75	—	2.78
	—	2.75	—	2.78

Table 7.11: Na⁺...O₆ Distances for Axial Sodium Ions Between Planes G2 & G-2

The sodium ion that is located between the tetrad planes of G2 and G-2 satisfies a nearly perfect inverted-bipyramid coordination geometry. This central sodium ion, located at the interface between tetrad planes, is coordinated by eight equidistant carbonyl oxygen atoms, with an average value for the two independent sites between

coaxial pairs (A, B) and (C, D) of $2.75 \pm 0.02 \text{ \AA}$. The individual values of the $\text{Na}^+ \cdots \text{O}_6$ distances for the two coaxial tetraplex pairs are given in Table 7.11.

The ion situated between the G2 and G3 (and G-2 and G-3) tetrads is partially displaced towards the G3 plane. The individual distances of $\text{Na}^+ \cdots \text{O}_6$ coordination for the relevant sites in each of the four tetraplexes are detailed in Table 7.12. The average value for the $\text{Na}^+ \cdots \text{O}_6$ distances for the four independent sites are $2.89 \pm 0.07 \text{ \AA}$ and $2.60 \pm 0.07 \text{ \AA}$ for the G2 and G3 distances, respectively.

Plane of Ion Coordination	Tetraplex $\text{Na}^+ \cdots \text{O}_6$ Distances(\AA)			
	Tetraplex A	Tetraplex B	Tetraplex C	Tetraplex D
G2	2.92	2.78	2.96	2.78
	2.95	2.79	2.96	2.85
	2.97	2.86	2.98	2.87
	3.01	2.87	2.99	2.95
G3	2.63	2.53	2.61	2.49
	2.65	2.54	2.65	2.50
	2.69	2.54	2.66	2.54
	2.71	2.55	2.70	2.54

Table 7.12: $\text{Na}^+ \cdots \text{O}_6$ Distances for Axial Sodium Ions Between Planes G2 & G3

For the ion situated between the G3 and G4 (and G-3 and G-4) tetrads the displacement towards the G4 tetrad is even greater than that towards G3 for the ion located

between tetrad planes G2 and G3. The individual distances of $\text{Na}^+ \cdots \text{O}_6$ coordination for the axial sodium cations located between the G3 and G4 (and G-3 and G-4) tetrads are given in Table 7.13. The average values for the $\text{Na}^+ \cdots \text{O}_6$ distances for the four independent sites between the G3 and G4 tetrads is $3.32 \pm 0.10 \text{ \AA}$ for the G3 tetrad coordination, and $2.43 \pm 0.03 \text{ \AA}$ for the G4 tetrad coordination. The observed outward displacement of sodium ions away from the central tetrad pair G2 and G-2 tetrad step is probably the result of electrostatic repulsion between cations.

Plane of Ion Coordination	Tetraplex $\text{Na}^+ \cdots \text{O}_6$ Distances(\AA)			
	Tetraplex A	Tetraplex B	Tetraplex C	Tetraplex D
G3	3.10	3.25	3.21	3.20
	3.29	3.26	3.31	3.32
	3.33	3.27	3.35	3.33
	3.55	3.28	3.41	3.41
G4	2.40	2.41	2.39	2.40
	2.43	2.43	2.43	2.42
	2.46	2.46	2.45	2.45
	2.48	2.47	2.45	2.46

Table 7.13: $\text{Na}^+ \cdots \text{O}_6$ Distances for Axial Sodium Ions Between Planes G3 & G4

There is no sodium cation between the G4 and G5 (and G-4 and G-5) tetrad planes because, as the ion displacement continues towards the 3' end of the tetraplexes, the

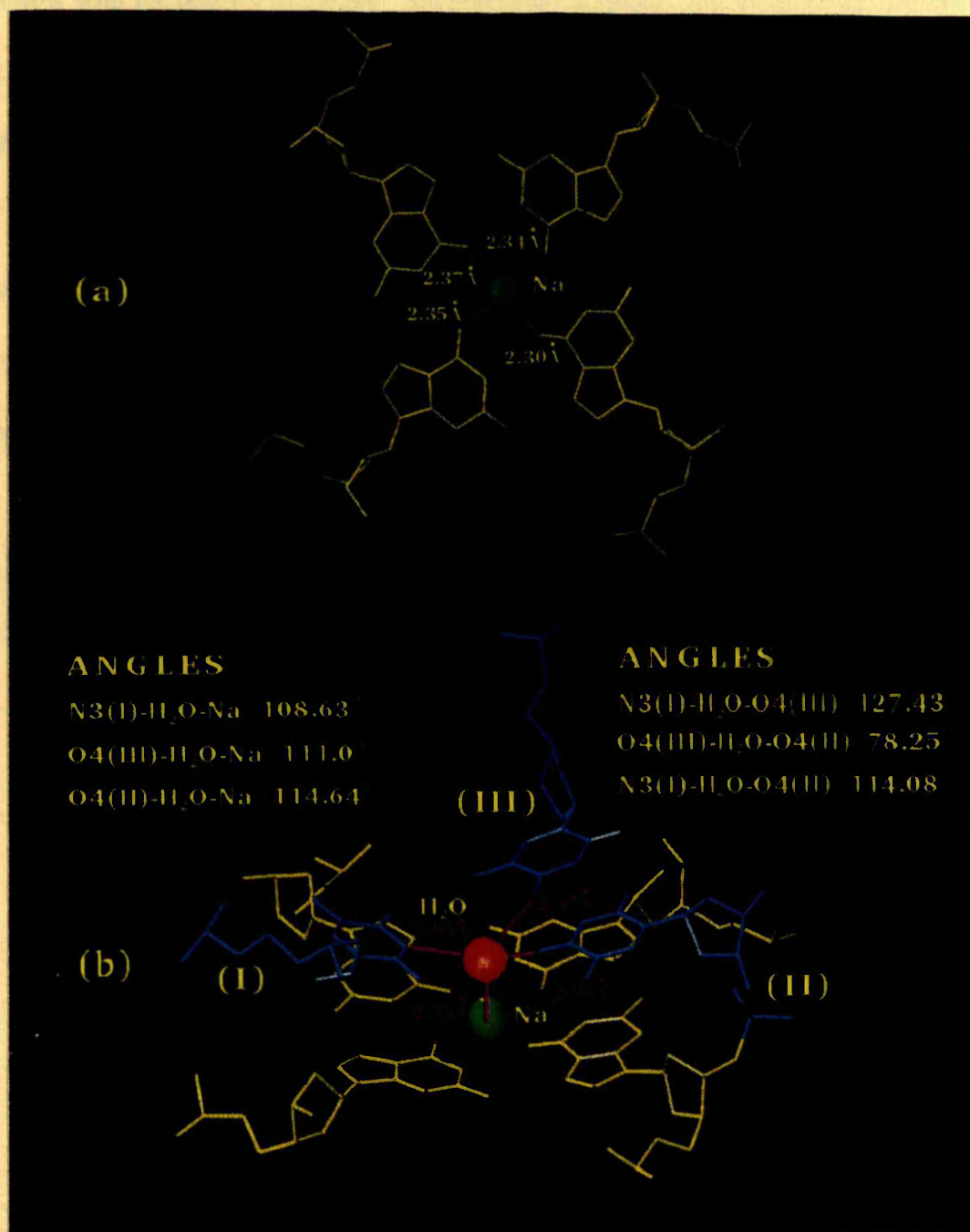


Figure 7.12: The Intra-Tetrad Sodium Cations

sodium ions actually lie in the plane of the tetrad formed by the four G5 guanine bases. This situation is illustrated in Figure 7.12(a) which shows a representative G5 tetrad in yellow with the density of the Na⁺ ion presented in green. The intra-tetrad sodium ions occupy the tetrad planes at G5 and G-5. There are four independent examples of the sodium ion coplanar with the G5 tetrad in tetraplexes A, B, C and D. Each Na⁺ ion is coordinated by four guanine O₆ atoms, depicted as broken yellow lines in Figure 7.12(a), with an average Na⁺...O₆ distance of 2.34±0.02Å, taken over the four independent sites. The individual values for the Na⁺...O₆ coordination distances are presented in Table 7.14.

Plane of Ion Coordination	Tetraplex Na ⁺ ...O ₆ Distances(Å)			
	Tetraplex A	Tetraplex B	Tetraplex C	Tetraplex D
G5	2.33	2.33	2.33	2.30
	2.34	2.34	2.34	2.34
	2.35	2.34	2.37	2.35
	2.38	2.36	2.38	2.37

Table 7.14: Na⁺...O₆ Distances for G5 Planar Axial Sodium Ions

The Na⁺ cation has an ionic radius of approximately 1.00Å based upon observation of the six coordinate ion in crystal structures, whilst the O₆ atom possesses a van der Waals radius of 1.40Å. Therefore, the observed Na⁺...O₆ distances for the coplanar ion in the G5 tetrad indicate that, whilst the Na⁺ cation fits into this intra-tetrad position, the fit is not optimal but is, in fact, a little tight. Because of its larger ionic radius of

1.50–1.60Å, it is unlikely that the potassium cation could adopt a position coplanar with a tetrad.

Table 7.15 summarizes the $\text{Na}^+ \cdots \text{Na}^+$ distances observed along the length of the coaxial tetraplex pairs (A, B) and (C, D) and also provides the individual temperature factors for the sodium cations. The temperature factor for the ion located in a near perfect inverted-bipyramid geometry between the G2 and G-2 tetrads is one of the lowest of the set, and the temperature factors for the other ions increase as they move towards the 3' of the tetraplex. Finally, the temperature factor is at its highest for the sodium ion located in the more exposed G5 tetrads. The distance separating sodium ions also increases as they move away from the central G2/G-2 chelation cage towards the 3' ends of the individual tetraplexes.

Ion Location	Coaxial Pair AB		Coaxial Pair CD	
	B(Å ²)	Na ⁺ ...Na ⁺ (Å)	B(Å ²)	Na ⁺ ...Na ⁺ (Å)
G5 Plane(A, C)	10.50		10.04	
		4.31		4.27
Between G4 & G3(A, C)	9.40		9.39	
		3.82		3.68
Between G3 & G2(A, C)	4.85		4.58	
		3.52		3.48
Between G2 & G-2(AB, CD)	5.15		4.48	
		3.45		3.54
Between G-2 & G-3(B, D)	7.60		7.53	
		3.66		3.82
Between G-3 & G-4(B, D)	9.96		10.30	
		4.35		4.27
G-5 Plane(B, D)	10.09		10.13	

Table 7.15: $\text{Na}^+ \cdots \text{Na}^+$ Distances and Temperature Factors for Axial Sodium Ions

The crystal structure of the parallel-stranded tetraplex formed by the oligonucleotide d(TG₄T) reported here thus provides direct crystallographic evidence to support both the proposals that the Na⁺ cation binds within the tetrad planes (Howard and Miles (1982); Williamson *et al.*, (1989)) and within eight coordinate tetrad cages (Sundquist and Klug (1989)). It is, however, only the ion located between and not within the coaxially stacked tetraplexes that adopts a near perfect inverted-bipyramid coordination geometry, and this therefore begs the question of where the Na⁺ cations would be located in an isolated, unstacked tetraplex? It is conceivable that the tetraplexes stack even in solution, but, if they do not, with no ion located in a position like the Na⁺ cation found between planes G2 and G-2, it might be reasonably expected that different ion locations might be found, perhaps involving co-planar sodium cations at both ends of the tetraplexes. The resolution and relevance of these questions depend a great deal on whether parallel-stranded tetraplexes are formed *in vivo* and, if they are, on the nature of their surrounding environment.

The co-planar sodium ion is coordinated, in addition to the four O₆ atoms, by a water molecule that is bound at the fifth position. The presence of this pentacoordinated sodium geometry in Na⁺ stabilised guanine tetrad structures was suggested by Bouhoutsos-Brown *et al.*, (1982) on the basis of NMR studies of guanosine 5'-monophosphate and has been observed in crystallographic studies of a biologically active analogue of the small molecule antamanide(II) complexed with sodium (Karle (1974)), in which very similar Na⁺...O distances are observed. This is illustrated for the tetraplex structure in Figure 7.12(b). The guanine nucleotides, which originate from the G5 tetrad of tetraplex D, are shown in yellow. The Na⁺ cation and water

oxygen are displayed in their electron density, with the sodium density coloured green and the oxygen density coloured red. The average distance between the sodium cation and axial water is $2.53 \pm 0.03 \text{ \AA}$, and this value is consistent with the radii for Na^+ and molecular oxygen mentioned above. The axial water molecules form hydrogen-bonds with three thymine bases in the figure, and the thymine residues are coloured in blue.

Thymine residue (I) originates from the 3' end of the A tetraplex in the symmetry cell related by a positive translation along the *b*-axis. The thymine nucleotide (III) also originates from the same cell as residue (I), but is donated from the 3' end of tetraplex C. The final thymine residue, labelled (II) in Figure 7.12(b), belongs to the same D tetraplex which also contains the guanine residues coloured in yellow. Thymine residue (I) coordinates the water molecule by means of the hydrogen belonging to the endocyclic N_3 atom of the base, whilst residues (II) and (III) coordinate the water via the exocyclic O_4 atom in each case. Assuming that the water hydrogens are therefore involved in the coordination of the O_4 atoms of the thymine residues (II) and (III), this leads to coordination distances, for the example in Figure 7.12(b), of $\text{N}_3\text{-H(I)} \cdots \text{O(H}_2\text{O)}$ [2.93 \AA], $\text{O}_4\text{(II)} \cdots \text{H-O(H}_2\text{O)}$ [2.95 \AA] and $\text{O}_4\text{(III)} \cdots \text{H-O(H}_2\text{O)}$ [2.87 \AA], measured directly between the nitrogen and oxygen atomic positions. The minimum and maximum distances for $\text{N-H} \cdots \text{O}$ hydrogen-bonds are 2.74 \AA and 3.07 \AA , respectively (Saenger (1984)). Likewise, the minimum and maximum distances for $\text{O-H} \cdots \text{O}$ hydrogen-bonds are 2.60 \AA and 3.05 \AA , respectively, indicating that the coordination distances observed in the coordination of the axial water by the thymine residues (I), (II) and (III) are sensible. The coordination of the axial water molecule by the three thymine residues leads to the formation of a distorted tetrahedral geometry,

as witnessed by the angles listed in Figure 7.12(b).

A similar arrangement of thymine nucleotides coordinating the axial water is in evidence for the other three examples contained within the unit cell. For the water molecule coordinating the coplanar Na^+ cation within the G5 tetrad of tetraplex B, the distorted tetrahedral coordination is accomplished by thymine residues originating from the B tetraplex itself, the 3' end of tetraplex A in the cell related by combined $+a$ and $+b$ axial translations, and a 3' residue from the C tetraplex in the symmetry cell related by combined $+b$ and $-c$ axial translations. The thymine from the B tetraplex coordinates the water via the exocyclic O_4 atom of the base, whilst the forementioned symmetry related thymine utilises the exocyclic O_4 and endocyclic N_3 hydrogen atom, respectively.

For tetraplex C the axial water is coordinated by the O_4 atom of a thymine originating from the C tetraplex itself, the O_4 atom of a 3' residue of the D tetraplex in the symmetry cell related by combined $+a$ and $-b$ axial translations, and the N_3 hydrogen atom of a 3' residue from the B tetraplex from the symmetry cell related by combined axial translations of $+c$ and $-b$. Finally, the axial water of tetraplex A is coordinated by the O_4 atom of a thymine residue of the A tetraplex itself, the O_4 atom of a 3' residue of the B tetraplex originating from the symmetry cell related by combined $+a$ and $-b$ axial translations, and a water molecule. In this case it is possible that the non-axial water molecule exists in the model in place of the expected third thymine residue, probably donated from a symmetry molecule, but which is one of the five missing thymine residues in the final model, presumably due to high thermal disorder.

The interdigitation of tetraplexes in adjacent cells, described above, accounts in

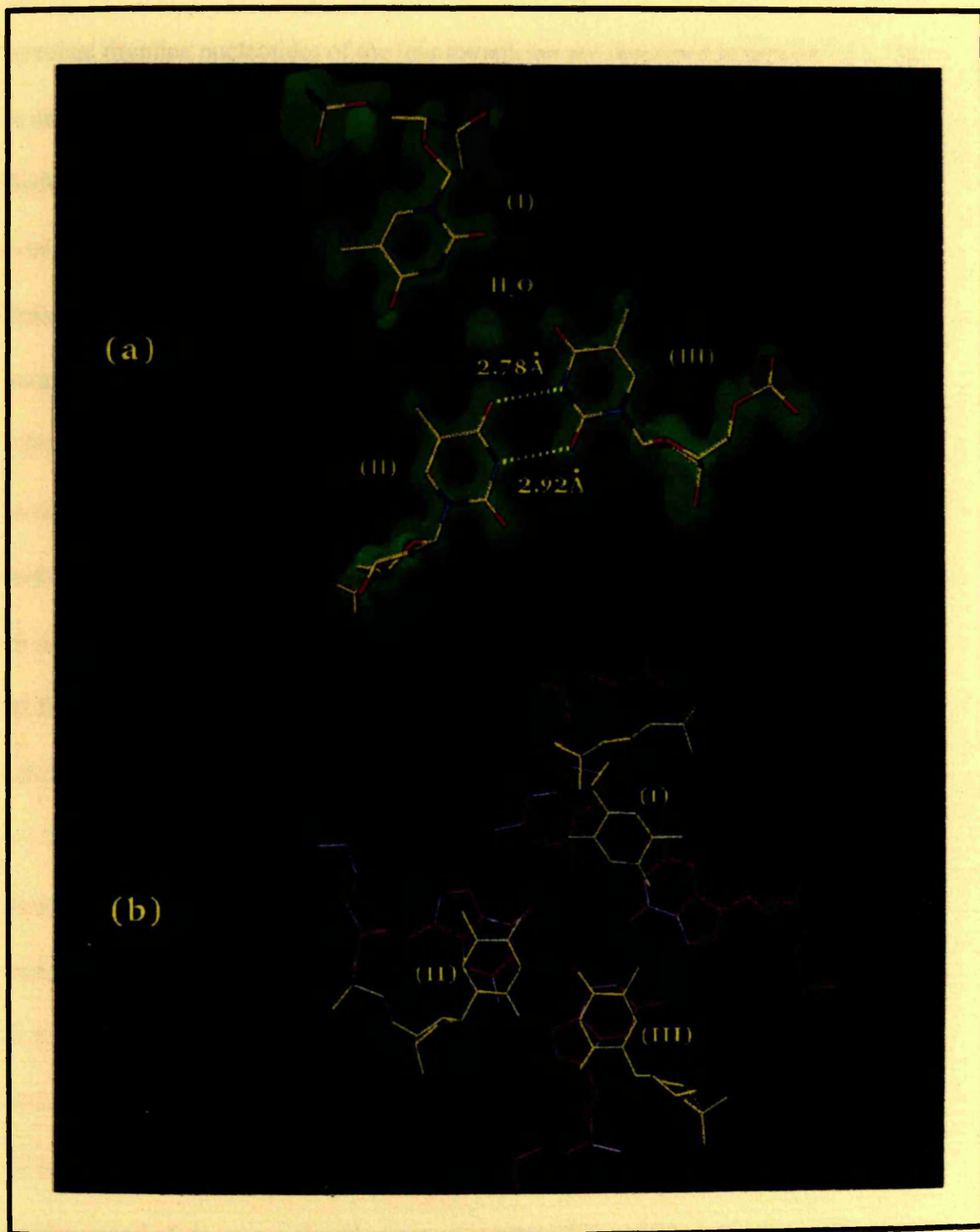


Figure 7.13: The Longitudinal, Interdigitating 3' Thymine Nucleotides

part for the remarkably high crystalline order of the tetraplex crystal structure in its formation of favourable lattice contacts. Other examples of contacts mediated by the terminal thymine nucleotides of the four tetraplexes are described in section 7.11. The arrangement of the three thymine residues which coordinate the axial water molecule is further illustrated in Figure 7.13. Figure 7.13(a) illustrates the three thymine nucleotides which coordinate the axial water in tetraplex D. The thymines are depicted using the default colours and the $(3F_o - 2F_c)$ electron density, contoured at 1.5σ , is shown in the semi-transparent mode in green. The three labels for the thymine residues, (I), (II) and (III), correspond to the usage in Figure 7.12. The thymine residues are not equidistant in relation to each other. The nucleotide labelled (I) is a further distance removed from residues (II) and (III) than the latter two residues are from each other. It is interesting to note that the density for the linkage between the sugar and phosphate of residue (I) is discontinuous in the figure, providing evidence of the conformational disorder that affects some of the thymine residues making intermolecular contacts.

The (II) and (III) thymine residues are well-defined and approach each other closely enough to form a striking intermolecular T•T base-pair mismatch. Residue (II) originates from the D tetraplex whilst residue (III) extends from the symmetry related C tetraplex, as previously described. The well-resolved T•T base-pair consists of two hydrogen-bonds. One of the hydrogen-bonds involves the exocyclic O_4 atom of residue (II) and the endocyclic N_3 hydrogen atom of residue (III), whilst the other is comprised of the exocyclic O_2 atom of residue (III) and the hydrogen atom of the endocyclic N_3 atom of residue (II). The hydrogen-bond lengths are 2.78\AA for the $O_4(\text{II}) \cdots \text{H}-N_3(\text{III})$ hydrogen-bond, and 2.92\AA for the $O_2(\text{III}) \cdots \text{H}-N_3(\text{II})$ hydrogen-

bond. Both values fall comfortably within the minimum and maximum distances, measured directly from the N and O atoms, for this type of hydrogen-bonding interaction i.e. 2.74Å and 3.07Å, respectively.

There are three other similar examples of this T•T base-pair mismatch interaction for the thymine residues coordinating the axial waters of tetraplexes A, B and C. The discovery of these intermolecular T•T base-pair mismatches is significant since they remain quite a rare occurrence in nucleic acid structure studies. Efforts to crystallize the intramolecular T•T mismatched pair in a duplex DNA structure have so far met without success (see Hunter (1992)). Only one other example of an intermolecular T•T mismatch has been observed in the sequence d(CGCGCGTTTTTCGCGCG) (Chattopadhyaya *et al.*, (1988)), which forms a hairpin structure where, as for the d(TG₄T) tetraplexes, a symmetry related molecule is positioned so that two thymines are close enough to pair.

The thymine residues (I), (II) and (III) also approach the G5 guanine tetrad closely enough, at approximately 3.4Å, to facilitate stacking interactions between themselves and the G5 tetrad. This situation is illustrated in Figure 7.13(b), where the guanine residues are painted purple and the thymine nucleotides yellow. The general principles governing stacking interactions are discussed in section 7.10.

The thymine residue (I) straddles the gap between two guanine residues and, therefore, there is no effective overlap of the aromatic rings. However, there is a close intermolecular contact between the O_{4'} atom of the thymine sugar and the five-membered ring of a guanine heterocycle. This kind of interaction is probably as a result of the relatively low electronegativity of the O_{4'}, which encourages close approach to the π -

electronic system, and which has been observed in nucleoside and nucleotide structures (Saenger (1984)). The thymine base (I) also stacks such that the electronegative O₂ and O₄ atoms are positioned above the electropositive C₆ and N₁ hydrogens of adjacent guanine nucleotides. All of these electrostatic interactions may contribute a little to the stability of the thymine nucleotide conformation.

