163

Protein-DNA Recognition: The Importance of the Structure of DNA and its Components

by

Leslie William Tari

A Thesis
Submitted to the Faculty of Graduate Studies in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Department of Chemistry
University of Manitoba
Winnipeg, Manitoba

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

ISBN 0-612-13524-1

Canada

Name ____

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

CHEMISTRY, PHYSICAL — SUBJECT TERM

0 4 9 4 — SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture	0729
Art History	0377
Cinema	0900
Dance	0378
Fine Arts	0357
Information Science	0723
Journalism	0391
Library Science	0399
Mass Communications	0708
Music	0413
Speech Communication	0459
Theater	0465

EDUCATION

General	0515
Administration	0514
Adult and Continuing	0516
Agricultural	0517
Art	0273
Bilingual and Multicultural	0282
Business	0688
Community College	0275
Curriculum and Instruction	0727
Early Childhood	0518
Elementary	0524
Finance	0277
Guidance and Counseling	0519
Health	0680
Higher	0745
History of	0520
Home Economics	0278
Industrial	0521
Language and Literature	0279
Mathematics	0280
Music	0522
Philosophy of	0998
Physical	0523
Psychology	0525
Reading	0535
Religious	0527
Sciences	0714
Secondary	0533
Social Sciences	0534
Sociology of	0340
Special	0529
Teacher Training	0530
Technology	0710
Tests and Measurements	0288
Vocational	0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language	
General	0679
Ancient	0289
Linguistics	0290
Modern	0291
Literature	
General	0401
Classical	0294
Comparative	0295
Medieval	0297
Modern	0298
African	0316
American	0591
Asian	0305
Canadian (English)	0352
Canadian (French)	0355
English	0593
Germanic	0311
Latin American	0312
Middle Eastern	0315
Romance	0313
Slavic and East European	0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy	0422
Religion	
General	0318
Biblical Studies	0321
Clergy	0319
History of	0320
Philosophy of	0322
Theology	0469

SOCIAL SCIENCES

American Studies	0323
Anthropology	
Archaeology	0324
Cultural	0326
Physical	0327
Business Administration	
General	0310
Accounting	0272
Banking	0770
Management	0454
Marketing	0338
Canadian Studies	0385
Economics	
General	0501
Agricultural	0503
Commerce-Business	0505
Finance	0508
History	0509
Labor	0510
Theory	0511
Folklore	0358
Geography	0366
Gerontology	0351
History	
General	0578
Ancient	0579
Medieval	0581
Modern	0582
Black	0328
African	0331
Asia, Australia and Oceania	0332
Canadian	0334
European	0335
Latin American	0336
Middle Eastern	0333
United States	0337
History of Science	0585
Law	0398
Political Science	
General	0615
International Law and Relations	0616
Public Administration	0617
Recreation	0814
Social Work	0452
Sociology	
General	0626
Criminology and Penology	0627
Demography	0938
Ethnic and Racial Studies	0631
Individual and Family Studies	0628
Industrial and Labor Relations	0629
Public and Social Welfare	0630
Social Structure and Development	0700
Theory and Methods	0344
Transportation	0709
Urban and Regional Planning	0999
Women's Studies	0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture	
General	0473
Agronomy	0285
Animal Culture and Nutrition	0475
Animal Pathology	0476
Food Science and Technology	0359
Forestry and Wildlife	0478
Plant Culture	0479
Plant Pathology	0480
Plant Physiology	0817
Range Management	0777
Wood Technology	0746
Biology	
General	0306
Anatomy	0287
Biostatistics	0308
Botany	0309
Cell	0379
Ecology	0329
Entomology	0353
Genetics	0369
Limnology	0793
Microbiology	0410
Molecular	0307
Neuroscience	0317
Oceanography	0416
Physiology	0433
Radiation	0821
Veterinary Science	0778
Zoology	0472
Biophysics	
General	0786
Medical	0760

EARTH SCIENCES

Biogeochemistry	0425
Geochemistry	0996
Geodesy	0370
Geology	0372
Geophysics	0373
Hydrology	0388
Mineralogy	0411
Paleobotany	0345
Paleoecology	0426
Paleontology	0418
Paleozoology	0985
Palynology	0427
Physical Geography	0368
Physical Oceanography	0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences	0768
Health Sciences	
General	0566
Audiology	0300
Chemotherapy	0992
Dentistry	0567
Education	0350
Hospital Management	0769
Human Development	0758
Immunology	0982
Medicine and Surgery	0564
Mental Health	0347
Nursing	0569
Nutrition	0570
Obstetrics and Gynecology	0380
Occupational Health and Therapy	0354
Ophthalmology	0381
Pathology	0571
Pharmacology	0419
Pharmacy	0572
Physical Therapy	0382
Public Health	0573
Radiology	0574
Recreation	0575
Speech Pathology	0460
Toxicology	0383
Home Economics	0386

PHYSICAL SCIENCES

Pure Sciences

Chemistry	
General	0485
Agricultural	0749
Analytical	0486
Biochemistry	0487
Inorganic	0488
Nuclear	0738
Organic	0490
Pharmaceutical	0491
Physical	0494
Polymer	0495
Radiation	0754
Mathematics	0405
Physics	
General	0605
Acoustics	0986
Astronomy and Astrophysics	0606
Atmospheric Science	0608
Atomic	0748
Electronics and Electricity	0607
Elementary Particles and High Energy	0798
Fluid and Plasma	0759
Molecular	0609
Nuclear	0610
Optics	0752
Radiation	0756
Solid State	0611
Statistics	0463

Applied Sciences

Applied Mechanics	0346
Computer Science	0984
Engineering	
General	0537
Aerospace	0538
Agricultural	0539
Automotive	0540
Biomedical	0541
Chemical	0542
Civil	0543
Electronics and Electrical	0544
Heat and Thermodynamics	0348
Hydraulic	0545
Industrial	0546
Marine	0547
Materials Science	0794
Mechanical	0548
Metallurgy	0743
Mining	0551
Nuclear	0552
Packaging	0549
Petroleum	0765
Sanitary and Municipal	0554
System Science	0790
Geotechnology	0428
Operations Research	0796
Plastics Technology	0795
Textile Technology	0994

PSYCHOLOGY

General	0621
Behavioral	0384
Clinical	0622
Developmental	0620
Experimental	0623
Industrial	0624
Personality	0625
Physiological	0989
Psychobiology	0349
Psychometrics	0632
Social	0451

PROTEIN-DNA RECOGNITION:

THE IMPORTANCE OF THE STRUCTURE OF DNA AND ITS COMPONENTS

BY

LESLIE WILLIAM TARI

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Abstract

The primary focus of this dissertation is on the importance of the structure of DNA, and its individual components, in the protein-DNA recognition process. The research described herein centers on two distinct areas of investigation. In the first, the crystal structure of a DNA fragment contained in a regulatory region of the *E. coli* genome has been determined, and its importance in the context of site specific recognition has been inferred.

The existence of distinct families of DNA-binding proteins, which employ related structural motifs for recognition, suggests the involvement of common sequence elements in DNA regulatory regions. One sequence that is represented frequently in functionally important sites involving protein-DNA interactions is GTG/CAC. Using x-ray crystallography, the structure of the DNA hexamer d(CGGTGG)/d(CCACCG), a part of the interior operator (O_I, nucleotides +44 to +49) of the *gal* operon, co-crystallized with spermine, was determined to a resolution of 2.5 Å. Analysis of the structure reveals a number of distortions from canonical B-DNA, including a novel non-Watson-Crick hydrogen-bonding scheme between adenine and thymine in the GTG region. The ability of this sequence to adopt unusual conformations in its GTG region may be the critical factor that confers sequence selectivity to the binding of Gal repressor. In addition, this is the first conclusive example of a crystal structure of spermine with native B-DNA, providing insight into the mechanics of polyamine:DNA binding, as well as possible explanations for the biological action of spermine.

In the second series of studies, nucleosides and nucleotides modified at the sugar moiety have been synthesized in an effort to structurally characterize novel oligonucleotides. These molecules have been studied in the context of protein recognition, and as potential antisense constructs, anti-viral agents and enzymatically resistant molecular probes. The major goal in the design of antisense oligonucleotides is the creation of enzymatically stable oligonucleotide analogues that form stable duplexes with their natural complements. While extensive studies have been carried out on oligonucleotides modified in their bases and phosphate backbones, sugar modifications have received little attention. In this study, 1-(2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranosyl) thymine was used as a nucleoside substitute in the synthesis of the dimer ApT*(2'-deoxyadenosine-(3'-6')-[1-(2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranosyl) thymine] phosphate). Comparative CD studies indicate that the global structural parameters of ApT* in solution are isomorphous with those of natural ApT. ApT*, relative to natural ApT, was also found to possess an increased resistance to degradation by nucleases. The results indicate that oligonucleotides end-capped with 2',3'-dideoxy-β-D-erythro-hex-2'-enopyranose nucleosides may be effective antisense agents. The crystal structure of a precusor molecule, 1-(2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranosyl) thymine was also determined and compared with that of the active conformer of AZT, in an effort to assess its potential activity against HIV.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my gratitude to my supervisor, Professor Anthony S. Secco, for making the graduate school experience a fun and fruitful one. The high standards he works by, as well as the example he has set, certainly serve to inspire.

I would like to thank the members of my advisory and/or examining committee, Dr. John Westmore, Dr. Peter Loewen and Dr. Frank Hawthorne for their helpful input. I am grateful to Dr. Stephen Evans for the advice he has given me, and for providing the lab with his molecular modeling program SETOR. Dr. James Charlton, also on my committee, has been particularly helpful, assisting me in matters ranging from computer operating systems to organic synthesis.

I express my sincere gratitude to Dr. Krishan Sadana for teaching me the tricks of the trade regarding the synthesis of oligonucleotides, and to Dr. Frank Hruska, for always having an open office door and for providing insightful advice about virtually anything I needed help with.

I am also grateful to Dr. Loren Williams for his generous assistance with x-ray data collection at M. I. T.

I am certainly indebted to my parents, Les and Susan, for all of the sacrifices they made to put me through school and for encouraging me to pursue science. I extend special thanks to Joanne, for her unyielding patience and commitment.

I gratefully acknowledge the Natural Sciences and Engineering Research Council of Canada, and the University of Manitoba for financial support throughout my graduate studies.

Chapter I

Introduction

1 Significance and Background

1.1 The Importance of DNA Structure in Biology

The double helical structure of DNA proposed by Watson and Crick in 1953[1] provided a unifying concept to molecular biology. It led to the discovery of the genetic code and the understanding of how that code is utilized in the creation of usable biological products. Also, it provided a simple mechanism for the transmission of genetic information by complementary hydrogen bonding of base-pairs, and an explanation of mutations as a result of errors in the reading of the genetic code. In the time since the discovery of the double helix, these ideas have been explored in depth, and much has been learned about the complexity of the biological mechanisms controlling these events.

A crucial biological process, inherent to all living systems, is the conversion of genetic information into usable biological products, *viz.* the transcription of genetic code for protein synthesis. The polymerases, repressors and activators involved in transcription initiate their action by binding to specific, short base-pair sequences of an effectively infinite molecular chain of DNA. However, before specific binding can occur, the proteins must *locate* and *recognize* these targets.

When considering even a simple organism such as *E. coli*, the magnitude of this problem becomes apparent. The *E. coli* genome usually contains single operator and promoter sites for each operon. It also contains approximately 10^7 base pairs of DNA. Comparatively, each cell contains a much smaller number of copies of each particular regulatory protein (*i.e.* ~10 *lac* repressor molecules, ~5000 RNA polymerase molecules, *etc.*). This implies that the DNA double-helix possesses sequence dependent features which regulatory proteins are able to recognize and bind to with a tremendous degree of specificity. However, to date, it is not understood exactly how DNA is able to confer sequence specificity to the binding of regulatory proteins. One of the major goals of this dissertation is to provide insight into a potentially critical recognition factor, which was discovered in the crystal structure of d(CGGTGG)/d(CCACCG), a sequence of DNA contained in a regulatory region of the *E. coli* genome.

1.11 The Mechanics of the Regulatory Process: The *gal* Operon

The most elementary regulatory mechanisms used in bacteria obey the following rule: *A system is turned "on" when it is needed and "off" when it is not needed.* Since almost all reaction processes in living cells are catalyzed by enzymes, a cell can effectively regulate the activities of its metabolic pathways by altering the amounts of available enzymes in response to environmental factors. Cells conserve a great deal of energy and vastly improve their efficiencies by synthesizing only the necessary enzymes, a process which is initiated when a certain molecule or group of molecules called **inducers** appear in the cell. When a given set of proteins are not required, the expression of the genes coding these proteins is prevented by **repressor** proteins, which block transcription by binding to specific regions of the genomic DNA called **operator** sites. The inducers promote expression by conformationally altering the repressor proteins, so that they are no longer able to bind to the operator sites, or by interacting with a positive regulator protein, like the **cyclic AMP receptor protein (CRP)** found in *E. coli.* bacteria, and converting it into a form which binds to a genomic region called a **promoter**. The CRP bound promoter facilitates the binding of RNA polymerase, and hence, provides a starting point for transcription. A section of genomic DNA containing a collection of operators and promoters, and the sequences coding the enzymes whose expression they regulate are referred to as an **operon**. Operons may be defined as **constitutive** and **non-constitutive** systems; constitutive systems continue to express their genes at a significant level, even when the operon is "off", while in non-constitutive systems, gene expression is almost completely suppressed in the "off" state. The sequence of DNA which is the focus of this study is contained in a regulatory region of the *E. coli* genome called the *gal* operon, a constitutive operon which may be the new paradigm replacing *lac* for the mechanisms of regulation of gene expression in prokaryotic systems.

The *gal* operon exhibits a dual role in cellular metabolism: **a)** it is responsible for the catabolism of the sugar galactose, which the cell utilizes as an alternative energy source when glucose is not available, and **b)** one of the structural genes encoded in the operon converts uridinediphosphoglucose to uridinediphosphogalactose (in the absence of exogenous galactose), which must be available at all times, as it is a precursor in the synthesis of the *E. coli* cell wall. The three structural genes in the *gal* operon encode the enzymes galactokinase (*galK*), galactose-1-phosphate uridyltransferase (*galT*) and uridine diphosphogalactose 4'-epimerase (*galE*) (each cistron is approx. 1100 base-pairs long[2]). The mechanism of catabolism by the three *gal* enzymes is summarized in Fig. 1. Overall, the concerted action of the enzymes convert galactose to glucose 1-phosphate, which is utilized as an energy source in *E. coli* bacteria.

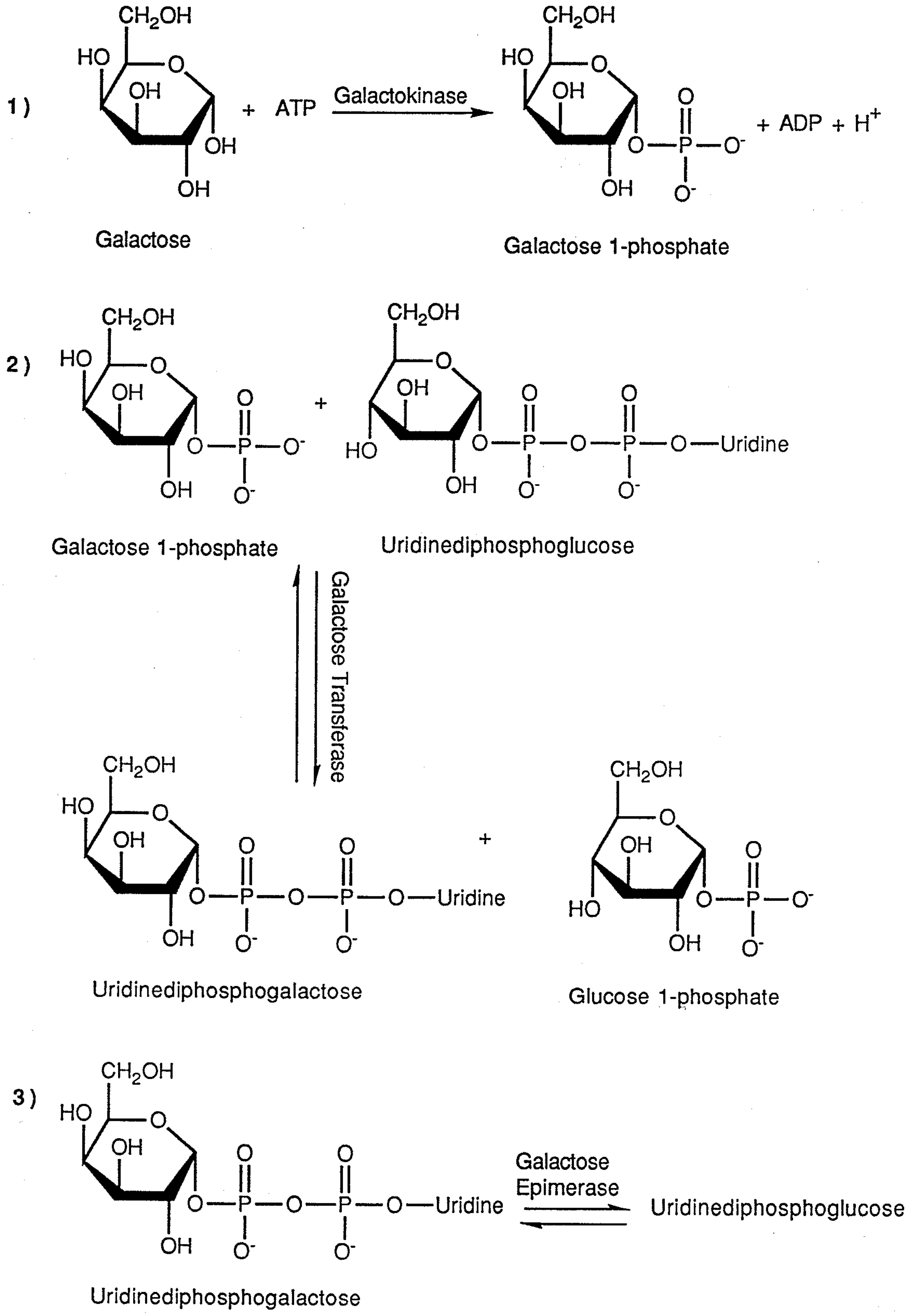


Fig. 1 Reaction scheme showing the galactose catabolism pathway.

A schematic representation of the *gal* operon is provided in Fig. 2. Regulation of the operon occurs at two levels[3], with primary repression occurring in the *absence* of galactose, mediated by the Gal repressor, GalR, and a second level of regulation operating in the *presence* of galactose, which is mediated by cyclic AMP (adenosine monophosphate) and CRP (*i.e.* glucose concentration) and a second repressor protein, called the Gal isorepressor, GalS. GalR binds to two operator sites, O_E (exterior operator), located upstream of the promoter sites and O_I (interior operator), located downstream within the *galE* gene, which are separated by 114 base-pairs. The repressors bound to O_E and O_I bend their corresponding DNA segments and interact with each other to form a closed loop complex (Fig. 3), blocking transcription from the overlapping promoter sites P_1 and P_2[4]. At the primary level, galactose induces transcription by interacting with GalR and altering its conformation, so that GalR no longer recognizes its binding sites, O_E or O_I. The inactivation of GalR in the presence of galactose induces gene expression from P_1 and P_2 by a 10- to 15-fold increase, depending on the secondary factors described above. GalS functions in a similar manner to GalR, repressing transcription by binding to O_I and O_E, but only when GalR is inactivated. GalS attenuates expression of the *gal* genes to a much smaller extent than GalR, since it binds with a lesser affinity to the DNA operator sites, and also because GalR epistatically suppresses the inactivation of GalS by galactose[3], for reasons, and by a mechanism, that are not fully understood. It is interesting to note that this prevention of complete galactose-induced inactivation of GalS actually suppresses transcription of the *gal* genes by a 2- to 4-fold amount (*i.e.* galactose induces transcription by a factor of more than 30 in the *absence* of GalS!). In contrast to the secondary negative control exerted by GalS, a secondary positive control is exercised by the cAMP-CRP complex. Cyclic-AMP-CRP binding switches transcription initiation between start sites from S_1 and S_2 entirely to S_1[4] (by covering the P_2 promoter site when it binds). In high glucose environments, cAMP synthesis is inhibited[5], so that levels of available cAMP-CRP are low. In the absence of cAMP-CRP, no allosteric control is exercised on the binding of RNA polymerase, which then binds with equal facility to the promoter sites P_1 and P_2. This has the effect of attenuating the expression of the *gal* genes, because an unusually large proportion of transcripts initiated at S_2 are aborted at the tri- to hexa-nucleotide stage[6]. This second layer of control provides the *gal* operon with the remarkable ability to modulate its expression in response to two separate factors, galactose and glucose.

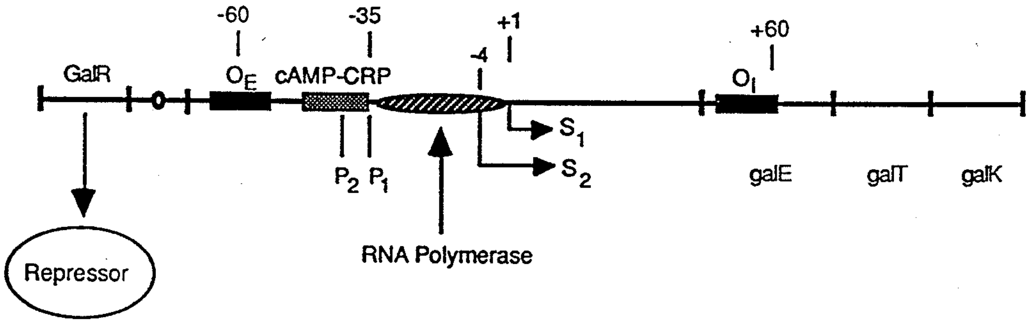


Fig. 2 The *gal* operon of *E. coli* and the unlinked *galR* gene. P_1 and P_2 are overlapping promoters, differentially regulated by cAMP-CRP complex. S_1 and S_2 represent the start points of transcription for P_1 and P_2 respectively. Transcription starts from either of S_1 or S_2 in the absence of cAMP-CRP, but is only stimulated from S_1 in its presence, since cAMP-CRP binds to and blocks the P_2 site. Each transcript is polycistronic, encoding E, T, and K (epimerase, transferase, kinase). Gal repressor protein, the *galR* gene product, interacts with two sites in the operon, O_E and O_I, to inhibit transcription. Both O_E and O_I, whose centers of symmetry are 114 base pairs apart, function in repression of both promoters. (Not drawn to scale).

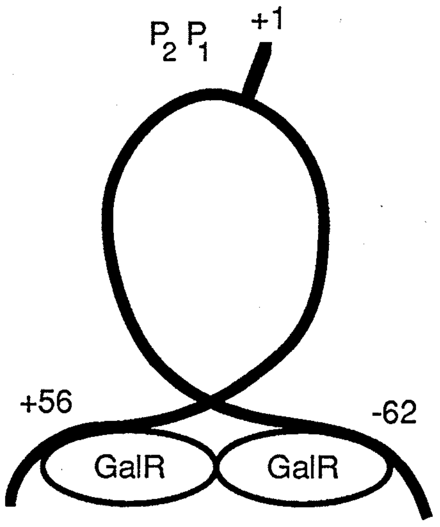


Fig. 3 Gal repressors bound to O_E and O_I bend the corresponding DNA segments and interact with each other to achieve repression. +1 refers to the start site of transcription from the P_1 promoter. (Not drawn to scale).

The concerted effect of the various controlling mechanisms on the expression of the *gal* enzymes is summarized in Table 1. The multiple levels of control are required because the enzymes of the *gal* operon constitute an amphibolic system. They are needed for degradation of galactose as a carbon and energy source as well as for the biosynthesis of complex carbohydrates in cell walls and cell membranes. Physiologically, these two processes are unrelated. The dual negative control of *gal* provides *E. coli.* with the ability to adapt to both catabolic and anabolic requirements of the galactose pathway. However, one fundamental aspect of the regulatory process in *gal* and the other known regulator systems is not understood: *What sequence specific property of DNA confers the enormous degree of specificity to the binding of regulatory proteins required for biological function*? To gain insight into this problem, the crystal structure of a critical region of the interior operator (O_I) of the *gal* operon has been determined, and some intriguing structural features have been discovered in this region which may be pivotal in the recognition of O_I by Gal repressor. Before describing the structure, some of the basic principles of DNA structure will be first discussed.

Galactose	Glucose	Rate of Expression
Low	High	Low
Low	Low	Low
High	High	Moderate
High	Low	High

Table 1 Table summarizing the effects of cellular concentrations (qualitative) of galactose and glucose on the rate of expression of the *gal* genes. Galactose concentration is the primary "switch" for transcription, since it disables the GalR and GalS repressors. Glucose modulates expression to a lesser extent, by altering the level of cAMP. At low glucose concentration cAMP synthesis is induced, increasing the level of cAMP-CRP, which, when bound to P_1, allosterically enhances the initiation of transcription.

1.2 DNA Structure and Nomenclature

1.21 Bases, Nucleosides and Nucleotides

Deoxyribonucleic acids are long, fibre-like polymers made up of a linear array of monomers called **nucleotides**. Nucleotides are constructed from three components: a nitrogen heterocyclic **base**, a 2'-deoxyribose **sugar**, and a **phosphate** residue. There are four types of bases found in DNA, which belong to the monocyclic **pyrimidine** family or the bicyclic **purine** family. The bases and their atomic numbering schemes are shown in Fig. 4.

The bases are joined from N(9) in purines and N(1) in pyrimidines to C1' of the 2'-deoxy-D-ribose sugar, to form the four unique **nucleosides**. The nucleosides are always in the β-form, with the base on the same side of the sugar as C5' (Fig. 5).

The nucleotides, which are the phosphate esters of nucleosides (Fig. 6), form the basic building blocks of DNA. Successive nucleotide units in an oligonucleotide chain are joined by phosphodiester linkages between O3' of one nucleotide and O5' of the next. Oligonucleotide sequences are written in the 5' to 3' direction, using only the abbreviated names for the bases. For example, the oligonucleotide dimer 2'-deoxyadenosine-(3'-5')-2'-deoxythymidine phosphate in abbreviated form is d(ApT), or simply d(AT).

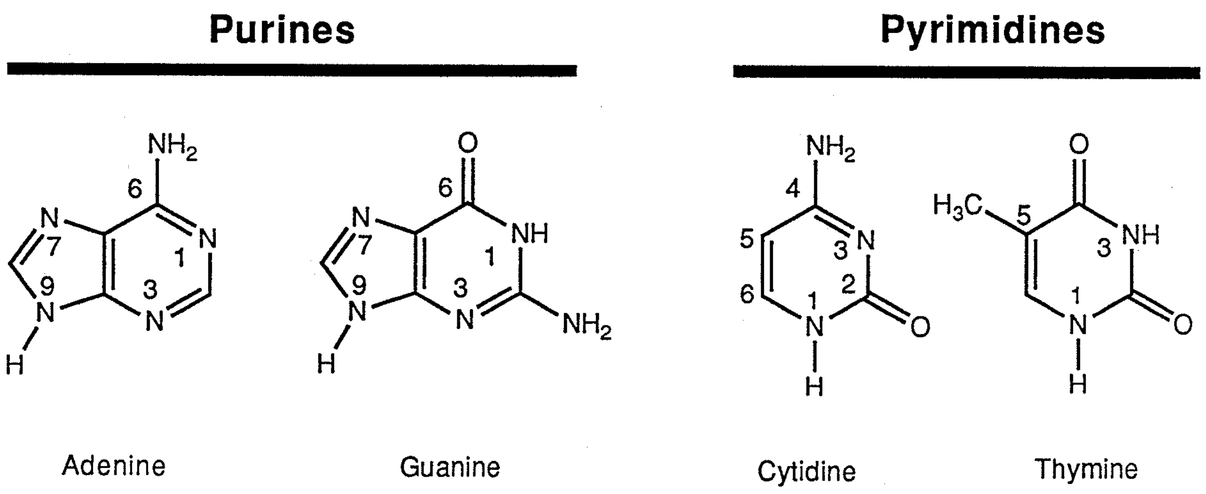


Fig. 4 Structures of the purine and pyrimidine bases found in DNA, in their major tautomeric forms. The common abbreviated names for the bases are (A) adenine, (G) guanine, (C) cytidine and (T) thymine.

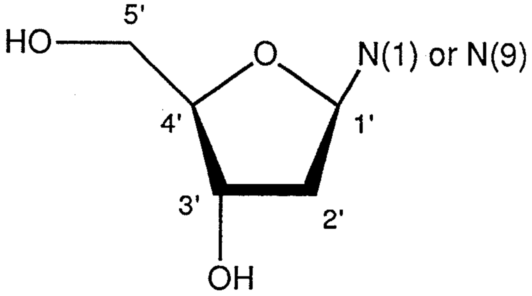
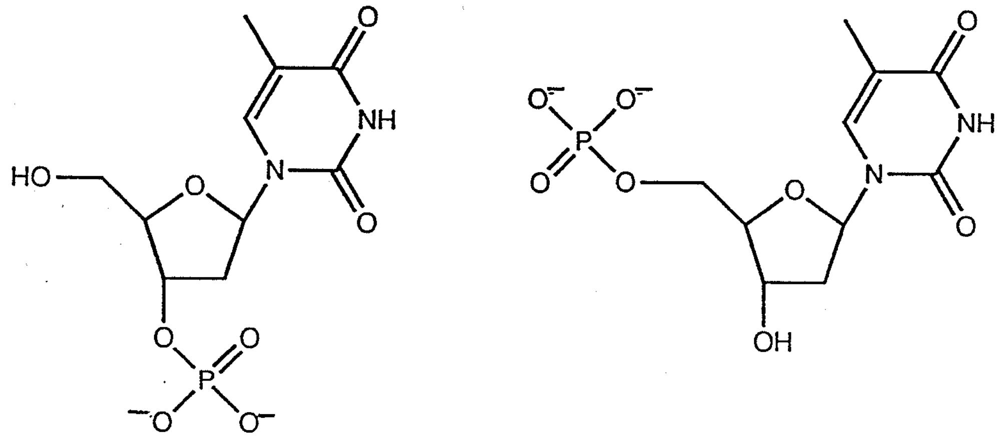
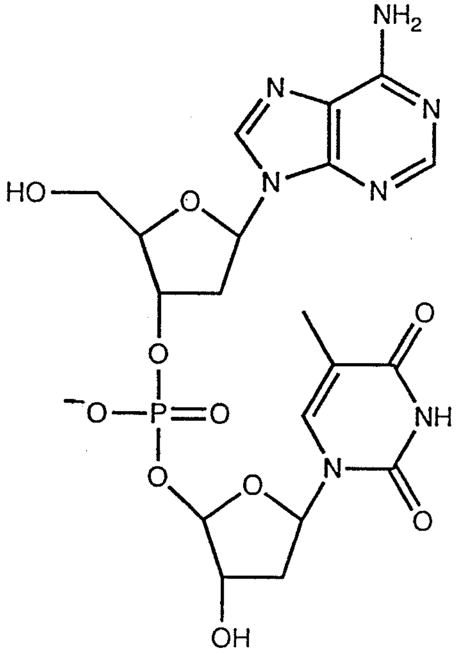


Fig. 5 Structure of a deoxynucleoside showing the atomic labeling convention. The four nucleosides found in DNA are named deoxyadenosine, deoxyguanosine, deoxycytidine and deoxythymidine, respectively.



2'-deoxythymidine 3'-phosphate
d(Tp)

2'-deoxythymidine 5'-phosphate
d(pT)



2'-deoxyadenosine-(3'-5')-2'-deoxythymidine phosphate
d(AT)

Fig. 6 Examples of nucleotides and oligonucleotides at neutral pH.

1.22 Nucleotide Geometry

The overall geometry of the DNA double-helix is a consequence of the conformational preferences of its individual nucleotide units. The geometry of nucleotides is described by the phosphate backbone torsion angles α, β, γ, δ, ε and ζ, the furanose ring torsion angles θ_0 to θ_4, and χ for the glycosidic bond (Fig. 7). However, because there is a strong interdependence between many of these torsion angles, nucleotides can be more simply described using only four sets of parameters: the sugar pucker, the *syn-anti* conformation about the N-glycosidic bond, the conformation about the exocyclic C4'-C5' bond, and the geometries of the C-O and P-O ester bonds.

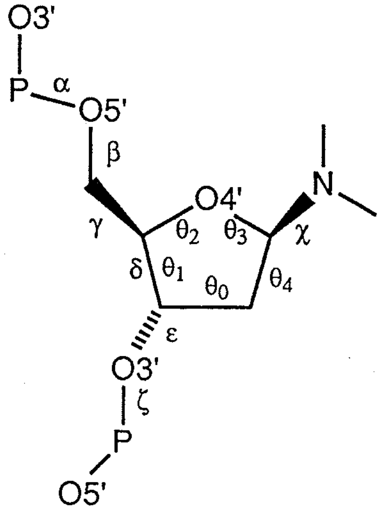


Fig. 7 IUPAC torsion angle notation for polynucleotide chains.

1.221 Sugar Pucker

The furanose sugar ring is generally non planar, and can be puckered in an envelope (E) conformation with four coplanar atoms and the fifth atom out of plane by 0.5 Å, or in a twist (T) conformation which has adjacent out-of-plane atoms oppositely disposed about the plane made by the other three (Fig. 8).

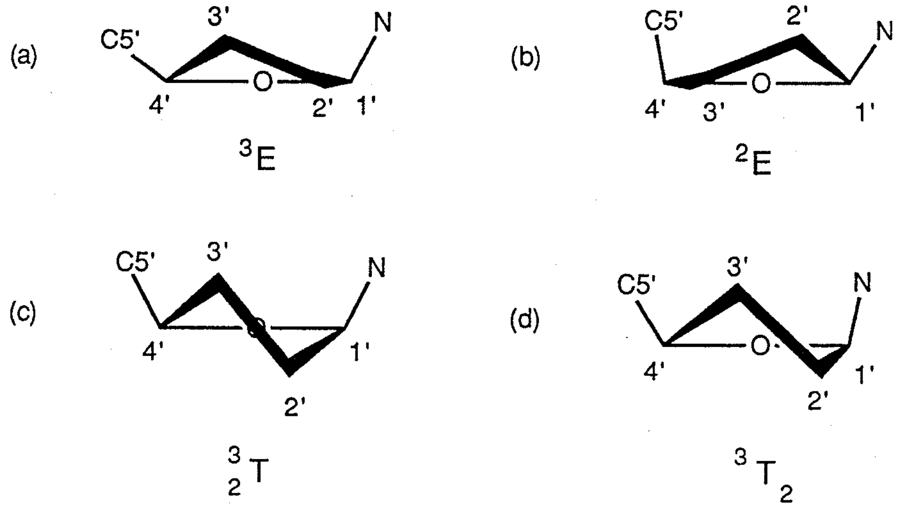


Fig. 8 Examples of some sugar puckering modes, viewed with the plane made up of atoms C4', O4' and C1' perpendicular to the plane of the page. (a) Envelope C3'-*endo*, ^{3}E. (b) Envelope C2'-*endo*, ^{2}E. (c) Symmetrical twist C2'-*exo*-C3'-*endo*, ${}^{3}_{2}T$. (d) Asymmetric twist, with major C3'-*endo* and minor C2'-*exo*, $^{3}T_{2}$.

Atoms displaced from these planes on the same side as C5' are called *endo;* those on the opposite side are called *exo.* Since the transition between E and T forms is generally a facile process, statistically it is expected that puckered conformations should be predominately asymmetric. Thus, an abbreviated notation has been devised to distinguish between large deviations from planarity (major puckering) and small deviations from planarity (minor puckering). Using this notation, superscripts for *endo* atoms and subscripts for *exo* atoms precede or follow the letters E or T for major and minor puckering respectively.

An elegant treatment of puckering in furanose sugars, using the concept of pseudorotation, was devised by Altona and Sundaralingam[7]. Using the pseudorotation concept, sugar puckers are treated in terms of two parameters: the degree of pucker (or maximum torsion angle), ν_{max}, and the phase angle, *P*. Using the endocyclic torsion angles θ_0 to θ_4 (see Fig. 7), *P* is calculated from equation (1):

$$\tan P = \frac{(\theta_2 + \theta_4) - (\theta_1 + \theta_3)}{2 \bullet \theta_0 \bullet (\sin 36^\circ + \sin 72^\circ)} \qquad (1)$$

By convention, $P = 0^\circ$ is defined such that θ_o is maximally positive, corresponding to the symmetrical twist, ${}^{3}_{2}T$. The mirror image pucker, ${}^{2}_{3}T$, has a phase angle $P = 180^\circ$ (Fig. 9). At every phase angle *P*, the sum of the five torsion angles is always zero. The maximum out-of-plane pucker, ν_{max}, is given by:

$$\nu_{max} = \frac{\theta_0}{\cos P} \qquad (2)$$

In 2'-deoxyribonucleotides, potential energy barriers limit the pseudorotation and lead to preferred puckering modes. There are two ranges of pseudorotation phase angles that are found to be preferred: C3'-*endo* at $0^\circ \leq P \leq 36^\circ$ (in the "north" of the cycle, or N) and C2'-*endo* at $144^\circ \leq P \leq 190^\circ$ ("south", or S). Deoxyribose puckering preferences are described in more detail in Fig. 10.

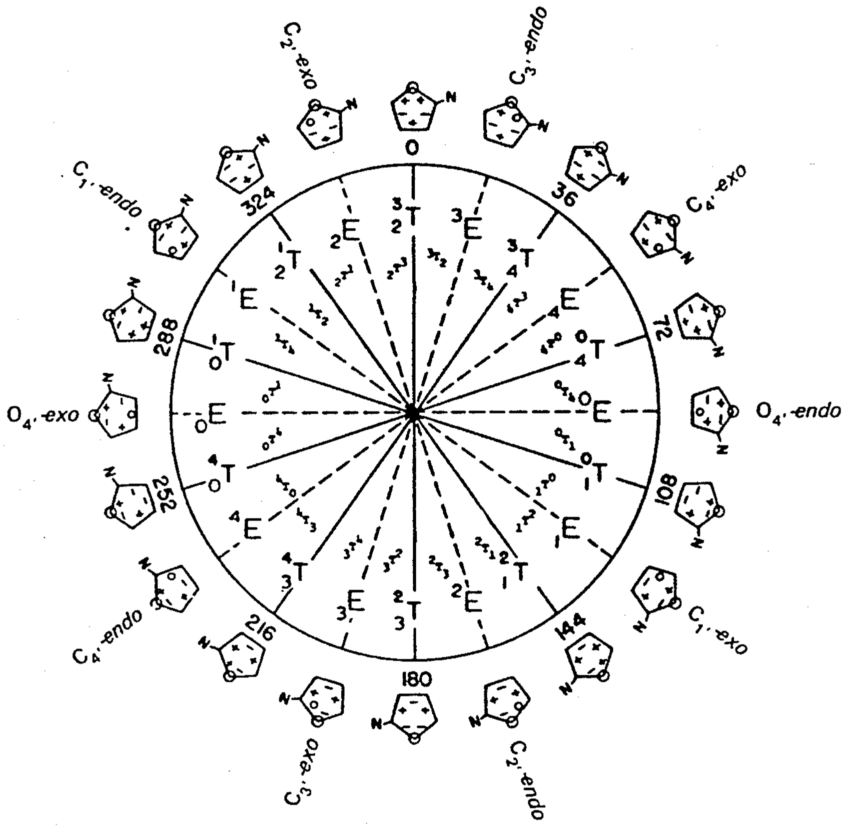


Fig. 9 Pseudorotation cycle of the furanose ring in nucleosides. Values of the phase angle *P* are given in multiples of 36°. Envelope and twist forms alternate every 18°. After rotation by 180° the mirror image of the starting position is found. On the periphery of the cycle, the signs of the endocyclic torsion angles are indicated as (+) positive, (-) negative or (0) for an angle of 0°. From Ref. 7.

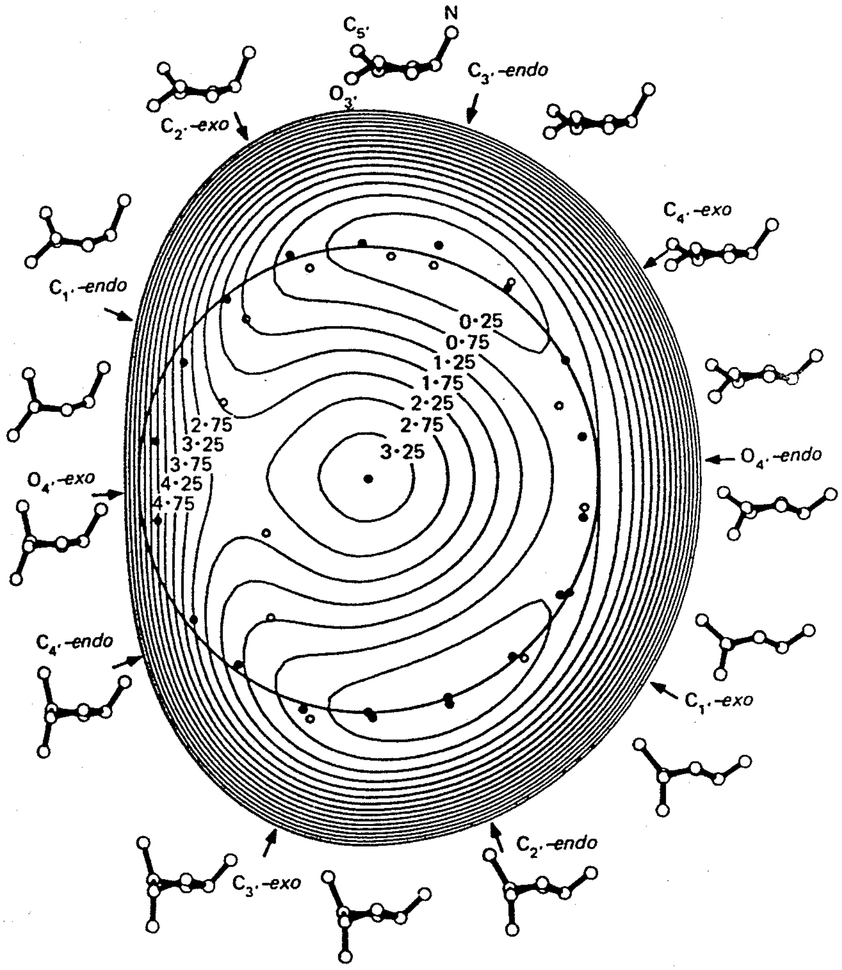


Fig. 10 An energy map of 2'-deoxyribose conformations as a function of P. $P = 0^{o}$ is at the top of the circle, and increases incrementally in a clockwise fashion. The puckering amplitude which is given by ν_{max}, (Eqn. 2) increases radially from the central dot representing a planar 2'-deoxyribose. The ring represents the path of true pseudorotation with $\nu_{max} = 0.4$ Å; filled circles describe energies calculated when all five torsion angles θ_n are constrained to the path of pseudorotation; empty circles correspond to only θ_1 constrained and other torsion angles relaxed. The largest deviations occur at $P \approx 270^{o}$ (O4'-*exo* puckering). Drawn molecules show relaxed geometries and belong to nearest empty circles in the energy plot. Numbers in the diagram represent energies (kcal/mole) above global minimum. From Ref. 8.

1.222 *Syn-anti* Conformation of the Base

In nucleotides, the planes of the bases are almost perpendicular to the planes of their respective sugars, and they typically bisect the O4'-C1'-C2' angle. Therefore, principally two regions of conformational space are occupied by bases, which are described by the glycosidic torsion angle χ: the *syn* and *anti* conformations. The torsion angle χ is defined by atoms O4'-C1'-N9-C4 in purine nucleotides, and by O4'-C1'-N1-C2 in pyrimidine nucleotides (Fig. 11).

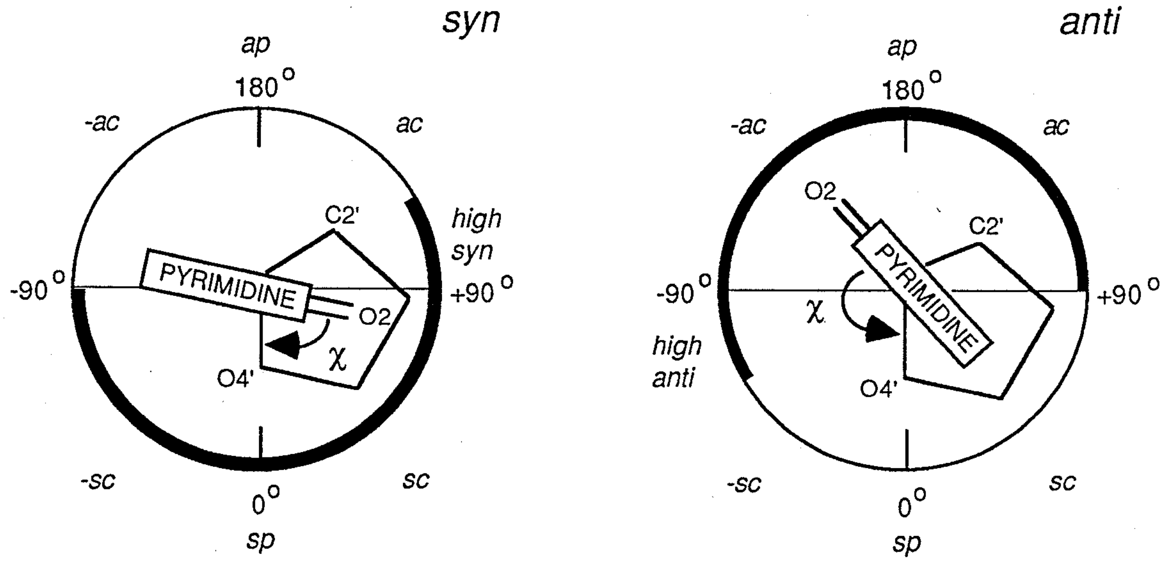


Fig. 11 IUPAC definition of *syn* and *anti* conformational ranges shown for a pyrimidine nucleoside (darkened portions of the circles). χ is defined by the atoms O4'-C1'-N1-C2, and the base is towards the viewer. The *high-syn* (+*ac*) conformation overlaps with the *anti* region and the *high-anti* (-*sc*) conformation overlaps with part of *syn*.

Generally, both pyrimidine and purine nucleotides show a preference for the *anti* conformation, since it moves the bulkier portions of the bases (O2 of pyrimidines and N3 in purines) away from the sugar. χ is also strongly correlated with sugar pucker (again, mainly for steric reasons), which, for pyrimidine nucleosides, falls in the range $-180^{\circ} \leq \chi \leq -138^{\circ}$ for C3'-*endo*, and the range $-144^{\circ} \leq \chi \leq 115^{\circ}$ for C2'-*endo* [9]. There is one important exception to the general preference for *anti*-forms. NMR[10], CD[11] and X-ray diffraction studies[12,13] show that guanine prefers to adopt the *syn*-form in mononucleotides, in alternating d(CpG) tract oligomers and in Z-DNA. This is supported by classical potential energy[14] and quantum mechanical calculations[15], which suggest that this effect arises from a favorable electrostatic interaction between the N2 amine of guanine and the 5'-phosphate anion, associated with polarization of one of the nitrogen non-bonding electrons towards the purine ring.