Thymine residues (II) and (III) exhibit a more extensive overlap with the guanine heterocycles in comparison with residue (I). Dipole-induced interactions would appear to play the major role in this case, with the C₄=O₄ and exocyclic methyl groups superimposed over the aromatic system of the adjacent guanine for residue (II), and the C₂=O₂ and C₄=O₄ groups lying marginally over the aromatic system of the guanine below for residue (III). There would also appear to be favourable electrostatic interactions involving the electropositive C₆ and electronegative O₄ of residue (II) and the electronegative N₃ and electropositive C₆ of the adjacent guanine. Likewise, for residue (III), electrostatic interactions may involve the electronegative O₂ and O₄ of the thymine and the electropositive N₉ and C₆ of the guanine below. It is worth noting that there are some apparently unfavourable interactions as well: oxygens of the carbonyls in residues (II) and (III) appear to be in register with the O₄ of the guanines below.

It may be the case that the apparent lack of favourably dipole-induced dipole interactions for residue (I) contributes significantly to the signs of conformational disorder in the electron density in Figure 7.13(a). Additionally, this analysis may possibly help to explain the apparent absence of the third thymine coordinating the axial water of tetraplex A, which is occupied instead by a water molecule, as discussed above.

7.9 Hydration of the Tetraplex

Crystals of nucleic acids, like crystal of proteins, contain a considerable amount of solvent, generally in the region of 50% of the total volume of the crystal (Dock-Bergeon and Moras (1992)). As a result, the considerable hydration of the nucleic acid is a key factor in the difficulty encountered in obtaining crystals which diffract to high-resolution. However, when crystals of a suitable order can be obtained, the opportunity presents itself for experimental observation of nucleic acid hydration *in vitro*, with subsequent speculation upon the *in vivo* relevance of the patterns observed.

Detailed accounts of nucleic acid hydration are given by Saenger (1984, 1987) and Westhof (1988), whilst a brief overview of recent key results is contained in Berman (1991, 1994). In DNA, distinct hydration patterns are observed around the sugar-phosphate backbone as well as in the major and minor-grooves. In their study of the water structure reported in a number of DNA structures, Saenger *et al.*, (1986) proposed that in mixed-sequence DNA the hydration of the backbone is related to the global conformation. In both A- and Z-DNA a continuous chain of water molecules links the phosphate oxygens along the backbone. The water molecules are shared between adjacent phosphate groups. In the B-form structures there are more water molecules situated around each phosphate and virtually no water bridges between adjacent phosphates. These observations are explained by consideration of the distances separating the phosphate oxygens in the various global conformations. In A- and Z-DNA the distances are approximately 5.4Å and 4.6Å, respectively, and are thus short enough to be bridged by a single water molecule. In B-form DNA the corresponding

phosphate oxygen distances are about 6.5 Å or more, too large for a single water bridge to be formed.

The hydration of the nucleic acid helix appears to be influenced by the width of the grooves and the sequence of the structure. A regular pattern of groove hydration was first observed by Drew and Dickerson (1981) for the B-form dodecamer d(CGCGAATTCGCG). The groove hydration was observed in the central AATT region of the dodecamer where a string of well-ordered waters forms a single spine of hydration which is hydrogen-bonded to nitrogen and oxygen atoms lining the narrow minor-groove, which is very narrow in this region. As the groove widens at the CGCG ends, the regular arrangement of waters is disrupted. The recent structure of the B-form decamer d(CCAACGTTGG) (Privé *et al.*, (1991) was shown to have a spine of hydration in the centre of the minor-groove as well as two ribbons at the end of the helix where the groove widens. As mentioned above, it has been proposed, for the original d(CGCGAATTCGCG) dodecamer structure, that the water spine was specific to the AATT sequence, and that if the sequence included a guanine residue then the presence of the N₂ groups of the guanine base would break the spine. Recent results, based on molecular dynamics calculations, demonstrate that spines can exist in regions containing guanine bases (see Berman (1991) and references within). In A-DNA structures, where the major-groove is deep and narrow whilst the minor-groove is shallow and wide (Saenger (1984)), some sequence specific hydration is observed in the major-groove. A variety of hydration patterns are observed in the A-form major-groove, including strings of waters linking the phosphates on adjacent strands and fused pentagons involving water oxygen atoms and the functional groups of the bases. These

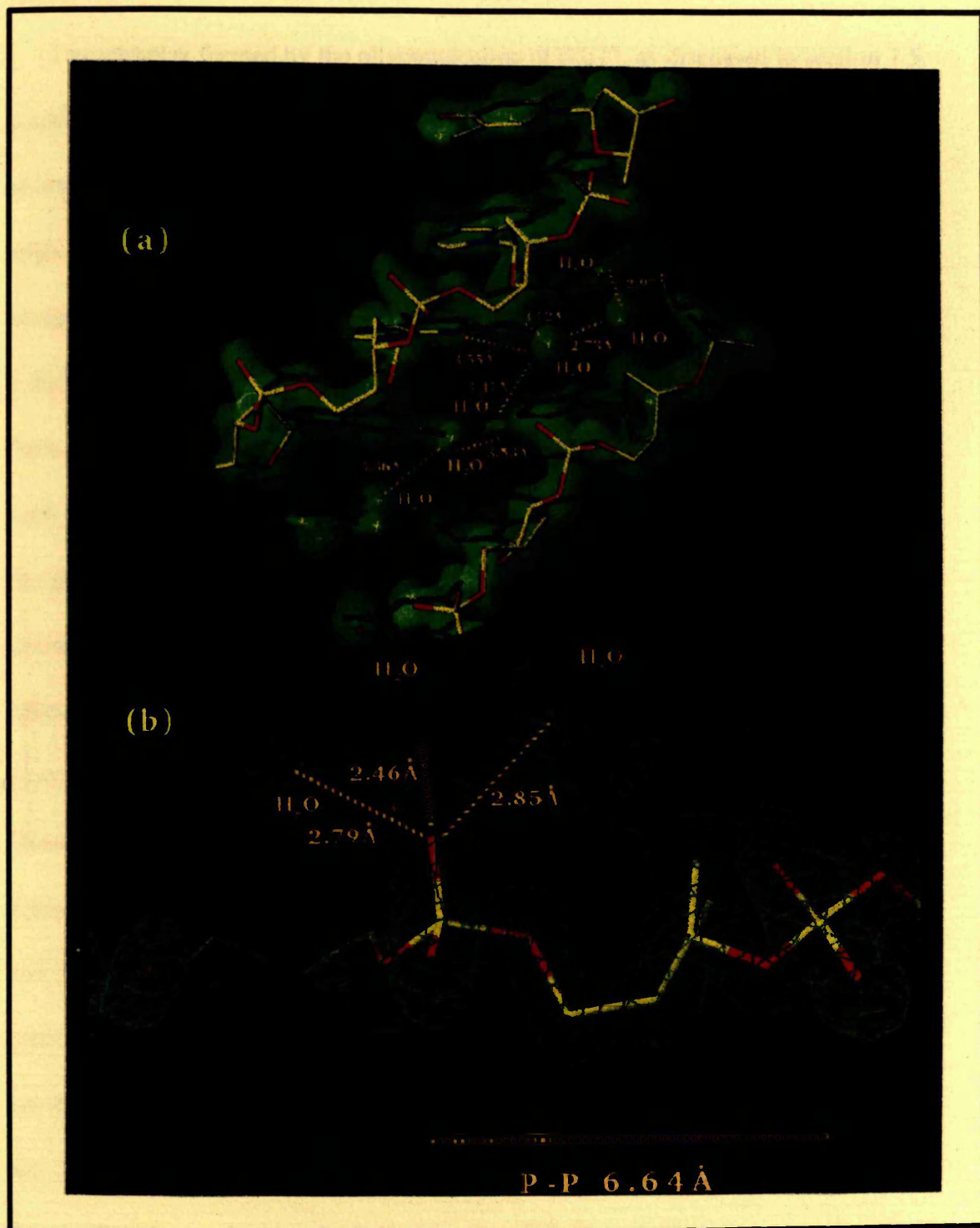


Figure 7.14: Hydration of the Tetraplex

patterns are covered extensively in Saenger (1984, 1987), Westhof (1988), Kennard and Hunter (1989, 1991) and Berman (1991, 1994).

The tetraplex formed by the oligonucleotide d(TG₄T), as discussed in section 7.5, consists of a parallel orientation of strands which generate four equivalent grooves of essentially minor-groove character, containing the N₂, N₃ and C₈ atoms. The edge of the guanine bases that would define the major-groove in duplex DNA lies on the interior of the tetraplex. The relatively narrow grooves, which vary between 2.30Å and 3.25Å in width, are favourable binding sites for water molecules, which form hydrogen-bonds with the exposed N₂ amino groups, the heterocyclic N₃ and each other, forming a well-organized hydration network. The spine of hydration in the groove is illustrated in Figure 7.14(a). Two adjacent strands are shown in the default colouring, with the (3F_o - 2F_c) electron density, contoured at 1.5σ, shown in green in the semi-transparent form. The spine of hydration runs up the centre of the groove, with the oxygen atoms depicted as yellow asterisks with their density in green. Distances between water molecules and possible hydrogen-bonds are shown as broken lines. The minimum and maximum distance for the N...O hydrogen-bonds are 2.74Å and 3.07Å, respectively. In Figure 7.14(a), possible hydrogen-bonds between waters and the exposed heterocyclic N₃ atoms (3.55Å) and the N₂ amino groups (3.12Å) are indicated. The minimum and maximum distances for the O...O hydrogen-bonds are 2.60Å and 3.05Å, respectively, and water-water distances are depicted in the figure. The observed groove hydration for the d(TG₄T) tetraplex is supported by recent molecular dynamics results (reviewed in Berman (1991)). Perhaps the major factor in the formation of the tetraplex hydration is not the base sequence but the relatively

narrow width of the grooves. Chuprina *et al.*, (1991), in their recent molecular dynamics study, have proposed that the width of the groove is more important than the sequence in determining the architecture of the water around it.

The separation between adjacent phosphates on the same strand in the tetraplex structure is characteristic of B-form DNA, with an average value in the region of $6.63 \pm 0.14 \text{ \AA}$, as discussed in section 7.5. The majority of the sugar-puckers, likewise, adopt southern conformations as in canonical B-DNA structures, as detailed in section 7.4. The hydration of the tetraplex backbone therefore corresponds to the pattern commonly observed in B-form DNA, whereby the water molecules cluster around each individual phosphate, but form virtually no water bridges between adjacent phosphates on the same strand. An example of this pattern is shown in Figure 7.14(b). A segment of a single sugar-phosphate backbone is depicted in the default colouring. The electron density, in green, is represented in the conventional chickenwire. No waters are found bridging the neighbouring phosphates which, in the figure, have a separation of 6.64 \AA . Three water molecules are shown coordinating one of the phosphate oxygens of the phosphate group on the left of the figure, with O...O coordination distances of 2.79 \AA , 2.46 \AA and 2.85 \AA , well within the theoretical expectations. This pattern of backbone hydration is observed along the length of each individual tetraplex strand.

7.10 Stacking of the Guanine Tetrad Planes

In addition to the base-base associations via hydrogen-bonding, the stability of nucleic acid helical structures is also generated by the formation of base-stacking interactions.

This normally takes place such that base planes are stacked at the van der Waals distance of approximately 3.4 Å and parallel to each other. This arrangement arises as a result of vertical rather than horizontal interactions, and recurring stacking patterns in various structural studies suggest that base–stacking interactions are specific (Saenger (1984)). In particular, it is frequently found that the polar substituents -NH_2 , =N- or =O of one base is superimposed over the aromatic system of the adjacent base, and that this manner of stacking specificity even overrides hydrogen–bonding effects. Overall, a survey of stacking patterns in base, nucleoside and nucleotide crystal structures (Bugg *et al.*, (1971)) suggested that forces between permanent dipoles are only of minor importance for the stabilisation of base stacks. Rather it appears that dipole–induced dipole interactions play the major role, with the permanent dipole, predominantly in C=O or C-NH_2 groups, superimposed over the π –electronic system of the adjacent base. Purine–purine stacks are the most stable, and this is a key factor in the stability of aggregates of guanine and its derivatives, including the d(TG₄T) tetraplex structure.

In addition to permanent dipoles, π –electron systems and dipole–induced dipole moments, another operative force influencing base–stacking interactions is hydrophobic bonding. This is particularly relevant in the case of macromolecular nucleic acid structures where the encouraged interaction between nucleic acid bases significantly contributes to secondary and tertiary structure stabilisation. Another contributory influence in conveying stacking specificity is London dispersion forces, which arises from the instantaneous asymmetry of the electronic charge distributions within atomic groups. As a result, instantaneous dipoles created in one group of atoms polarize the electronic system of neighbouring atoms or molecules, thus inducing parallel dipoles

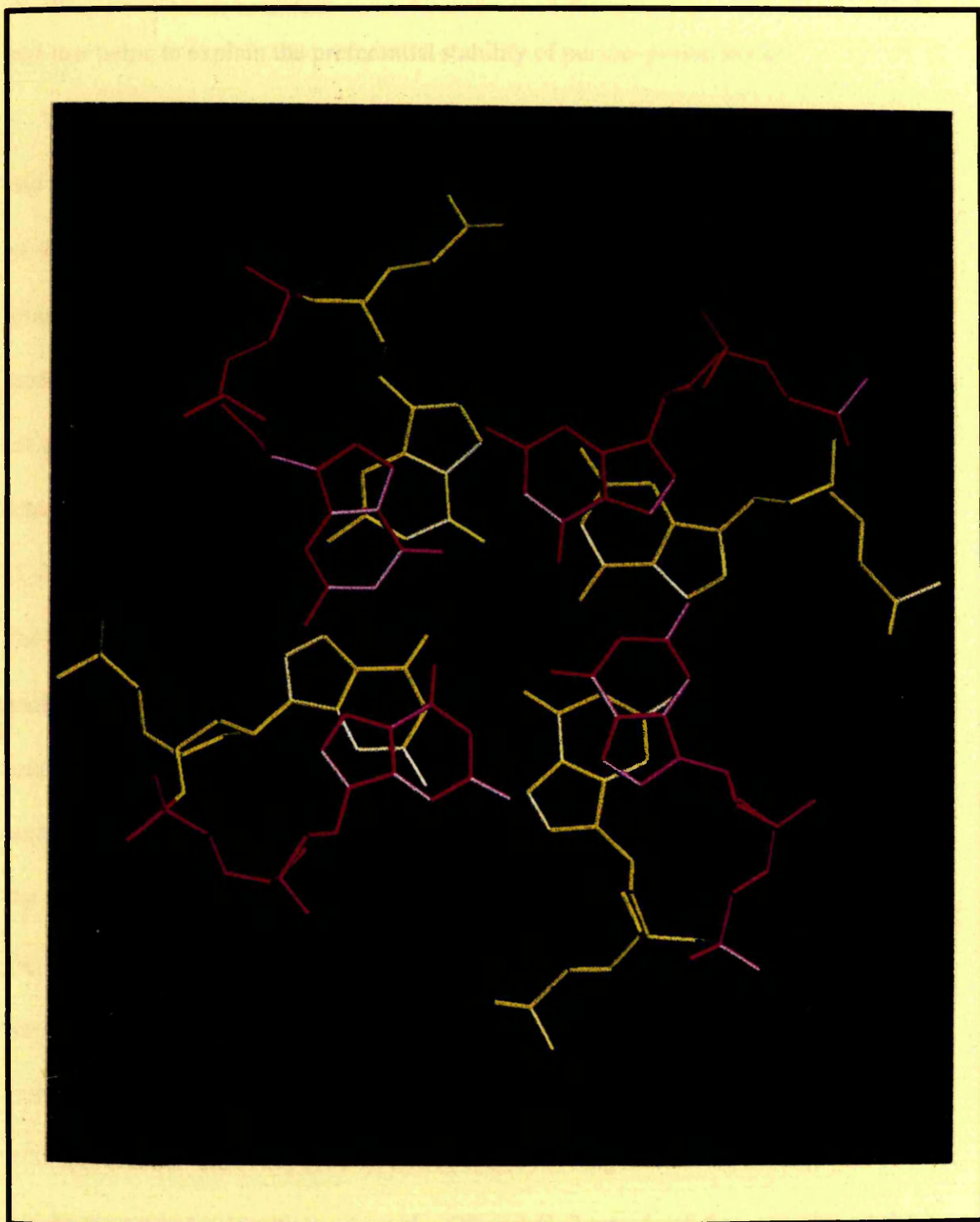


Figure 7.15: Guanine Tetrad Stacking Within a Tetraplex

which attract each other. London dispersion and permanent dipoles combine and result in significant effects which are more pronounced in purine than in pyrimidine bases, and this helps to explain the preferential stability of purine–purine stacks.

Stacking interactions involving the three thymines which coordinate the axial water and the G5 guanine tetrad have been discussed in section 7.8. Two distinct patterns of stacking arrangement involving guanine tetrads stacks are observed in the crystal structure, one intramolecular and the other intermolecular. The intramolecular stacking arrangement is illustrated in Figure 7.15. This arrangement is found for the adjacent tetrad planes G2 and G3, G3 and G4 and G4 and G5. Figure 7.15 depicts the guanine tetrad stacking within one tetraplex, at tetrads G3 and G4. The G3 tetrad, in yellow, is situated in the figure below the G4 tetrad, in purple. For all the tetrad steps within the four individual tetraplexes in the crystal unit cell the adjacent guanine bases stack such that N₂ of one is sandwiched between the N₇ and N₃ of the tetrads above and below. Since the protons of N₂ are electropositive but N₇ and N₃ are electronegative, it seems likely that the favourable overlap of partial charges contributes to the stability of the stacking conformation. In this arrangement there is little in the way of overlap of the aromatic rings. This pattern of stacking is in contrast to that observed in foldback tetraplex structures (Kang *et al.*, (1992); Smith and Feigon (1992); Macaya *et al.*, (1993); Wang *et al.*, (1993)), and is consistent with the NMR structural studies of d(TG₄T) (Aboul-ela *et al.*, (1994)) and d(T₂G₄T) (Wang and Patel (1993a)).

At the tetraplex interface where the G2 and G–2 tetrads of the coaxially stacked tetraplex pairs (A, B) and (C, D) meet, a very different stacking arrangement is apparent. This is illustrated in Figure 7.16, where the G2 and G–2 tetrads are coloured yellow and

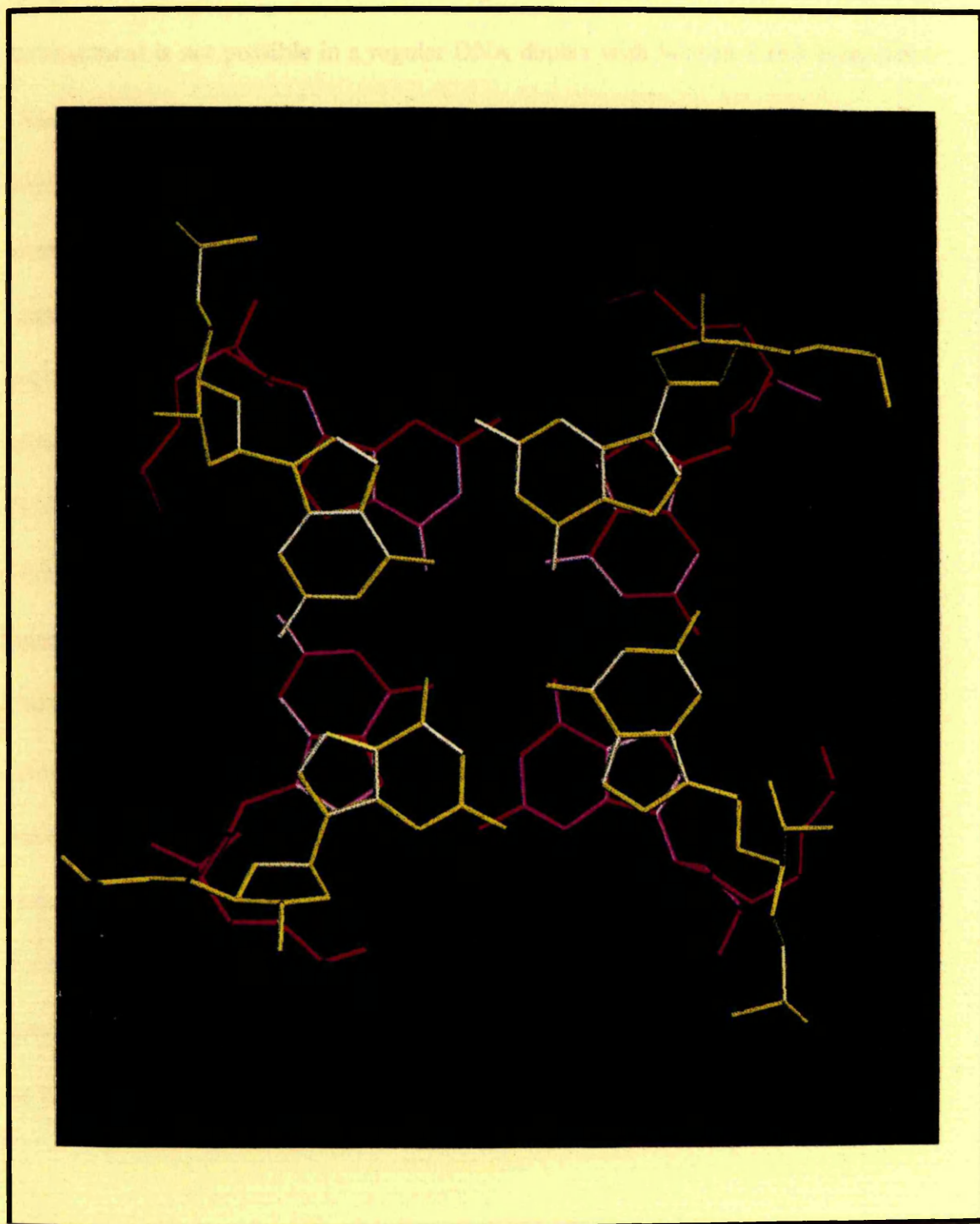


Figure 7.16: Stacking at the Interface Between Tetraplexes

purple. In this arrangement there is extensive overlap of the five-membered rings of the guanine bases, but the six-membered rings have no effective overlap. This stacking arrangement is not possible in a regular DNA duplex with Watson–Crick base-pairs (Saenger (1984)). However, an analogous interaction occurs at positions where the glycosyl bond changes from *syn* to *anti* for the outer two guanine quartets in the anti-parallel tetraplex structures (Smith and Feigon (1993)). Despite the fact that there appear to be unfavourable overlap of the electronegative O₆ atoms at the interior of the tetrads, and the electronegative N₂ atoms at the exterior of the tetrads, the fact that the tetraplexes studied here selected this distinctive stacking arrangement might suggest that it is energetically preferred. The principles governing the stability of stacking arrangements, outlined above and discussed in more detail in Saenger (1984), do not readily explain the stability conferred by the overlap of the five-membered rings, where the vertex constituting an atomic position of one base lies midway above the bond of another base. It seems likely that the interaction is hydrophobic and involves favourable interaction of dipoles and the π -electron system of the two rings. Quantum chemical calculations may shed some light on the matter, but empirical ultra-high-resolution studies of the tetraplex, leading to electron deformation density (EDD) analysis, if possible, might provide a clear experimental explanation. EDD studies are discussed in Chapter 8, section 8.2.2.

7.11 Arrangements of the Thymine Nucleotides

The high crystalline order of the tetraplex structure may be attributed to favourable lattice contacts. In addition to the 5' end-to-end stacking of tetraplex pairs within the cell, a variety of longitudinal and lateral contacts involving the thymine residues contribute to this stability. These contacts involve 27 of the 32 thymine residues which ought to occupy the unit cell. As discussed in Chapter 6, even after anisotropic refinement, there was no clear density for the five missing thymine bases, presumably because of high thermal disorder, and these were excluded from the final model. The five missing thymines of the structure are distributed such that one is absent from the 3' end of tetraplexes A, B, C and D. The fifth residue is missing from the 5' end of tetraplex A.