1.223 The Orientation About the C4'-C5' Bond

In oligonucleotides, the rotation about the C4'-C5' bond (defined by the torsion angle γ), determines the positioning of the 5'-phosphate group relative to the sugar and base. The orientation about γ populates three main conformational ranges, which are all staggered (Fig. 12). However, these three conformational ranges, *+sc*, *-sc* and *ap* are not uniformly populated, showing a strong dependence on the presence of a 5'-phosphate[16-18], sugar pucker and base[19-22]. In purine nucleosides, the *+sc* and *ap* conformers occur with equal frequency in both C2'-*endo* and C3'-*endo* puckers, while *-sc* is only rarely observed with C2'-*endo*. Pyrimidine nucleosides show a strong preference exclusively for *+sc*, regardless of sugar pucker (the *+sc* conformer is stabilized by hydrogen bonding interactions which are described later). The predominance of the *+sc* and *ap* conformations, when compared to the *-sc* conformation, results from the *gauche* effect[23]. The *gauche* effect directs torsion angles of groups X-C-C-Y, where X and Y are electronegative substituents like OH or F, into + or - *gauche* (*+sc* or *-sc*) and tends to avoid *ap*. Thus, +sc and *ap* conformations of γ are favored, since they correspond to the *-sc* and *+sc* conformations of the torsion angle O5'-C5'-C4'-O4' (Fig. 12).

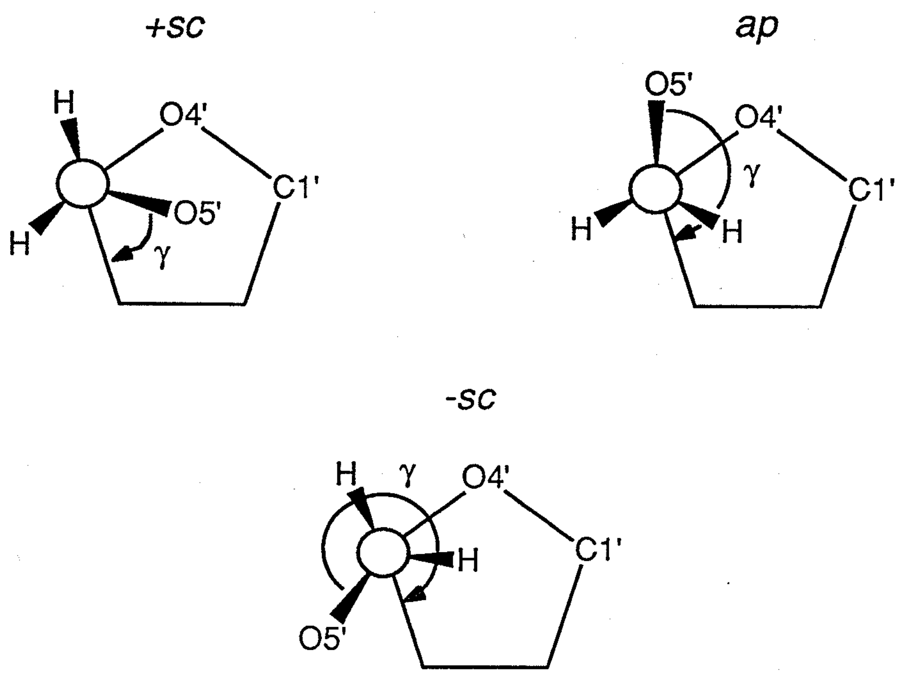


Fig. 12 The three conformations observed in oligonucleotides around the torsion angle γ, viewed down the C5'-C4' bond. The *+sc* conformation, which places the phosphate directly over the ribose sugar, occurs the most frequently. In nucleosides and nucleotides, the *-sc* conformation, which corresponds to an *ap* conformation around the torsion angle O5'-C5'-C4'-O4', is rarely observed due to the *gauche* effect.

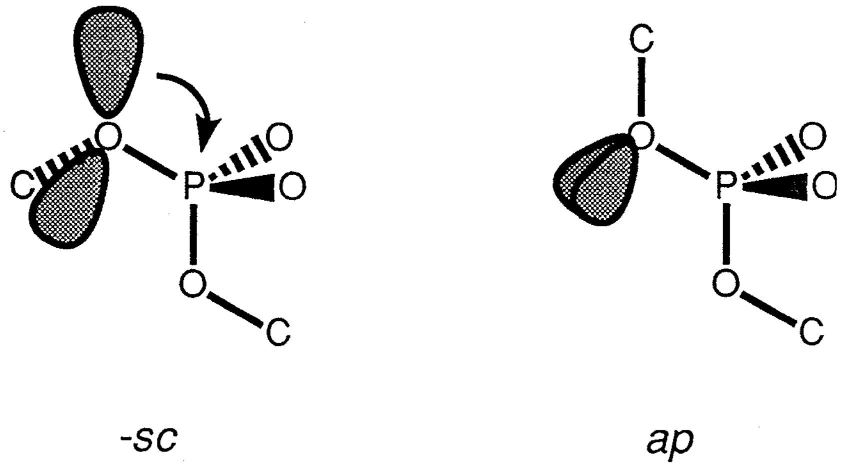
In nucleotides, the *+sc* conformation dominates for both purines and pyrimidines and is independent of sugar pucker. In both nucleosides and nucleotides, the *+sc* conformer is stabilized by hydrogen bonding interactions between C6(H) pyrimidine or C8(H) purine and O5'. The increased selectivity for *+sc* conformations of γ in nucleotides arises from the fact that nucleotides realize stronger base-O5' hydrogen bonding interactions than nucleosides, because the phosphate attached to O5' renders this atom more electronegative than O5'-H in nucleosides. Crystallographic evidence for the existence of hydrogen bonds between base protons and O5' has been observed in the crystal structures of UpA[24,25] and 5-aminouridine[26] (for a review which discusses C-H hydrogen bonding, see Ref. 27).

1.224 Orientation About the C-O and P-O Ester Bonds

The C-O and P-O ester bonds play a crucial role in determining inter-nucleotide geometry, and thus, global oligonucleotide structure. Severe steric hindrance between the phosphate group and sugar moiety restricts the C-O torsion angles, β (C5'-O5') and ε (C3'-O3'), to the *ap*, and *ap* to *-ac* ranges respectively. NMR studies[28] of several dinucleotide phosphates indicate that ε is related to sugar pucker, moving from *ap* to *-ac*, as sugar pucker changes from C3'-*endo* to C3'-*exo*. Inter-nucleotide base-stacking, which is a primary force in determining the structure of oligonucleotides, also directly influences ε[29].

Rotations about the P-O ester bonds, α (P-O5') and ζ (O3'-P) are also restrained, particularly in right-handed helical oligonucleotides, since the formation of the right-handed Watson-Crick double-helical conformation is only possible when α is in the *-sc* range, and ζ lies in the *-sc* to *-ap* range[30]. Theoretical studies on dimethyl phosphate[31] and ethyl methyl phosphate[32], suggest that the orientation of P-O bonds is dictated primarily by the *gauche* effect, showing energy minima at C-O-P-O_{ester} torsion angles in the *+sc*, *+sc* and *-sc*, *-sc* ranges. In a C-Ö-P-O_{ester} system, with two lone pairs of electrons at Ö and an adjacent polar bond $P^{\delta+}–O^{\delta-}_{ester}$, the *+sc* and *-sc* conformations place one of the lone pairs of electrons *ap* to the adjacent polarized P-O bond, where it can donate electrons into the bond (Fig. 13). In the all-*trans* orientation of the C-O-P-O-C chain, the lone-pair orbitals are not anti parallel to the polar P-O_{ester} bonds, a state which is 7 kcal/mol higher in energy than the *+sc* and *-sc* conformations.

When nucleotides are built into polynucleotides, the allowed conformational space for α and ζ narrows considerably, and the *+sc*, *+sc* conformation is forbidden for steric reasons[33,34]. Polynucleotides only adopt the *-sc* conformations in α and *-sc* to *-ap* conformations in ζ (corresponding to a right-handed helix) to avoid unfavorable non-bonded internucleotide contacts (except in rare cases; *i.e.* Z-DNA (left-handed) forming sequences, where changes in other torsion angles force α and ζ into different ranges). Steric arguments dictate[30] that allowed values of ζ in right-handed Watson Crick DNA depend intimately upon sugar pucker; in B-DNA (C2'-*endo* pucker), ζ adopts conformations towards the *-ac* range, while in A-DNA (C3'-*endo* pucker), the *-sc* range is preferred.



-sc *ap*

Fig. 13 Description of the *gauche* effect. The *-sc* conformation of the C-O-P-O group is favored because it places an oxygen lone pair *ap* to the polarized P-O bond, which optimizes electron donation to the polar bond. In the *ap* conformation, the orbitals and the adjacent P-O bond are in *+sc* and *-sc* positions and electron transfer is diminished.

1.23 The Structure of DNA

The primary structure of DNA has each nucleoside joined by a phosphodiester bond from its 5'-oxygen to the 3'-hydroxyl of one neighbor and by a second phosphodiester from its 3'-hydroxyl to the 5'-hydroxyl of its other neighbor (Fig. 6 shows this for a dimer, d(ApT)). Thus, the uniqueness of a DNA strand lies solely in the sequence of the bases. In right-handed double helical DNA (the focus of this dissertation) two single strands of DNA align in an anti parallel fashion (*i.e.* 5'-3' to 3'-5') and interact *via* a series of complementary hydrogen bonds, coiling into a double helix which has the phosphates on the outside, with the hydrogen-bonded base-pairs on the inside stacked upon one another, like coins in a roll.

1.231 Hydrogen Bonding in DNA

Right-handed, double-helical DNA utilizes highly specific hydrogen-bonding motifs to establish the fidelity of DNA transcription and translation. The base-pairing pattern in DNA is called **Watson-Crick pairing**. In the Watson-Crick scheme, adenine pairs with thymine, and guanine with cytosine in complementary oligonucleotide strands. The imino N-H and amino NH_2 groups of the bases act as hydrogen-bond donors, while the sp^2-hybridized electron pairs on the keto oxygens and on the ring nitrogens act as hydrogen-bond acceptors (Fig. 14). Hydrogen-bonds are largely electrostatic in nature, with a +0.2e charge on the hydrogens, and a -0.2e charge on the oxygens and nitrogens. However, hydrogen-bonding interactions differ from pure electrostatic interactions in that they show a strong directional dependence, with donor-H-acceptor angles of 160°-180° providing optimal strength[35], which falls off rapidly outside that angular range. This stringent geometric requirement of hydrogen-bonds is responsible for the high degree of sequence specificity in duplex formation. Interaction energies of Watson-Crick base-pairs, calculated by quantum mechanical methods[36], suggest that A:T base-pairs, which are stabilized by two hydrogen-bonds, are much weaker than C:G, base-pairs, which are stabilized by three (the calculated interaction energies in kcal/mole are -7.00 for A:T and -16.79 for C:G).

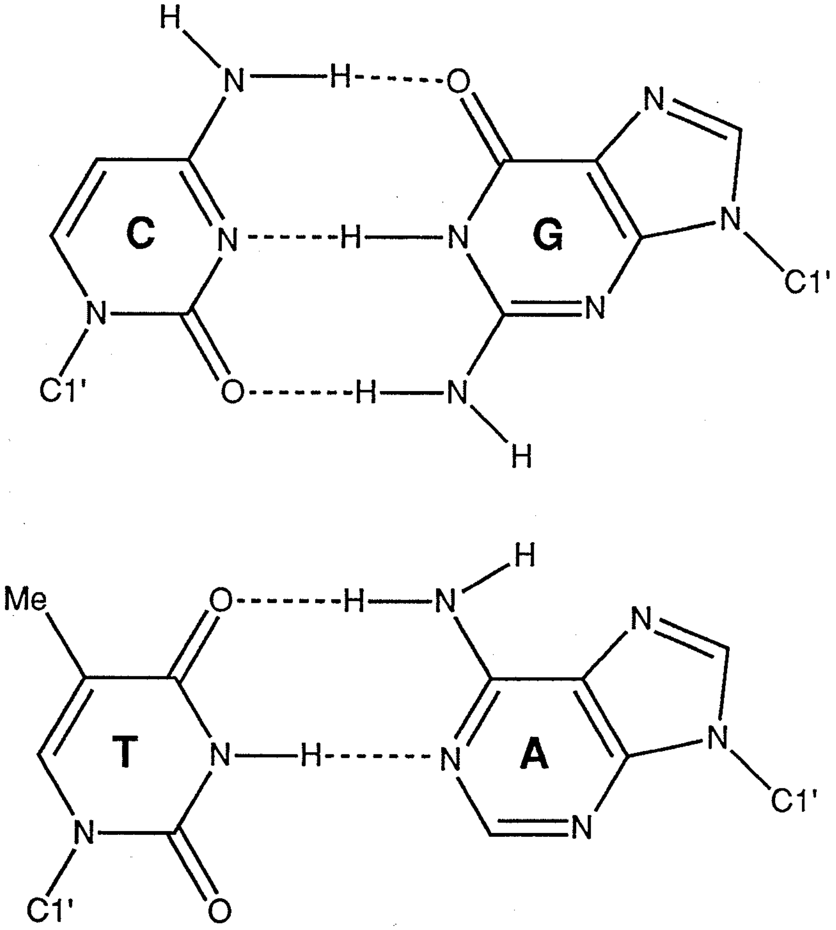


Fig. 14 Watson-Crick base-pairing for C:G (top) and T:A (bottom). Hydrogen-bonds are depicted as dashed lines.

While Watson-Crick base-pairing is the dominant pattern in DNA, other schemes have been observed in the solid state[37]. However, all of the non-Watson-Crick base-pairings in DNA which have been observed are mismatches; *i.e.* either they involve chemically modified bases, or they pair A with a base other than T or G with a base other than C. Since the emphasis of this study is on possible protein recognition factors in unmodified DNA, base mismatches will not be discussed further here.

1.232 Base Stacking Interactions

In aqueous solution, the hydrophobic effect is the major driving force which shapes the secondary structure of DNA. When a non-polar substance is dissolved in water, it disrupts the hydrogen-bonding network between water molecules, which cluster in an ordered manner around the surface of the solute molecule so as to orient their protons away from the hydrophobic surface. Therefore, the water molecules at the water:solute interface lose entropy, which is an energetically unfavorable situation. To minimize this loss of entropy, non-polar molecules in aqueous solution tend to aggregate, minimizing their surface area to volume ratio, thus releasing a greater number of ordered water molecules into the disordered bulk water of the surroundings. In DNA, the minimization of hydrophobic surface area is achieved by maximizing the contact between the hydrophobic surfaces of the bases *via* the formation of a double-helix, with the hydrophilic phosphates on the outside and the hydrophobic base-pairs on the inside, stacked upon one another. However, while hydrophobic interactions are primarily responsible for inducing base-stacking, other forces, namely dipolar and London dispersion forces, provide additional stabilization, and more importantly, convey sequence specificity to the base-stacking process.

London dispersion forces are attractive forces between proximal atoms that arise from electrokinetic interactions between the atoms. At any instant, the electronic charge distribution within atomic groups is asymmetric due to electron cloud fluctuations. These instantaneous dipoles created in one group of atoms polarize the electronic system of neighboring atoms or molecules, thus inducing parallel dipoles which attract each other. Dispersion forces are additive and decrease with the sixth power of distance. They are independent of temperature and increase with the product of the polarizabilities of the interacting molecules[38]. In addition, bases possess permanent dipoles, which, combined with dispersion forces, lead to appreciable sequence dependent variations in stacking energies (Table 2).

1.233 The Two Distinct Conformations of Right-Handed Double-Helical DNA -- A-DNA and B-DNA

Studies of DNA structure in solution and in the solid state (comprehensively reviewed in Ref. 40) have revealed that DNA is capable of a high degree of structural polymorphism, and that global DNA structure is intrinsically related to the extent of hydration of the double-helix. Two distinct conformations of right-handed double-helical DNA have been identified: The A-form, which occurs at low humidity (high salt), and the B-form, which occurs at high-humidity and is the prevailing conformation *in vivo.* A- and B-form DNA adopt globally regular structures which are described by a range of characteristic features. All double helical DNA possess characteristic

Stacked Dimers		Stacking Energies (kcal/mol dimer)
↑ C-G G-C ↓		-14.59
↑ C-G A-T ↓	↑ T-A G-C ↓	-10.51
↑ C-G T-A ↓	↑ A-T G-C ↓	-9.81
↑ G-C C-G ↓		-9.69
↑ G-C G-C ↓	↑ C-G C-G ↓	-8.26
↑ T-A A-T ↓		-6.57
↑ G-C T-A ↓	↑ A-T C-G ↓	-6.57
↑ G-C A-T ↓	↑ T-A C-G ↓	-6.78
↑ A-T A-T ↓	↑ T-A T-A ↓	-5.37
↑ A-T T-A ↓		-3.82

Table 2 Total stacking energies (kcal/mol dimer) for the ten possible dimers with B-DNA geometry obtained from quantum mechanical calculations[39]. Arrows indicate the 5' to 3' direction.

grooves, whose dimensions depend implicitly upon DNA conformation. Grooves form because the two glycosidic bonds branch off from the same side of the helix and because base-pairs are displaced from the helix axis. Thus, the outer envelope of the double-helix is not cylindrically smooth, but displays two grooves of different widths and depths. The **minor groove** is on the O2 (pyrimidine) or N3 (purine) side of the base-pair while the **major groove** is on the opposite side (Fig. 15a). The dimensions of the major and minor grooves are used to describe the overall shape of the DNA. The global parameters, vertical **rise** per base-pair and base-pair **twist**, define the pitch of the helix. Sideways **inclination** of the base-pairs permits the separation of the bases along the helix axis to be smaller than the van der Waals distance, 3.4 Å, yielding a shorter, fatter helix. A detailed description of how DNA structure has been parameterized is provided in Fig. 15.

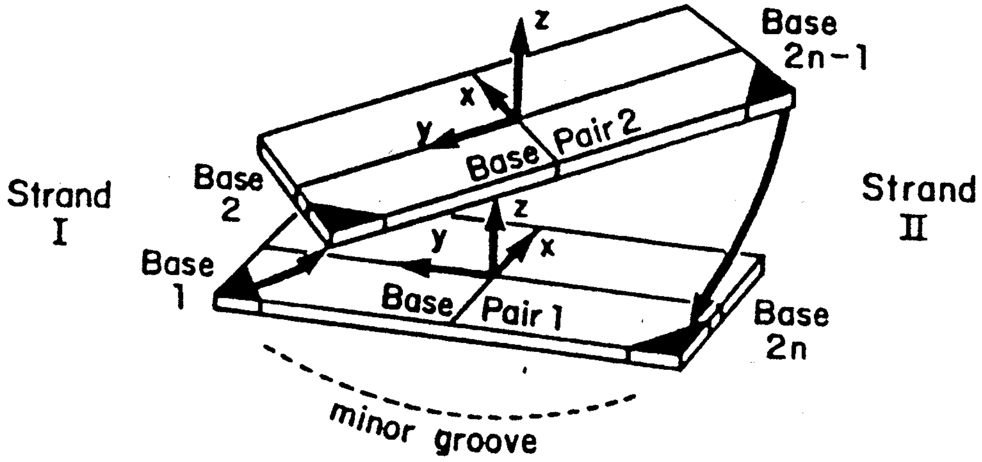


Fig. 15a Definitions of local reference axes (x,y,z) at the first two base-pairs of an n-base-pair double helix. View is into the minor groove. Shaded corners locate attachments of bonds to sugar C1' atoms. Curved arrows indicate 5'-to-3' direction of each backbone strand. Bases along strand I are numbered from 1 to n in a 5'-to-3' direction, and bases along strand II are numbered from n+1 to 2n, also in a 5'-to-3' direction. Base-pairs are numbered from 1 to n, in agreement with the bases of strand I. From Ref. 41.

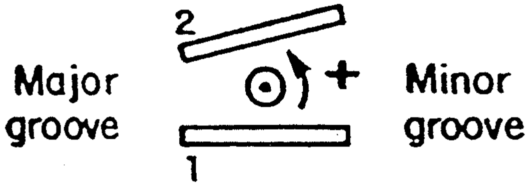


Fig. 15b Positive roll opens the angle between base-pairs towards the minor groove. View is along the long base-pair axis from strand I to strand II, (along -y). Curved arrow represents the rotation of base 2 relative to base 1, about the y axis. From Ref. 41.

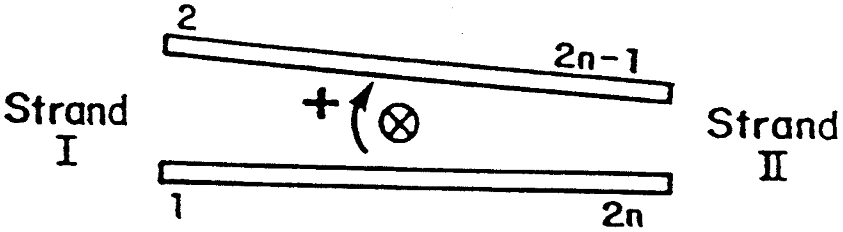


Fig. 15c Positive tilt opens the angle between base-pairs towards strand I. View is from the minor groove side (along +x). Curved arrow represents rotation of base-pair 2 relative to base-pair 1, about the x axis. From Ref. 41.

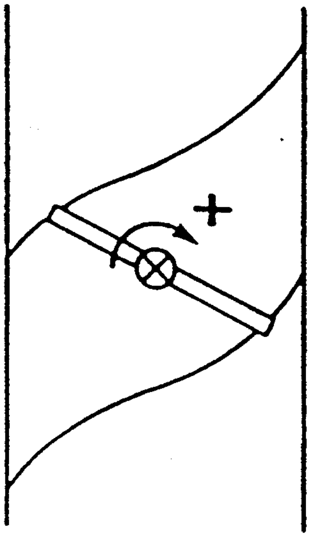


Fig 15d Schematic view of the minor groove of an A-DNA helix, showing the positive inclination of one base-pair. From Ref. 41.

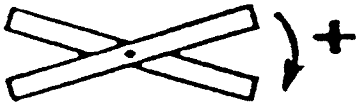
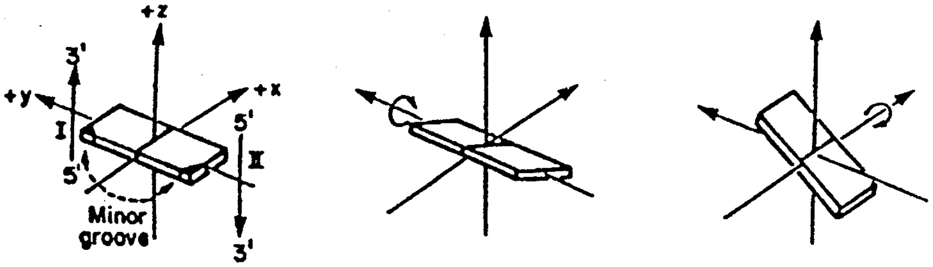


Fig. 15e Illustration of positive propeller twist, viewed down the long axis of a base-pair. The usual propeller twists in A- and B-DNA are negative in this convention, which is used because it is consistent with sign definitions for torsion angles. From Ref. 41.

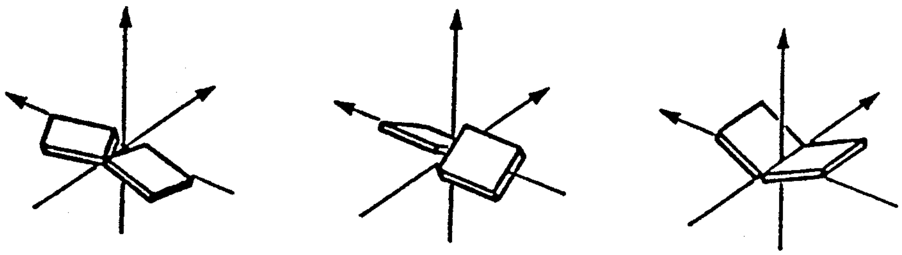
ROTATION



Coordinate frame

Tip (θ)

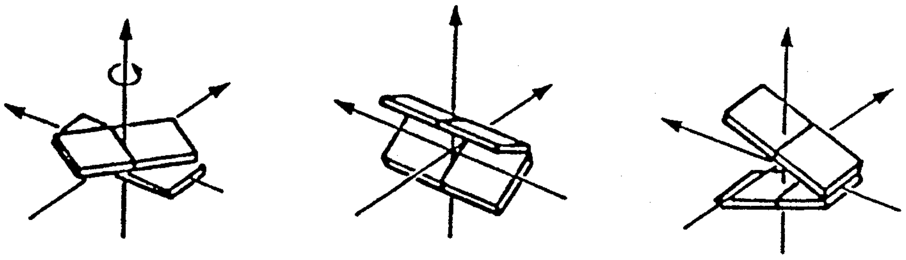
Inclination (η)



Opening (σ)

Propeller twist (ω)

Buckle (κ)



Twist (Ω)

Roll (ρ)

Tilt (τ)

Fig. 15f Definitions of various rotations involving two bases of a pair (upper two rows) or two successive base-pairs (bottom row). In the top row the motions of the two bases are coordinated, and in the middle row their motions are opposed. Columns at the left, center and right describe rotations about the z, y and x axes, respectively. The standard coordinate frame is defined at the upper left. From Ref. 41.

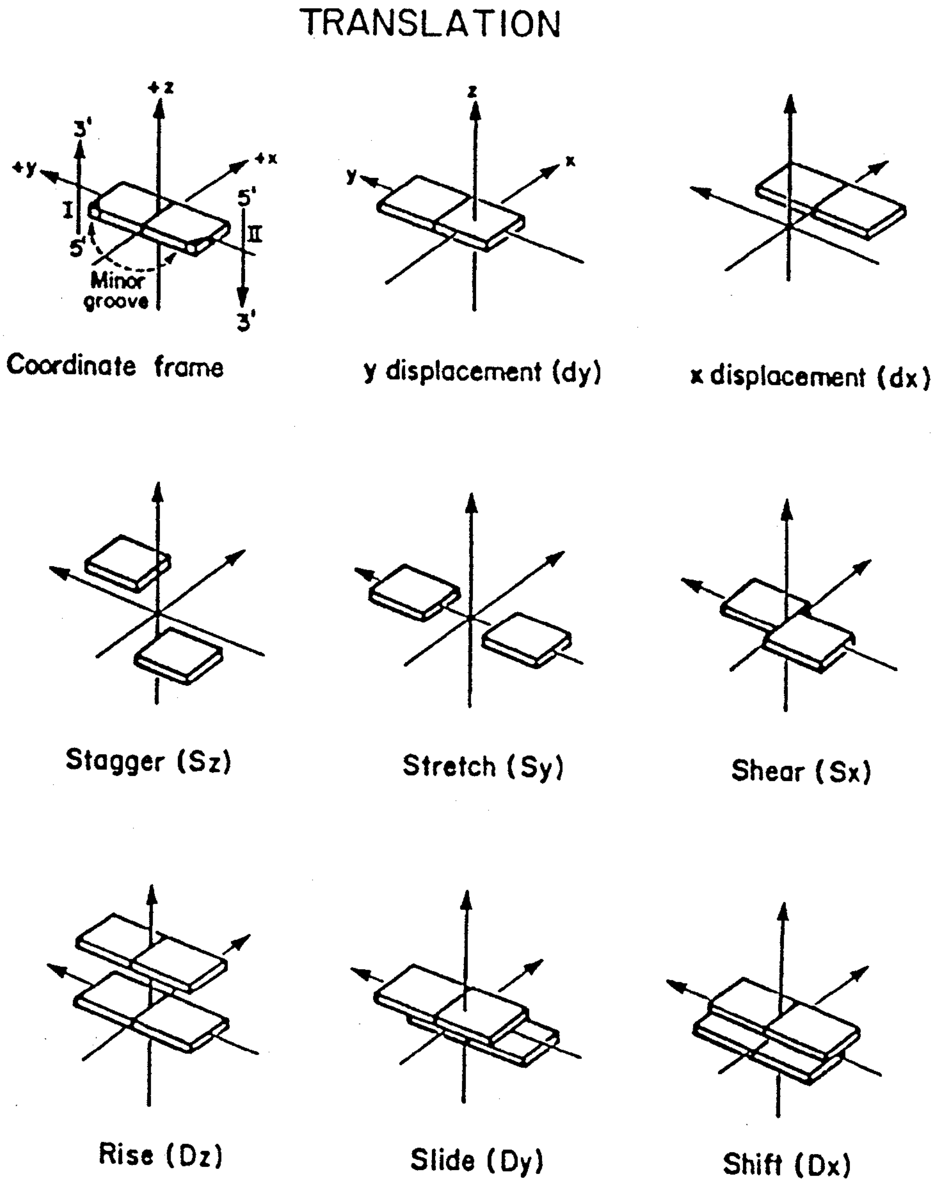


Fig. 15g Definitions of various translations involving two bases of a pair (upper two rows) or two successive base-pairs (bottom row). In the top row the motions of the two bases are coordinated, and in the middle row their motions are opposed. Columns at the left, center and right describe translations along the z, y and x axes, respectively. The standard coordinate frame is defined at the upper left. From Ref. 41.

1.234 A-DNA

The general anatomy of A-DNA follows the Watson-Crick model with anti-parallel, right-handed double-helices. The planes of the sugar rings are parallel to the helix axis, and the phosphate backbone is on the outside of a cylinder of about 24 Å in diameter. The bases are displaced 4.5 Å away from the helix axis, into the minor groove (negative x displacement), creating a 3 Å hollow core down the helix axis, a shallow minor groove, and a deep major groove. There are 11 bases in each helical turn of 28 Å, corresponding to a vertical rise per base-pair of 2.56 Å (Table 3). To maintain the normal van der Waals separation of 3.4 Å between stacked bases, base-pairs roll away from each other by 10°-15° (see Fig. 15b), and successive base-pairs slide (in the negative direction) away from one another. The concerted action of roll and slide causes base-pairs to incline by approx. 20° (Fig. 15d), macroscopically resulting in a shortened helix, which still maintains the mandatory separation of the stacked bases. This sort of helical contraction, inherent in A-form helices, occurs when DNA is in a low humidity (typically less than ~75%) environment, where the amount of available water is insufficient to fully hydrate the individual phosphates in the backbone. The DNA conformationally adjusts in low humidity environments by globally adopting C3'-*endo* sugar puckers, which closes the distance between adjacent intrastrand phosphates to 5.9 Å (Fig. 16). The shorter distance between adjacent phosphates allows water molecules to bridge them, so that overall, less water is required to hydrate the phosphate backbone of the DNA. The sugar-backbone torsion angle conformations adopted by A-DNA are shown in Table 4, and the structure of a canonical A-form double-helix is depicted in Fig. 17.

1.235 B-DNA

The B-form of the DNA double-helix is favored in humid environments (≥ 95% relative humidity), and is thus the predominant conformation adopted by DNA *in vivo*. B-DNA adopts a longer, narrower helical conformation than A-DNA, with a rise per base-pair of 3.4 Å (Table 3), and 10 base-pairs per full turn of the helix (which corresponds to a 34 Å helical pitch, and a 36° base-pair twist). The helical diameter is about 20 Å. Bases stack above their neighbors on the same strand and are perpendicular to the helix axis, which they straddle. The sugar puckers populate the south-eastern quadrant of the pseudorotational cycle (*i.e.* C2'-*endo* to O4'-*endo*), which causes a relatively large 7.0 Å separation of the intrastrand phosphates (Fig. 16). The sugar-backbone torsion angle conformations adopted by B-DNA are shown in Table 4, and the structure of a canonical B-form double-helix is depicted in Fig. 18.

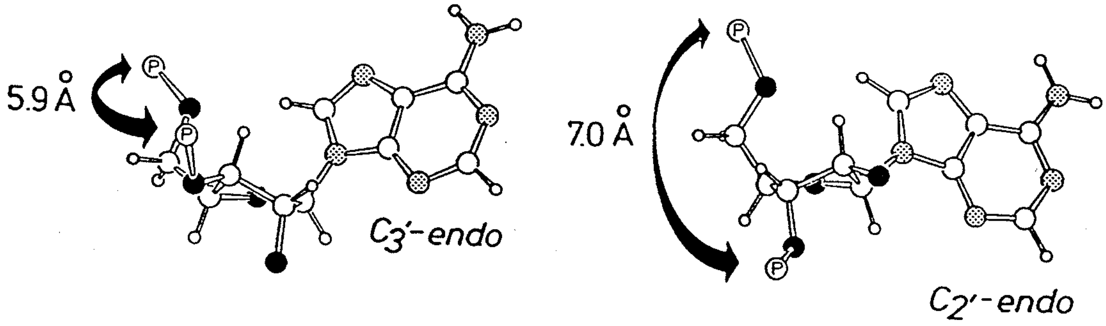


Fig. 16 Nucleotides in C3'-*endo* (left) and C2'-*endo* (right) conformations as observed in A- and B-type polynucleotide helices. Phosphate-phosphate distances are indicated in Ångstroms. From Ref. 40.

Table 3 Average Helix parameters for A- and B-DNA (From Ref. 42)

Structure Type	Helix Sense	Base-Pairs per turn	Twist per bp (deg)	Displacement of helix axis (Å)	Rise per bp (Å)	Inclination (deg)	Sugar Pucker	Groove Width (Å)		Groove Depth (Å)	
								minor	major	minor	major
A-DNA	R	11	32.7	4.5	2.56	20	C3'-*endo*	11.0	2.7	2.8	13.5
B-DNA	R	10	36.0	-0.2 to -1.8	3.3-3.4	-1 to -6	C2'-*endo*	5.7	11.7	7.5	8.8

Table 4 Average Torsion Angles (°) for DNA Helices (From Ref. 43)

Structure Type	α	β	γ	δ	ε	ζ	χ
A-DNA	-50	172	41	79	-146	-78	-154
B-DNA	-41	136	38	139	-133	-157	-102

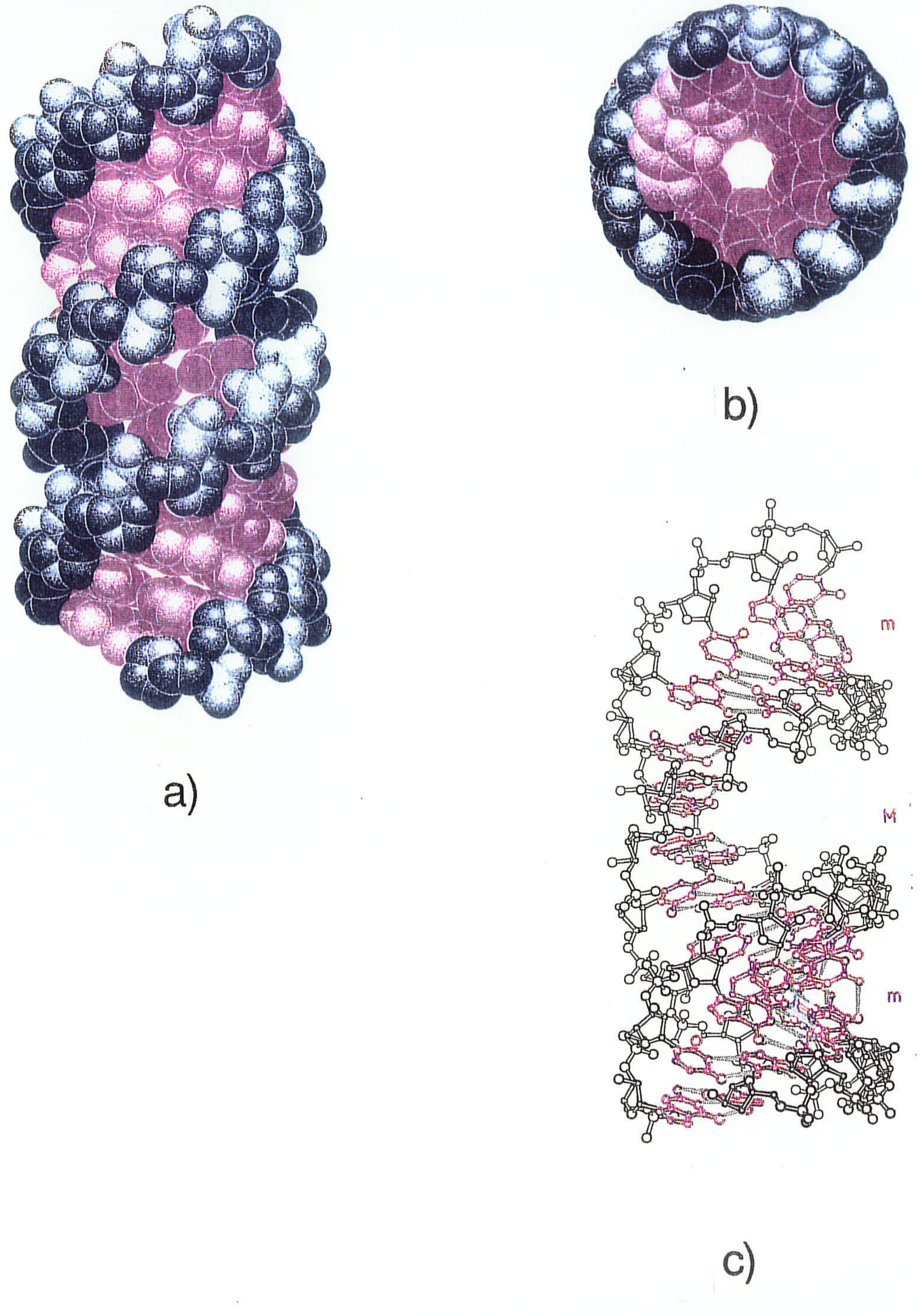


Fig. 17 **a)** A space filling model of A-DNA, viewed parallel to the helix axis. **b)** The same model rotated by 90° so that the view is down the helix axis, showing the 3 Å hole in the center. **c)** To show the relative dimensions of the grooves in an A-form helix, an A-RNA double helix has been tilted by 32°. This view shows the depth of the major groove (M) and the shallowness of the minor groove (m). All diagrams are from Ref. 40.

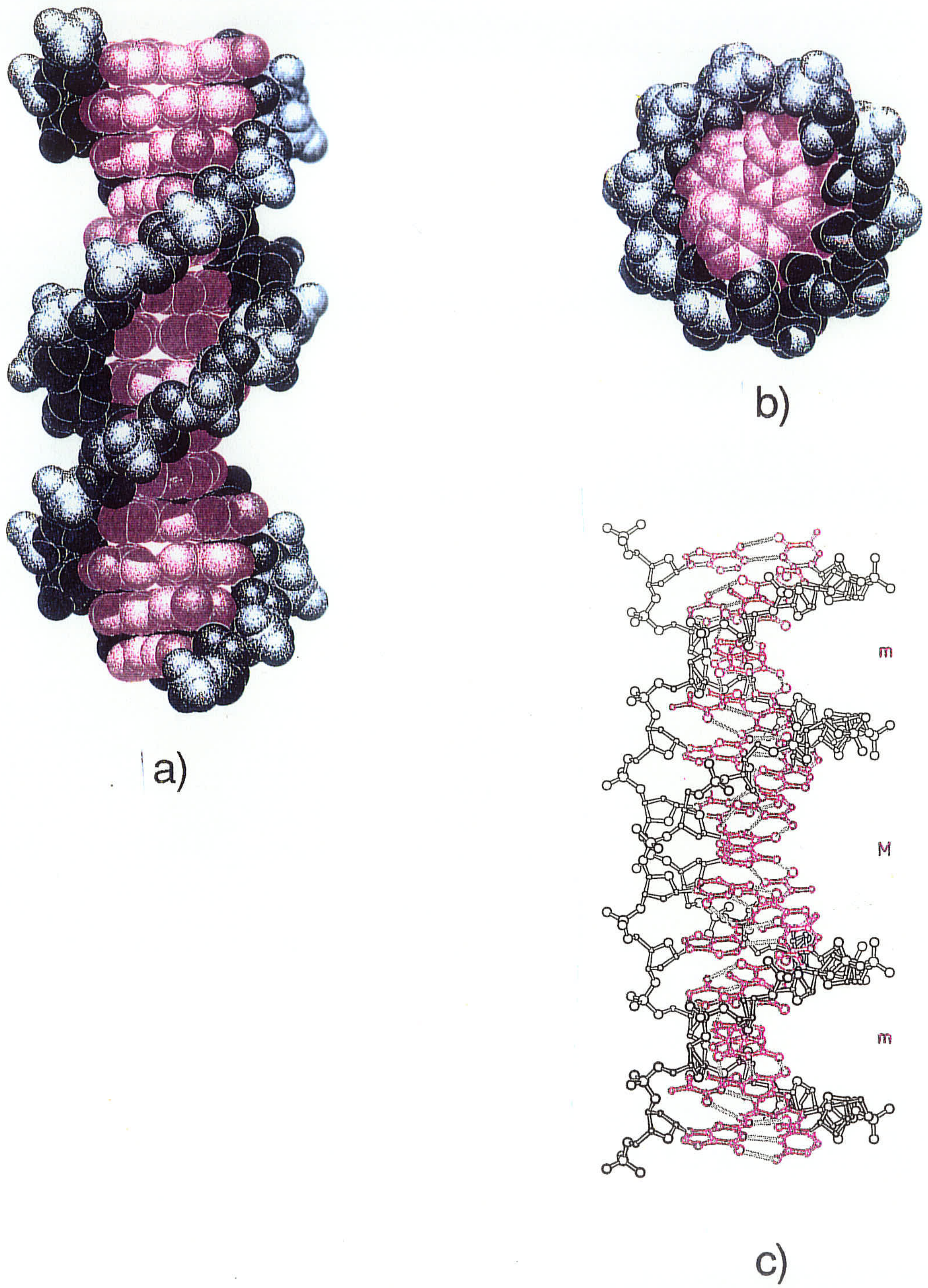


Fig. 18 **a)** A space filling model of B-DNA, viewed parallel to the helix axis. **b)** The same model rotated by 90° so that the view is down the helix axis. **c)** To show the relative dimensions of the major groove (M) and the minor groove (m) in a B-form helix, the model has been tilted by 30°. From Ref. 40.

1.3 Why Study the DNA Fragment, d(CGGTGG)/d(CCACCG)?

No simple code has been found correlating DNA sequence with the binding of regulatory proteins[44-46]. However, distinct families of DNA-binding proteins, employing related structural motifs for recognition, suggest that common sequence elements exist in DNA regulatory regions. The duplex trimer GTG/CAC (hereafter referred to as just GTG) appears approximately five times more frequently than statistically expected in functionally important sites of protein-DNA interaction, suggesting a possible role in the overall regulatory process[47,48]. In complexes of regulatory proteins bound to their consensus sequences the DNA is bent at GTG sites[49-51]. NMR spectra of the *lac* operator sequence[47,48] and physico-chemical results of *cis*- and *trans*-diammine-dichloroplatinum(II) binding to DNA sequences containing GTG[52,53] reveal striking helical perturbations in the region of the GTG trimer. Furthermore, it has been suggested that specific features must be inherent to the trimer GTG to result in the anomalously short lifetime observed by NMR for the T·A base pair in this sequence[54].

The unusual conformational deformability of, and structural features associated with, GTG sequences may significantly contribute to sequence specific protein-DNA molecular recognition processes involving these sequences. To study possible structural anomalies at a GTG site in a DNA control region (while avoiding any attenuating influence molecular symmetry may have on conformational freedom) a non-self-complementary sequence corresponding to a fragment of the *gal* operon containing this trimer has been investigated. Although many DNA control regions that are the target sites of regulatory proteins exhibit pseudo-dyad symmetry in their sequences, it has been demonstrated that repressor-operator contacts in the *lac* system are not symmetric across the pseudo two-fold operator sequence[55], thus highlighting the need for structure determinations of asymmetric or non-self-complementary DNA fragments. The similarity between the *lac* and *gal* repressor and operator sequences suggests that the asymmetry in contacts in the *lac* system may also be observed in the *gal* system.

A review of the Nucleic Acid Database[37] indicates that less than five percent of all DNA crystal structures compiled to date have resulted from studies on non-self-complementary sequences. One reason for the paucity of structural data from non-self-complementary sequences is the increased difficulty in preparation of crystalline samples suitable for x-ray diffraction experiments relative to more commonly studied self-complementary sequences, *viz.* the additional steps of synthesizing and purifying a second strand, and the problem in achieving a crystallization solution devoid of excess single strand contaminant. Compounding the difficulty of growing suitable crystals of non-self-complementary sequences is the inability of these sequences to take advantage of the molecular two-fold symmetry that often facilitates the crystallization of self-complementary strands. With the exception of a nonamer, the few non-self-complementary sequences that have been reported are dodecamers which crystallized as B-DNA duplexes.

B-DNA, the most biologically relevant conformation of DNA, is the best candidate for structural studies into DNA molecular recognition features. Analysis of crystal packing in unmodified B-DNA structures can help characterize possible molecular recognition features by illustrating which sites interact and how they are used in the aggregation of surrounding molecules. However, the family of unmodified B-DNA crystal structures is fairly limited, being represented exclusively by dodecamers and decamers: dodecamer packing modalities most often lead to the space group $P2_12_12_1$ and less commonly to R3, whereas decamer structures are evenly distributed among the space groups $P2_12_12_1$, C2 and P6. Given the individual packing preferences for these duplex lengths, it is reasonable to expect that other duplex lengths that adopt the B-form may display additional types of packing arrangements and, thereby, permit further understanding of the types of intermolecular contacts available for molecular recognition processes.

While structures of duplexes less than ten base pairs in length are common choices for study in other forms of DNA, *viz.* A- and Z-DNA, their under representation in the B family is, likely, a consequence of the difficulty in obtaining suitable crystals (a problem recognized in the report of a self-complementary phosphorothioate hexamer[56] whose structure has been solved, but whose unmodified sequence diffracted only to 8 Å resolution). Difficulties in crystallization of shorter strands in the B-form without the aid of stabilizing drugs may result from a greater variety of, but less discriminating (or possibly sequence dependent) contacts B-type duplexes present to surrounding molecules compared with A- or Z-type duplexes. This explanation is consistent with the fact that B-DNA crystals are generally weaker scatterers of x-rays than A- or Z-DNA crystals. This weaker scattering suggests that B-DNA yields less uniform (and/or, possibly, less rigid) crystal packing arrangements. Limiting crystal structure studies of B-DNA to dodecamers and decamers may restrict the packing types to just those arrangements that will accommodate these strand lengths.

Reported here, is the 2.5 Å crystal structure of the GTG-containing, non-self-complementary hexamer d(CGGTGG)/d(CCACCG), which is part of the interior operator of the *gal* operon (O_I, nucleotides +44 to +49)[57]. The sequence, novel among oligonucleotide crystal structures, is the first example of an unmodified hexamer crystallizing as B-DNA. Furthermore, the structure analysis reveals a new packing arrangement for B-DNA duplexes.