Longitudinal contacts are made with molecules in the neighbouring cells, resulting in the formation of a pseudocontinuous helix that runs along the length of the crystal. This interdigitation of tetraplexes in adjacent crystal cells at the 3' ends, resulting in the coordination of the axial waters of the four tetraplexes and the formation of intermolecular T•T base-pairs, has been discussed in section 7.8. These arrangements account for 11 of the 27 observed thymine residues. The remaining 16 thymine residues are described and discussed in this section. Some of the thymine nucleotides are conformationally disordered, as also discussed in section 7.8, although multiple conformations have not yet been modelled for the structure. For all of the thymine nucleotides discussed, none of them form a well-defined thymine tetrad of the kind found in the NMR study of r(UG₄U) (Cheong and Moore (1992)).

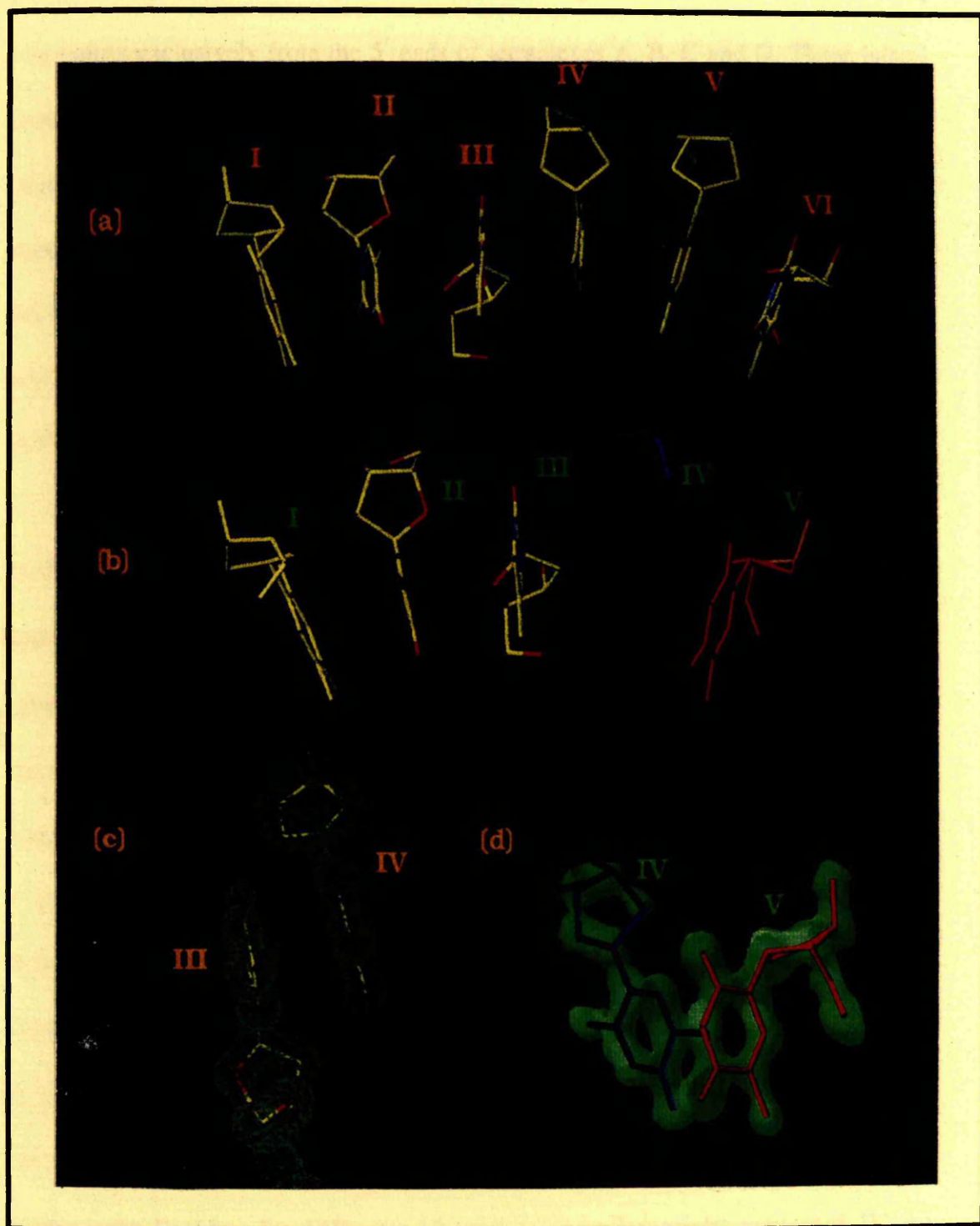


Figure 7.17: The Lateral Stacking of the 5' Thymine Nucleotides

In addition to the longitudinal interactions already described, there are lateral interactions between the tetraplexes that are made by the stacking of thymine bases originating exclusively from the 5' ends of tetraplexes A, B, C and D. These lateral interactions are such that, at the *bc* face between symmetry related tetraplexes, thymine residues for the unit cell and symmetry related cells stack in two arrangements, the appearance of which is rather analogous to a pile of coins. One arrangement contains six thymine bases, whilst the other consists of five. The stacking thymine residues, which click together rather like two locks between the adjacent unit cells, can be viewed in Figure 7.3, and are illustrated in more detail in Figure 7.17.

Figure 7.17(a) depicts the thymine residues of the six-membered stack, and Figure 7.17(b) the residues of the five-membered stack. In Figure 7.17(a) the residues II, III and VI, depicted in the default colours, originated from the same unit cell, whilst the residues I, IV and V, coloured in yellow, originate from the neighbouring cell related to the first by a negative translation along the unit cell *a*-axis. Residues I and III extend from the A tetraplexes; residues IV and VI from the C tetraplexes; and residues II and V from the B and D tetraplexes. In Figure 7.17(b) the residues II and III, depicted in the default colours, originate from the same cell as residues II, III and VI in Figure 7.17(a), whilst residue I, depicted in yellow, originates from the same cell as residues I, IV and V in the same figure. Residue IV, coloured blue, and V, coloured red, originate from the adjacent cell related by combined unit cell axial translations of $+c$ and $-a$, and $+c$, respectively. Residues I and III extend from the C tetraplex, whilst residues II, IV and V extend from the D, B and A tetraplexes, respectively. Figure 7.17 thus describes an intricate intermeshing of symmetry related residues which clearly assist in establishing

the high crystalline order of the tetraplex structure. Figure 7.17(c) and (d) presents the residues III and IV from Figure 7.17(a), and residues IV and V from Figure 7.17(b), respectively. In each case the stacking thymines are shown in $(3F_o - 2F_c)$ electron density contoured at 1.5σ , using both the chickenwire and semi-transparent modes of representation of the density. The aromatic π -electron system for the pyrimidine bases are clearly well-defined, however, there is evidence of conformational disorder in the deoxyribose rings, as witnessed in the discontinuity of the density at this contouring level.

The six- and five-membered thymine stacks are stabilised by specific stacking arrangements, as discussed generally in section 7.10. Figure 7.18(a)–(e) illustrates the individual stacking arrangements of the members of the six thymine stack. For residues I and II, the N_1 of residue I and the N_3 of residue II lie within the aromatic π -electron system of the other base. The interaction of point charges is evident in the superposition of the electronegative N_3 of residue I and the electropositive C_2 of residue II. The residues II and III, shown in Figure 7.18(b), stack such that the N_1 of residue III and the methyl group of residue II appear over the π -electron system, but there is no apparent superposition of electrostatic point charges. The residues III and IV stack in such a way that the N_3 atom and $C_4=O_4$ keto substituent of residue III, and the $C_4=O_4$ keto substituent of residue IV, lie over the π -electron system. In addition, the electropositive C_2 of residue III is superimposed upon the electronegative C_5 of residue IV. In Figure 7.18(d), the IV and V bases stack such that there is extensive overlap of the aromatic rings, and the electropositive C_2 and electronegative N_3 of residue IV are superimposed upon the electronegative N_3 and electropositive C_4 of residue V,

approximately equally, in Figure 7.18, the large positive charge on the C4 atom of the thymine molecule is superimposed upon the C4 atom and methyl substituents of molecule V.

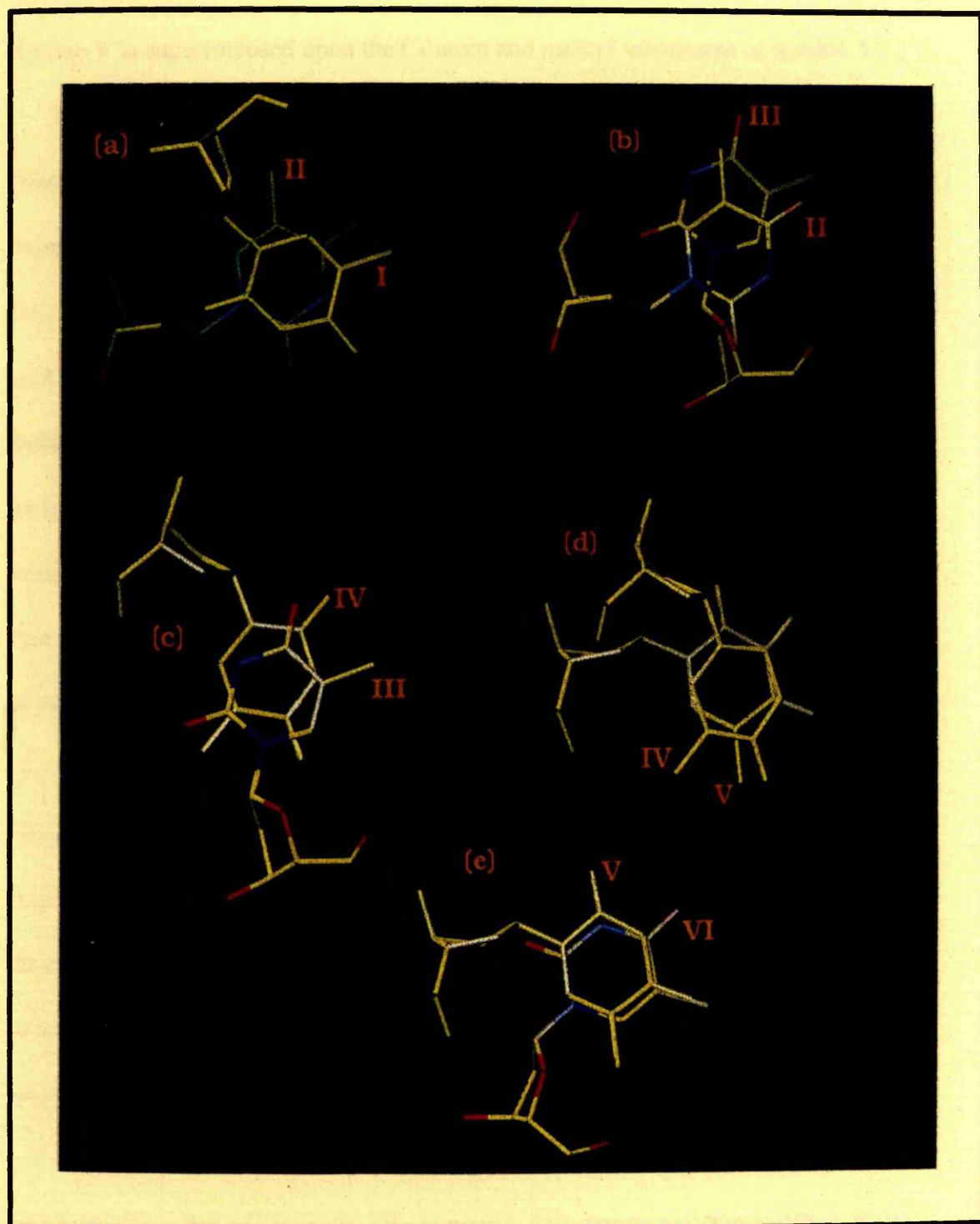


Figure 7.18: Stacking Patterns of the Six-Membered Thymine Stack

respectively. Finally, in Figure 7.18(e), the bases overlap thoroughly such that there are extensive point charge interactions. In particular, the $C_4=O_4$ keto substituent of residue V is superimposed upon the C_5 atom and methyl substituent of residue VI.

The stacking arrangements of the thymines of the five-membered stack are illustrated in Figure 7.19(a)–(d). The six-membered rings of residues I and II are extensively overlapped in a manner similar to the five-membered rings of the G2 and G–2 tetrads, discussed in section 7.10 and illustrated in Figure 7.16. The rings overlap such that the vertex constituting an atomic position lies approximately midway along a bond of the other base. It appears likely, as indicated earlier, that this interaction may be hydrophobic and involve some favourable interaction of dipoles and the π -electron system of the two rings. Residue II and III overlap in a manner practically identical to that observed in Figure 7.17(b). In this case the weakly electronegative methyl carbon of residue II appears superimposed above the electropositive C_4 ring atom of residue III. In Figure 7.19(c), there is little overlap of the aromatic ring systems of residues III and IV. The $C_4=O_4$ keto substituent of residue III lies over the ring of residue IV, with the electropositive C_4 atom over the electronegative C_5 atom of residue IV. The residues IV and V overlap in a manner similar to that already seen in Figure 7.18(e), although, in this case, there is not such an apparent superposition of electrostatic point charges.

The discussion thus far has accounted for 22 of the 27 thymine residues included in the final model of the unit cell contents. The remaining five residues make a number of same cell and symmetry related intermolecular contacts, largely involving the coordination of phosphate oxygens by means of $N_3-H \cdots O$ hydrogen-bonds, within



Figure 7.19: Stacking Patterns of the Five-Membered Thymine Stack

the allowed range of 2.74Å to 3.07Å. In Figure 7.2, it can be seen that two of the thymine residues make intermolecular contacts with tetraplexes of the same cell in the channel separating coaxially stacked tetraplex pairs. A 5' residue of tetraplex B bends back to coordinate tetraplex D, whilst a 5' residue of tetraplex D extends to coordinate tetraplex A. Both thymines behave in a similar manner. The B tetraplex residue is illustrated in Figure 7.20(a), where the molecular model is shown in the chickenwire electron density. The 5' residue actually coordinates a phosphate group of its own tetraplex with a $N_3 \cdots O$ distance of 2.94Å, although it can be seen by the distance indicated that the D tetraplex is only marginally further away. A number of base–water distances, of approximately 3Å magnitude, are also indicated in the figure.

In addition to the T•T base–pairs already discussed in section 7.8, another mismatch, in this case of the G•T variety, is also observed in the structure. This intermolecular mismatch is formed by a 3' thymine residue of the D tetraplex not involved in the coordination of the axial waters, and is illustrated in Figure 7.20(b). The thymine bases reaches across the *ac* face of the unit cell to make a symmetry contact with a guanine of the A tetraplex in the cell related by a positive axial translation along the *b*-axis of the cell. The guanine residue in question is shown in red, and the $O_2 \cdots H-N_2$ (2.95Å) and $N_3-H \cdots N_3$ (2.98Å) hydrogen–bonds, where the thymine atoms are quoted first, are depicted by broken lines. The electron density for the thymine base is poorer than that observed for the other thymine residues illustrated in this chapter, particularly with regard to the density for the glycosyl linkage, which is discontinuous at 1.5 σ contouring. This evidence of conformational disorder, and the hydrogen–bonding distances, indicate that this intermolecular mismatch is relatively weak. The

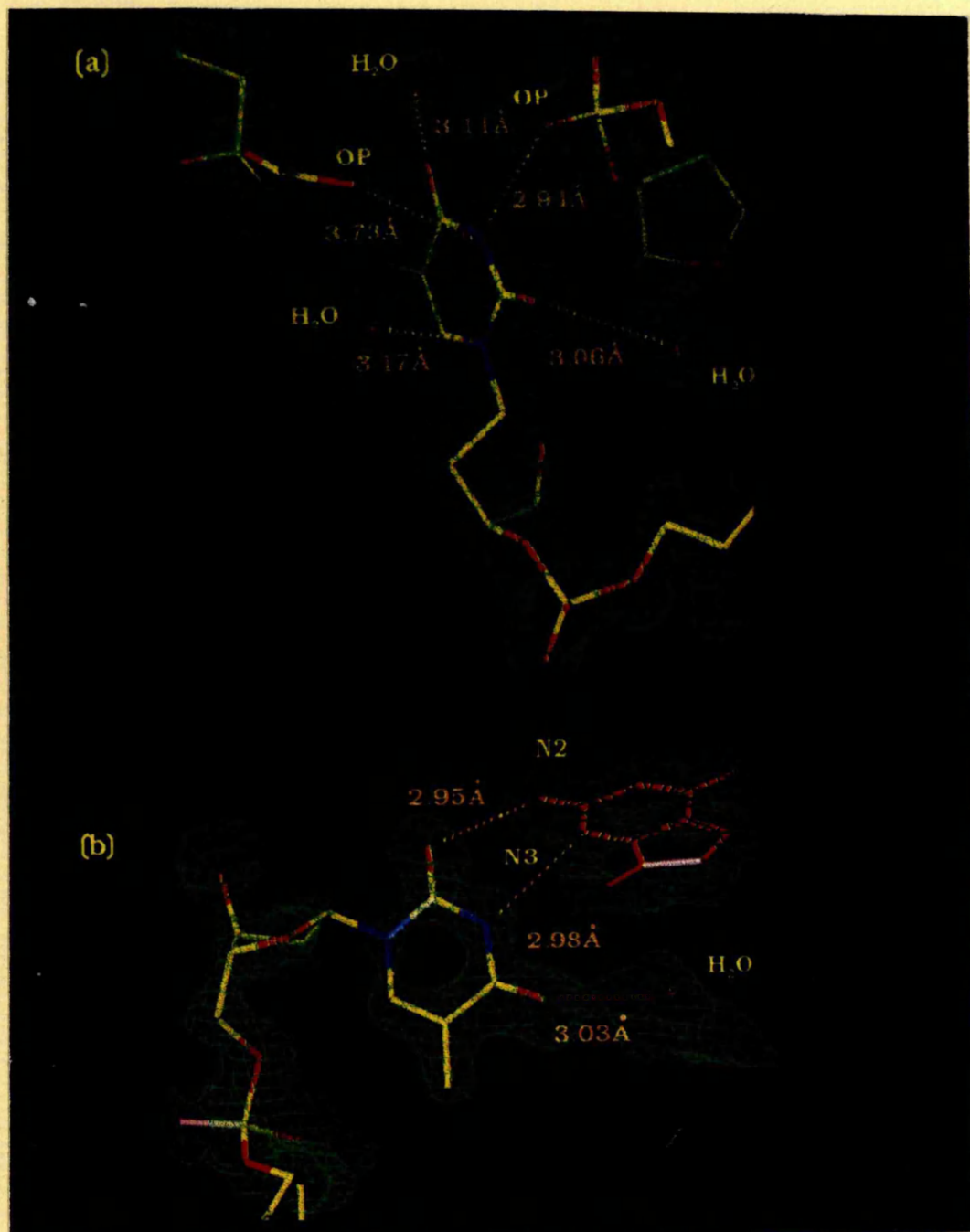


Figure 7.20: Other Intermolecular Thymine Contacts – I

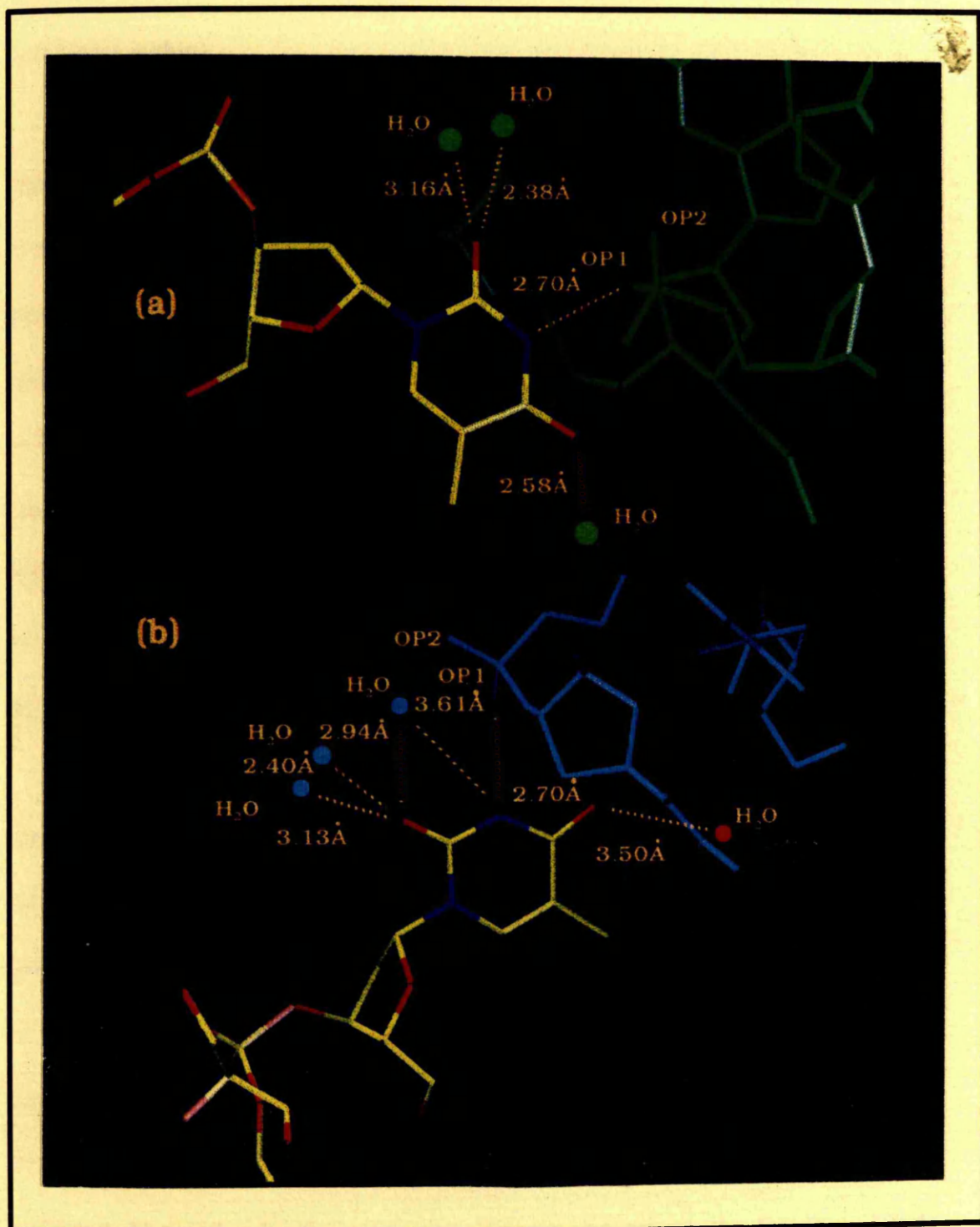


Figure 7.21: Other Intermolecular Thymine Contacts – II

purine–pyrimidine G•T pairing is not as rare as the T•T pairing also encountered in the structure, and has been observed in a number of other cases (see Hunter (1992) and references within).

The remaining two thymine residues make symmetry related contacts involving the coordination of phosphate group oxygens. In Figure 7.21(a), a 5' residue from the D tetraplex, depicted in the default colours, is shown making a $N_3-H \cdots O$ hydrogen-bonding contact with a phosphate oxygen (O_{P1}) from the B tetraplex in the adjacent cell related by a positive translation along the c -axis of the unit cell. In Figure 7.21(b), a 5' residue from the B tetraplex, again displayed in the default colours, is shown making a similar $N_3-H \cdots O$ hydrogen-bonding contact with the C tetraplex, coloured blue, from a neighbouring cell related by a negative translation along the c -axis of the crystal. Water molecules are represented in Figure 7.21 by green, blue and red dots, and a number of base–water distances are indicated in the figure.