Also described in this dissertation is the key role that the polyamine, spermine, plays in DNA association. This is the first conclusive example of a crystal structure of spermine complexed with native B-DNA. A spermine molecule was tentatively identified bound to a B-DNA dodecamer[58], but it did not refine in a well behaved manner. Spermine, $(NH_3(CH_2)_3NH_2(CH_2)_2)_2^{4+}$, is a member of the family of aliphatic, polycationic ligands present in essentially all prokaryotic and eukaryotic cells that interact with anionic cellular components such as membranes[59,60] and nucleic acids. Through their interactions with DNA, polyamines play an important role in a diverse array of fundamental biological processes (for reviews see Refs. 61-64). Polyamines are necessary for normal cell growth and differentiation[61]. *In vivo* a correlation is found between intracellular polyamine concentration and nucleic acid synthesis, and *in vitro* polyamines enhance both transcription and translation[65,66]. Polyamines also stabilize duplex DNA against thermal denaturation[67-69], condense DNA in chromatin[70] into compact structures, stabilize triplex DNA[71], and induce B-DNA to Z-DNA and B-DNA to A-DNA transitions[63,72]. An additional biological function undertaken by polyamines that is well documented is their ability to increase the fidelity of type II restriction endonucleases[73,74]. However, mechanistic details describing how polyamines accomplish this are still unavailable. In this dissertation a possible explanation for this phenomenon is provided. A second aspect of polyamine:DNA interaction that remains obscure is the specificity of polyamine binding. Such binding may occur through charge-charge interactions in a dynamic, non-specific manner, driven by the release of bound ions[69,75-77], as formulated in the theory of Manning[78]. Alternatively, a static association may occur where polyamines utilize sequence dependent structural properties of DNA in the binding process, employing direct or water mediated hydrogen bonding interactions with DNA bases and phosphate oxygens, and van der Waals interactions with hydrophobic regions of DNA[79-84]. The results of this structural study support the latter theory of association, with spermine molecules associating with specific regions of the DNA *via* water mediated contacts with backbone and base atoms in the major groove, which appear to stabilize some intriguing structural anomalies in the DNA duplex.

Chapter II

Experimental Section

Synthesis, Purification, Crystallization and Crystal Structure Determination of d(CGGTGG)/d(CCACCG)

2.1 Synthesis and Purification

The two strands of the hexamer, d(C(1)-G(2)-G(3)-T(4)-G(5)-G(6)) and d(C(7)-C(8)-A(9)-C(10)-C(11)-G(12)) (where the numbers indicate base position), were purchased from the Regional DNA Synthesis Laboratory, University of Calgary. For crude purification of the respective single-strands, and for separation of excess single-stranded material from annealed duplex, anion-exchange high pressure liquid chromatography (HPLC) was used. All HPLC was done on a Varian LC Star 9010 solvent delivery system coupled with a Varian LC Star 9050 UV/vis detector. For anion exchange chromatography, the system was fitted with a VYDAC 304OL54 analytical column, which incorporates a weak, DEAE-type anion exchange function into a thin polymer film covering the surface of a silica matrix. To protect the column, the system was also fitted with a teflon solvent filter and a 2 cm guard column, which was simply a shorter version of the main separating column. All samples were injected into the system using a 500 μL injector loop. The solvents used as eluants were: A) 80% 0.125 M sodium phosphate - 20% acetonitrile (v/v, pH 6.0), and B) 80% 0.375 M sodium phosphate - 20% acetonitrile (v/v, pH 6.0). Acetonitrile was added to the eluants to compensate for the slightly hydrophobic nature of the ion exchange material. The eluant flow rate, and the optimized solvent program used for purification are detailed in Table 5. Once optimized, purification of the respective single strands, and the separation of excess single-stranded DNA from double-stranded DNA were accomplished in a single step. HPLC profiles of each single-strand and of annealed duplex are shown in Fig. 19.

Time (min.)	% A	% B
0-25	100	0
32	40	60
34	40	60
39	100	0
41	100	0

Table 5 Solvent program for the crude purification of d(CGGTGG)/d(CCACCG) by anion-exchange HPLC. The flow rate used was 2.0 mL/min. (A) is 80% 0.125 M sodium phosphate - 20% acetonitrile (v/v, pH 6.0), and (B) is 80% 0.375 M sodium phosphate - 20% acetonitrile (v/v, pH 6.0). Levels of (A) and (B) were varied linearly between steps in the program.

DNA samples were prepared and purified as follows: **1)** Two 5-6 mM solutions of the individual DNA strands were prepared by dissolving 4.0-5.0 mg of each strand, respectively, in distilled, deionized water (dH_2O, 400 μL). **2)** Before injection into the HPLC, volumes of the single-strand solutions corresponding to approximately equi-molar amounts of each strand were combined in an Eppendorf vial, stirred, and the volume of the new solution containing both strands was made up to 500 μL with dH_2O, and allowed to sit for ~ 5.0 min. to anneal the strands (waiting for 5 min. did not seem critical, as the complementary strands annealed almost immediately). **3)** Samples were injected into the HPLC using a 500 μL syringe, and the peaks corresponding to the hexamer DNA duplex in successive runs were collected in 10 mL polypropylene test tubes and stored in a freezer.

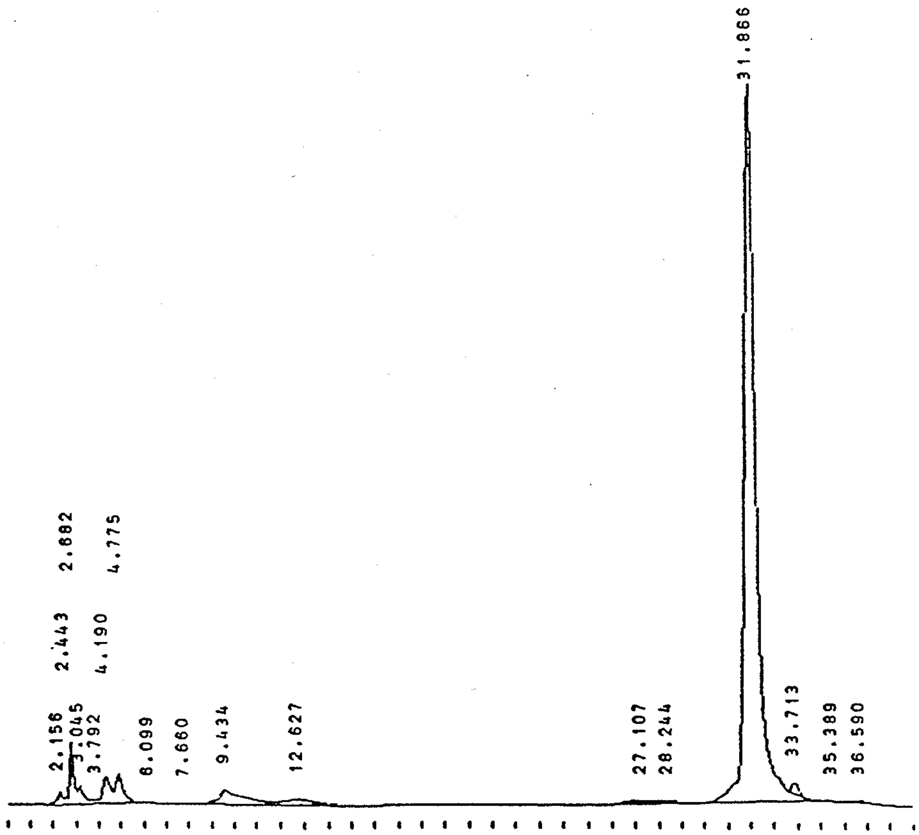


Fig. 19a Anion-exchange HPLC trace of a 50 µM solution of d(CGGTGG) in dH_2O, using the solvent program from Table 5.

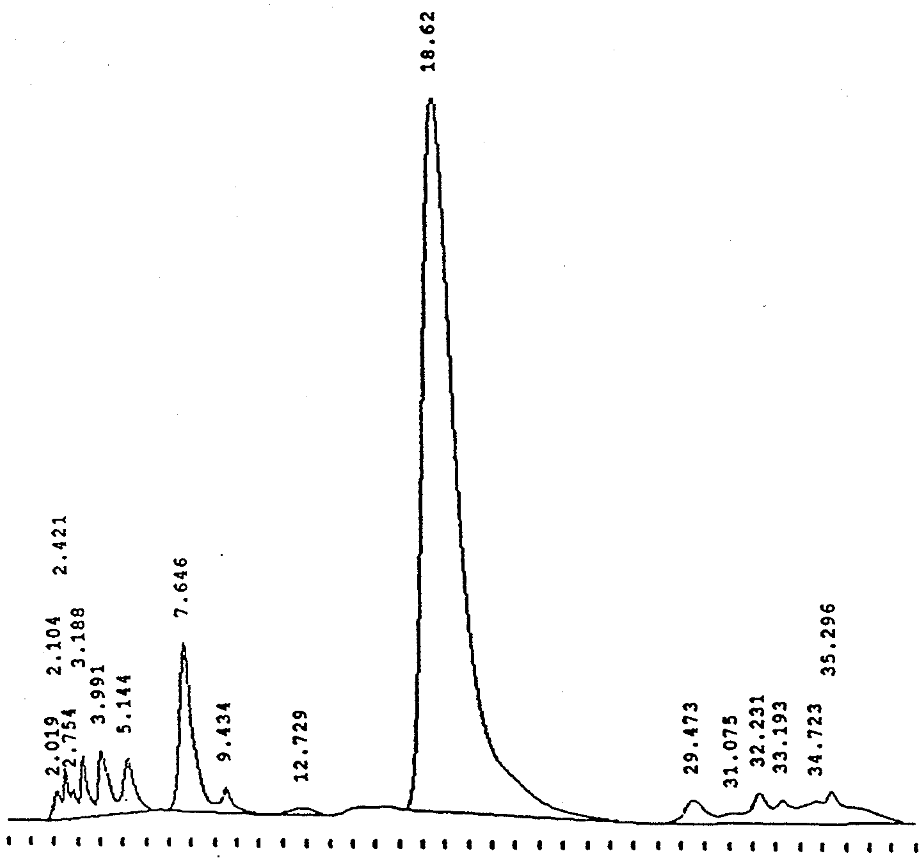


Fig. 19b Anion-exchange HPLC trace of a 50 μM solution of d(CCACCG) in dH_2O, using the solvent program from Table 5.

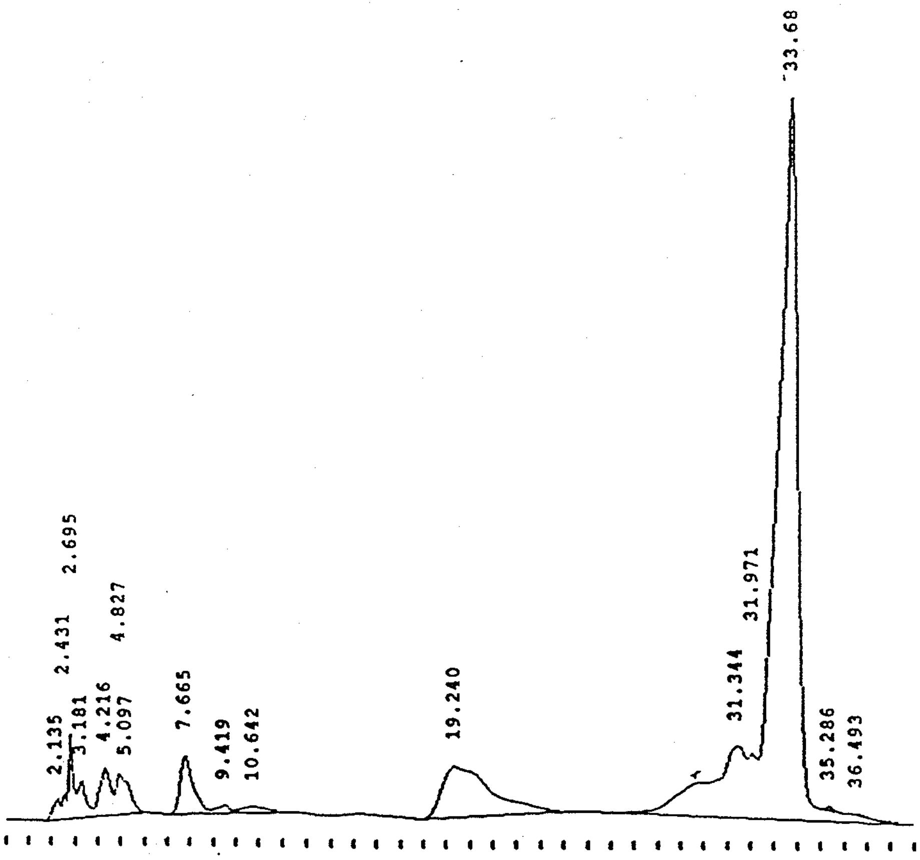


Fig. 19c Anion-exchange HPLC trace of a 50 μM solution of the annealed duplex, d(CGGTGG)/d(CCACCG) in dH_2O (with d(CCACCG) slightly in excess), using the solvent program from Table 5.

The volumes of each of the single-strand stock solutions required to yield equi-molar amounts of each strand (and thus no large excess of one strand after annealing) were determined empirically by titrating one strand against the other on the HPLC. Mixtures of the single-strand solutions with volume ratios were employed that maximally suppressed the respective single-stranded peaks on the HPLC trace. In addition, these mixtures were always made to contain a slight excess of the d(CCACCG) strand, since the elution time of d(CCACCG) differed greatly from the elution times of both d(CGGTGG) and the annealed duplex, respectively (the d(CGGTGG) strand and the duplex had similar elution times, and were difficult to resolve). In a single run, the most material that could be purified without a loss in resolution was about 0.5 mg (0.3 μmoles) of each single-strand. After all of the material was processed, test-tubes containing the purified duplex DNA solution were pooled and concentrated to a 3-4 mL volume on a Savant Speed Vac Concentrator, connected to a TRIVAC high-vacuum pump and two sequential dry-ice:ethanol cold-traps. To avoid problems caused by the precipitation of sodium phosphate (from the HPLC eluants), further concentration of the DNA solution was not attempted. Desalting of the DNA solution was carried out using Millipore Sep-Pak C-18 cartridges (one for every 2 mL of DNA solution), and the following protocol:

1) The cartridge was washed with 10 mL HPLC grade acetonitrile, followed by a wash with 10 mL dH_2O.

2) 2 mL of 0.5 M ammonium acetate were added to the cartridge.

3) Using a 1 mL polypropylene syringe, the DNA solution was added to the cartridge, and the eluted material was collected and re-injected on to the cartridge three times to assure quantitative binding of the DNA to the cartridge.

4) The cartridge was washed with 10 mL dH_2O to remove all traces of salt.

5) The DNA was eluted using five 1 mL fractions of a 60% HPLC grade MeOH:40% dH_2O (v/v) solution.

6) The eluant containing the desalted DNA was concentrated on a Speed Vac Concentrator to a fine, white powder (about 6.5 mg were obtained).

The purity of the DNA after anion-exchange HPLC was 70%-80% (determined by analysis of a small amount of the purified sample by anion exchange HPLC using the solvent program from Table 5). Further purification of the DNA was required, since it was later found that crystals did not grow from samples of DNA that contained more than 5% impurity (determined by HPLC) mainly in the form of an excess of one of the single strands. A second round of purification of the DNA using anion-exchange HPLC was then attempted, with little success. More effective, however, was reverse-phase HPLC, carried out using a Hamilton PRP-1 semi-preparative column, fitted with a teflon filter and guard column. Samples were loaded on to the column *via* a 500 μL injector loop. The solvents used as eluants were A) 0.1 M triethyl ammonium acetate (TEAAC) (pH 7.0), and B) HPLC grade acetonitrile. The flow rate and solvent program used for purification are shown in Table 6. A sample HPLC run is shown in Fig. 20. The semi-pure DNA was prepared and purified as follows: **1)**

The DNA (6.5 mg) was dissolved in 1 ml dH_2O. **2)** 50 μL aliquots were taken from this solution, diluted to 500 μL with dH_2O, and injected. **3)** Peaks corresponding to the double-stranded DNA were collected in 10 mL polypropylene test tubes, pooled, concentrated on a Speed Vac Concentrator down to about a 1 mL volume, and desalted by the same protocol as shown for the DNA obtained from the initial anion-exchange purification. The desalted DNA was concentrated on a Speed Vac Concentrator to a fine, white powder, which was greater than 95% pure (this is evidenced by the HPLC trace of a sample of the pure DNA, shown in Fig. 21). The final yield of pure DNA was 3.1 mg. The measured UV absorption characteristics of the hexamer duplex in dH_2O are; λ_{max}, 256 nm and ε_{256}, 8.0 x 10^4 L x mol^{-1} x cm^{-1}. The measured ε_{260} of the hexamer is 7.8 x 10^4 L x mol^{-1} x cm^{-1}. This corresponds well with the ε_{260} value of 8.1 x 10^4 L x mol^{-1} x cm^{-1}, derived from empirical calculations[85] on a hexamer duplex with an identical sequence to the DNA used in this study.

Time (min.)	% A	% B
0-5	90	10
10	70	30
15	50	50
20	90	10
22	90	10

Table 6 Solvent program for the final purification of d(CGGTGG)/d(CCACCG) by reverse-phase HPLC. The flow rate used was 2.0 mL/min. (A) is 0.1 M triethyl ammonium acetate (pH 7.0), and (B) is HPLC grade acetonitrile. Levels of (A) and (B) were varied linearly between steps in the program.

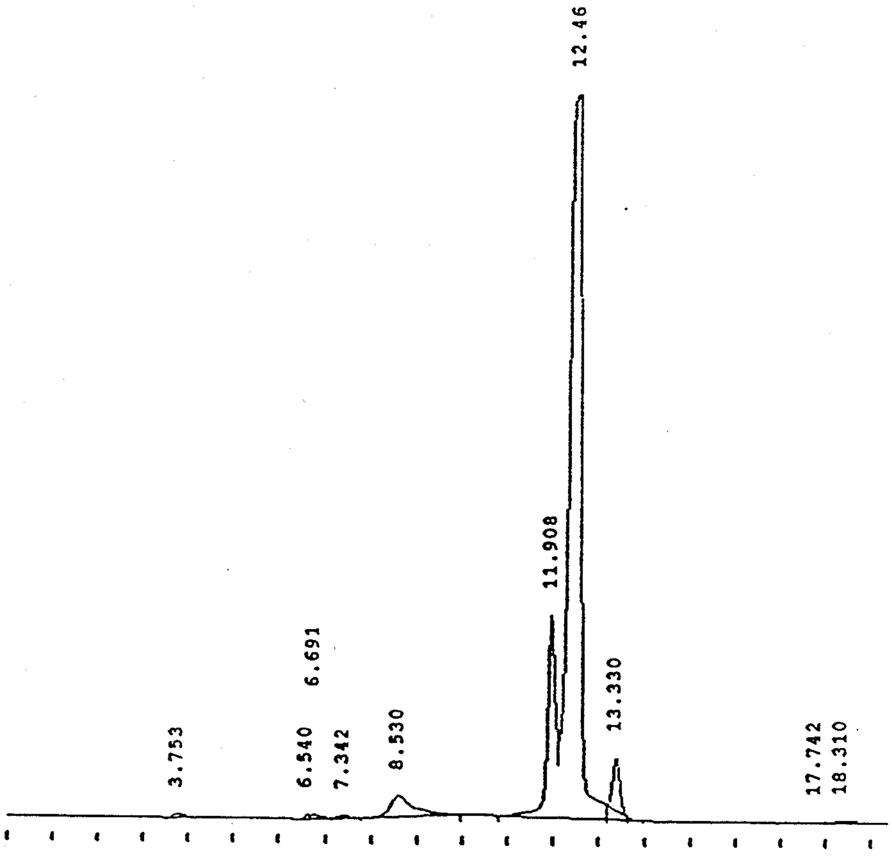


Fig. 20 Reverse-phase HPLC trace of a 200 μM solution of the semi-pure annealed duplex, d(CGGTGG)/d(CCACCG) in dH_2O, using the solvent program from Table 6.

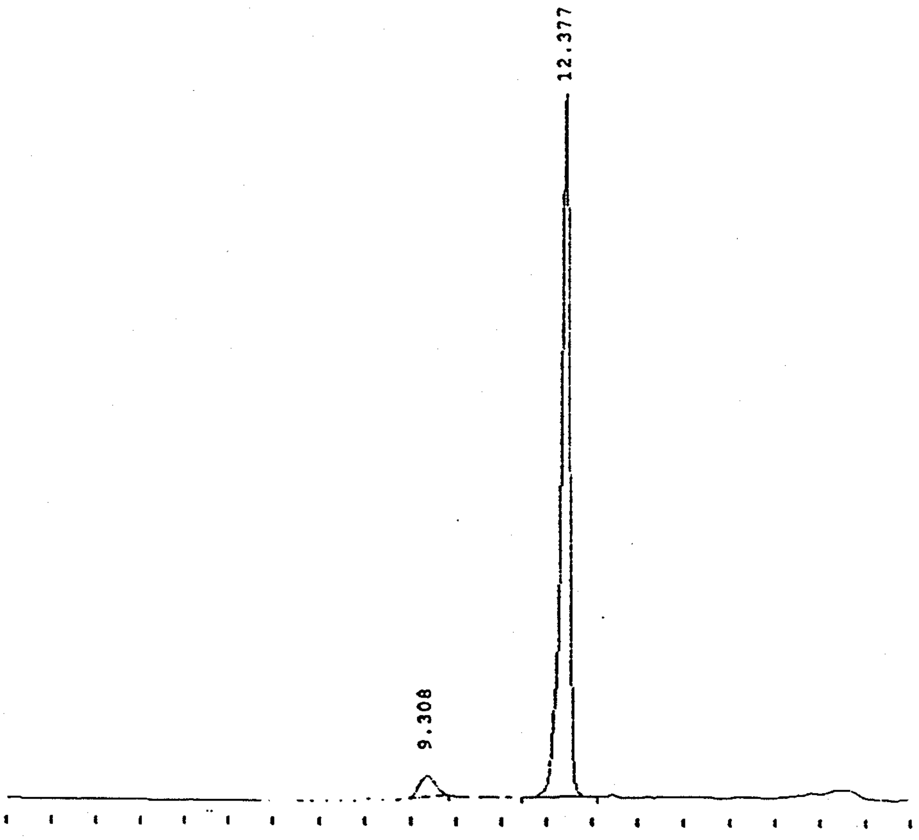


Fig. 21 Reverse-phase HPLC trace of a 20 μM solution of the fully purified DNA hexamer, d(CGGTGG)/d(CCACCG) in dH_2O, using the solvent program from Table 6. The area under the impurity peak is less than 5% of the area of the DNA peak.

2.2 Crystal Growth

The purified DNA, d(CGGTGG)/d(CCACCG) (hereafter referred to as DHEX), was crystallized using the hanging drop vapor diffusion method. In this method, a hanging drop containing an aqueous solution of DNA, various co-factors and a water soluble, non-volatile precipitant is suspended in a sealed chamber over a reservoir filled with a solution that is identical in all respects to the droplet solution (minus the DNA), except for the precipitant concentration, which is made higher in the reservoir (Fig. 22).

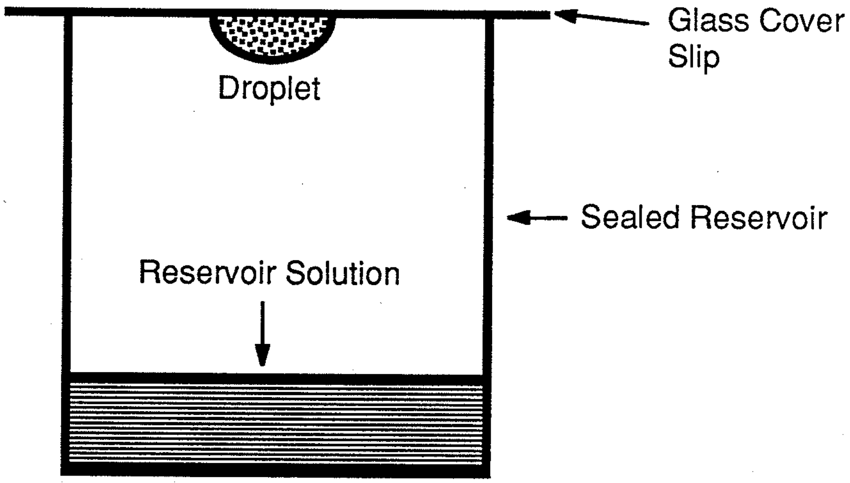


Fig. 22 A schematic diagram of a vapor diffusion cystallization setup. A DNA-containing droplet, with about 1/100 the volume of the reservoir solution is suspended from a glass cover slip and sealed in the reservoir. The concentration of precipitant is higher in the reservoir solution, causing the diffusion of water out of the droplet, leading to precipitation/crystallization of the DNA in the droplet.