It is perhaps interesting to note a resemblance between the manner of packing of the d(TG₄T) tetraplexes in the unit cell and the concept of quasi-equivalence. The idea of quasi-equivalence was conceived by Caspar and Klug (1962) in order to explain the packing of protein subunits to form the spherical shells of icosahedral viruses. In order to construct the virion shells with all of the protein units identical, it was necessary to drop the strict mathematical requirement that all contacts be equivalent, whilst retaining the essential physical principle. The basic assumption, therefore, was that the virus shell is held together by the same type of bonds throughout, but that these bonds may be deformed in slightly different ways in the different, non-symmetry related environments. In other words, the contacts are quasi-equivalent, within a tolerated

degree of deformation.

This concept is analogous to what occurs within the tetraplex crystal cell. The unit cell contains, approximately, four identical guanine tetraplexes which are held together by a variety of intermolecular thymine nucleotide contacts. The conformations adopted by the 3' and 5' thymine nucleotides are very similar in most cases, but are not identical, and vary significantly depending on the contacts being made. This can be appreciated by reference to Figures 7.2, 7.3, 7.10 and 7.17. The ability of the thymine residues to adopt these similar but distinct intermolecular contacts yields the non-crystallographic symmetry. This illustrates one of the distinct advantages of the structure crystallizing without the imposition of the exact translational symmetry constraints, for the terminal residues can search unhindered for the optimal intermolecular symmetry contacts which thus establish the high degree of crystal order.

7.12 Location and Environment of the Calcium Cations

Nucleic acids contain four different potential sites for the binding of metal ions. These are the negatively charged phosphate oxygen atoms, the ribose hydroxyls (in RNA but not DNA), the base ring nitrogens and the exocyclic base keto groups (Saenger (1984)). Phosphate groups can, in general, bind to a variety of metal ions belonging to the alkaline, alkaline earth and transition metal series to form salt-like complexes between positively charged metal ions and negatively charged phosphate oxygens. The base ring nitrogens and the sugar hydroxyls are not relevant to this discussion, but are discussed in detail in Saenger (1984). It has already been demonstrated in section 7.8 of

this chapter how the base keto substituents of the guanines of the G₄ tetrads, which are not commonly involved in metal complexation, coordinate the alkaline metal sodium cation along the length of the interior of the coaxially stacked tetraplexes.

Calcium ions were included in the crystallization liquor. As discussed in Chapter 2, divalent cations, in particular Mg²⁺, have been known for some time to be involved in the stabilisation of nucleic acid structure and also of their packing in the crystal lattice (Dock-Bregeon and Moras (1992)). The crystals of the d(G₄T₄G₄) anti-parallel quadruplex were grown in the presence of Mg²⁺, although ion positions and environment were not recorded in the paper of Kang *et al.*, (1992).

A total of nine well-defined Ca²⁺ cations were located in the crystal structure of the d(TG₄T) parallel-stranded tetraplex. Eight of the calcium cations are depicted in Figure 7.22(a). The (3F_o - 2F_c) electron density in this figure is contoured at 5σ such that the density for the sodium and calcium cations is visible in green. The perspective in Figure 7.22(a) is such that the view is down the helical axis of the two coaxial tetraplex pairs occupying the unit cell. The two sodium ions visible in the figure are those which occur between the G2 and G-2 tetrads, as illustrated in Figure 7.11. Two guanine G2 tetrads, one from each of the coaxial pairs, are visible in the figure.

In Figure 7.22(a), four of the calcium cations lie in a plane coplanar with a sodium ion at the centre of the plane for each of the tetraplex pairs. These eight calcium ions mediate bridging contacts between the tetraplexes A and B, and C and D. These contacts incorporate the phosphate oxygens of the tetraplexes and thereby assist in stabilising the crystal packing arrangement. A ninth calcium ion, not shown in Figure 7.22(a), can be viewed to the left of tetraplex A, midway down the length of the tetraplex, in Figure

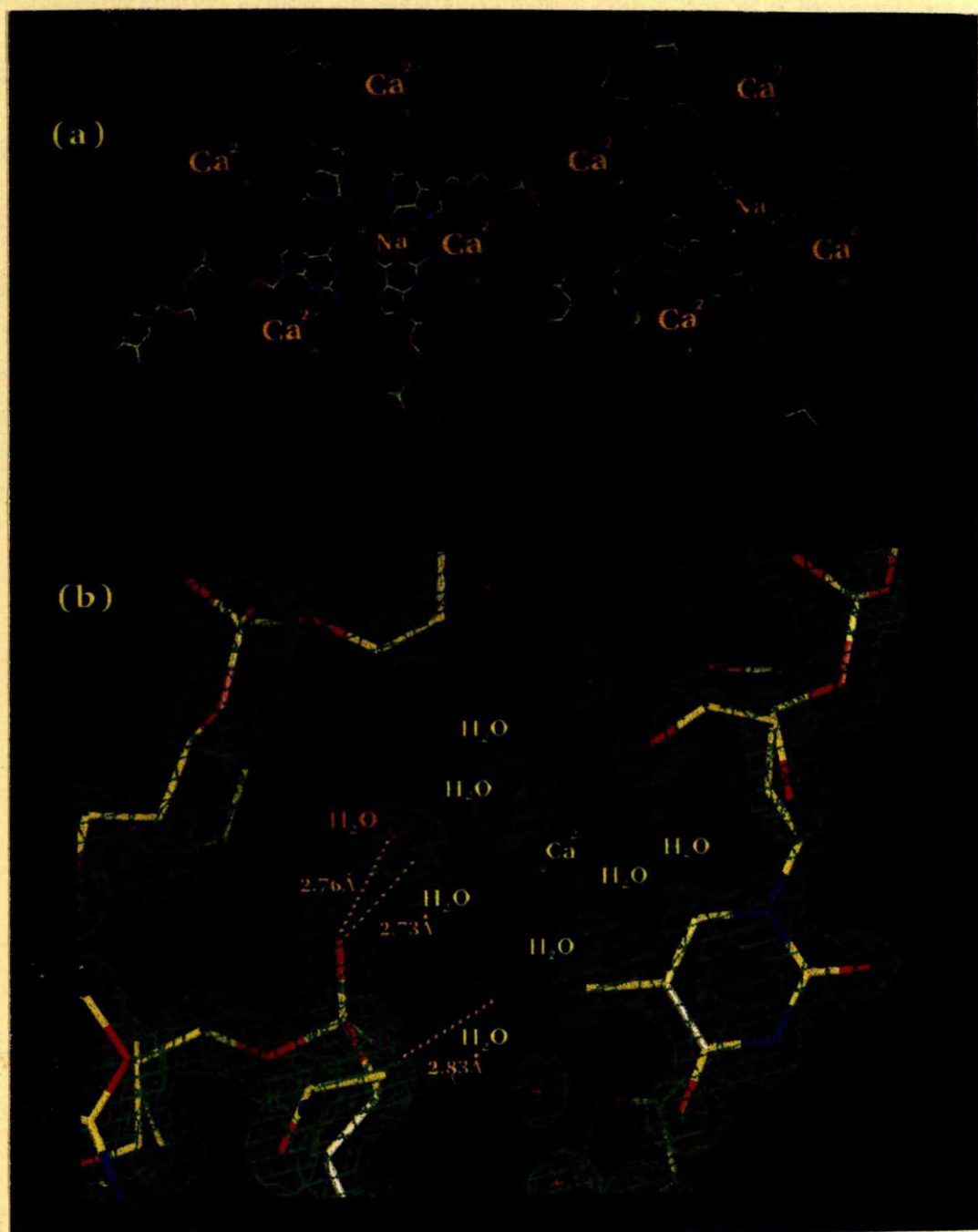


Figure 7.22: Location and Environment of the Ca^{2+} Cations

7.2. This additional calcium cation would appear to contribute to the non-equivalence of the two tetraplex pairs. Considering that DNA crystal structures ought to contain as many positive charges as there are negatively charged phosphates, all of the counterions are clearly not accounted for in the final structure model. This situation is not unusual and, therefore, it is highly likely that a number of the water molecules modelled in the structure could in fact be salt ions. No evidence was found in the electron density for the polyamine spermine which was also include in the crystallization liquor.

Ca ²⁺ Ion	B(Å ²)	Atomic Coordinates			Ca ²⁺ ··· O(H ₂ O)						
		(X)	(Y)	(Z)	1	2	3	4	5	6	7
1	7.92	8.559	12.881	16.638	2.33	2.39	2.39	2.41	2.43	2.45	2.50
2	9.44	16.576	15.251	27.462	2.42	2.42	2.44	2.44	2.46	2.46	2.47
3	7.65	-2.589	15.023	24.371	2.39	2.39	2.40	2.42	2.45	2.48	2.52
4	7.82	5.494	17.319	35.170	2.37	2.39	2.41	2.43	2.46	2.46	2.47
5	9.52	1.889	4.567	-10.761	2.37	2.40	2.42	2.43	2.47	2.50	2.50
6	9.22	9.830	6.771	0.152	2.39	2.40	2.40	2.41	2.48	2.49	2.56
7	8.36	-9.298	6.793	-3.078	2.36	2.39	2.40	2.40	2.42	2.47	2.48
8	7.07	-1.360	8.831	7.836	2.38	2.40	2.43	2.43	2.45	2.47	2.53
9	11.17	-3.351	-5.092	-11.538	2.40	2.42	2.43	2.46	2.48	2.48	2.52

Table 7.16: Ca²⁺ Location, Temperature Factor and Ca²⁺ ··· O(H₂O) Distances

Divalent cations have been observed in a number of cases to stabilise the packing of crystal structures (reviewed in Kennard and Hunter (1989, 1991); Joshua-Tor and Sussman (1993)). This effect has been observed in tRNAs, as well as in A- and B-helical structures, and especially in structures of Z-DNA oligonucleotides at high-

resolution. In these structural examples, one Mg^{2+} cation is typically coordinated by six oxygen atoms, one of which is a phosphate oxygen of the backbone, the others belonging to water molecules. A number of sites also form intermolecular contacts. The calcium ions in the d(TG₄T) tetraplex structure do not, as might be expected, participate in direct contacts with the phosphate backbone, but rather interact with the backbone by means of water molecules. Each of the nine observed calcium cations possesses a sharply defined hydration sphere consisting of seven waters. The $\text{Ca}^{2+} \cdots \text{O}$ distances for the waters of the nine spheres vary between 2.33 Å and 2.56 Å. Table 7.16 summarizes the positions of the calcium cations, expressed in terms of their atomic coordinates in the final model, their individual temperature factors and the $\text{Ca}^{2+} \cdots \text{O}$ distances for all of the 63 individual waters making up the nine spheres.

An example of one of the hydrated calcium cations is given in Figure 7.22(b). The nucleic acid is painted in the default colouring scheme, whilst the electron density is rendered in the conventional chickenwire form in green. The calcium cation, and the seven oxygen atoms of the waters constituting the hydration sphere, are labelled in yellow. Two of the waters are depicted making hydrogen-bonding contacts with the oxygens of the phosphate group in the foreground of the figure. The $\text{O} \cdots \text{O}$ distances, presented as broken lines, are 2.73 Å and 2.83 Å, comfortably within the expected limits for the $\text{O} \cdots \text{O}$ hydrogen-bond. An additional water molecule, which does not belong to the calcium cation hydration sphere, is labelled in orange. This water molecule also coordinates one of the oxygens of the forementioned phosphate group with a $\text{O} \cdots \text{O}$ hydrogen-bonding distance of 2.76 Å.

This concludes the description and discussion of the experimental work undertaken

in the derivation of the crystal structure of the parallel-stranded tetraplex formed by the oligonucleotide d(TG₄T) in the presence of Na⁺ cations, and the description of the structure itself. A general discussion of the contents of this thesis, as well as a speculative presentation of future projects stemming from the results of this work, is presented in Chapter 8.

If the theory of cybernetics is by itself to oust all metaphysical concepts – including the concepts of soul, of life, of value, of choice, of memory – which until recently served to separate the machine from man, it must conserve the notion of writing, trace, *grammè*[written mark], or grapheme, until its own historico-metaphysical character is also exposed. Even before being determined as human (with all the distinctive characteristics that have always been attributed to man and the entire system of significations that they imply) or nonhuman, the *grammè* – or the *grapheme* – would thus name the element. An element without simplicity. An element, whether it is understood as the medium or as the irreducible atom, of the arche-synthesis in general, of what one must forbid oneself to define within the system of oppositions of metaphysics, of what consequently one should not even call *experience* in general, that is to say the origin of *meaning* in general.

Jacques Derrida, *Of Grammatology*

Whenever I attempt to “analyse” a text which has given me pleasure, it is not my “subjectivity” I encounter but my “individuality”, the given which makes my body separate from other bodies and appropriates its suffering or its pleasure: it is my body of bliss I encounter. And this body of bliss is also *my historical subject*; for it is at the conclusion of a very complex process of biographical, historical, sociological, neurotic elements (education, social class, childhood configurations, etc.) that I control the contradictory interplay of (cultural) pleasure and (non-cultural) bliss, and that I write myself as a subject at present out of place, arriving too soon or too late (this *too* designating neither regret, fault, nor bad luck, but merely calling for a *non-site*): anachronic subject, adrift.

Roland Barthes, *The Pleasure of the Text*

Please note carefully that Moses did not include among God’s Ten Commandments: ‘Thou shalt not lie!’ That’s no accident! Because the one who says, ‘Don’t lie!’ has first to say, ‘Answer!’ and God did not give anyone the right to demand an answer from others. ‘Don’t lie!’ ‘Tell the truth!’ are words which we must never say to another person in so far as we consider him our equal. Perhaps only God has that right, but He has no reason to resort to it since He knows everything and does not need our answers.

Milan Kundera, *Immortality*

My propositions serve as elucidations in the following way: anyone who understands me eventually recognizes them as nonsensical, when he has used them – as steps – to climb up beyond them. (He must, so to speak, throw away the ladder after he has climbed up it.)

Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*

Chapter 8

Discussion

In this chapter it is my intention to summarize the reasons for this project, the results obtained and ideas for future research work relating to the structural study, by means of X-ray crystallography, of parallel-stranded guanine tetraplexes.

8.1 Overview of the Thesis Contents

Chapter 1, in addition to providing a brief summary of the key principles governing nucleic acid structure, presents a review of the current understanding of the structures of DNA quadruplexes formed on the basis of the cyclic guanine tetrad. Guanine base-tetrad formation was first proposed to explain the self-ordered helical structures formed by guanylic acid (Gellert *et al.*, (1962)), and was later invoked to explain the unusual structures formed by other guanosine containing polymers and mononucleotides (Tougard *et al.*, (1973); Arnott *et al.*, (1974); Zimmerman *et al.*, (1975); Zimmerman (1976); Mariani *et al.*, (1976)). Interest in four-stranded DNA structures adopted by sequences rich in guanine bases was rekindled in recent years by the dis-

covery of higher-order structures in *Oxytricha* telomere sequences (Oka and Thomas (1987)). It was shown that *Oxytricha* macronuclear, telomeric DNA could, when purified, slowly associate *in vitro* under high salt conditions without protein present. Studies on the *Tetrahymena* telomere sequence d(T₂G₄)₄ (Henderson *et al.*, (1987)) presented evidence of the formation of an unusual conformation *in vitro*. These experiments probed the nature of the unusual structures formed using gel migration, visible light spectroscopy, ³¹P and ¹H NMR, nuclease digestion and a variety of biochemical techniques.

In addition to chromosomal telomeres, sequences based on oligoguanine repeats are found at a number of elements that mediate recombination events. An example of these are immunoglobulin switch regions. Sen and Gilbert (1988) demonstrated, using complementary DNA strands, electrophoretic mobility studies and N₇ methylation utilising dimethyl sulphate, that sequences located in recombination hotspots for immunoglobulin diversity formed parallel-stranded quadruple helical structures *in vitro*. Sequences with a propensity for the formation of guanine tetraplex structures have been found to exist in a number of genes. It has also been suggested that guanine tetraplex formation could be the basis of the dimerization of the HIV RNA genome (Sundquist and Heaphy (1993)), whilst the oligonucleotide d(T₂G₄T₂) has been found to inhibit HIV viral-mediated cell fusion (Wyatt *et al.*, (1994)). Perhaps the most convincing evidence for a biological role for guanine tetraplex structures is the demonstration of selective binding by a variety of cellular proteins.

Two predominant classes of unusual structures are formed by guanine-rich DNA. Both are based upon tetraplex formation cohered by cyclic guanine tetrads. One class

of structures involves the folding back of a repetitive sequence, or the formation of a hairpin dimer, resulting in anti-parallel strands and the presence of both *anti* and *syn* glycosidic angles for the guanine nucleotides within a strand (Henderson *et al.*, (1990); Sundquist and Heaphy (1993); Sundquist and Klug (1989); Williamson *et al.*, (1989)). The second class of structures is formed by the association of four independent strands in which all of the glycosidic angles are thought to be *anti* (Sen and Gilbert (1988, 1990)). The parallel-stranded tetraplexes of the second class are similar to the conformations originally observed in fibre studies.

The formation of these structures depends on the presence of group I metal ions, in particular the cations sodium and potassium. A variety of studies have explored the roles that specific ions, base sequence and other factors could play in the formation of structures of unusual conformation based upon the cyclic guanine tetrad. In addition to biochemical, optical and calorimetric studies, high-resolution NMR investigations have revealed a wealth of detail with regard to structures of the foldback form, in particular the alternating pattern of *anti* and *syn* glycosidic angles within the individual guanine nucleotides on a given strand. The only crystallographic determination of a guanine quadruplex structure to date has been of the foldback class, formed by hairpin dimers of the *Oxytricha* telomeric sequence d(G₄T₄G₄) in the presence of the K⁺ ion, to 2.5 Å resolution (Kang *et al.*, (1992)). It has been demonstrated by NMR analysis that the oligonucleotide d(TG₄T) forms a parallel-stranded guanine tetraplex in solution (Aboul-ela *et al.*, (1992)). The high-resolution crystal structure of the parallel-stranded tetraplex of the oligonucleotide d(TG₄T), in the presence of the Na⁺ cation, offers the opportunity to study in detail the conformational characteristics of

this class of DNA structure. In addition, the exceptionally high-resolution of the study makes possible a clear insight into the position and coordination of the Na^+ cations essential for the tetraplex structure to form.

It is a necessary precondition for an X-ray crystallographic study of a macromolecule that suitable crystals are available. The search for crystals of the oligonucleotide d(TG₄T) formed the starting point of the experimental study. Considerable advances in synthetic chemistry have resulted in the plentiful supply of oligonucleotides for study in recent years. As a consequence of this, the number of high-resolution X-ray crystallographic studies of DNA oligonucleotides has grown rapidly in the last fifteen years. However, oligonucleotide sequences frequently fail to crystallize despite extensive efforts on the part of investigators. Prior to this project a number of other researchers have attempted, without success, to produce crystals of guanine-rich oligonucleotides with a view to undertaking X-ray crystallographic studies. This fact may be reflected in part by the observation that, previous to the structure of the d(TG₄T) tetraplex, only one other crystal structure of a guanine tetraplex has been reported (Kang *et al.*, (1992)).

The oligonucleotide 5' – TpGpGpGpGpT – 3' was synthesised by Dr. Alastair Murchie on a 1 μmol scale using β -cyanoethylphosphoramidite chemistry, whereafter it was purified utilising NENSORBTM columns. Tetraplex formation was induced by annealing the samples in an insulated waterbath at 70°C. The samples were then concentrated in the presence of Na^+ and quantified by UV absorption, ready for crystallization trials. As an additional precaution, the integrity of the samples was verified using MALDI-TOF mass spectrometry. The mass spectrometry analysis was

carried out by Dr. Uwe Piele.

Initially, crystallization trials were investigated at room temperature, with the crystallization trays stored in insulated boxes in order to regulate the environment of the trays. The conventional hanging drop method was employed in all crystallization trials. Trials were attempted using a variety of buffers and precipitating agents. Small crystals, not large enough for X-ray analysis, were produced at an early stage. However, in general, these crystals were commonly of a poor shape and usually redissolved. This observation may have been as a consequence of the relative instability of the tetraplex at this ambient temperature, with a dynamic equilibrium existing between the tetraplex and single-stranded DNA (Aboul-ela *et al.*, (1992)). It was observed, however, that the divalent Ca^{2+} cation was conducive to crystal formation, unlike other divalent cations investigated in trials. Furthermore, the polyamine spermine appeared to be essential in promoting crystal growth.

When facilities for storage at 4°C became available, attention was focused on trials at this temperature. The quality of the microcrystals visibly improved after this temperature transition. The alcohol MPD appeared to be the most effective precipitating agent, whilst the Ca^{2+} cation and the polyamine spermine both appeared to be essential for crystal growth. A pH range of 6.5–7.5 was optimal. The buffers Na–Cacodylate/HCl and Tris/HCl were used in the pH ranges 6.6–7.2 and 7.2–7.5, respectively, with little variation in the character of the crystals produced. Crystals remained too small for diffraction studies until NaCl was also introduced as a precipitating agent. Up until this point NaCl concentration in the starting drop was kept constant at 50 mM, with a proportional increase in NaCl concentration in the well. This measure produced larger

crystals which grew over a period of 2–3 weeks and diffracted to at least 2Å resolution, but were hopelessly twinned. The conditions were optimised by varying the concentration of NaCl in the starting drop in the range 130–180 mM, again proportionally increasing the concentration in the well. The resultant crystals grew in a triclinic habit within 2–3 weeks to a maximum size of 300x200x200 µm. Diffraction was observed to 1.6Å on a R–AXIS IIC phosphorimager area detector at the University of York. Twinned crystals, which still tended to occur, could, in some cases, be split to yield two crystals for diffraction studies. The crystals were characterised using 2° oscillation images and found to have the space group P1 with cell parameters $a = 28.76\text{\AA}$, $b = 35.47\text{\AA}$, $c = 56.77\text{\AA}$, $\alpha = 74.49^\circ$, $\beta = 77.64^\circ$, $\gamma = 89.73^\circ$.

Some manner of intrinsic twinning was characteristic of all crystals grown. Even crystals which appeared to be single exhibited characteristics of intrinsic twinning when diffracted, giving rise to diffraction spots which possessed neighbouring satellite peaks. The extent of this crystal defect varied and usually appeared to be rather minor. However, concern arose that the presence of satellite peaks in the diffraction pattern might lead to problems in the subsequent processing of the data. For this reason the decision was taken to collect the X-ray diffraction data using the R–AXIS IIC phosphorimager area detector in York University, since this allowed the use of the interactive processing package *DENZO* (Otwinowski (1993a; Gewirth (1993))). Since *DENZO* is an interactive program it allowed the satellite peaks to be ignored by permitting the user definition of the integration box for the diffraction spots. This facility proved to be useful in dealing with some of the more difficult diffraction data-sets, although Dr. Peter Moody demonstrated that the processing package *MOSFLM* (Leslie

(1990)) could handle data where the twinning problem was not too pronounced. Data collection on the R-Axis IIC was, however, fast and convenient nonetheless, with a dry-air crystal cooler, supplied by Dr. Peter Moody, available for the d(TG₄T) crystals grown at 4°C.

The native crystals diffracted very strongly and appeared to have a long life in the X-ray beam. X-ray diffraction data were collected using the oscillation method at 4°C. Data-sets were collected from two crystals with monochromated radiation from a Cu (1.543 Å) source and from one crystal using a Mo (0.711 Å) source. Using the copper radiation, 180° of data were collected for each crystal in a series of 90 2° oscillation images. Strong diffraction spots were observed to a resolution of 1.6 Å at the edge of the image plate. Inspection of the diffraction images revealed that certain important base-stacking reflections were overloaded. In order to compensate for this the same image was recollected using the same exposure time but with the generator at half-power. Once a molybdenum X-ray source became available, data were collected from a single crystal using 1.7° oscillations. Spots could be observed at the plate edge at a maximum resolution of 1.00 Å. A total of 173.5° of data were collected before the crystal perished dramatically. Whilst the merged data were 92.9% complete to 1.0 Å, the decision was taken to discard the data in the range 1.0–1.2 Å since the value for $\bar{I}/\sigma(\bar{I})$ fell below three within this range. The final, truncated native data-set was 97.7% complete in the resolution range 10.00–1.20 Å, containing 65,264 individual reflections, with 80.2% of the data with $F \geq 3\sigma(F)$ and 97.6% with $F \geq 2\sigma(F)$.

Once the native data had been collected, the native Patterson function and the self-rotation function could be analysed. The Patterson function contained two interesting

features which gave some indication of a likely packing arrangement within the cell. A strong peak was evident at approximately (0, 0, 0.5) at a height slightly less than half that of the origin. The presence of this peak led to initial fears that the crystal cell had perhaps been misindexed. The unit cell was therefore verified by use of the NRCVAX cell reduction program *CREDUC* (Gabe *et al.*, (1989)), and by mounting a crystal on a Rigaku AFC5R 4-circle diffractometer. In the later case, the data collection software was set to search for individual peaks and to subsequently calculate the smallest cell of the highest symmetry using the Delauney reduction procedure (AFC5R SOFTWARE PROGRAM). It was therefore concluded that the strong peak at approximately (0, 0, 0.5) arose from a noncrystallographic translation parallel with the *c*-axis of the cell. This translation also manifested itself as a systematic attenuation of intensities for reflections with odd-numbered *l* indices.

A series of ten vector streaks, characteristic of nucleic acid base-stacking within a helix, were also observed in the Patterson along the direction of the *v*-axis, thus indicating the direction of a helical axis. Self-rotation calculations revealed the presence of noncrystallographic two- and four-fold axes coincident at approximately 11° off the *b*-axis, thus coincident with the axis defining the orientation of the strong base-stacking transform. The combined observations gathered from the Patterson and self-rotation function analyses led to the proposal of a model for the unit cell contents in which two tetraplexes stacked end-to-end along a common axis, orientated at about 11° to the *b*-axis of the cell, related to another coaxially stacked pair by a noncrystallographic translation of approximately (0, 0, 0.5). Whilst these analyses, in particular the self-rotation function, indicated that a tetraplex was formed by the oligonucleotide

d(TG₄T) in the cell, they did not resolve the question of whether the tetraplex was parallel- or anti-parallel-stranded. The relative polarity of the stacked tetraplexes also remained unresolved. For this proposed model a total of twelve base-pair steps would be anticipated if the terminal thymine bases formed T₄ tetrads, although this seemed unlikely on the basis of NMR evaluations (Aboul-ela *et al.*, (1992); Wang and Patel (1993a)). The length of the unit cell *b*-axis (35.5 Å) is more representative of ten tetrads steps, and ten vector planes were observed in the native Patterson. This suggested that some intermeshing of the tetraplexes, involving the thymine nucleotides, was occurring.

Once the crystallization and data collection conditions had been optimised and the crystals properly characterised, the next step in the structure determination was the solution of the phase problem. Two possible avenues of investigation presented themselves for the d(TG₄T) crystal structure – the multiple isomorphous replacement and molecular replacement methods. Since an experimental search model for molecular replacement was not initially available, attention was directed first of all towards the search for heavy-atom derivative crystals. The majority of the time spent searching for derivatives was concentrated on obtaining crystals of the synthetic oligonucleotide d(U_{Br}G₄T). Initially, both d(U_{Br}G₄T) and d(U_IG₄T) derivative syntheses were screened, but attention was soon concentrated on the former, due principally to its relative abundance. The oligonucleotide d(TG₄T-O-PO₂=S) was also investigated, as it was anticipated that the thiol group might act as a site for ligation of mercury. All nucleotides were synthesised by Dr. Alastair Murchie and analysed using MALDI-TOF mass spectrometry by Dr. Uwe Piele. Intensive trials of the bromo-derivative were carried out at room temperature and 4°C under red light to minimise the risk of photolysis. Whilst

displaying a preference for cooler conditions and the presence of spermine and the Ca^{2+} cation, the bromo-derivative failed to yield crystals of the size and quality of the native material. The largest crystals that could be grown possessed a maximum dimension of $60\text{ }\mu\text{m}$ and exhibited a distinct hexagonal habit, thus indicating a different space group from the native form. Small, hexagonal crystals were also observed using the native material in the earlier room temperature trials. The $\text{d}(\text{TG}_4\text{T-O-PO}_2=\text{S})$ derivative did not yield crystalline material of any kind and mass spectrometry analysis revealed the oligonucleotide, the product of an exploratory synthesis, was contaminated. An interesting observation was the fact that the sulphur of the thiol group exchanged with oxygen when exposed to water, as had been suspected by Dr. Alastair Murchie.

Since the trials involving synthetic oligonucleotides had proved unsuccessful, the decision was taken to attempt a number of heavy-atom soaks when data collection time became available on the R-AXIS IIC at York University. Due to the observed stability of the native crystals, heavy-atom soaks were carried out at 4°C in the mother liquor. Soaks were attempted using the reagents K_2PtCl_4 , $\text{Pb}(\text{CH}_3\text{CO}_2)_2$, $\text{UO}_2(\text{CH}_3\text{CO}_2)_2$, EuCl_3 and CsCl_2 . The volatile liquids I_2 and MeI were also added to the capillaries of premounted crystals in York and allowed to diffuse before removal of the liquid and subsequent data collection.

It was hoped that the cesium cation might replace the Na^+ ions stabilising the tetraplex, but, probably due to the size of the cesium ion, the soaked crystals suffered heavy cracking. The best cesium soak diffraction data exhibited highly mosaic diffraction spots and could not be processed. The strontium ion has been shown to stabilise guanine tetraplexes better than either the K^+ or Na^+ cations (Chen (1992)), whilst elec-

trophoretic gel mobility experiments in the laboratory of Professor David Lilley have shown Sr^{2+} to be effective in tetraplex formation of d(TG₄T). This information was not available at the time of the derivative trials, and so strontium was not investigated as a possible derivative.

Data-sets on all the derivative soaks were collected using the R-AXIS IIC to 2.0Å, 1.8Å or 1.6Å. A total of 180° of data were collected for the $\text{Pb}(\text{CH}_3\text{CO}_2)_2$ and I_2 soaks, whilst 360° of data were collected for the K_2PtCl_4 and MeI soaks with a view to recording a possible anomalous diffraction signal in each case. The EuCl_3 soak exhibited weak diffraction and was severely radiation sensitive, whilst the $\text{UO}_2(\text{CH}_3\text{CO}_2)_2$ soak diffracted weakly, due largely to its small size, and perished after 40° of data had been collected. All of the derivative data-sets, with the marginal exception of the uranium derivative soak, scaled well with the native data, giving R_{iso} index values of about 0.1 or less. The calculated isomorphous and anomalous difference Patterson syntheses were all featureless, indicating that the derivative soaks had been fruitless. The calculation of difference density maps using the refined model for the unit cell contents and the forementioned derivative data-sets confirmed this observation.

The decision could have been made to further pursue the generation of heavy-atom derivatives for investigation. With regards to the purely synthetic oligonucleotides this may have involved further exploration of the crystallization conditions for the bromo-uracil synthesis, or a switch in emphasis towards the iodo-uracil synthesis or further phosphothiol derivatisations. Thought might also have been given to devising new syntheses, such as attaching the labelled uracil base to the 3' rather than the 5' end of the oligonucleotide. Heavy-atom soaking experiments, had they been explored, might

have focused on trials involving other heavy-atom reagents as well as testing new buffering conditions for the heavy-atom soaks themselves. However, a search model for molecular replacement study became available based upon the results of the NMR characterisation of the tetraplex in solution (Aboul-ela *et al.*, (1992)). Despite the fact that the NMR investigation indicated that the terminal thymines were disordered in solution, the initial test structure contained these residues modelled as T₄ tetrads. Since NMR study of r(UG₄U) had presented evidence for the formation of analogous U₄ tetrads (Cheong and Moore (1992)), it was therefore thought that thymine tetrads might possibly form within the crystal structure.

A number of parallel-stranded guanine tetraplex models were generated on the basis of the conformation of the original NMR model. These were then incrementally over- and underwound to generate a series of molecular structures. Molecular replacement cross-rotation and translation function studies were undertaken using the package *AMoRe* (Navaza (1992)). The optimal model yielded a set of highest peaks related by non-crystallographic symmetry and consistent with the packing models suggested by analysis of the Patterson and self-rotation functions. The results of this analysis illustrated that the individual tetraplexes were interpenetrating at the 5' ends within the crystal and at the 3' ends at the unit cell interface. In each case the thymine tetrads were completely clashing in the same area of space within the cell.

Further molecular replacement calculations were carried out using a search model derived by stripping out all of the thymine residues, since it now appeared likely that thymine tetrads were not occurring within the crystal structure. Again, translation function solutions were found nearly identical as those for the complete tetraplex

model, and consistent with the observations of the Patterson and self-rotation analyses. Rigid-body refinement computations using *AMoRe*, and then later *XPLOR* (Brünger (1993)), demonstrated that, without the packing restraint presented by the thymine nucleotides, the tetraplexes could move within the cell. This led to two similar models for the position and orientation of four tetraplexes within the cell, distinguished by the distance separating the tetraplexes in each of the two coaxial pairs, both of which were consistent with the results of the packing analysis. None of the packing models thus derived could be distinguished on the basis of the crystallographic residual and, therefore, it was apparent that further refinement would be necessary to resolve this ambiguity.

Following on from rigid-body refinement, the derived models for the unit cell contents were initially refined using simulated annealing with *XPLOR* (Brünger (1993)). In particular, the slow-cooling protocol (Brünger *et al.*, (1990)) was employed. In all cases only restraints to ensure base planarity were used. All of the packing models tested attained R-indices slightly greater than 0.30 at the culmination of the simulated annealing procedure. This procedure was typically repeated, in each case, before a minimum for the refinement was reached. Inspection of the $(3F_o - 2F_c)$ and $(F_o - F_c)$ electron density syntheses at this stage provided some evidence that the packing model which involved the tetraplex pairs within the cell forming a pseudocontinuous helix, with their 5' G₄ tetrads overlapping in both cases, was correct, as shown in Figure 5.12. However, at this resolution and value for the crystallographic residual, this observation was inconclusive due to the poor quality of the electron density observed. Furthermore, it was impossible to glean any information regarding the conformation of the thymine

nucleotides.

It was therefore necessary to further refine the packing models. At this stage the Automated Refinement Procedure, *ARP* (Lamzin and Wilson (1993)), was deployed. *ARP* reiteratively defines and updates macromolecular structure, utilizing other *CCP4* source code. The models were refined, in turn, using the unrestrained *ARP* procedure, in which the program updates all atoms in the structure with the application of unrestrained least-squares minimization, and restrained *ARP*, in which the program is coupled with the least-squares minimization programs *PROTIN* and *PROLSQ* (Hendrickson and Konnert (1980); Hendrickson (1985)). In the latter application, a modified stereochemical dictionary for nucleic acids was used. Since *ARP* depends on high-quality, high-resolution data, all of the data in the range 10.00–1.20 Å was used. The application of unrestrained *ARP* refinement quickly resolved the ambiguity in the tetraplex packing model, even when applied to the incorrect molecular replacement solution, confirming that the model in Figure 5.12 was correct. The definition and connectivity of the electron density from the unrestrained refinement procedure was extremely good. This allowed the modelling, following a number of iterative refinement applications, of 27 of the 32 thymine nucleotides, as well as 14 sodium cations and 9 calcium cations, both included in the crystallization liquor. When no more of the model could be located, the cell contents were reiteratively refined, with isotropic temperature factors, using the restrained *ARP* procedure, in which H₂O structure was defined and updated.

Further refinement, including anisotropic temperature factors and the introduction of riding hydrogen atoms, was then undertaken using *SHELXL-93* (Sheldrick (1993)).

The structure was refined, initially, using the CGLS algorithm and, finally, using block-matrix least-squares to facilitate the output of standard deviations on individual bond lengths and bond angles. The final structure has an overall deviation from ideality of 0.016 Å in bond lengths and 1.61° in bond angles. This model, which has an R-index of 12.4% for all the structure factor amplitude data from 10.0–1.2 Å, contains 2,453 non-hydrogen atoms, consisting of 4 d(TG₄T) parallel-stranded tetraplexes (i.e. 16 strands of the d(TG₄T) oligonucleotide), 9 calcium ions, 14 sodium ions and 514 water molecules. The free R-factor index (Brünger (1992)) for the CGLS anisotropic refinement was also calculated and was 0.1758 for 1 out of 10 of the reflections being excluded from the refinement procedure. Even after anisotropic refinement, there was no clear density for the five missing thymine bases, presumably because of high thermal disorder, and these were excluded from the final model.

In Chapter 7 the crystal structure of the parallel-stranded tetraplex formed by the oligonucleotide d(TG₄T) in the presence of sodium ions is reported. The hexanucleotide d(TG₄T) forms the anticipated four-stranded tetraplex. This tetraplex structure has been solved to one of the highest resolutions ever achieved for a nucleic acid. Four tetraplexes occupy the irreducible repeating unit of the crystal. The strands of the tetraplex form a right-handed helix and are aligned parallel to each other. The cell is composed of two pairs of tetraplexes, where members of a pair stack with opposite polarity. The spacing between successive guanine tetrads within a tetraplex is uniform. Furthermore, the same spacing is observed at the interface where stacked tetraplexes join, giving the impression of a continuous helix. Within the tetraplex, the strands associate by means of the expected hydrogen-bonding between N_{1H} and O₆, and

between N_{2H} and N₇. The guanine tetrads exhibit a high degree of planarity and are connected to the backbones by glycosyl bonds that are exclusively *anti*.

Most of the guanine sugars exhibit a B-DNA character, having a pseudorotation angle of $158.5 \pm 11.3^\circ$ and an average phosphate-phosphate separation of $6.63 \pm 0.14 \text{ \AA}$. However, at the tetraplex interface all of the G2 sugars, as defined in Figure 7.1, switch to a conformation resembling that of A-DNA, with an average pseudorotation angle of $45.6 \pm 8.0^\circ$ and a phosphate separation of $6.93 \pm 0.15 \text{ \AA}$. This switch is apparently required to avoid steric clash at the tightly packed interface. Elsewhere in the molecule, where the sugars are uniformly like B-DNA, the helical structure has a repeat corresponding to 10.4 tetrads per turn. However, RMS comparisons of the tetraplexes, on the basis of the guanine tetrads only, illustrates that all four tetraplexes are equivalently distinct. This suggests that the tetraplex is a flexible structure and that this flexibility is utilised in the packing of four tetraplexes in the dense unit cell. Apart from the discrepancy at the 5' end, these tetraplex structures are in reasonable agreement with that determined from the same molecule in solution (Aboul-ela *et al.*, (1992, 1994)).

At the stacking interface between tetraplexes there is extensive overlap of the five-membered rings of the guanine bases, but the six-membered rings have no effective overlap. An analogous interaction occurs at positions where the glycosyl bond changes from the *syn* to the *anti* conformation in anti-parallel tetraplex structures. That the tetraplexes studied here selected this distinctive stacking arrangement might suggest that it is energetically preferred. For all other tetrad steps within the tetraplex, the adjacent guanine bases stack such that the N₂ of one is sandwiched between the N₇ and N₃ of the tetrads above and below. As the hydrogens of N₂ are electropositive,

but N_7 and N_3 are electronegative, it seems likely that the favourable overlap of partial charges contributes to the stability of the stacking conformation.

Seven sodium ions were observed per stacked pair of tetraplexes, including one located at the tetraplex interface. Most of the ions are located between consecutive tetrads and are coordinated by the tetrad's eight buried O_6 atoms. The sodium ion that is located between the tetrad planes of the stacked coaxial pairs satisfies a nearly perfect inverted-bipyramid coordination geometry. As one moves towards the 3' of each tetraplex the ions between tetrads are displaced further towards the 3' tetrad, until there is no sodium ion located between the final two tetrad planes since, at the 3' end of the tetraplex, the sodium ions actually lie in the plane of the tetrad formed by the four G5 bases. This coplanar sodium ion is coordinated by the four O_6 atoms and by a water molecule that is bound at the fifth position. A similar pentacoordinate sodium geometry has been observed in crystallographic studies of small molecules (Karle (1974)). The axial water molecule forms hydrogen-bonds with three thymine bases. These bases are donated from the tetraplex itself and from two symmetry related tetraplexes. The thymine bases coordinate the water molecule, forming a slightly distorted tetrahedral geometry. The thymine from the tetraplex forms an unusual intermolecular T•T base-pair with one of the symmetry related thymines. Because of its large ionic radius, it is unlikely that potassium could adopt a similar position coplanar with a tetrad.

The parallel orientation of strands in the tetraplex generates four equivalent grooves of essentially minor-groove character, containing the N_2 , N_3 and C_8 atoms. The edge of the guanine bases that would define the major-groove in duplex DNA lie on the interior of the tetraplex. The relatively narrow grooves are favourable binding sites for

water molecules, which form hydrogen-bonds with the exposed N_2 amino groups, the heterocyclic N_3 and each other, forming a well-organized hydration network. Whereas in A-DNA the shorter distances between adjacent phosphates facilitates the formation of water bridges with the free phosphate oxygens, in B-DNA this distance is too large. This is clearly visible in the tetraplex structure where waters cluster around phosphate oxygens but do not form bridges between adjacent phosphates.

The high crystalline order of the tetraplex may be attributed to favourable lattice contacts. In addition to the 5' end-to-end stacking of the tetraplex pairs within the cell, there are also a variety of contacts made with tetraplexes of neighbouring cells, mediated by the thymine nucleotides. The tetrahedral coordination of the axial water by three nucleotides is one such example. There are lateral interactions between the tetraplexes that are made by the stacking of thymine bases originating from the 5' termini of several molecules. These produce two distinct stacks of thymines at the *bc* face of the unit cell. A six-thymine stack consists of three bases from one cell and three from the neighbouring cell. A five-thymine stack at the edge of the *bc* face consists of thymines from four cells which meet at this edge. Two of the thymine bases make cross-contacts between the coaxial pairs in the cell involving coordination of the phosphate oxygens. Another two make similar contacts but with symmetry related molecules. One thymine makes a symmetry related contacts that results in yet another mismatch, in this case an intermolecular G•T base-pair. Some thymine bases clearly appear conformationally disordered. No well-defined thymine tetrad of the kind formed by the uracil bases in the r(UG₄U) structure (Cheong and Moore (1992)) was observed.

Calcium ions were included in the crystallization solution and nine ions were located. Each of these ions is well-ordered and each possesses a well-defined hydration sphere consisting of seven water molecules. Eight of the calcium cations located form a plane with the sodium ions coaxially bound between stacked tetraplexes, and therefore appear to stabilise the stacking arrangement of the tetraplexes within the cell. The Ca^{2+} cations do not contact the phosphate groups of the DNA, as might be expected, but rather interact by means of the surrounding water molecules of the hydration spheres.

8.2 Future Work

Following on from the high-resolution determination of the crystal structure of the parallel-stranded tetraplex formed by the oligonucleotide d(TG₄T), a number of different avenues of further research possibly lie ahead. The discussion below presents a brief summary of a few of these likely projects, some of which are already underway.

8.2.1 Higher Resolution X-Ray Structural Studies

In the study presented in this thesis, the structure of the parallel-stranded guanine tetraplex formed by the oligonucleotide d(TG₄T) has been determined to an upper resolution limit of 1.2 Å. Thus this structural determination presents one of the highest resolutions ever achieved for a nucleic acid. Dickerson (1992) presents an extensive review of nucleic acid structures solved by X-ray crystallographic means. In this review, the highest resolution achieved for an A-helical oligonucleotide is 1.5 Å for the octamer ATGCGCAT, whilst the corresponding maximal resolution for a B-

helical oligonucleotide is 1.3Å for the decamer CCAAGATTGG which contains a G●A base-pair mismatch. Thus the structural determination of d(TG₄T) to 1.2Å resolution presents the highest resolution yet achieved for an oligonucleotide of either the A- or B-DNA morphology.

In fact, the only nucleic acid structures to have been determined to a higher resolution than the d(TG₄T) tetraplex belong to the left-handed Z-helical oligonucleotide family. Specifically speaking, these sequences are ^{5me}CGTA^{5me}CG, CGCGTG, CGC-GCG and CGCGCG (in presence of spermine), which have been determined to resolutions of 1.2Å, 1.0Å, 1.0Å and 0.9Å, respectively (see Dickerson (1992) and references within). Resolutions comparable to these Z-DNA sequences have also been attained with a couple of irregular helical oligonucleotide structures.

Experiments conducted at the Daresbury synchrotron by Dr. Ben Luisi have demonstrated that crystals of the d(TG₄T) parallel-stranded tetraplex can diffract to a resolution limit of at least 0.88Å at 4°C. Subsequent data collection experiments at the Hamburg imaging plate by Dr. Kathryn Phillips have succeeded in extending the resolution of the native data to an upper limit of 0.9Å at 100K. Thus, purely on the basis of 4°C temperature studies, the potential clearly exists to establish the d(TG₄T) tetraplex structure as not only the highest resolution structural determination of any oligonucleotide of the A- or B-form, but also as one of a handful of structures which stand at the current threshold of X-ray diffraction determination of nucleic acid structure.

Interest exists in possible experimental techniques which might assist in extending the resolution of the d(TG₄T) crystals even further. Perhaps the most obvious approach in this regard is the successful application of flash-freezing techniques to macromolec-

ular crystals. This technique involves the slowing down of the thermal motions of the atoms in the crystal structure by instantaneous freezing of the sample using liquid nitrogen, and can often lead to an extension of the observed diffraction as a result. A temperature of 100K, obtained by bathing the sample with a stream of cold nitrogen gas, is maintained. It is necessary that this temperature be maintained within a small range for the duration of the data collection. The trickiest aspect of this technique when applied to macromolecules is that the crystals must be kept ice-free, a difficulty exacerbated by the fact that macromolecular crystals contain a considerable volume of solvent. Commonly, it is necessary to transfer the crystal to a soaking medium containing a suitable antifreeze reagent such as glycerol, an operation which can often hamper the integrity of the crystal itself. Fortunately, the d(TG₄T) crystals are stable in a mother liquor which contains a relatively high concentration of the alcohol MPD which imparts antifreeze properties to the liquor, and thus may assist in the successful application of the flash-freezing technique to the tetraplex crystals.

Other techniques may be considered potentially useful in extending the resolution of the d(TG₄T) tetraplex crystals. A similar effect to the lowering of temperature during data collection is achieved when the molecules are held firmly in the crystal lattice by strong intermolecular forces such as hydrogen-bonds. As has been detailed in this study, the exceptionally high-resolution diffraction observed for the d(TG₄T) crystals can be attributed to an extensive arrangement of intermolecular contacts mediated by the thymine nucleotides. However, many of the observed thymine bases show signs of thermal disorder, whilst five of the 32 nucleotides could not be located in the final structure, presumably as a result of thermal disorder. If it were possible to somehow

optimise this network of intermolecular contacts then the observed diffraction limit for the crystals might improve by an appreciable order of magnitude. A possible approach to this problem might be the controlled heating and slow-cooling of a crystal, prior to a diffraction experiment, in the anticipation that the thymine bases might achieve a more stable energy minimum, improving the intermolecular contacts and thus the crystal diffraction properties. A combination of both forementioned techniques might well prove advantageous, particularly if stable conformations could be achieved for the five missing thymine nucleotides.