Over time, water diffuses out of the droplet and into the reservoir to equalize the concentrations of precipitant in the droplet and reservoir. Dehydration of the droplet increases the concentrations of precipitant, DNA and co-factors, and eventually leads to supersaturation of the droplet in DNA (and co-factors), and to precipitation. The formation of DNA crystals and not disordered precipitates upon supersaturation depends critically on a number of factors, including rate of diffusion, purity of the DNA and co-factors, relative concentrations of co-factors and DNA, temperature, molecular properties of the DNA and the co-factors, pH and ionic strength to name a few. The exact combination of these factors required to yield diffraction quality DNA crystals must be determined empirically, by screening a wide range of conditions. Conditions are screened by varying the relative concentrations of DNA, co-factors and precipitant, as well as the droplet:reservoir precipitant gradient, over a reasonably fine grid. If certain conditions produce crystals, the crystals may be too small or of insufficient quality for diffraction measurements. In this case, a finer grid is screened around the conditions that produced the crystals, or the co-factors are changed to increase the chances of obtaining crystals of diffraction quality. In this section, the crystallization of DHEX will be described in detail, including the selection of co-factors and starting conditions, the range of conditions sampled, and the experimental details.

Crystals of DHEX were grown from droplets containing aqueous solutions of DHEX, a precipitant and co-factors. The precipitant used was 2-methyl-2,4-pentanediol (MPD). MPD is an ideal precipitant because it is non-volatile and water soluble, but sufficiently hydrophobic so as not to solvate DNA.

The co-factors used in DHEX crystallization droplets were:

1) Sodium cacodylate (dimethyl arsenate) as a buffer. Buffer solution is required to constrain the pH of the droplets to biologically reasonable values (around 7.0). Media that are too acidic or basic can alter the ionic form of the bases, or actually chemically degrade the DNA. Sodium cacodylate is the most commonly used buffer for DNA crystallizations because it is a stable compound that only weakly interacts with DNA. Also, it has a favorable buffering strength in the pH ranges of 5 to 7 (pK_a=6.19), and strong anti-bacterial activity.

2) Magnesium chloride as a counter-ion to the negatively charged DNA. Cations are essential for maintaining the electrostatic neutrality of nucleic acids in solution. Cations play an important role in the crystallization process, since they both stabilize DNA against thermal denaturation[67], and facilitate the condensation of DNA into compact structures[70].

3) The polyamine, spermine (see Section 1.3). Spermine, like all polycations, stabilizes compact nucleic acid structures. However, spermine, with four positive charges, differs from metal cations in its ability to interact simultaneously with several duplexes. This capacity is maximized both by the polymorphic nature of the interactions (electrostatic, hydrogen bonding and hydrophobic) of spermine and by the length (approximately 15 Å). Spermine is able to mediate interactions between DNA molecules in ways that are unlike effects of other cations, and in the case of DHEX, was critical to the success of crystallization.

Before setting up crystallization trials, an approximate measure of the solubility of DHEX in a representative crystallizing solution had to be determined, so that appropriate starting conditions could be chosen. The reagents used in the solubility test and all crystallization setups were; Certified A. C. S. grade magnesium chloride hexahydrate (Fisher), spermine tetrahydrochloride ((N,N'-bis[3-aminopropyl]-1,4-butanediamine)tetrahydrochloride, Sigma), Cacodylic acid, sodium salt hydrate (sodium cacodylate, Aldrich), Gold label 2-methyl-2,4-pentanediol (MPD, Aldrich) and distilled, deionized H_2O (dH_2O, obtained locally). The solution chosen for the DHEX solubility test contained; 1 mM DHEX, 10 mM $MgCl_2 \cdot 6H_2O$, 5 mM Spermine, 30 mM sodium cacodylate, and 15% v/v MPD. A 5 μL droplet of this solution was placed on a silanized glass cover slip (25 mm x 25 mm, Fisher), weighed, and observed under a microscope at 30x magnification while uncovered. After about 10 minutes, a brownish precipitate began to appear, indicating that the saturation point of the droplet solution had been reached. At this point, the droplet was immediately re-weighed. From the change in weight, the final concentrations of the reagents in the droplet were calculated. From this test, it was determined that saturation of a 2 mM DHEX crystallizing solution occurs between 30% and 35% MPD (v/v) at room temperature. Therefore, for the initial set of trial crystallizations, a reservoir MPD concentration of 35% (v/v) was used. Since crystals of nucleic acids are frequently unstable at room temperature, all crystallizations were equilibrated at 4° C. Lower temperatures also slow the equilibration process and nucleation kinetics, which can increase the chances of forming larger, more ordered crystals[86].

All crystallizations were set up on plastic Linbro tissue culture 24-well plates, with 1.7 x 1.6 cm wells (3.5 mL capacity). Droplets were suspended on Fisher 2.5 x 2.5 cm glass cover-slips, which were silanized prior to use with a dimethyldichlorosilane solution. The glass cover-slips with the droplets were sealed over their respective wells with Dow Corning high vacuum silicone grease. In each crystallization trial, 5 μL droplets were suspended over 500 μL reservoir solutions. Droplets were made by mixing an appropriate volume of a prefabricated "drop stock" containing a mixture of MPD, $MgCl_2$, sodium cacodylate, spermine and dH_2O, which, when mixed with a solution of DHEX in dH_2O, yielded the desired concentration of all components in the droplet. Concentrated solutions of spermine, sodium cacodylate, *etc.* were used to make the drop stock, and the correct proportions calculated using a spreadsheet program. Inherently, it is extremely difficult to precisely measure and work with volumes on the order of 1-5 μL, which was necessitated in this experiment by the small droplet sizes used. Therefore, a much larger volume of drop stock for each set of conditions was made than was actually required in a droplet (0.5 mL of drop stock was typically made for each set of conditions, using Pipetman pipettes), to maximize precision and accuracy. MPD, which is highly viscous, was always added last to the drop stock, and the entire solution was subsequently drawn back into the pipette tip and re-dispersed several times, to insure that the MPD was quantitatively transferred. Since water rapidly evaporates from droplets exposed to the air (saturation can occur in 5-10 minutes), it was imperative to minimize such exposure. Therefore, when a series of droplets were made on a single plate, the separate drops and reservoirs were made and sealed quickly and sequentially. Crystallization set-ups were done at room temperature and transferred into a 4° C cold-room as quickly as possible. The initial crystallization conditions screened, and the sampling intervals for each reagent, are described in Fig. 23.

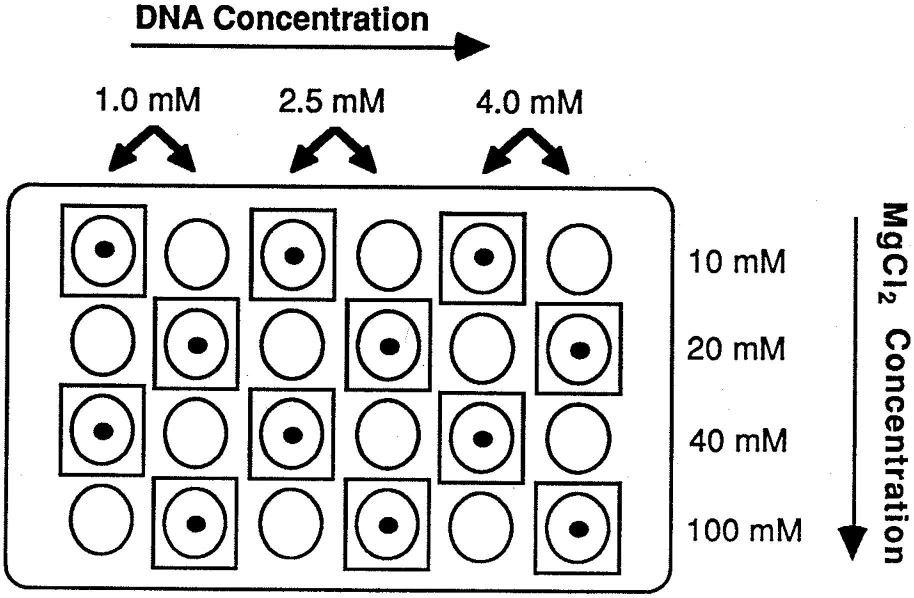


Fig. 23 A depiction of a 24-well crystallization plate. The reservoirs are shown as circles, glass cover-slips as squares and crystallization droplets as small darkened circles. The view is from the top. Twelve crystallization trials are contained on a single plate. DHEX and $MgCl_2$ concentrations were varied as shown. All droplets (5 μL) contained 30 mM sodium cacodylate buffer (pH 7.0) and 7% (v/v) MPD. Spermine concentration was varied from plate to plate, in increments of 1 mM, from 0 mM to 5 mM. The reservoirs with droplets were filled with 500 μL 35% (v/v) MPD solutions, which contained sodium cacodylate, spermine and $MgCl_2$ in the same concentrations as in their respective droplets.

The initial crystallization trials were stored in a cold-room at 4° C, and were checked on a weekly basis. After 1 week, there was evidence of precipitate formation in most of the droplets. After 3-4 weeks, no change in the appearance of the precipitated material in the droplets was observed, but the droplets continued to shrink, which indicated that the saturation points of the droplets was being reached well before the MPD concentration in the droplets reached 35% (v/v). Most of the droplets contained a brown, cloudy precipitate.

However, long, hexagonal barrel-like crystals did form in droplets with concentrations in the ranges 1-2.5 mM DHEX, 10-20 mM $MgCl_2$ and 5 mM spermine. Confirmation that the material in these droplets was crystalline was obtained by examining the crystals in the drops under polarized light. The barrel-like fragments uniformly extinguished the polarized light when rotated in the beam, at 90° intervals. However, the crystals were too small to be used for the acquisition of x-ray diffraction data; the largest crystals had dimensions of about 0.05 mm x 0.04 mm x 0.1 mm. These larger crystals appeared in the droplets with concentrations of 1 mM and 2.5 mM DHEX, 10 mM $MgCl_2$ and 5 mM spermine. It was observed that the crystals were thermally unstable, re-dissolving at room temperature in less than a minute. Thus, all subsequent work with and handling of DHEX crystals was done in a low temperature environment (the 4° C cold-room). In an attempt to grow larger crystals, two strategies were employed. Firstly, a narrower range of droplet conditions were sampled on a finer grid, around the conditions which yielded the larger crystals in the initial trials. Secondly, since it was observed that the droplets were reaching saturation before the MPD concentration had reached 35% (v/v), the MPD concentration in the reservoir was lowered to 25% (v/v). DHEX concentration was varied from 1 mM to 3 mM, in increments of 1 mM, $MgCl_2$ concentration was varied from 7.5 mM to 15 mM in increments of 2.5 mM, and spermine concentration was varied from 4 mM to 6 mM in increments of 1 mM. The MPD concentration in all the droplets was increased to 10% (v/v), to further lessen the MPD concentration gradient between the droplet and reservoir. Sodium cacodylate buffer (pH 7.0) concentrations in all droplets were held at 30 mM, and all crystallizations were equilibrated at 4° C. Large showers of small crystals appeared in many of the droplets within 2-3 weeks and did not grow larger with time. However, in the droplet containing 2 mM DHEX, 10 mM $MgCl_2$

and 5 mM spermine, only a few small crystals were observed after the two week period, and these continued to grow for 6 months. After 6 months growth ceased, and the droplet contained two large diffraction quality crystals of similar size (0.1 x 0.1 x 0.40 mm^3 and 0.15 x 0.15 x 0.3 mm^3, respectively) with well-formed, hexagonal barrel-like habits. A number of crystallizations were then set up using conditions identical to those which yielded the large crystals. These set-ups also yielded large crystals of similar dimensions, and over the course of three years about a dozen such crystals were required to determine unit cell and space group, to perfect x-ray data collection techniques, and finally, to collect the diffraction data which led to the solution of the crystal structure of DHEX. It should be noted that subsequent experiments using seeding techniques and variable droplet sizes did not improve crystal quality or size. In fact, seeding droplet solutions with DHEX microcrystals usually produced only a shower of small, unusable crystals. The crystals used for data collection were mounted in sealed glass capillaries (0.2 mm or 0.3 mm in diameter), and surrounded with small amounts of mother liquor. The mounting of all crystals was done in a cold room at 4° C, using a locally developed mouth pipetting device. Photographs of DHEX crystals are shown in Fig. 24.

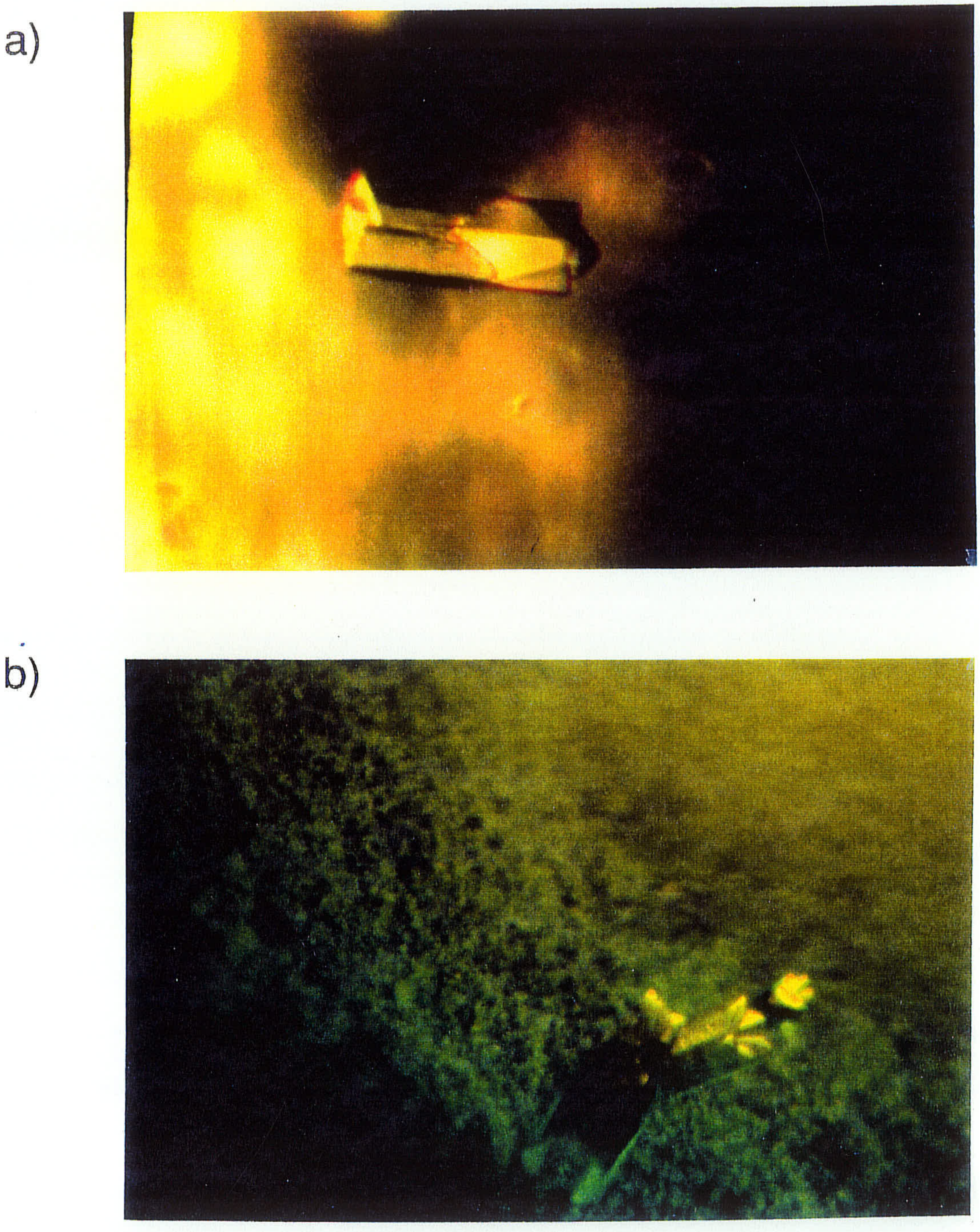


Fig. 24 Photographs of crystals of DHEX. In **a)**, a screw dislocation results in the formation of a "unicorn's horn" at the end of the crystal. In **b)**, the crystallographic symmetry manifests itself macroscopically, in the formation of ridges at the end of the crystal, which are related by a six-fold axis.

2.3 Acquisition of X-ray Diffraction Data

2.31 The Diffraction of X-rays from Crystals

When x-rays strike an atom, the undulating electric field component of the electromagnetic waves induces a sympathetic oscillation in the atom's electrons (the electrons oscillate with the same frequency as the impinging radiation). The oscillating electrons, in turn, coherently re-emit the radiation. When monochromatic x-rays strike a collection of atoms in a molecule or crystal, the re-emitted or "scattered" radiation from the respective atoms forms an interference pattern called a diffraction pattern. The diffraction pattern is the mathematical inverse of the electron density distribution of the diffracting object, and is related to it by a **Fourier transform** (3).

$$F(b) = \int_V f(r) e^{2\pi i (b \bullet r)} dV \qquad (3)$$

f(r) is a function of real space and F(b), the Fourier transform, is a function of reciprocal space (*i.e.* if r represents a vector in real space with length y, then b is a vector in reciprocal space with dimensions proportional to 1/y). From the reciprocal space function F(b), the original real space function f(r) can be regenerated by taking the **inverse Fourier transform** of F(b) (4).

$$f(r) = \int_{V^*} F(b) e^{-2\pi i (b \bullet r)} dV^* \qquad (4)$$

In both equations (3) and (4), integration with respect to V or V^* has to be calculated over the whole space, which is direct space in (3) and reciprocal space in (4).

A mathematical property of Fourier transform pairs of paramount importance in crystallography involves the transformation properties of products of functions, like F(b)•G(b) (5).

$$FT^{-1}[F(b)\bullet G(b)] = f(r)^{\frown}g(r) \tag{5}$$

F(b) and G(b) are functions of reciprocal space, and FT^{-1} represents the execution of an inverse Fourier transform operation on the product of the two functions. The result of taking the Fourier transform of a product function is not simply the product of the respective transforms of F(b) and G(b), f(r)•g(r), but instead, it is the **convolution** of f(r) and g(r) (a convolution operation is represented by the symbol $\frown$). The convolution of two functions is represented by the following integral;

$$C(u) = f(r)^{\frown}g(r) = \int_S f(r)\bullet g(u-r)dr \tag{6}$$

where S is the r space. The integrand is a function of u and r (u is a vector which translates g(r) with respect to f(r)), while the integral is only a function of u. Pictorial examples of convolutions are provided in Fig. 25.

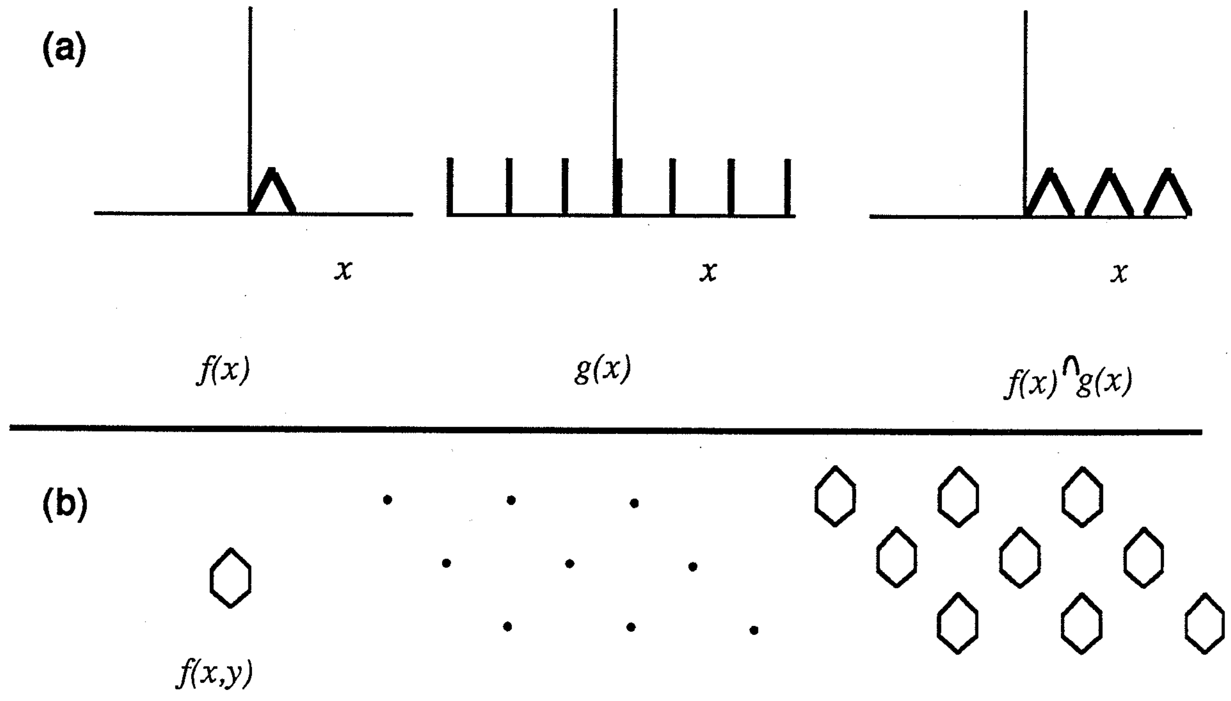


Fig. 25 (**a**) The convolution of the function f(x) with a one dimensional lattice of Dirac $\delta(x-a)$ delta functions. (**b**) The convolution of the function f(x,y) with a two dimensional lattice. A crystal is simply a convolution of a three-dimensional lattice with the asymmetric portion of a molecule or group of molecules.

The transformation properties of products of functions and convolutions of functions can be summarized in two statements: 1) *The Fourier transform of a product of two functions is equal to the convolution of the respective Fourier transforms of the two functions.* 2) *The Fourier transform of a convolution of two functions is equal to the product of the respective Fourier transforms of the two functions.* Therefore, since a crystal is a convolution of a molecular fragment and a three-dimensional periodic lattice, the diffraction pattern from a crystal is the product of the Fourier transform of the molecular fragment and the Fourier transform of the lattice (which is called the *reciprocal lattice*). The relationship between a crystal and its resultant diffraction pattern is illustrated in Fig. 26.