8.2.2 Electron Deformation Density Studies

It is, however, absolutely essential that the observed diffraction of the d(TG₄T) crystals be extended considerably if electron deformation density (EDD) studies of the tetraplex are to become viable in the future. The EDD represents the change in the electron density that results from the deformation of the spherical charge distribution around the free atoms when chemical bonds are made (reviewed in Angermund *et al.*, (1985)). Since it is the electrons in the valence shells that are mainly redistributed during the deformation, it is these which contribute most to the EDD.

The spherically symmetrical electron density distribution for a free atom can be accurately calculated by *ab initio* theoretical methods. The sum of all the electron densities of the spherical atoms placed at bonding distances from one another is known as the promolecule density, and the molecule so formed the promolecule. The difference between the experimentally determined total electron density, which describes the equilibrium state, and the promolecule density, which corresponds to a hypothetical

distribution of electrons, is the EDD. A positive deformation density at a position in a molecule signifies that there is an increase in electron density over and above the sum of the electron densities of the free, spherically symmetrical atoms. Conversely, a negative deformation density represents a reduction in the corresponding electron density. In other words, the EDD depicts the manner in which the valence shell electrons density distribution is polarised in the structure.

Electron deformation density studies of a nucleic acid, such as the d(TG₄T) tetraplex, would provide a novel experimental insight into the distribution of the electron density within a nucleic acid structure and the role it plays in determining the structure and stability of the morphology adopted. Experimental evidence of this type would provide unique empirical insights into the role of the polarization of the electron density in determining sugar–phosphate backbone conformation, base–pairing interactions and, in particular, base–stacking interactions, which play a key role in helical nucleic acid stability. Various theoretical models have been devised to account for the forementioned interactions (Saenger (1984)) but, due to the limits experienced in empirical investigations of structure, direct experimental evidence relating to the electron density polarity in nucleic acid helices has never yet been derived.

It is not unreasonable to propose the d(TG₄T) tetraplex structure as a possible candidate for EDD study of nucleic acids, although one has to admit a healthy degree of reservation. Whilst the objective in a conventional X-ray structure determination is to obtain as small a residual electron density as possible between the model and the observed diffraction data, the ultra–high–resolution studies required for EDD analysis demand much more rigorous precision. EDD experiments on small molecules typically

require data to a diffraction limit of 0.3\AA , and the aim is therefore to locate accurately all the atoms and to optimize all experimental parameters. Thus, the normal criteria for the estimation of the accuracy of an X-ray structure determination, as discussed in Chapter 6, are not sufficient. Therefore, an EDD study demands a much greater experimental and computational effort. Typically, symmetry equivalents need to be measured under different experimental conditions in order to eliminate instrumental errors and detect possible double reflections; high precision liquid nitrogen freezing is essential; crystals must be of excellent quality and have good morphology; all enclosing crystal faces must be indexed and measured to correct for absorption corrections; and, if the study is carried out in a capillary under an inert gas, an additional correction for the absorption of the glass is essential.

Clearly, the application of such criteria in the study of a macromolecule crystal presents great difficulty. Furthermore, it is not feasible to expect diffraction of the magnitude of 0.3\AA from nucleic acid crystals such as those obtained for the d(TG₄T) oligonucleotide. However, should methods, such as those previously mentioned, succeed in extending the diffraction of the d(TG₄T) crystals to a limit of approximately 0.7\AA then, with rigorous attention to experimental procedure, it is conceivable that valuable information may be derived from EDD studies. Such information could prove to be valuable and insightful in establishing new experimental evidence for understanding the principle forces governing nucleic acid structure.

8.2.3 Structural Studies of Ribonucleicacid Tetraplexes

The vast majority of studies on guanine tetraplex structures, including the one presented here, have concentrated on the formation of quadruplexes or tetraplexes by DNA oligonucleotides. Comparatively little work has been done on the formation of guanine tetrad structures by ribonucleicacid oligonucleotides. A notable exception to this is the work of Cheong and Moore (1992), discussed in Chapter 1, on the RNA analogue of $d(TG_4T)$, $r(UG_4U)$, using NMR methods.

In conjunction with attempts to derive crystals of the $d(TG_4T)$ tetraplex, some effort was invested in crystallization trials using the RNA oligonucleotide $r(UG_4U)$. As discussed briefly in Chapter 2, conditions were arrived at which allowed the growth of small crystals of the RNA sequence, but these were inadequate for X-ray diffraction study. Whilst efforts in pursuit of RNA crystals were suspended once the study of the $d(TG_4T)$ tetraplex became more intensive, this work has been continued by Dr. Ben Luisi and Dr. Kathryn Phillips. Working from the best conditions available, attempts to enlarge the crystals have proved successful, and it might be reasonably anticipated that crystals suitable for X-ray diffraction might soon be available.

The study of the RNA form of the tetraplex is an exciting prospect for a number of reasons. Primarily, interest resides in the fact that the RNA tetraplex adopts the A-helical nucleic acid morphology and not the B-helical morphology, as observed in the DNA tetraplex. Thus, a comparative study of how the adoption of the A-form affects such parameters as base-stacking overlaps and the location and coordination of the stabilising cation(s), relative to the B-form, is clearly of interest. Another compelling

structural motive for this study is the opportunity to study, by X-ray crystallography, the U₄ tetrad detected by Cheong and Moore (1992) in their original study. Such a tetrad, apparently requiring that each uracil base possess two hydrogen-bonds, one with each neighbouring base, has never been observed before and is therefore of considerable interest. Finally, studies of the RNA tetraplex may provide valuable information that could prove to be insightful in understanding the nature of the guanine tetrad structure that it has been proposed mediates dimerization of the RNA genome of HIV-1 (Sundquist and Heaphy (1993)), should this structure be parallel or anti-parallel stranded.

8.2.4 The Phosphorothioate Tetraplex

Phosphorothioate analogues of nucleotides are powerful tools to investigate enzyme mechanisms which involve reaction at phosphorous (Saenger (1984)). In these analogues one unesterified phosphate oxygen is replaced by sulphur and the resultant group is termed a phosphorothioate group. Whilst the phosphorous atom in a normal phosphate mono- or di-ester (as in DNA or RNA) is prochiral, the phosphorothioate phosphorous is chiral. Subsequently, unlike the normal phosphate group, reaction with the phosphorothioate group can reveal whether the reaction proceeds with inversion or conservation of the stereochemical configuration at the phosphorous.

The predominant structural concerns when dealing with thiophosphate groups are related to the nature of the sulphur-oxygen substitution. The P-S bond, with a length of 1.8Å, is approximately 0.3Å longer than the P-O bond length, whilst the van der Waals radii of sulphur, at 1.85Å, exceeds that of oxygen by 0.45Å. Both of these

features add with the result that the sulphur atom protrudes outwards by about 0.8Å in comparison to the oxygen atom. The electronegativity of sulphur and oxygen differ significantly with the consequence that, in the thiophosphate group, the negative charge is not equally distributed between the two atoms. In fact the negative charge resides preferentially on oxygen, leaving the sulphur almosts neutral and the P–S bond with a predominantly double–bond character.

In a recent study, Wyatt *et al.*, (1994) found that the phosphorothioate oligonucleotide T₂G₄T₂ was a powerful inhibitor of HIV infection *in vitro*. The oligonucleotide was selected by the combinatorial screening of a library of phosphorothioate oligonucleotides that contained all possible octanucleotide sequences. The ability of the oligonucleotide to inhibit cell–to–cell and virion–to–cell fusion and spread of HIV infection was determined using an acute HIV–1 infection assay which measures protection from HIV–induced cytopathic effects. Circular dichroism analysis revealed that the T₂G₄T₂ phosphorothioate oligonucleotide formed a parallel–stranded tetrameric guanosine–quartet backbone. The use of competition studies involving antisera specific for the protein gp120 determined that the tetraplex bound the protein, and that this binding specifically involved attachment to the V3 loop of the protein. The protein gp120 is the major envelope protein of the HIV virus and it is the binding of gp120 to its cellular receptor which initiates virus and cell membrane fusion. Whilst the mechanism for membrane attachment is poorly understood, it is known that the V3 loop, a hypervariable region which maintains a high percentage of positively charged amino acids across all HIV strains, is essential.

Wyatt and colleagues speculate in their paper that the basis of the inhibition of HIV

infection by the $T_2G_4T_2$ phosphorothioate tetraplex may lie in its presentation of a rigid, compact structure with a high thio-anionic charge density. The provision of a densely charged anionic surface may hence facilitate the observed strong interaction with the cationic V3 loop. The researchers speculate further that, although the V3 loop is a hypervariable region, the functional requirement for cationic residues in the V3 loop may limit the virus's ability to evolve resistance to dense polyanionic inhibitors.

The results of Wyatt *et al.*, (1992) and the derivation of the high-resolution crystal structure of the parallel-stranded tetraplex formed by the oligonucleotide d(TG₄T), reported here, suggest an exciting avenue of research in HIV therapeutics. Compounds derived from the G-quartet structural motif now appear to be potential candidates for use in anti-HIV chemotherapy. However, the structural mechanism for the inhibition of HIV infection may still require further elucidation, for it may turn out that the specific handedness of the phosphorothioate groups has a role to play in possible mechanisms. Wyatt *et al.*, (1992) in their proposition of an anionic scaffold do not address this problem. Whilst their proposition may turn out to be correct, especially given the hypervariable nature of the V3 loop region, the results remain incomplete in this regard at this time.

Clearly the resolution of this question requires a significant volume of investigation. Even assuming that the thymine nucleotides, being disordered, play no role in the inhibition of HIV transmission, this still leaves approximately $(2^5)^3$ distinct stereochemical isomers of the guanine-quartet tetraplex structure. Inclusion of the thymine residues results in an even greater number of distinct stereochemical isomers of the $T_2G_4T_2$ tetraplex structure. Investigation of individual stereoisomers may reveal exceptional

potency for certain stereoisomers with regards to the inhibition of HIV transmission and, perhaps, a reduced or possibly minor effectiveness for others. If this were to be the case then a modelling or crystallographic study of the relevant tetraplexes(s) binding to the V3 loop would provide a much more specific basis for rational drug design of therapeutic reagents. It is however likely, for the foreseeable future at least, that the method of screening and amplifying random sequences discussed above (so called "irrational" drug design), for its speed and effectiveness, will predominate in this particular area of medical research.

8.2.5 Tetraplex-Protein Complexes

Perhaps the most persuasive evidence for a biological role for guanine tetraplex structures is the demonstration of selective binding by a variety of cellular proteins, as discussed in Chapter 1. Research interest has focused in this area in recent years and at least two proteins have been identified, in recent reports, which specifically recognise parallel-stranded G-quartet structures.

The yeast *KEM1* gene product has been shown to be a nuclease specific for the parallel-stranded G4 tetraplex structure (Liu *et al.*, (1993); Liu and Gilbert (1994)). This protein specifically recognises the parallel-stranded tetraplex form and cleaves contiguous single-stranded regions at the preceeding 5' terminus. A role has been suggested for this nuclease activity and the guanine tetrad tetraplex DNA structure in meiosis (Liu and Gilbert (1994)). An abundant protein called *Tetrahymena* G4-binding protein has also been shown to specifically bind parallel-stranded G4-DNA (Schierer and Henderson (1994)). Following on from the determination of the parallel-stranded

G4-DNA tetraplex formed by d(TG₄T), the X-ray crystal structures of tetraplex-protein complexes formed with the above proteins constitutes the most obvious next step in tetraplex structural studies generally, and in investigating the possible biological functions of the parallel-stranded tetraplex structure in particular.

8.3 Conclusion

In summary, the tetraplex structure reported here has been solved to one of the highest resolutions ever achieved for a nucleic acid. This species is parallel-stranded, but the information on metal ion coordination and tetrad-tetrad interactions that has been obtained at this resolution may provide insight that is relevant to all guanine tetraplexes, including both parallel and anti-parallel foldback structures. Such structural detail will likely be important in the analysis of tetraplex-protein interactions and thus in understanding the biological function of these structures. In addition to tetraplex protein interactions, the structure offers a number of exciting possibilities for further research: ultra-high-resolution X-ray diffraction structural studies; comparisons with the RNA tetraplex form mediating HIV genome dimerization; and in understanding the structural basis of the potent inhibition of HIV infection conferred by the phosphorothioate parallel-stranded tetraplex.

When employed correctly, the Technology Game functions as a subtle form of psychological blitzkrieg. The types of technology presented are unimportant. What is important is the speed at which the new technologies supercede the old. This applies to consumerism on all levels, from individuals to whole governments. The technology referred to can vary from digital watches to super-cooled computer memories. From VCR's to PET scanners. From the newest in personal vibrators to satellite-based particle beam weapons.

As each wave of technology is released, it must be accompanied by a demand for new skills, new language. Consumers must constantly update their ways of thinking, always questioning their understanding of the world. Going back to old ways, old technology is forbidden. There is no past, no present, only an endless future of inadequacy.

It makes no difference if the technologies presented are beneficent or even functional. It is advisable, however, that they go fast, make a lot of noise and come in a variety of decorator colors.

Richard Kadrey, *Genocide*

Natural diversity . . . strengthened by cultural diversity, which allows mankind to better adapt to a variety of life conditions and to better use the resources of the world. In this area, however, we are now threatened with monotony and dullness. The extraordinary variety which humans have put into their beliefs, their customs and their institutions is dwindling every day. Whether people die out physically or become transformed under the influence of the model provided by industrial civilizations, many cultures are disappearing. If we do not want to live in a world covered with a single technological, pidgin-speaking, uniform way of life – that is, in a very boring world – we have to be careful. We have to use our imagination better.

François Jacob, *The Possible and the Actual*

The inferno of the living is not something that will be; if there is one, it is what is already here, the inferno where we live every day, that we form by being together. There are two ways to escape suffering it. The first is easy for many: accept the inferno and become such a part of it that you can no longer see it. The second is risky and demands constant vigilance and apprehension: seek and learn to recognize who and what, in the midst of the inferno, are not inferno, then make them endure, give them space.

Italo Calvino, *Invisible Cities*

Subjectivity is always schizophrenic in technoculture. Speaking subjectively of subjectivity, of course! It's always lived in a double sense. On the one hand your body is a processed world, processed as in sampler music – the language of aliasing, of condensation, of syncopation, of displacement, of speed-up and slow-down, all pretty much digitally recorded. That's the normal language by which we live in TV culture, in consumer culture, in our jobs and our music. Subjectivity now is fully ironic, fully ambivalent, fully paradoxical and contradictory. The technocratic specialist practices mechanical forgetfulness. That is, they manage to so engross themselves in data work that they lose sight of the ability to think deeply about what it means to be a human being and to engage in social relationships outside the imperatives of the technostucture. That's pure mechanical forgetfulness. Or the corporative elite engage in suicidal nihilism, which is to say that they make of their bodies what Nietzsche called a site of conscious experimentation, and they make of their minds a site of experimental cruelty, creating ribald games of mindfucking cruelty. As a teacher I have found that many of my students lack those options and in fact refuse those options and they're caught in these cruel situations. Some commit suicide, and others become slackers. The mood was caught very accurately in the film *Slacker*. They live in this kind of interzone. The "Vague Generation" is what the Montreal writer Michael Boyce calls them. They live ambivalently with a kind of passionate commitment to politics – yet in many ways, with a kind of indifference. They view themselves as powerless, but also try to create interesting forms of art, at the same time dealing with the inner reality of having their options closed down.

Vaguers really combine ambivalence, indifference, paradox and ironic immersion as a life ethic. They're bored but fascinated; anxious, but pretty well at peace with themselves. Montreal, where I come from, has the highest suicide rate in the Western world. A lot of students throw themselves onto train tracks. There is a suicide a week on the Montreal subways. Typically, it's young people that commit suicide and that bespeaks a sense of phenomenal depression – a sense of bleakness and closed-downness of experience. It's almost unimaginable. And there is no one helping because the technostucture just moves ahead. The powers that be just turn with complete indifference to this. Montreal is ahead of Germany. Germany used to be ahead. Ahead in the race to virtuality.

Arthur Kroker, *Codes of Privilege*, Mondo 2000, No.11

*posed little time
so I'll say
'good night'*

References

1. Arnold, R. M. (1974) The effect of the 1973-74 oil crisis on the U.S. economy.



References

- Aboul-ela, F., Murchie, A.I.H. and Lilley, D.M.J. (1992) NMR study of parallel-stranded tetraplex formation by the hexadeoxynucleotide d(TG₄T). *Nature* **360**, 280–282.
- Aboul-ela, F., Murchie, A.I.H., Norman, D.G. and Lilley, D.M.J. (1994) Solution Structure of a Parallel-stranded Tetraplex Formed by d(TG₄T) in the Presence of Sodium Ions by Nuclear Magnetic Resonance Spectroscopy. *J. Mol. Biol.* **243**, 458–471.
- Ahmed, S., Kintanar, A. and Henderson, E. (1994) Human telomeric C strand tetraplexes. *Nature Struct. Biol.* **1**, 83–88.
- Akman, S.A., Lingeman, R.G., Doroshov, J.H. and Smith, S.S. (1991) Quadruplex DNA Formation in a Region of the tRNA gene supF Associated with Hydrogen Peroxide Mediated Mutations. *Biochemistry* **30**, 8648–8653.
- Anderson, D.J., Lee, P., Levine, K.L., Sang, J., Shah, S.A., Yang, O.O., Shank, P.R. and Linial M.L. (1992) Molecular Cloning and Characterization of the RNA Packaging Defective Retrovirus SE21Q1b. *J. Virol.* **66**, 204–216.

- Angermund, K., Claus, K.H., Goddard, R. and Krüger, C. (1985) High-Resolution X-ray Crystallography – An Experimental Method for the Description of Chemical Bonds. *Angew Chem. Int. Edn. Engl.* **24**, 237–247.
- Arndt, U.W. (1986) X-ray Position-Sensitive Detectors. *J. Appl. Cryst.* **19**, 145–163.
- Arndt, U.W. and Wonacott, A.J., eds. (1977) *The Rotation Method in Crystallography*. North Holland, Amsterdam.
- Arnott, S., Chandrasekaran, R. and Marttila, C.M. (1974) Structures for Polyinosinic Acid and Polyguanylic Acid. *Biochem. J.* **141**, 537–543.
- Balagurumoorthy, P., Brahmachari, S.K., Mohanty, D., Bansai, M. and Sasisekharan, V. (1992) Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res.* **20**, 4061–4067.
- Ballard, J.G. (1985) Introduction to the French Edition of *Crash*(1974). In *Crash*, pp. 5–9. Triad/Panther Books, London.
- Bang, I. (1910) Untersuchungen über die Guanylsäure. *Biochem. Z.* **26**, 293–311.
- Barthes, R. (1975) *The Pleasure of the Text*. Hill and Wang, New York.
- Baudrillard, J. (1983) The Precession of Simulacra. In *Simulations*, pp. 1–80. Semiotext(e), New York.
- Baudrillard, J. (1990) *Cool Memories: 1980–1985*. Verso, London.

- Baudrillard, J. (1993) The Orders of Simulacra. In *Symbolic Exchange and Death*, pp. 50–86. Sage Publications, London.
- Benjamin, W. (1973) *Illuminations*. Fontana, London.
- Berman, H.M. (1991) Hydration of DNA. *Curr. Opin. Struct. Biol.* **1**, 423–427.
- Berman, H.M. (1994) Hydration of DNA: take 2. *Curr. Opin. Struct. Biol.* **4**, 345–350.
- Bernal, J.D. (1926) On the interpretation of X-ray, single crystal rotation photographs. *Proc. Roy. Soc. London A* **113**, 117–164.
- Biessmann, H., Carter, S.B. and Mason, J.M. (1990) Chromosome ends in *Drosophila* without telomeric DNA sequences. *Proc. Natl. Acad. Sci. USA* **87**, 1758–1761.
- Bieth, E., Gabus, C. and Darlix, J.-L. (1990) A study of the dimer formation of Rous sarcoma virus RNA and of its effect on viral protein synthesis *in vitro*. *Nucleic Acids Res.* **18**, 119–127.
- Blackburn, E.H. (1990) Telomeres and Their Synthesis. *Science* **249**, 489–490.
- Blackburn, E.H. and Gall, J.G. (1978) A Tandemly Repeated Sequence at the Termini of the Extrachromosomal Ribosomal RNA Genes in *Tetrahymena*. *J. Mol. Biol.* **120**, 33–53.
- Blackburn, E.H. and Szostak, J.W. (1984) The Molecular Structure of Centromeres and Telomeres. *Annu. Rev. Biochem.* **53**, 163–194.

- Blow, D.M. and Crick, F.H.C. (1959) The Treatment of Errors in the Isomorphous Replacement Method. *Acta Cryst.* **12**, 794–802.
- Blundell, T.L. and Johnson, L.N. (1976) *Protein Crystallography*. Academic Press, London.
- Borzo, M., Detellier, C., Laszlo, P. and Paris, A. (1980) ^1H , ^{23}Na and ^{31}P NMR Studies of the Self-Assembly of the 5' – Guanosine Monophosphate Dianion in Neutral Aqueous Solution in the Presence of Sodium Cations. *J. Am. Chem. Soc.* **102**, 1124–1134.
- Bouhoutsos–Brown, E., Marshall, C.L. and Pinnavaia, T.J. (1982) Structure-Directing Properties of Na^+ in the Solution Ordering of Guanosine 5'–Monophosphate. Stoichiometry of Aggregation, Binding to Ethidium, and Modes of Na^+ Complexation. *J. Am. Chem. Soc.* **104**, 6576–6584.
- Brown, T. and Kennard, O. (1992) Structural basis of DNA mutagenesis. *Curr. Opin. Struct. Biol.* **2**, 354–360.
- Brown, W.R.A. (1989) Molecular cloning of human telomeres in yeast. *Nature* **338**, 774–776
- Brünger, A.T. (1991) Simulated Annealing in Crystallography. *Annu. Rev. Phys. Chem.* **42**, 197–223.
- Brünger, A.T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475.

- Brünger, A.T. (1993) *XPLOR: A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, London.
- Brünger, A.T., Karplus, M. and Petsko, G.A. (1989) Crystallographic Refinement by Simulated Annealing: Application to Crambin. *Acta Cryst.* **A45**, 50–61.
- Brünger, A.T., Krukowski, A. and Erickson, J.W. (1990) Slow-Cooling Protocols for Crystallographic Refinement by Simulated Annealing. *Acta Cryst.* **A46**, 585–593.
- Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) Crystallographic *R* Factor Refinement by Molecular Dynamics. *Science* **235**, 458–460.
- Bugg, C.E., Thewalt, U.T. and Marsh, R.E. (1968) Base Stacking in Nucleic Acid Components: The Crystal Structures of Guanine, Guanosine and Inosine. *Biochem. Biophys. Res. Commun.* **33**, 436–440.
- Bugg, C.E., Thomas, J.M., Sundaralingam, M. and Rao, S.T. (1971) Stereochemistry of Nucleic Acids and Their Constituents. X. Solid-State Base-Stacking Patterns in Nucleic Acid Constituents and Polynucleotides. *Biopolymers* **10**, 175–219.
- Burroughs, W.S. (1987) *The Ticket That Exploded*. Paladin Books, London.
- Calladine, C.R. and Drew, H.R. (1992) *Understanding DNA: The Molecule & How It Works*. Academic Press, London.
- Calvino, I. (1979) *Invisible Cities*. Pan Books, London.

- Caspar, D.L.D. and Klug, A. (1962) Physical Principles in the Construction of Regular Viruses. In *Cold Spring Harbour Symposia on Quantitative Biology* (Frisch, L., ed), **XXVII**, pp. 1–24. Long Island Biological Association, Cold Spring Harbour, L. I., New York.
- Chantot, J.F. and Guschlbauer, W. (1972) Mechanism of gel formation by guanine nucleosides. *Jerus. Symp. Quant. Chem. Biochem.* **4**, 205–214.
- Chattopadhyaya, R., Ikuta, S., Grzeskowiak, K. and Dickerson, R.E. (1988) X-ray structure of a DNA hairpin molecule. *Nature* **334**, 175–179.
- Chen, F.-M. (1992) Sr^{2+} Facilitates Intermolecular G-Quadruplex Formation of Telomeric Sequences. *Biochemistry* **31**, 3769–3776.
- Cheong, C. and Moore, P.B. (1992) Solution Structure of an Unusually Stable RNA Tetraplex Containing G- and U-Quartet Structures. *Biochemistry* **31**, 8406–8414.
- Chevrier, B., Dock, A.C., Hartmann, B., Leng, M., Moras, D., Thuong, M.T. and Westhof, E. (1986) Solvation of the Left-handed Hexamer d(5BrC-G-5BrC-G-5BrC-G) in Crystals Grown at Two Temperatures. *J. Mol. Biol.* **188**, 707–719.
- Chuprina, V.P., Heinemann, U., Nurislamov, A.A., Zielenkiewicz, P., Dickerson, R.E. and Saenger, W. (1991) Molecular dynamics simulation of the hydration shell of a B-DNA decamer reveals two main types of minor-groove hydration depending on groove width. *Proc. Natl. Acad. Sci. USA* **88**, 593–597.
- Coffin, J.M. (1979) Structure, Replication, and Recombination of Retrovirus Genomes: Some Unifying Hypotheses. *J. Gen. Virol.* **42**, 1–26.

- Cram, D.J., Ho, S.P., Knobler, C.B., Maverick, E. and Trueblood, K.N. (1986) Host-Guest Complexation. 38. Cryptahemispherands and Their Complexes. *J. Am. Chem. Soc.* **108**, 2989–2998.
- Crick, F.H.C. (1989) *What Mad Pursuit*. Weidenfeld and Nicholson, London.
- Crick, F.H.C. and Magdoff, B.S. (1956) The Theory of the Method of Isomorphous Replacement for Protein Crystals. I. *Acta Cryst.* **9**, 901–908.
- Cross, S.H., Allshire, R.C., McKay, S.J., McGill, N.I. and Cooke, H.J. (1989) Cloning of human telomeres by complementation in yeast. *Nature* **338**, 771–774.
- Crowther, R.A. (1972) The Fast Rotation Function. In *The Molecular Replacement Method: A Collection of Papers on the Use of Non-Crystallographic Symmetry* (Rossman, M.G., ed), pp. 173–175. Gordon and Breach, New York.
- Crowther, R.A. and Blow, D.M. (1967) A Method of Positioning a Known Molecule in an Unknown Crystal Structure. *Acta Cryst.* **23**, 544–548.
- Darlix, J.-L., Gabus, C., Nugeyre, M.-T., Clavel, F. and Barré-Sinoussi, F. (1990) *Cis* Elements and *Trans*-acting Factors Involved in the RNA Dimerization of the Human Immunodeficiency Virus HIV-1. *J. Mol. Biol.* **216**, 689–699.
- Debord, G. (1977) *The Society of the Spectacle*. Black and White, Detroit.
- DeLillo, D. (1986) *White Noise*. Pan Books, London.
- Derrida, J. (1976) *Of Grammatology*. John Hopkins University Press, Baltimore.

- Diamond, R. (1971) A Real-Space Refinement Procedure for Proteins. *Acta Cryst.* **A27**, 436–452.
- Dickerson, R.E. (1992) DNA Structure from A to Z. In *Methods In Enzymology* (Lilley, D.M.J. and Dahlberg, J.E., eds), **211**, pp. 67–111. Academic Press, London.
- Dickerson, R.E. and Drew, H.R. (1981) Structure of a B-DNA Dodecamer II. Influence of Base Sequence on Helix Structure. *J. Mol. Biol.* **149**, 761–786.
- Dock-Bregeon, A.C. and Moras, D. (1992) Crystallization of nucleic acids and co-crystallization of proteins and nucleic acids. In *Crystallization of Nucleic Acids and Proteins: A Practical Approach* (Ducruix, A. and Giegé, R., eds), pp. 145–174. The Practical Approach Series, IRL Press, Oxford.
- Donohue, J. (1956) Hydrogen-Bonded Helical Configurations of Polynucleotides. *Proc. Natl. Acad. Sci. USA* **42**, 60–65.
- Drew, H.R. and Dickerson, R.E. (1981) Structure of a B-DNA Dodecamer III. Geometry of Hydration. *J. Mol. Biol.* **151**, 535–556.
- Egli, M., Gessner, R.V., Williams, L.D., Quigley, G.J., van der Marel, G.A., van Boom, J.H., Rich, A. and Frederick, C.A. (1990) Atomic-resolution structure of the cellulose synthase regulator cyclic diguanylic acid. *Proc. Natl. Acad. Sci. USA* **87**, 3235–3239.
- Eisenberg, D. and Crothers, D. (1979) Principles of Spectroscopy. In *Physical Chemistry with Applications to the Life Sciences*, pp. 520–522. The Benjamin/Cummings Publishing Company Inc., California.

- Emery, H.S. and Weiner, A.M. (1981) An Irregular Satellite Sequence is Found at the Termini of the Linear Extrachromosomal rDNA in *Dictyostelium discoideum*. *Cell* **26**, 411–419.
- Fang, G. and Cech, T.R. (1993a) The β Subunit of *Oxytricha* Telomere–Binding Protein Promotes G–Quartet Formation by Telomeric DNA. *Cell* **74**, 875–885.
- Fang, G. and Cech, T.R. (1993b) *Oxytricha* telomere–binding protein: DNA–dependent dimerization of the α and β subunits. *Proc. Natl. Acad. Sci. USA* **90**, 6056–6060.
- Feyerabend, P. (1987) *Farewell to Reason*. Verso, London.
- Feyerabend, P. (1988) *Against Method: Outline of an Anarchistic Theory of Knowledge*. Verso, London.
- Fitzgerald, P.M.D. (1991) Molecular replacement. In *Crystallographic Computing: From Chemistry to Biology* (Moras, D., Podjarny, A.D. and Thierry, J.C., eds), **5**, pp. 333–347. Oxford University Press, Oxford.
- Fox, G.C. and Holmes, K.C. (1966) An Alternative Method of Solving the Layer Scaling Equations of Hamilton, Rollett and Sparks. *Acta Cryst.* **20**, 886–891.
- Freer, S.T. (1985) Classic ($F_o - F_c$) Fourier Refinement. In *Methods In Enzymology* (Wyckoff, H.W., Hirs, C.H.W. and Timasheff, S.N., eds), **115**, pp. 235–237. Academic Press, London.
- French, S. and Wilson, K. (1978) On the Treatment of Negative Intensity Observations. *Acta Cryst.* **A34**, 517–525.

- Fujii, S., Wang, A.H.-J., van der Marel, G., van Boom, J.H. and Rich, A. (1982) Molecular structure of (m5dC-dG)₃: the role of the methyl group on 5-methyl cytosine in stabilizing Z-DNA. *Nucleic Acids Res.* **10**, 7879-7892.
- Fujinaga, M. and Read, R.J. (1987) Experiences with a New Translation-Function Program. *J. Appl. Cryst.* **20**, 517-521.
- Gabe, E.J., Le Page, Y., Charland, J.-P., Lee, F.L. and White, P.S. (1989) NRCVAX – an interactive program system for structure analysis. *J. Appl. Cryst.* **22**, 384-387.
- Gaffney, B.L., Wang, C. and Jones, R.A. (1992) Nitrogen-15-Labeled Oligodeoxynucleotides. 4. Tetraplex Formation of d[G(¹⁵N⁷)GTTTTTGG] and d[T(¹⁵N⁷)GGGT] Monitored by ¹H Detected ¹⁵N NMR. *J. Am. Chem. Soc.* **114**, 4047-4050.
- Gall, J., ed. (1986) *Molecular Biology of Ciliated Protozoa*. Academic Press, New York.
- Gandour, R.D., Fronczek, F.R., Gatto, V.J., Minganti, C., Schultz, R.A., White, B.D., Arnold, K.A., Mazzocchi, D., Miller, S.R. and Gokel, G.W. (1986) Solid-State Structural Chemistry of Lariat Ether and BiBLE Cation Complexes: Metal Ion Identity and Coordination Number Determine Cavity Site. *J. Am. Chem. Soc.* **108**, 4978-4088.
- Gehring, K., Leroy, J.-L. and Guéron, M. (1993) A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* **363**, 561-565.
- Gellert, M., Lipsett, M.N and Davies, D.R. (1962) Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci. USA* **48**, 2013-2018.

- Gewirth, D. (1993) *The DENZO Manual. A Description of the program DENZO: An Oscillation Data Processing Program for Macromolecular Crystallography* by Zbyszek Otwinowski. Department of Molecular Biophysics and Biochemistry, Yale University and the Howard Hughes Medical Institute, New Haven, CT 06511, USA. Special Sigler Lab Edition.
- Giacovazzo, C., Monaco, H.L., Viterbo, D., Scordari, F., Gilli, G., Zanotti, G. and Catti, M. (1992) *Fundamentals of Crystallography* (Giacovazzo, C., ed). International Union of Crystallography, Oxford University Press, New York.
- Giraldo, R. and Rhodes, D. (1994) The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.* **13**, 2411–2420.
- Gottschling, D.E. and Zakian, V.A. (1986) Telomere Proteins: Specific Recognition and Protection of the Natural Termini of Oxytricha Macronuclear DNA. *Cell* **47**, 195–205.
- Gray, A. (1987) *Lanark: A Life in 4 Books*. Paladin Books, London.
- Green, D.W., Ingram, V.M. and Perutz, M.F. (1954) The structure of haemoglobin IV. Sign determination by the isomorphous replacement method. *Proc. Roy. Soc. London A* **225**, 287–307.
- Greider, C.W. (1990) Telomeres, Telomerase and Senescence. *BioEssays* **12**, 363–369.
- Greider, C.W. and Blackburn, E.H. (1985) Identification of a Specific Telomere Terminal Transferase Activity in Tetrahymena Extracts. *Cell* **43**, 405–413.

- Greider, C.W. and Blackburn, E.H. (1987) The Telomere Terminal Transferase of *Tetrahymena* is a Ribonucleoprotein Enzyme with Two Kinds of Primer Specificity. *Cell* **51**, 887–898.
- Greider, C.W. and Blackburn, E.H. (1989) A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature* **337**, 331–337.
- Groebe, D.R. and Uhlenbeck, O.C. (1988) Characterization of RNA hairpin loop stability. *Nucleic Acids Res.* **16**, 11725–11735.
- Guo, Q., Lu, M. and Kallenbach, N.R. (1992a) Adenine Affects the Structure and Stability of Telomeric Sequences. *J. Biol. Chem.* **267**, 15293–15300.
- Guo, Q., Lu, M., Marky, L.A. and Kallenbach, N.R. (1992b) Interaction of the Dye Ethidium Bromide with DNA Containing Guanine Repeats. *Biochemistry* **31**, 2451–2455.
- Gupta, G., Garcia, A.E., Guo, Q., Lu, M. and Kallenbach, N.R. (1993) Structure of a Parallel-Stranded Tetramer of the *Oxytricha* Telomere DNA Sequence dT₄G₄. *Biochemistry* **32**, 7098–7103.
- Hammond-Kosack, M.C.U. and Docherty, K. (1992) A consensus repeat sequence from the human insulin gene linked polymorphic region adopts multiple quadruplex DNA structures in vitro. *FEBS Letters* **301**, 79–82.
- Harada, Y., Lifchitz, A., Berthou, J. and Jolles, P. (1981) A Translation Function Combining Packing and Diffraction Information: An Application to Lysozyme (High-Temperature Form). *Acta Cryst.* **A37**, 398–406.

- Haraway, D. (1985) A Manifesto for Cyborgs: Science, Technology and Socialist Feminism in the 1980s. *Socialist Review* **80**, 65–107.
- Hardin, C.C., Henderson, E., Watson, T. and Prosser, J.K. (1991) Monovalent Cation Induced Structural Transitions in Telomeric DNAs: G–DNA Folding Intermediates. *Biochemistry* **30**, 4460–4472.
- Hardin, C.C., Watson, T., Corrigan, M. and Bailey, C. (1992) Cation–Dependent Transition between the Quadruplex and Watson–Crick Hairpin Forms of d(CGCG₃GCG). *Biochemistry* **31**, 833–841.
- Hardin, C.C., Corrigan, M., Brown II, B.A. and Frederick, L.N. (1993) Cytosine–Cytosine Base–Pairing Stabilizes DNA Quadruplexes and Cytosine Methylation Greatly Enhances the Effect. *Biochemistry* **32**, 5870–5880.
- Harley, C.B., Futcher, A.B. and Greider, C.W. (1990) Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458–460.
- Henderson, E., Hardin, C.C., Walk, S.K., Tinoco Jr., I. and Blackburn, E.H. (1987) Telomeric DNA Oligonucleotides Form Novel Intramolecular Structures Containing Guanine–Guanine Base Pairs. *Cell* **51**, 899–908.
- Henderson, E.R. and Blackburn, E.H. (1989) An Overhanging 3' Terminus Is a Conserved Feature of Telomeres. *Mol. Cell. Biol.* **9**, 345–348.
- Henderson, E.R., Moore, M. and Malcolm, B.A. (1990) Telomere G–Strand Structure and Function Analysed by Chemical Protection, Base Analogue Substitution, and Utilization by Telomerase in Vitro. *Biochemistry* **29**, 732–737.

- Hendrickson, W.A. (1985) Stereochemically Restrained Refinement of Macromolecular Structures. In *Methods In Enzymology* (Wyckoff, H.W., Hirs, C.H.W. and Timasheff, S.N., eds), **115**, pp. 252–270. Academic Press, London.
- Hendrickson, W.A. and Konnert, J.H. (1980) Incorporation of Stereochemical Information into Crystallographic Refinement. In *Computing In Crystallography* (Diamond, R., Rameseshan, S. and Venkatesan, K., eds), pp. 13.01–13.23. Indian Academy of Sciences, Bangalore, India.
- Hendrickson, W.A. and Ward, K.B. (1976) A Packing Function for Delimiting the Allowable Locations of Crystallized Macromolecules. *Acta Cryst.* **A32**, 778–780.
- Hicke, B.J., Celander, D.W., MacDonald, G.H., Price, C.M. and Cech, T.R. (1990) Two versions of the gene encoding the 41–kilodalton subunit of the telomere binding protein of *Oxytricha nova*. *Proc. Natl. Acad. Sci. USA* **87**, 1481–1485.
- Hilbers, C.W., Haasnoot, C.A.G., de Bruin, S.H., Joordens, J.J.M., van der Marel, G.A. and van Boom, J.H. (1985) Hairpin formation in synthetic oligonucleotides. *Biochemie* **67**, 685–695.
- Hingerty, B.E., Brown, R.S. and Klug, A. (1982) Stabilization of the Tertiary Structure of Yeast Phenylalanine tRNA by $[\text{Co}(\text{NH}_3)_6]^{3+}$: X-Ray Evidence for Hydrogen Bonding to Pairs of Guanine Bases in the Major Groove. *Biochim. Biophys. Acta* **697**, 78–82.
- Hobza, P. and Sandorfy, C. (1987) Nonempirical Calculations on All the 29 Possible DNA Base Pairs. *J. Am. Chem. Soc.* **109**, 1302–1307.

- Hoogsteen, K. (1959) The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Cryst.* **12**, 822–823.
- Howard, F.B., Frazier, J. and Miles, H.T. (1977) Stable and Metastable Forms of Poly(G). *Biopolymers* **16**, 791–809.
- Howard, F.B. and Miles, H.T. (1982) Poly(inosinic acid) Helices: Essential Chelation of Alkali Metal Ions in the Axial Channel. *Biochemistry* **21**, 6736–6745.
- Hunter, W.N. (1992) Crystallographic Studies of DNA Containing Mismatches, Modified and Unpaired Bases. In *Methods In Enzymology* (Lilley, D.M.J. and Dahlberg, J.E., eds), **211**, pp. 221–231. Academic Press, London.
- Ishida, T., Doi, M., Ueda, H., Inoue, M. and Sheldrick, G.M. (1988) Structural Studies of the Interaction Between Indole-Derivatives and Biologically Important Aromatic-Compounds. 20. Specific Ring Stacking Interaction on the Tryptophan 7-Methylguanine System: Comparative Crystallographic Studies of Indole-Derivatives 7-Methylguanine Base, Nucleoside, and Nucleotide Complexes. *J. Am. Chem. Soc.* **110**, 2286–2294.
- Jacob, F. (1989) *The Possible and the Actual*. Penguin Books, London.
- Jensen, L.H. (1985) Overview of Refinement in Macromolecular Structure Analysis. In *Methods In Enzymology* (Wyckoff, H.W., Hirs, C.H.W. and Timasheff, S.N., eds), **115**, pp. 227–234. Academic Press, London.

- Jin, R., Gaffney, B.L., Wang, C., Jones, R.A. and Breslauer, K.J. (1992) Thermodynamics and structure of a DNA tetraplex: A spectroscopic and calorimetric study of the tetramolecular complexes of d(TG₃T) and d(TG₃T₂G₃T). *Proc. Natl. Acad. Sci. USA* **89**, 8832–8836.
- Jones, A., Bartels, K. and Schwager, P. (1977) Refinement of Crystal Orientation Parameters. In *The Rotation Method in Crystallography* (Arndt, U.W. and Wonacott, A.J., eds), pp. 105–117. North Holland, Amsterdam.
- Jones, T.A. (1978) A Graphics Model Building and Refinement System for Macromolecules. *J. Appl. Cryst.* **11**, 268–272.
- Jones, T.A. (1985) Interactive Computer Graphics: FRODO. In *Methods In Enzymology* (Wyckoff, H.W., Hirs, C.H.W. and Timasheff, S.N., eds), **115**, pp. 157–171. Academic Press, London.
- Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in these Models. *Acta Cryst.* **A47**, 110–119.
- Joshua-Tor, L. and Sussman, J.L. (1993) The coming of age of DNA crystallography. *Curr. Opin. Struct. Biol.* **3**, 323–335.
- Junghans, R.P., Boone, L.R. and Skalka, A.M. (1982) Retroviral DNA H Structures: Displacement–Assimilation Model of Recombination. *Cell* **30**, 53–62.
- Kabsch, W. (1988) Automatic Indexing of Rotation Diffraction Patterns. *J. Appl. Cryst.* **21**, 67–71.

- Kabsch, W. (1993) Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J. Appl. Cryst.* **26**, 795–800.
- Kadrey, R. (1989) Genocide. In *Semiotext(e) SF* (Rucker, R., Wilson, P.L. and Wilson, R.A., eds), pp. 159–166. Semiotext(e)/Autonomedia, New York.
- Kang, C.-H., Zhang, X., Ratliff, R., Moyzis, R. and Rich, A. (1992) Crystal structure of four-stranded *Oxytricha* telomeric DNA. *Nature* **356**, 126–131.
- Karle, I.L. (1974) The Conformation of the Sodium Complex of a Biologically Active Analog of Antamanide in the Crystalline State. *Biochemistry* **13**, 2155–2162.
- Katinka, M.D. and Bourgain, F.M. (1992) Interstitial telomeres are hotspots for illegitimate recombination with DNA molecules injected into the macronucleus of *Paramecium primaurelia*. *EMBO J.* **11**, 725–732.
- Kennard, O. and Hunter, W.N. (1989) Oligonucleotide structure: a decade of results from single crystal X-ray diffraction studies. *Quarterly Rev. Biophys.* **22**, 327–379.
- Kennard, O. and Hunter, W.N. (1991) Single-Crystal X-Ray Diffraction Studies of Oligonucleotides and Oligonucleotide–Drug Complexes. *Angew Chem. Int. Edn. Engl.* **30**, 1254–1277.
- Kim, J., Cheong, C. and Moore, P.B. (1991) Tetramerization of an RNA oligonucleotide containing a GGGG sequence. *Nature* **351**, 331–332.

- Kim, S.H., Sussman, J.L., Suddath, F.L., Quigley, G.J., McPherson, A., Wang, A.H.J., Seeman, N.C. and Rich, A. (1974) The General Structure of Transfer RNA Molecules. *Proc. Natl. Acad. Sci. USA* **71**, 4970–4974.
- Klobutcher, L.A., Swanton, M.T., Donini, P. and Prescott, D.M. (1981) All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proc. Natl. Acad. Sci. USA* **78**, 3015–3019.
- Kohwi, Y. and Kohwi-Shigematsu, T. (1988) Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. *Proc. Natl. Acad. Sci. USA* **85**, 3781–3785.
- Kroker, A. (1993) Codes of Privilege: Arthur Kroker Interviewed by Sharon Grace. *Mondo 2000* **11**, 60–67.
- Kundera, M. (1991) *Immortality*. Faber and Faber, London.
- Kung, H.-J., Hu, S., Bender, W., Bailey, J.M., Davidson, N., Nicholson, M.O. and McAllister, R.M. (1976) RD-114, Baboon, and Woolly Monkey Viral RNAs Compared in Size and Structure. *Cell* **7**, 609–620.
- Ladner, J.E., Finch, J.T., Klug, A. and Clark, B.F.C. (1972) High-resolution X-ray Diffraction Studies on a Pure Species of Transfer RNA. *J. Mol. Biol.* **72**, 99–101.
- Lamzin, V.S. and Wilson, K.S. (1993) Automated Refinement of Protein Models. *Acta Cryst.* **D49**, 129–147.

- Lattman, E.E. and Love, W.E. (1970) A Rotational Search Procedure for Detecting A Known Molecule In a Crystal. *Acta Cryst.* **B26**, 1854–1857.
- Laughlan, G., Murchie, A.I.H., Norman, D.G., Moore, M.H., Moody, P.C.E., Lilley, D.M.J. and Luisi, B. (1994) The High-Resolution Crystal Structure of a Parallel-Stranded Guanine Tetraplex. *Science* **265**, 520–524.
- Lee, J.S. (1990) The stability of polypurine tetraplexes in the presence of mono- and divalent cations. *Nucleic Acids Res.* **18**, 6057–6060.
- Leonard, G.A. and Hunter, W.N. (1993) Crystal and Molecular Structure of dCGTA-GATCTACG at 2.25Å Resolution. *J. Mol. Biol.* **234**, 198–208.
- Leslie, A.G.W. (1990) Molecular Data Processing. In *Crystallographic Computing – From Chemistry to Biology* (Moras, D., Podjarny, A.D. and Thierry J.D., eds), **5**, pp. 50–61. Oxford University Press, Oxford.
- Levis, R.W. (1989) Viable Deletions of a Telomere from a Drosophila Chromosome. *Cell* **58**, 791–801.
- Leyner, M. (1990) I Was an Infinitely Hot and Dense Dot. In *My Cousin, My Gastroenterologist*, pp. 3–8. Harmony Books, New York.
- Linial, M.L. and Miller, A.D. (1990) Retroviral RNA packaging – sequence requirements and implications. *Curr. Opin. Microbiol. Immunol.* **157**, 125–152.
- Lipps, H.J. (1980) *In Vitro* aggregation of the gene-sized DNA molecules of the ciliate *Stylonychia mytilus*. *Proc. Natl. Acad. Sci. USA* **77**, 4104–4107.