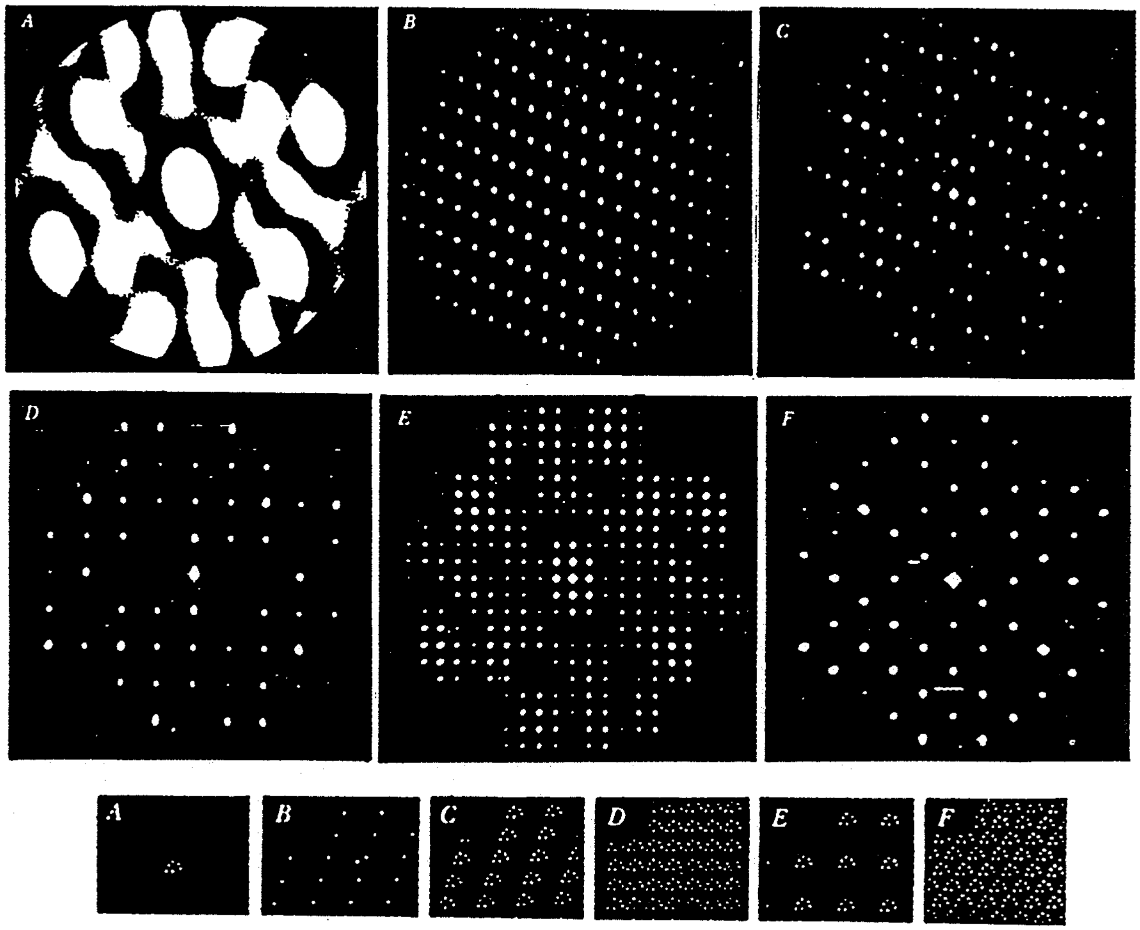


Fig. 26 Diffraction patterns obtained by shining a laser light source on different object masks which represent molecules, lattices and crystals. Diffraction patterns are shown in the upper panel, and the object masks which created them in the lower panel. The different masks are of: **a)** a molecule, **b)** a lattice, **c)** a crystal (convolution of the molecule and lattice). The diffraction pattern from the "crystal" in c) is simply the reciprocal lattice (diffraction pattern b)), sampled at points in the molecular transform in a) that are non-zero. (*i.e.* the product of the respective Fourier transforms of the molecule and lattice). In **d)**, **e)** and **f)** diffraction patterns from masks with a variety of lattice sizes are shown, to illustrate the reciprocity between the dimensions of the diffracting object and the diffraction pattern. Diffraction patterns from larger lattices are more finely sampled. From Ref. 87.

To derive the diffraction pattern from the electron density, a number of simplifications can be made. Since crystals are made up of three-dimensional periodic lattices of molecules, only a very small portion of the crystal is unique. The unique portion of the crystal is contained in the **unit cell**. The unit cell is one of the "bricks" from which the crystal is built; it is always chosen as the ensemble of lattice points comprising the smallest box with the highest metric symmetry. Since the unit cell possesses intrinsic symmetry, only a portion of it is unique. This unique portion of the unit cell is called the **asymmetric unit**. Therefore, to derive the diffraction pattern that would be generated from a given crystal, only the atomic contents of the asymmetric unit need be known. The equation describing the diffraction pattern from a crystal is obtained by taking the Fourier transform of the electron density in the crystal, using an expression which is analogous to equation 3);

$$F(b) = \sum_{j=1}^{N} f_j e^{2\pi i(b \bullet r)} \qquad (7)$$

At a given point in reciprocal space described by the reciprocal lattice vector b, F(b) is a complex function called the **structure factor**, which describes the amplitude and phase of the scattered wave giving rise to the diffraction spot located at b. The electron density in (7) is assumed to consist of a series of N discrete point atoms in the unit cell. The "scattering factor", f_j, for each atom is proportional to the number of electrons in the atom. The integral in equation (3) is replaced by a sum in (7) because the diffraction pattern is only non-zero at reciprocal lattice points. Each reciprocal lattice point in a diffraction pattern represents a discrete spatial periodicity defined by its **Miller indices**, h, k and l, as shown in Fig. 27.

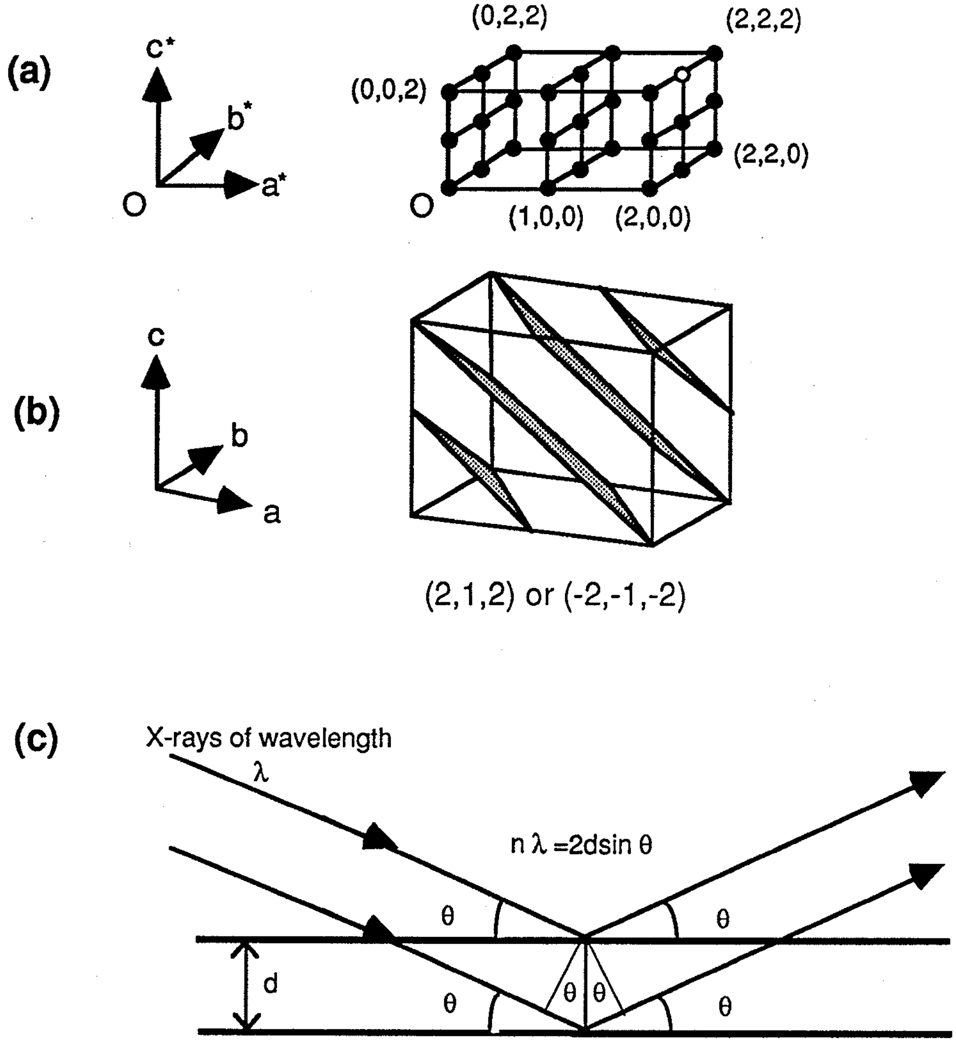


Fig. 27 a) A portion of the reciprocal lattice derived from an orthorombic unit cell. The Miller indices, (h,k,l) of the diffraction spots are shown. The (2,1,2) spot is drawn as a hollow circle. Each of the reciprocal lattice points represent different spatial periodicities in direct space (lattice planes), which correspond to their Miller indices. b) The inverse Fourier transform of the (2,1,2) diffraction spot is a series of lattice planes, which cut the unit cell h times along *a*, k times along *b* and l times along *c*. c) The diffraction process is geometrically analogous to the "reflection" of x-rays from these lattice planes. Constructive interference between waves, and hence diffraction, only occurs when the x-rays strike a given set of planes at certain angles, which are given by Bragg's Law (shown in the Figure).

The explicit form of the structure factor equation is:

$$F(hkl) = \sum_{j=1}^{N} f_j e^{-(B\sin^2\theta)/\lambda^2} e^{2\pi i(hx+ky+lz)} \qquad (8)$$

where h, k and l are the Miller indices of the diffraction spot, and x, y and z are the fractional coordinates along the direct lattice axes *a*, *b*, and *c* of each atom in the asymmetric unit. An extra exponential term has been added to (8) to account for the attenuation of the scattered x-rays resulting from thermal vibration of atoms in the lattice. This extra term corrects for the isotropic (spherical) thermal vibration of the atoms, with $B=8\pi^2\bar{u}^2$, where $\bar{u}$ is the root mean square amplitude of atomic vibration. With high quality x-ray data from crystals that diffract to high resolution, the isotropic thermal correction can be replaced by a tensor, which corrects not only for the extent of thermal motion of atoms, but also for the anisotropy in the thermal motion of individual atoms.

The electron density distribution in a crystal is obtained by taking the inverse Fourier transform of the diffraction pattern using equation (4), which, when written explicitly for this application, becomes:

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl) e^{-2\pi i(hx+ky+lz)} \qquad (9)$$

V is the volume of the unit cell, F(hkl) the structure factor equation for the reflection with Miller indices h, k and l, and x, y and z are fractional coordinates along the unit cell axes *a*, *b* and *c* respectively.

In single crystal x-ray diffraction experiments, the observed quantities are the structure factor *intensities*, $I(hkl) \propto F(hkl) \cdot F^*(hkl) = |F(hkl)|^2$ ($F^*(hkl)$ is the complex conjugate of $F(hkl)$). $I(hkl)$ provides only the magnitude of $F(hkl)$, and not the phase of this complex quantity. Therefore, electron density maps cannot be generated directly from the intensities. There is no way to determine the phases of structure factors experimentally, so that the phases must be determined from a crude model of the electron density in the crystal (called "solving the structure"). Several methods are used to derive initial models, depending on the presence or absence of heavy atoms, the number of atoms in the asymmetric unit and the diffracting resolution of the crystal. The methods used to solve the crystal structure of DHEX will be discussed in a later section. The experimental details of x-ray data collection from DHEX crystals are discussed in the next section.

2.32 Data Collection

While x-ray data sets were collected from a number of DHEX crystals, the information from two of the data sets provided most of the information used to solve the structure. The two intensity data sets, **A** and **B,** were collected on crystals from separate crystallizing drops, but grown under the conditions described in Section 2.2. Data set **A** was used primarily to determine the unit cell and diffraction symmetry, to determine the lifetime of the crystal in the x-ray beam and to narrow down, as much as possible, the space group of the DHEX crystals. Using a crystal measuring 0.15 x 0.15 x 0.25 mm^3, data set **A** was collected on a Rigaku AFC5R diffractometer at 4°C with 9 kW graphite monochromated CuK_α radiation (λ = 1.54 Å) in A. Rich's lab at M.I.T. All work on DHEX crystals prior to data collection was done in a cold-room at 4° C. The glass capillary containing the crystal was mounted in a brass pin using an epoxy resin. The brass pin was then inserted in a Huber goniometer head, and the goniometer head subsequently was mounted on the diffractometer under a 4°C nitrogen gas stream provided by a low-temperature device. The crystal was centered in the x-ray beam manually, using an optical microscope. Once centered in the beam, the orientation of the reciprocal cell axes in relation to the diffractometer axis system had to be determined, as well as the angular width of the reflections. Both the peak width and reciprocal cell orientation matrix were determined by randomly locating and centering on reflections using a search algorithm that executes a random "zig-zag" search of reciprocal space. From the intensity profiles of peaks found using the search algorithm (using ω scans), it was determined that the angular width of low angle reflections (3.5° - 4.5° on 2θ) were 0.5° - 0.6° at half-height. Once eight reflections were located using the search algorithm, they were indexed by the method of Sparks[88]. Using this method, the three shortest noncoplanar reciprocal vectors (found using the search algorithm) are selected to form a reciprocal basis set. The basis vectors are assigned systematically varied combinations of indices. Each combination of indices is used to generate an orientation matrix. Each new orientation matrix is used in conjunction with the angular positioning of the other reflections found in the search to generate indices for those spots. If the indices assigned to the basis vectors are correct, then the orientation matrix derived from these vectors is also correct, and the indices of the other reflections generated using that matrix will be integers (reciprocal lattice points, of course, must possess integer indices). If an orientation matrix generates non-integral indices for a reflection it is discarded. Once the reflections are correctly indexed, the unit cell can be determined from the orientation matrix. Using the Sparks method, the 8 reflections found in the search were indexed and the unit cell of crystal **A** was determined. The unit cell of crystal **A** was hexagonal, with dimensions $a = b =$ 54.51(0.07) Å, $c =$ 42.08(0.11) Å, $\alpha = \beta =$ 90°, $\gamma =$ 120° and $V=$ 108,300(359) Å^3. The standard deviations in the cell parameters were minimized by least squares refinement of the crystal orientation matrix and lattice constants with hexagonal symmetry constraints. To further minimize errors in the unit cell parameters, the orientation matrix was improved by alternating cycles of positional re-centering on reflections in the peak list and searches for new higher angle reflections. The best cell obtained for crystal **A** was: $a = b =$ 54.60(0.07) Å, $c =$ 42.30(0.05) Å, $\alpha = \beta =$ 90°, $\gamma =$ 120° and $V=$ 109,200(317) Å^3 (using 17 reflections between 3.5° and 6.5° on 2θ, that were well distributed in reciprocal space), and the orientation matrix that yielded this cell was used to locate reflections during data collection. Data were collected in the ranges 0 — h, -k — k, -l — l, out to a maximum 2θ value of 30°, corresponding to a maximum resolution of 2.95 Å. Data were collected using ω scans of width 0.7 + (0.3)tanθ degrees. The base scan width of 0.7° was selected on the basis of the peak widths found for the initial search reflections, and the tanθ term compensates for peak broadening as a function of θ. Reflections were scanned at 4° min.$^{-1}$, and were re-scanned up to four times to optimize counting statistics, if the intensity of a given reflection, I, was not at least five times above background, σ(I). $\sigma(I)=(SC + BL + BR)^{1/2}$, where SC=the total scan counts, and BL and BR are the left and right background counts. The intensities of three standard reflections with non-collinear reciprocal lattice vectors were measured at 150 reflection intervals throughout the data collection, to monitor crystal decay. Data collection was terminated when the intensities of the standard reflections had dropped to below 70% of their initial values. All data were corrected for crystal decay, Lorentz and polarization effects. The Lorentz factor takes into account the fact that, due to the geometry of data collection, some reflections are in a "reflecting position" for longer periods of time than others. The value depends on the scattering angle, and on the data collection method. The polarization factor corrects for the loss in the intensity of waves whose electric field vectors are normal to the plane from which they are reflected, relative to those waves whose electric field vectors are parallel to the planes. The polarization correction depends on the state of polarization of the incident x-ray beam and on the scattering angle of the diffracted beam. Ψ–scans of DHEX crystals in prior experiments showed small fluctuations in transmission coefficients (less than 10%), so that the application of an empirical absorption correction to the data from crystal **A** was deemed unnecessary. In total, 3203 reflections were collected from crystal **A**, with 841 observed above $\sigma(I)$. The intensity weighted reciprocal lattice showed 6/mmm symmetry, with R_{sym}=14.3% and a redundancy factor (average number of equivalents per unique reflection) of 4.3, for all data above $\sigma(I)$, yielding 269 unique reflections. The expression for R_{sym} is:

$$R_{sym} = \frac{\sum_{hkl}\sum_{i}\left|\langle I(hkl)\rangle - I(hkl)_i\right|}{\sum_{hkl}\sum_{i} I(hkl)_i} \tag{10}$$

where i is the number of symmetry equivalent reflections being averaged and $\langle I(hkl)\rangle$ is the average value of the symmetry equivalent I(hkl) values. In addition to the Laue symmetry of the diffraction pattern, all reflections 0,0,l, $l \neq 6n$, where n is an integer, were systematically absent, indicating a 6_1 or 6_5 screw axis parallel to the crystallographic *c* axis. The Laue symmetry of the diffraction pattern, in conjunction with the systematic absence along c^*, permitted assignment of the space group of the DHEX crystals to one of the two enantiomorphous space groups, $P6_122$ and $P6_522$. Structure solution and refinement later confirmed the correct space group as $P6_122$.

Data set **B** was collected in our laboratory on a Rigaku AFC6S diffractometer at 4°C with 2 kW graphite monochromated CuK_{α} radiation (λ = 1.54 Å), from a crystal of dimensions 0.20 x 0.20 x 0.35 mm^3. The crystal was kept at 4° C using a fixed nozzle low-temperature device, which cools a nitrogen gas stream by passage through a dewar filled with liquid nitrogen. The temperature is controlled by mixing room temperature gas with cooled gas, and is fine-tuned by a heater mounted upstream of the exit nozzle. Great care was taken to insure that the temperature of the crystals remained between 0° C and 4° C for the duration of the experiment. It should be noted that the temperature reading on the control console of the low temperature device was measured well upstream of the exit nozzle, and was generally about 15° C lower than actual temperature of the gas at the exit nozzle. Therefore, the temperature was always measured manually at the exit nozzle using a mercury thermometer. Mounting, centering, unit cell determination and optimization, peak width determination and data collection were carried out using methods already described for crystal **A**. Only the important differences between the data collection procedures for crystals **A** and **B** will be described. The unit cell was determined using 6 centered reflections, and the cell parameters subsequently refined using 10 reflections between 5.30° and 9.75° on 2θ, yielding as a final cell; $a = b = 54.45(0.06)$ Å, $c = 42.10(0.04)$ Å, $\alpha = \beta = 90°$ and $\gamma = 120°$ (V = 108112(279) $Å^3$.). The quality of crystal **B** was excellent, evidenced by the narrow symmetric intensity profiles of the search reflections (Fig. 28).

```
omega   c/step 0%                  50%              100%
               +---------+---------+---------+---------+
1.930      30. :   *     |         |         |         |
1.980      14. : *       |         |         |         |
2.030      22. :  *      |         |         |         |
2.080      21. :  *      |         |         |         |
2.130      26. :   *     |         |         |         |
2.180      21. :  *      |         |         |         |
2.230      27. :   *     |         |         |         |
2.280      24. :  *      |         |         |         |
2.330      21. :  *      |         |         |         |
2.380      31. :    *    |         |         |         |
2.430      24. :  *      |         |         |         |
2.480      21. :  *      |         |         |         |
2.530      17. : .       |         |         |         |
2.580      21. : *       |         |         |         |
2.630      18. : *       |         |         |         |
2.680      26. :   *     |         |         |         |
2.730      23. :  *<BKG  |         |         |         |
2.780      35. :   *     |         |         |         |
2.830      60. :       * |         |         |         |
2.880     127. :         |       * |         |         |
2.930     270. : CEN>    |         |         |       * |
2.980     281. :         |         |         |         *
3.030     267. :         |         |         |      *  |
3.080     177. :         |         |    *    |         |
3.130     102. :         |   *     |         |         |
3.180      33. :   *     |         |         |         |
3.230      28. :  *<BKG  |         |         |         |
3.280      35. :   *     |         |         |         |
3.330      38. :    *    |         |         |         |
3.380      30. :  *      |         |         |         |
3.430      16. :*        |         |         |         |
3.480      27. :  *      |         |         |         |
3.530      31. :   *     |         |         |         |
3.580      25. :  *      |         |         |         |
3.630      24. :  *      |         |         |         |
3.680      18. : *       |         |         |         |
3.730      27. :  *      |         |         |         |
3.780      21. : *       |         |         |         |
3.830      26. :  *      |         |         |         |
3.880      19. : *       |         |         |         |
3.930      27. :  *      |         |         |         |
               +---------+---------+---------+---------+
```

Fig. 28 An example of an ω scan profile of a reflection from crystal **B**. The value of ω is plotted on the abscissa, and the counts/step plotted on the ordinate.

The angular width of all scanned peaks was 0.19° to 0.25° in ω. Preliminary diffraction experiments revealed that DHEX crystals decay fairly rapidly in the x-ray beam; decay starts after about 3 days, and once started, renders the crystal useless for the acquisition of diffraction data about 1 hour after the onset of the decay. Since all the required preliminary crystallographic information was obtained from crystal **A**, crystal **B** was to be used to acquire high quality data that would be used for structure determination and refinement. To maximize data quality, a data collection strategy was employed which maximized the time spent scanning individual reflections, to increase the signal-to-noise ratio of the data, while minimizing the overall residence time of the crystal in the beam to avoid problems with crystal decay. ω scans were used in data collection, with a base scan width of 0.35° in ω, which was increased with θ in the same manner as for crystal **A**.

To minimize data collection time, only a unique set of diffraction data was collected. Since the Laue symmetry of DHEX crystals is 6/mmm, this corresponded to 1/24 of the reciprocal sphere of reflections, with indices $0 \leq h \leq k$, $0 \leq l$. Data were collected in shells, extending from low resolution, to a maximum resolution of 2.5 Å ($2\theta = 38°$). To optimize counting statistics, reflections were scanned at 2.0° min.$^{-1}$, and then rescanned up to eight times, or until the $I/\sigma(I)$ ratio was greater than 5. Data were corrected for Lorentz and polarization factors, but no corrections for absorption effects or time-dependent decay were required. Data set **B** was the better data set collected from a DHEX crystal, comprising 392 reflections with $F_O > \sigma(F_O)$ between 20-2.5 Å, (1233 unique data were collected, with 719 observed above $0\sigma(F_O)$ and 338 observed above $2\sigma(F_O)$) providing a ratio of 65 data per base pair. While this ratio is lower than anticipated, it is consistent with the inherent weaker scattering of B-DNA crystals (particularly non-self-complementary B-DNA) and is commensurate in percentage data observed with several other published structures[89]. Merging both data sets (R_{merge}=14.7%, calculated using an expression analogous to Eqn. (10) for equivalent reflections between data sets) yielded 414 reflections with $F_O > \sigma(F_O)$ observed between 20-2.5 Å. The slight variability in unit cell dimensions between crystals coupled with the difficulty in properly scaling data sets from the different size crystals militated against using the merged data set. The need for internal consistency and a data set absent of any averaging effects dictated that data set **B** be used for structure solution and refinement.

2.4 Structure Solution and Refinement

2.41 Primary Phasing

As previously discussed in section 2.31, the structure factor phases cannot be obtained experimentally. To determine an approximate set of initial phases, which can subsequently be improved using various refinement techniques, a crude model of the electron density in the crystal must be derived. While the electron density is not directly accessible from the structure factor intensities, all of the information about the electron density in the crystal is contained in the intensities. It is possible to extract the information present in the intensities by taking the Fourier transform of $F(hkl)\bullet F^*(hkl)=|F(hkl)|^2$ directly, using Eqn. (4);

$$P(u)=\int_{V^*} F(h)F^*(h)e^{-2\pi i(hu)}dV^* \qquad (11)$$

P(u) is called the **Patterson function**, u is a vector in real space, and h is a reciprocal space vector with indices h, k and l. From convolution theory, and

Eqn. (11), the following relationship is derived;

$$P(u) = FT^{-1}\left[F(h) \bullet F^{*}(h)\right] = \rho(r) \cap \rho(-r),$$

and from Eqn. (6);

$$P(u) = \int_{V} \rho(r) \bullet \rho(r-u)\, dV \qquad (12)$$

Evident from equation (12) is that the Patterson function, P(u), is the self-convolution of the electron density. P(u) is large for vector translations u between regions in the crystal of high electron density. In crystals of organometallic compounds, for example, the largest Patterson "peaks" correspond to vectors between the metal atoms. To solve the crystal structure of DHEX, the structure factor intensities and the Patterson map were used to orient and translate a DNA model in the unit cell, using methods described below.