- Lipps, H.J., Gruissem, W. and Prescott, D.M. (1982) Higher order DNA structure in macronuclear chromatin of the hypotrichous ciliate *Oxytricha nova*. *Proc. Natl. Acad. Sci. USA* **79**, 2495–2499.
- Liu, Z., Frantz, J.D., Gilbert, W. and Tye, B.-K. (1993) Identification and characterization of a nuclease activity specific for G4 tetrastranded DNA. *Proc. Natl. Acad. Sci. USA* **90**, 3157–3161.
- Liu, Z. and Gilbert, W. (1994) The Yeast *KEM1* Gene Encodes a Nuclease Specific for G4 Tetraplex DNA: Implication of In Vivo Functions for This Novel DNA Structure. *Cell* **77**, 1083–1092.
- Lu, M., Guo, Q. and Kallenbach, N.R. (1992) Structure and Stability of Sodium and Potassium Complexes of dT₄G₄ and dT₄G₄T. *Biochemistry* **31**, 2455–2459.
- Lundblad, V. and Szostak, J.W. (1989) A Mutant with a Defect in Telomere Elongation Leads to Senescence in Yeast. *Cell* **57**, 633–643.
- Luzzati, P.V. (1952) Traitement Statistique des Erreurs dans la Determination des Structures Cristallines. *Acta Cryst.* **5**, 802–810.
- Macaya, R.F., Schultze, P., Smith, F.W., Roe, J.A. and Feigon, J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci. USA* **90**, 3745–3749.
- Mariani, P., Mazabard, C., Garbesi, A. and Spada, G.P. (1989) A Study of the Structure of the Lyomesophases Formed by the Dinucleoside Phosphate d(GpG). An Approach by X-Ray Diffraction and Optical Microscopy. *J. Am. Chem. Soc.* **111**, 6369–6373.

- Marquet, R., Baudin, F., Gabus, C., Darlix, J.-L., Mougél, M., Ehresmann, C. and Ehresmann, B. (1991) Dimerization of human-immunodeficiency-virus (type 1) RNA: stimulation by cations and possible mechanism. *Nucleic Acids Res.* **19**, 2349–2357.
- Marx, K. and Engels, F. (1967) *The Communist Manifesto*. Penguin Books, London.
- McClintock, B. (1941) The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234–282.
- McLuhan M. (1987) *Understanding Media: The Extensions of Man*. Ark Paperbacks, London.
- Michel, F., Hanna, M., Green, R., Bartel, D.P. and Szostak, J.W. (1989) The guanosine binding site of the *Tetrahymena* ribozyme. *Nature* **342**, 391–395.
- Miles, H.T. and Frazier, J. (1978) Poly(I) Helix Formation. Dependence on Size-Specific Complexing to Alkali Metal Ions. *J. Am. Chem. Soc.* **100**, 8037–8038.
- Miller, M., Harrison, R.W., Wlodawer, A., Appella, E. and Sussman, J.L. (1988) Crystal structure of 15-mer DNA duplex containing unpaired bases. *Nature* **334**, 85–86.
- Momus. (1992) Cibachrome Blue. In *Lusts of a Moron: The Lyrics of Momus*, pp. 132–133. Black Spring Press, London.
- Monod, J. (1971) *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*. Knopf, New York.
- Morin, G.B. (1989) The Human Telomere Terminal Transferase Enzyme Is a Ribonucleoprotein That Synthesizes TTAGGG Repeats. *Cell* **59**, 521–529.

- Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. and Wu, J.-R. (1988) A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. USA* **85**, 6622–6626.
- Murchie, A.I.H. and Lilley, D.M.J. (1992) Retinoblastoma susceptibility genes contain 5' sequences with a high propensity to form guanine–tetrad structures. *Nucleic Acids Res.* **20**, 49–53.
- Murchie, A.I.H. and Lilley, D.M.J. (1994) Tetraplex folding of telomere sequences and the inclusion of adenine bases. *EMBO J.* **13**, 993–1001.
- Murchie, A.I.H., Aboul-ela, F., Laughlan, G., Norman, D.G., Luisi, B. and Lilley, D.M.J. (1995) Structure of Parallel–Stranded Guanine Tetraplexes. In *Nucleic Acids and Molecular Biology* (Eckstein, F. and Lilley, D.M.J., eds), **9** (In press). Springer–Verlag, Berlin, Heidelberg.
- Navaza, J. (1987) On the Fast Rotation Function. *Acta Cryst.* **A43**, 645–653.
- Navaza, J. (1992) AMoRe: A New Package for Molecular Replacement. In *Molecular Replacement: Proceedings of the CCP4 Study Weekend, 31 January – 1 February, 1992* (Dodson, E.J., Gover, S. and Wolf, W., eds), pp. 87–90. Science and Engineering Research Council Daresbury Laboratory, Daresbury, Warrington, WA4 4AD.
- Nietzsche, F.W. (1968) *The Will To Power*. Vintage Books, New York.
- Nietzsche, F.W. (1986) *Human, All Too Human: A Book for Free Spirits*. Cambridge University Press, Cambridge.

- Oka, Y. and Thomas Jr., C.A. (1987) The cohering telomeres of *Oxytricha*. *Nucleic Acids Res.* **15**, 8877–8898.

- Otwinowski, Z. (1993a) Oscillation data reduction program. In *Data Collection and Processing: Proceedings of the CCP4 Study Weekend, 29–30 January 1993* (Sawyer, L., Isaacs, I. and Bailey, S., eds), pp. 56–62. Science and Engineering Research Council Daresbury Laboratory, Daresbury, Warrington, WA4 4AD.

- Otwinowski, Z. (1993b) The SCALEPACK Manual. In *The DENZO Manual. A Description of the program DENZO: An Oscillation Data Processing Program for Macromolecular Crystallography by Zbyszek Otwinowski* (Gewirth, D., ed). Department of Molecular Biophysics and Biochemistry, Yale University and the Howard Hughes Medical Institute, New Haven, CT 06511, USA. Special Sigler Lab Edition.

- Panyutin, I.G., Kovalsky, O.I., Budowsky, E.I., Dickerson, R.E., Rikhirev, M.E. and Lipanov, A.A. (1990) G–DNA: A twice-folded DNA structure adopted by single-stranded oligo(dG) and its implications for telomeres. *Proc. Natl. Acad. Sci. USA* **87**, 867–870.

- Patterson, A.L. (1934) A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals. *Phys. Rev.* **46**, 372–376.

- Pearson, A.M., Rich, A. and Kreiger, M. (1993) Polynucleotide Binding to Macrophage Scavenger Receptors Depends on the Formation of Base quartet stabilized Four Stranded Helices. *J. Biol. Chem.* **268**, 3546–3554.

- Pedersen, C.J. (1988) The Discovery of Crown Ethers (Nobel Lecture). *Angew Chem. Int. Edn. Engl.* **27**, 1021–1027.
- Phillips, D.C., Rogers, D. and Wilson, A.J.C. (1950) Reliability index for centrosymmetric and non-centrosymmetric structures. *Acta Cryst.* **3**, 398–399.
- Pieleles, U., Zürcher, W., Schär, M. and Moser, H.E. (1993) Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. *Nucleic Acids Res.* **21**, 3191–3196.
- Pinnavaia, T.J., Marshall, C.L., Mettler, C.M., Fisk, C.L., Miles, H.T. and Becker, E.D. (1978) Alkali Metal Ion Specificity in the Solution Ordering of a Nucleotide, 5'-Guanosine Monophosphate. *J. Am. Chem. Soc.* **100**, 3625–3627.
- Pluta, A.F., Dani, G.M., Spear, B.B. and Zakian, V.A. (1984) Elaboration of telomeres in yeast: Recognition and modification of termini from *Oxytricha* macronuclear DNA. *Proc. Natl. Acad. Sci. USA* **81**, 1475–1479.
- Pluta, A.F. and Zakian, V.A. (1989) Recombination occurs during telomere formation in yeast. *Nature* **337**, 429–433.
- Prats, A.C., Sarih, L., Gabus, C., Litvak, S., Keith, G. and Darlix, J.L. (1988) Small finger protein of avian and murine retroviruses has nucleic acid annealing activity and positions the replication primer tRNA onto genomic RNA. *EMBO J.* **7**, 1777–1783.
- Price, C.M. and Cech, T.R. (1989) Properties of the Telomeric DNA-Binding Protein from *Oxytricha nova*. *Biochemistry* **28**, 769–774.

- Price, C.M., Skopp, R., Krueger, J. and Williams, D. (1992) DNA Recognition and Binding by the *Euplotes* Telomere Protein. *Biochemistry* **31**, 10835–10843.
- Privé, G.G., Yanagi, K. and Dickerson, R.E. (1991) Structure of the B–DNA Decamer C–C–A–A–C–G–T–T–G–G and Comparison with Isomorphous Decamers C–C–A–A–G–A–T–T–G–G and C–C–A–G–G–C–C–T–G–G. *J. Mol. Biol.* **217**, 177–199.
- Quinby, L. (1989) Panic Jeans. In *Panic Encyclopedia: The Definitive Guide to the Postmodern Scene* (Kroker, A., Kroker, M. and Cook, D., eds), pp. 131–136. Macmillan Education, London.
- Raghuraman, M.K. and Cech, T.R. (1989) Assembly and Self–Association of *Oxytricha* Telomeric Nucleoprotein Complexes. *Cell* **59**, 719–728.
- Raghuraman, M.K. and Cech, T.R. (1990) Effect of monovalent cation–induced telomeric DNA structure on the binding of *Oxytricha* telomeric protein. *Nucleic Acids Res.* **18**, 4543–4552.
- Raghuraman, M.K., Dunn, C.J., Hickie, B.J. and Cech, T.R. (1989) *Oxytricha* telomeric nucleoprotein complexes reconstituted with synthetic DNA. *Nucleic Acids Res.* **17**, 4235–4253.
- Richards, E.J. and Ausubel, F.M. (1988) Isolation of a Higher Eukaryotic Telomere from *Arabidopsis thaliana*. *Cell* **53**, 127–136.
- Rippe, K., Fritsch, V., Westhof, E. and Jovin, T.M. (1992) Alternating d(G–A) sequences form a parallel–stranded DNA homoduplex. *EMBO J.* **11**, 3777–3786.

- Robertson, J.M. and Woodward, I. (1937) An X-ray Study of the Phthalocyanines. Part III. Quantitative Structure Determination of Nickel Phthalocyanine. *J. Chem. Soc.* **219**, 219–230.
- Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F.C. and Klug, A. (1974) Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **250**, 546–551.
- Ross, A. (1991) *Strange Weather: Culture, Science and Technology in the Age of Limits*. Verso, London.
- Rossman, M.G. (1961) The Position of Anomalous Scatterers in Protein Crystals. *Acta Cryst.* **14**, 383–388.
- Rossman, M.G., ed. (1972) *The Molecular Replacement Method: A Collection of Papers on the Use of Non-Crystallographic Symmetry*. Gordon and Breach, New York.
- Rossman, M.G. (1979) Processing Oscillation Diffraction Data for Very Large Unit Cells with an Automatic Convolution Technique and Profile Fitting. *J. Appl. Cryst.* **12**, 225–238.
- Rossman, M.G. and Blow, D.M. (1962) The Detection of Sub-Units Within the Crystallographic Asymmetric Unit. *Acta Cryst.* **15**, 24–31.
- Saenger, W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.
- Saenger, W. (1987) Structure and Dynamics of Water Surrounding Biomolecules. *Ann. Rev. Biophys. Biophys. Chem.* **16**, 93–114.

- Saenger, W., Hunter, W.N. and Kennard, O. (1986) DNA conformation is determined by economics in the hydration of phosphate groups. *Nature* **324**, 385–388.
- Sarma, M.H., Luo, J., Umemoto, K., Yuan, R.D. and Sarma, R.H. (1992) Tetraplex Formation of d(GGGGGTTTTT): Proton NMR Study in Solution. *J. Biomol. Struct. Dyn.* **9**, 1131–1153.
- Schierer, T. and Henderson, E. (1994) A Protein from *Tetrahymena thermophila* That Specifically Binds Parallel-Stranded G4-DNA. *Biochemistry* **33**, 2240–2246.
- Schultze, P., Smith, F.W. and Feigon, J. (1994) Refined solution structure of the dimeric quadruplex formed from the *Oxytricha* telomeric oligonucleotide dGGGGTTTTGGGG. *Structure* **2**, 221–233.
- Sebeok, T. (1975) *How Animals Communicate*. Indiana University Press, Bloomington.
- Sen, D. and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364–366.
- Sen, D. and Gilbert, W. (1990) A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **344**, 410–414.
- Shakked, Z. (1991) The influence of the environment on DNA structures determined by X-ray crystallography. *Curr. Opin. Struct. Biol.* **1**, 446–451.
- Shampay, J., Szostak, J.W. and Blackburn, E.H. (1984) DNA sequences of telomeres maintained in yeast. *Nature* **310**, 154–157.

- Sheldrick, G.M. (1993) Institut für Anorganische Chemie, Universität Göttingen, Tammannstrasse 4, 37077 Göttingen, Germany.
- Shimizu, A. and Honjo, T. (1984) Immunoglobulin Class Switching. *Cell* **36**, 801–803.
- Shippen–Lentz, D. and Blackburn, E.H. (1989) Telomere Terminal Transferase Activity from *Euplotes crassus* Adds Large Numbers of TTTTGGGG Repeats onto Telomeric Primers. *Mol. Cell. Biol.* **9**, 2761–2764.
- Shippen–Lentz, D. and Blackburn, E.H. (1990) Functional Evidence for an RNA Template in Telomerase. *Science* **247**, 546–552.
- Sinha, N.D., Biernat, J., McManus, J. and Köster, H. (1984) Polymer support oligonucleotide synthesis XVIII^{1,2}: use of β -cyanoethyl-N,N-dialkylamino-/N-morpholino phosphoramidite of deoxynucleosides for the synthesis of DNA fragments simplifying deprotection and isolation of the final product. *Nucleic Acids Res.* **12**, 4539–4557.
- Smith, F.W. and Feigon, J. (1992) Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides. *Nature* **356**, 164–168.
- Smith, F.W. and Feigon, J. (1993) Strand Orientation in the DNA Quadruplex Formed from the *Oxytricha* Telomere Repeat Oligonucleotide d(G₄T₄G₄) in Solution. *Biochemistry* **32**, 8682–8692.
- Sterling, B. (1988) Twenty Evocations. *Mississippi Review* **47/48**, 122–129.
- Stout, G.H. and Jensen, L.H. (1968) *X-Ray Structure Determination.*, pp. 54. Macmillan, New York.

- Stuhlmann, H. and Berg, P. (1992) Homologous Recombination of Copackaged Retrovirus RNAs during Reverse Transcription. *J. Virol.* **66**, 2378–2388.
- Sundquist, W.I. (1991) The Structures of Telomeric DNA. In *Nucleic Acids and Molecular Biology* (Eckstein, F. and Lilley, D.M.J., eds), **5**, pp. 1–24. Springer-Verlag, Berlin, Heidelberg.
- Sundquist, W.I. and Heaphy, S. (1993) Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proc. Natl. Acad. Sci. USA* **90**, 3393–3397.
- Sundquist, W.I. and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342**, 825–829.
- Surridge, C. (1994) The results of thought and resolution. *Nature Struct. Biol.* **9**, 583.
- Sussman, J. (1991) Introduction to macromolecular refinement. In *Crystallographic Computing: From Chemistry to Biology* (Moras, D., Podjarny, A.D. and Thierry, J.C., eds), **5**, pp. 382–391. Oxford University Press, Oxford.
- Szostak, J.W. and Blackburn, E.H. (1982) Cloning Yeast Telomeres on Linear Plasmid Vectors. *Cell* **29**, 245–255.
- Tabor, C.W. and Tabor, H. (1976) 1,4-Diaminobutane(Putrescine), Spermidine, and Spermine. *Annu. Rev. Biochem.* **45**, 285–306.

- Taylor, R. and Kennard, O. (1982) Molecular Structures of Nucleosides and Nucleotides.
2. Orthogonal Coordinates for Standard Nucleic Acid Base Residues. *J. Am. Chem. Soc.* **104**, 3209–3212.
- Tougaard, P., Chantot, J.-F. and Guschlbauer, W. (1973) Nucleoside Conformations.
X. An X-Ray Fibre Diffraction Study of the Gels of Guanine nucleosides. *Biochem. Biophys. Acta* **308**, 9–16.
- Vattimo, G. (1988) *The End of Modernity: Nihilism and Hermeneutics in Post-Modern Culture*. Polity Press, Cambridge.
- Venczel, E.A. and Sen, D. (1993) Parallel and Antiparallel G-DNA Structures from a Complex Telomeric Sequence. *Biochemistry* **32**, 6220–6228.
- Virilio, P. (1989) The Last Vehicle. In *Looking Back on the End of the World* (Kamper, D. and Wulf, C., eds), pp. 106–119. Semiotext(e), New York.
- Virilio, P. (1991) *The Lost Dimension*. Semiotext(e)/Autonomedia, New York.
- Virilio, P. (1994) *The Vision Machine*. British Film Institute/Indiana University Press, London/Bloomington.
- Walmsley, R.W., Chan, C.S.M., Tye, B.-K. and Petes, T.D. (1984) Unusual DNA sequences associated with the ends of yeast chromosomes. *Nature* **310**, 157–160.
- Walsh, K. and Gualberto, A. (1992) MyoD Binds to the Guanine Tetrad Nucleic Acid Structure. *J. Biol. Chem.* **267**, 13714–13718.

- Wang, A.H.-J. (1992) Intercalative drug binding to DNA. *Curr. Opin. Struct. Biol.* **2**, 361–368.
- Wang, A.H.-J., Hakoshima, T., van der Marel, G., van Boom, J.H. and Rich, A. (1984) AT Base Pairs Are Less Stable than GC Base Pairs in Z-DNA: The Crystal Structure of d(m⁵CGTAm⁵CG). *Cell* **37**, 321–331.
- Wang, K.Y., McCurdy, S., Shea, R.G., Swaminathan, S. and Boltom, P.H. (1993) A DNA Aptamer Which Binds to and Inhibits Thrombin Exhibits a New Structural Motif for DNA. *Biochemistry* **32**, 1899–1904.
- Wang, Y. and Patel, D.J. (1992) Guanine Residues in d(T₂AG₃) and d(T₂G₄) Form Parallel-Stranded Potassium Cation Stabilized G-Quadruplexes with Anti Glycosidic Torsion Angles in Solution. *Biochemistry* **31**, 8112–8119.
- Wang, Y. and Patel, D.J. (1993a) Solution Structure of a Parallel Stranded G Quadruplex DNA. *J. Mol. Biol.* **234**, 1171–1183.
- Wang, Y. and Patel, D.J. (1993b) Solution structure of the human telomeric repeat d[AG₃(T₂AG₃)₃] G-tetraplex. *Structure* **1**, 263–282.
- Wang, Y., de los Santos, C., Gao, X., Greene, K., Live, D. and Patel, D.J. (1991a) Multinuclear Nuclear Magnetic Resonance Studies of Na Cation-stabilized Complex Formed by d(GGTTTTTCGG) in Solution – Implications for G-tetrad Structures. *J. Mol. Biol.* **222**, 819–832.

- Wang, Y., Jin, R., Gaffney, B., Jones, R.A. and Breslauer, K.J. (1991*b*) Characterization by ^1H NMR of glycosidic conformations in the tetramolecular complex formed by d(GGTTTTTGG). *Nucleic Acids Res.* **19**, 4619–4622.
- Watson, J.D. (1972) Origin of Concatemeric T7 DNA. *Nature New Biol.* **239**, 197–201.
- Watson, J.D. and Crick, F.H.C. (1953) A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738.
- Weiner, A.M. (1988) Eukaryotic Nuclear Telomeres: Molecular Fossils of the RNP World? *Cell* **52**, 155–157.
- Weisman-Shomer, P. and Fry, M. (1993) QUAD, a Protein from Hepatocyte Chromatin That Binds Selectively to Guanine-rich Quadruplex DNA. *J. Biol. Chem.* **268**, 3306–3312.
- Westhof, E. (1988) Water: An Integral Part of Nucleic Acid Structure. *Ann. Rev. Biophys. Biophys. Chem.* **17**, 125–144.
- Williamson, J.R. (1993) Guanine quartets. *Curr. Opin. Struct. Biol.* **3**, 357–362.
- Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1989) Monovalent Cation-Induced Structure of Telomeric DNA: The G-Quartet Model. *Cell* **59**, 871–880.
- Wilson, A.J.C. (1949) The Probability Distribution of X-ray Intensities. *Acta Cryst.* **2**, 318–321.
- Wilson, A.J.C. (1950) Largest likely values for the reliability index. *Acta Cryst.* **3**, 397–398.

- Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.E. (1980) Crystal structure analysis of a complete turn of B-DNA. *Nature* **287**, 755–758.
- Wittgenstein, L. (1971) *Tractatus Logico-Philosophicus*. Routledge Kegan and Paul, London.
- Wonacott, A.J. (1977) Geometry of the Rotation Method. In *The Rotation Method in Crystallography* (Arndt, U.W. and Wonacott, A.J., eds), pp. 75–103. North Holland, Amsterdam.
- Wyatt, J.R., Vickers, T.A., Roberson, J.L., Buckheit Jr., R.W., Klimkait, T., DeBaets, E., Davis, P.W., Rayner, B., Imbach, J.L. and Ecker, D.J. (1994) Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion. *Proc. Natl. Acad. Sci. USA* **91**, 1356–1360.
- Yu, G.-L., Bradley, J.D., Attardi, L.D. and Blackburn, E.H. (1990) *In Vivo* alteration of telomere sequences and senescence caused by mutated *Tetrahymena* telomerase RNAs. *Nature* **344**, 126–132.
- Zahler, A.M. and Prescott, D.M. (1988) Telomere terminal transferase activity in the hypotrichous ciliate *Oxytricha nova* and a model for replication of the ends of linear DNA molecules. *Nucleic Acids Res.* **16**, 6953–6972.
- Zahler, A.M. and Prescott, D.M. (1989) DNA primase and the replication of the telomeres in *Oxytricha nova*. *Nucleic Acids Res.* **17**, 6299–6317.
- Zahler, A.M., Williamson, J.R., Cech, T.R. and Prescott, D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature* **350**, 718–720.

- Zakian, V.A. (1989) Structure and Function of Telomeres. *Annu. Rev. Genet.* **23**, 579–604.
- Zakian, V.A., Runge, K. and Wang, S.-S. (1990) How does the end begin? Formation and maintenance of telomeres in ciliates and yeast. *Trends Genet.* **6**, 12–16.
- Zanotti, G. (1992) Protein Crystallography. In *Fundamentals of Crystallography* (Giacovazzo, C., ed), pp. 535–597. International Union of Crystallography, Oxford University Press, New York.
- Zimmerman, S.B. (1976) X-Ray Study by Fibre Diffraction Methods of a Self-aggregate of Guanosine-5'-Phosphate with the Same Helical Parameters as Poly(rG). *J. Mol. Biol.* **106**, 663–672.
- Zimmerman, S.B., Cohen, G.H. and Davies, D.R. (1975) X-Ray Fibre Diffraction and Model-building Study of Polyguanylic Acid and Polyinosinic Acid. *J. Mol. Biol.* **92**, 181–192.