From the unit cell dimensions and space group, the volume of the asymmetric unit was calculated to be 9008 Å^3. From the value of 1300-1600 Å^3/base-pair, commonly observed in A- or B-DNA oligonucleotide crystals, it was apparent that the asymmetric unit contained one full hexamer duplex. The most prominent feature on the native Patterson map (Fig. 29) was a column of more than nine flat, elliptical peaks, spaced by 3.4 Å, and inclined by approximately 26° with respect to the crystallographic *c* axis, with the *xy* projection of the column bisecting the crystallographic *a* and *b* axes. On the basis of the Patterson map alone, the following assertions regarding crystal structure were made: 1) The crystal is made up of stacked B-form double helices; 2) The orientation of helix axis of the duplex in the asymmetric unit is coincident with the direction of the column of Patterson peaks; 3) Fulfillment of 1) and 2) require that the duplex in the asymmetric unit be positioned approximately 1.7 Å away from the dyad at [*x*, 1-*x*, 5/12] in $P6_122$ or the dyad at [*x*, 1-*x*, 1/12] in $P6_522$, so that the two-fold symmetry operation would stack a second hexamer on the first, with an interhelix separation of 3.4 Å.

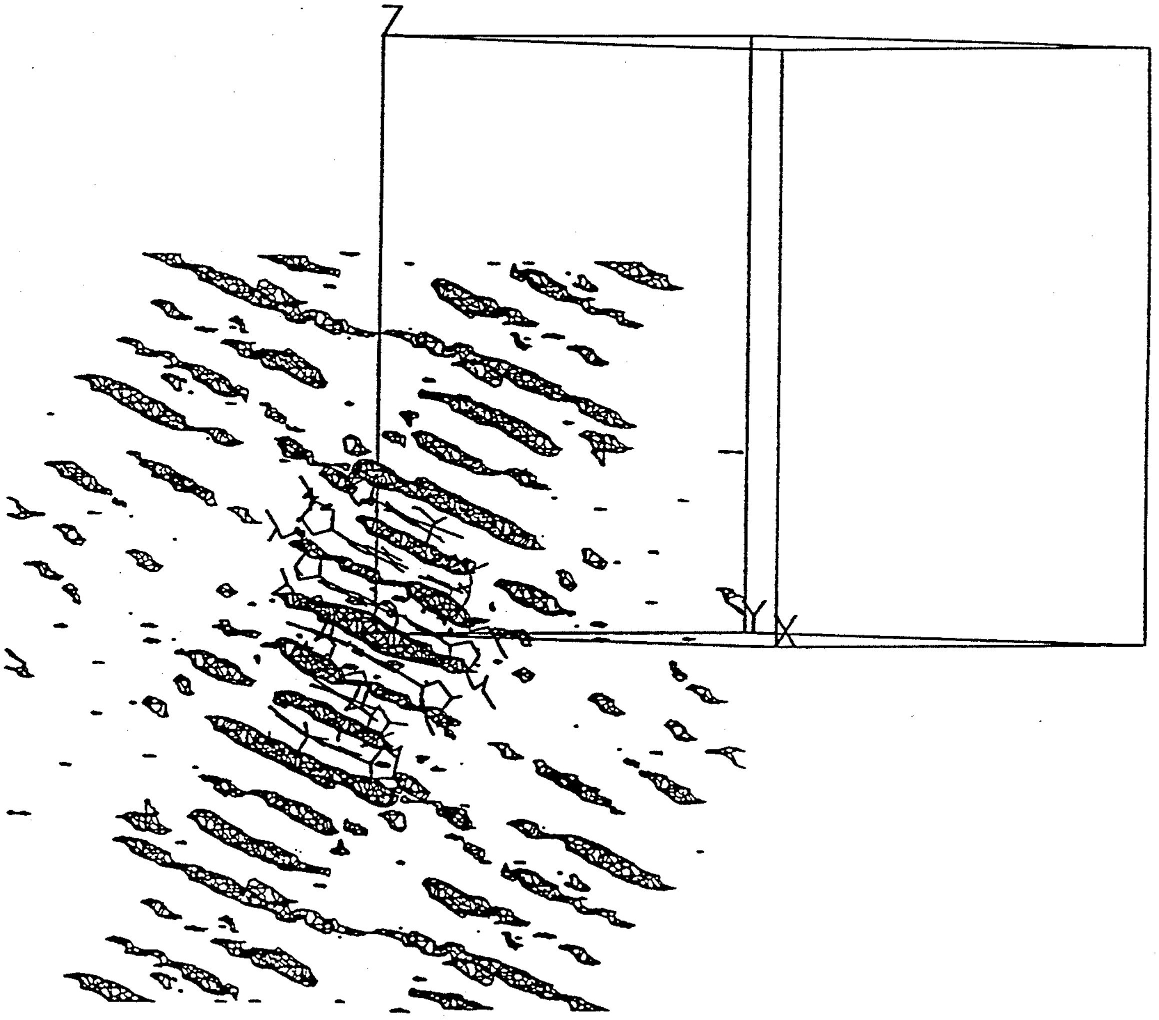


Fig. 29 A 40 Å x 40 Å x 40 Å section of the Patterson map generated using 6.0 Å - 3.0 Å data (above $2.0\sigma(F_o)$) from data set **B**. This map clearly shows the "stacking peaks", spaced by 3.4 Å, which are characteristic of B-DNA. The orientation of the helix axis of the DNA, which coincides with the orientation of a line drawn through the stacking peaks, is clearly visible in this view. The map is shown superimposed on the unit cell, with *c* (labeled "Z") in the plane of the page and *a* (labeled "X")towards the viewer. For convenience of visualizing the stacking distance and orientation, a canonical B-DNA model with the DHEX sequence, and center of mass coincident with the origin is also shown. The model was oriented using the Rossmann and Blow rotation function.

Using the X-PLOR program suite[90], the helix axis of a standard B-DNA model with the proper sequence, adapted from the fiber model of Arnott[43] was first oriented using the rotation function of Rossmann and Blow[91], with data greater than $2.0\sigma(F_O)$ in the range 6.0 - 3.0 Å. Using the Rossmann-Blow method, a B-DNA model with the correct sequence was generated and positioned with its center of mass at the origin of a model unit cell and its helix axis parallel to the *c* axis of the model cell. The axes of the model unit cell were mutually orthogonal (with axial lengths of 60 Å), and the space group symmetry of the cell was P1. The rotational search was carried out by looking for agreement between Patterson functions calculated using the model crystal and the observed data. The function used to evaluate the agreement is;

$$R(C) = \int_V P_1(x) \bullet P_M(Cx)dV \qquad (13)$$

where P_1 and P_M are the Patterson functions derived from the observed data (the "target crystal") and model crystal, respectively, and C is a rotation operator that rotates the coordinate system of P_M with respect to P_1. The volume of integration is spherical, and centered at the origin. The radius of the sphere of integration is chosen to coincide with the maximum molecular dimension of the model. If there exists a sufficient degree of isomorphism between the model structure and the crystal structure, a maximum in the rotation function R(C) indicates a potential orientation of the search molecule in the target crystal. Since the rotation function depends on overlap of Patterson peaks corresponding to intramolecular vectors, the model unit cell axial lengths are made large enough (at least 2 times the largest molecular dimensions) and the model cell symmetry is set to P1, so that the Patterson peaks generated from the search model exclusively represent vectors between intramolecular features. Also, as mentioned above, problems with interference from intermolecular vectors are avoided by setting the radius of integration in Eqn. (13) to the largest molecular dimension of the search molecule. Rotations were performed using the Eulerian angles θ_1, θ_2 and θ_3, which correspond to sequential rotations of the model in a fixed orthogonal system around the *z*, *x* and *z* axes, respectively By convention, a positive rotation corresponds to a clockwise rotation of the model. A search of the entire rotational asymmetric unit $0° \leq \theta_1 \leq 360°$, $0° \leq \theta_2 \leq 90°$ and $0° \leq \theta_3 \leq 60°$, in 1° intervals, revealed a prominent, broad solution, between 2.5σ and 2.7σ above background, at $\theta_2 = 26°\text{-}27°$ and $\theta_3 = 35°\text{-}45°$. With θ_1 held constant, the angular spread in θ_2 and θ_3 corresponds to errors in atomic positions of less than 1 Å in the regions of the search model furthest from the origin. The orientation about θ_1, however, was less well resolved. The angular spread in θ_1 was 25°, and with a central value of θ_1=100°. Since "streaking" of rotation function peaks in θ_1 and θ_3 is commonly observed at low θ_2 values when searching with Eulerian angles, a rotation function was performed using the pseudo-orthogonalized Eulerian angles[92], $\theta_+ = \theta_1 + \theta_3$, θ_2, and $\theta_- = \theta_1 - \theta_3$, which generally yields sharper, more well resolved peaks. A search using the pseudo-orthogonalized Eulerian angles provided the same solution, but did not improve the quality of the map. It is therefore likely that the ambiguity in θ_1 (which was set up in this system to represent the orientation of the helix around the helix axis) arose due to the unusual nature of the Patterson map from this structure. The intense "stacking peaks" dominating the Patterson map obscured the weaker features of the map necessary to provide information about the helical orientation of the molecule in the asymmetric unit.

Since the orientation of the model about the helix axis could not be accurately determined at this stage, the symmetry restrictions on possible translation solutions described above, as well as packing considerations were necessary to properly position the model in subsequent translation searches. Translation searches using the linear correlation coefficient between F_o^2 and F_c^2 as the target function were then carried out in the two possible enantiomorphous space groups, $P6_122$ and $P6_522$, with data from several resolution ranges. The expression for the linear correlation coefficient is;

$$C = \frac{\sum a \bullet b - \frac{(\sum a \bullet \sum b)}{N}}{\{[a^2 - \frac{(\sum a)^2}{N}]^{\frac{1}{2}}\} \bullet \{[\sum b^2 - \frac{(\sum b)^2}{N}]^{\frac{1}{2}}\}} \qquad (14)$$

where $a=|F_o|$ (observed structure factor moduli) and $b=|F_c|$ (calculated structure factor moduli from the model), or, $a=|F_o|^2$ and $b=|F_c|^2$ (structure factor intensities). N is the number of observations. The correlation coefficient runs from -1 to 1, with a coefficient of 1 indicating perfect parameter correlation, and a coefficient of -1 indicating perfect inverse parameter correlation. At the initial stages of the structure determination process, the correlation coefficient is superior to the commonly used R-factor (Eqn. 15) for assessing the correctness of a model.

The expression for the R-factor is;

$$R = \frac{\sum_{h} ||F_o(h)| - k|F_c(h)||}{\sum_{h} |F_o(h)|} \qquad (15)$$

where Fo(h) and Fc(h) are the observed and calculated structure factors, and k is a scale factor. It has been shown[93] that the theoretical value for R which would be obtained by using as a model the proper kind and number of atoms, randomly distributed in an acentric cell, is 0.59. It has also been shown that the square of the residual, $R' = \Sigma \, ||F_o(h)| - k||F_c(h)||^2$, is a linear function of the negative logarithm of the likelihood of an atomic model being correct, assuming that all observations are independent and normally distributed[94]. Therefore, since R and R' are closely related, the R-factor is relatively insensitive to errors (or changes) in the model when it is high (> 0.40), but becomes extremely sensitive to the correctness of the model when R drops (below about 0.30). The general trends in R-factor "sensitivity" are evident from a plot of $10^{-R'}$ *vs.* the likelihood that a structure is correct (Fig. 30). The correlation coefficient, C, changes linearly with the correctness of the model, and is thus extremely useful when R is high. However, when R drops below 0.25, the correlation coefficient approaches 1.0 and is no longer useful.

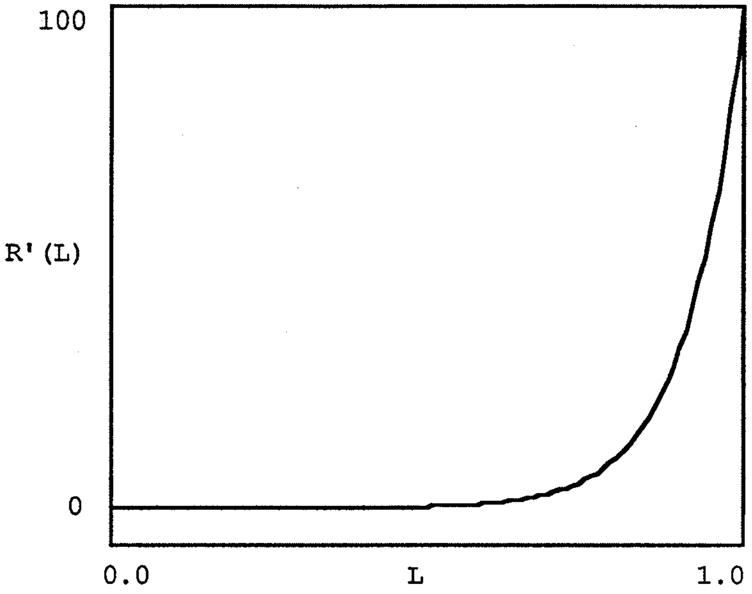


Fig. 30 A plot of $10^{-R'}$, where R' is the square of the residual, vs. L, the likelihood that the model from which the residual is calculated is correct. It is evident from this plot that when R' is high, it is an insensitive measure of changes in the validity of a model.

The translation solutions in $P6_122$ were always at least 3.3σ above the mean of the correlation coefficient maps, and among the top 30 peaks (for a search of the entire translational asymmetric unit; 0.0-1.0 along *a*, 0.0-1.0 along *b* and 0.0-5/12 along *c* (fractional coordinates)), whereas, the best solution in $P6_522$ was 2.0σ above the mean and not within the top 500 peaks. The solutions placed the model (in the only feasible positions from the standpoint of crystal packing), approximately 1.7 Å from the dyad axes at *c* = 5/12 and *c* = 1/12 in $P6_122$ and $P6_522$, respectively. To determine the correct orientation of the model around the helix axis, correlation coefficient searches using 7 - 3 Å data were made (Fig. 31), by rotating the correctly positioned model around its helix axis in each possible space group. The results in $P6_122$ were unambiguously superior to those in $P6_522$. Rigid body refinement (described in the next section) on the $P6_122$ solution, where the entire DNA model was treated as a rigid body, converged at an R-factor of 43.9% using 8.0 - 3.0 Å data above $2.0\sigma(F_o)$. After this point, refinement was continued using least squares techniques, simulated annealing and manual refitting of the model to difference electron density.

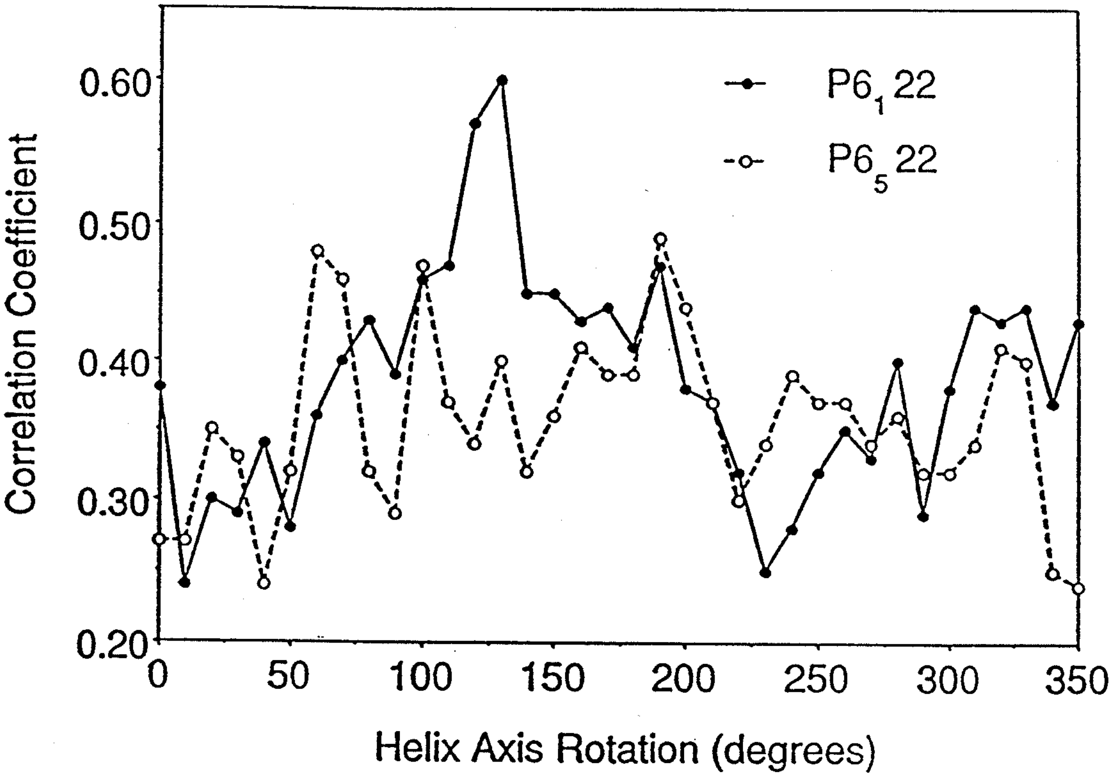


Fig. 31 To accurately determine the orientation of the B-DNA model around its helix axis, the linear correlation coefficient between F_o^2 and F_c^2 was plotted as a function of rotation of the model about its helix axis. Prior to the search, the model was positioned in the unit cell, with its helix axis inclined in accordance with the information provided by the Patterson map. In $P6_122$ a prominent solution, 2.63σ above background, appears corresponding to a rotation of the model by 130°. A finer rotation search between 120°-130° was then performed to precisely orient the model at 126°.

2.42 Refinement

After a crude model of the crystal structure has been determined, its fit against the observed data can be improved by various refinement techniques. The most commonly employed technique is the method of least squares. Least squares refinement involves deriving a set of linear equations in which the variables are the shifts from the trial parameters already obtained during primary phasing (*i.e.* x, y, z and temperature factors; B_{iso} in the case of isotropic, or six b parameters in the case of anisotropic). This is done by expanding the complex structure factor equation in a Taylor's series about the trial parameters, retaining only the first-derivative terms on the assumption that the shifts needed are sufficiently small that the terms involving second and higher order derivatives are negligible:

$$\Delta|F_c| = \sum_i (\partial|F_c|/\partial x_i)\Delta x_i + (\partial|F_c|/\partial y_i)\Delta y_i + (\partial|F_c|/\partial z_i)\Delta z_i + (\partial|F_c|/\partial B_i)\Delta B_i \quad (16)$$

The sum is over the number of observations, i. Using the linearized structure factors, the quantity that is minimized is the square of the residual, R', with a weighting function (w) applied to each term in the sum:

$$R'_w = \Sigma\, w||F_o(h)| - k|F_c(h)||^2 \qquad (17)$$

Unit weights were used for all refinements of DHEX.

Least squares refinement is only feasible on systems that are overdetermined. Crystals of small molecules which diffract to atomic resolution typically possess about 10 independent observations for each parameter that is refined. Thus, refinement of small molecule structures is generally a straightforward process when a good model is available. Macromolecular crystals, which can contain thousands of atoms in the asymmetric unit, do not, except in rare cases, diffract to atomic resolution. The lack of resolution, combined with an observation/parameter ratio which is often less than unity, makes the refinement of macromolecular structures by least squares intractable. The problems with refinement of macromolecular structures are ameliorated by using two strategies; **1)** the number of parameters can be lowered by dividing the structure into a set of discrete rigid bodies (the application of constraints), or, **2)** the x-ray data can be augmented by additional observations obtained from the known structural properties of portions of the macromolecule derived from small molecule crystallography and spectroscopic data (the application of restraints). When restraints are applied, the function that is minimized becomes

$$R = E + kX \tag{18}$$

where E represents the conformational energy and X the crystallographic term. The factor k controls the contribution of X to the total residual R. The restraint or energy term is of the form;

$$\begin{aligned} E = {} & \sum \frac{1}{2} K_b (b_i - b_o)^2 \text{(bonds)} + \sum \frac{1}{2} K_\tau (\tau_i - \tau_o)^2 \text{(bond angles)} \\ & + \sum K_\theta [1 + \cos(m\theta_i + \delta)] \text{ (torsion angles)} \\ & + \sum (Ar^{-12} + Br^{-6}) \text{ (non -bonded interactions) } + \ldots.. \end{aligned} \tag{19}$$

The four terms in Eqn. (19) describe bond length, valence angle, torsion angle, and non-bonded interactions. K_b is the bond stretching force constant and K_τ the bond angle bending force constant; K_θ is the torsional barrier and m and δ the periodicity and the phase of the barrier. A and B are the repulsive and the long-range non-bonded parameters, respectively. The summation extends to the i bonds, j valence angles, θ torsion angles and all non-bonded contacts. For the refinement of DHEX, all of the terms shown in Eqn. (19) were included, as well as an additional term which was required to maintain Watson-Crick hydrogen bonding between base-pairs. The weights of the empirical energy terms relative to each other and relative to the x-ray term must be chosen in such a way that everything is put on an appropriate scale: an overestimate of geometric restraints will produce a stereochemically perfect model with a high crystallographic R-factor; on the contrary, an underestimate of the same weight will result in a low R-factor, and a model with unreasonable bond lengths and angles. Energy restrained conjugate gradient least squares minimization using the CHARMM[95] potential (modified to maintain the planarity of the bases, and with intra- and inter-molecular electrostatic energy terms omitted) with 6.0-3.0 Å data above $2.0\sigma(F_o)$ was then carried out with X-PLOR and convergence was reached at an R-value of 26.9%. To this point, all base-pairs were strongly restrained to Watson-Crick geometry, and the molecule displayed sterically strained geometry in the vicinity of the T·A base-pair. Restraints on base-pairing were relaxed, and the refinement continued by simulated annealing. Simulated annealing is a molecular dynamics simulation, in which the atoms move in a force-field that includes a crystallographic term. Molecular dynamics of free atoms consists of solving Newton's equation of motion:

$$m_i \partial^2 x_i(t)/\partial t^2 = -\mathrm{grad}_x E_{tot}. \qquad (20)$$

E_{tot} is a potential energy function. To take into account the effect of the medium and the approximations used to calculate the total energy, dynamical effects can be better represented by a set of Langevin equations

$$m_i\partial^2 x_i(t)/\partial t^2 = -\mathrm{grad}_x E_{tot} + f_i(t) + -m_i b_i \partial x_i(t)/\partial t \tag{21}$$

where b_i is a frictional coefficient used to prevent atoms from moving away too much from their original positions, and $f_i(t)$ is a random force with Gaussian distribution and properties:

$$\begin{aligned} &\langle f_i(t)\rangle = 0, \\ &\langle f_i(t) f_i(0)\rangle = 2k_B T_o b_i m_i \delta(t). \end{aligned} \tag{22}$$

T_o is an artificial temperature and k_B is Boltzmann's constant. The simulation starts from an initial set of coordinates. Each atom is assigned a random velocity from a Maxwellian distribution corresponding to the temperature selected, and Eqn. (21) (or (20)) is integrated at a given temperature for a given time (usually fractions of a femtosecond). New velocities are then assigned, eventually at a new temperature, and the calculation continued. The simulation is normally performed over a period of a few picoseconds. The potential energy field the atoms move in, E_{tot}, is the sum of two terms:

$$E_{tot} = E_{emp} + E_{xray}. \tag{23}$$

E_{emp} represents an empirical energy function analogous to that defined by Eqn. (19), while E_{xray} is an "experimental" energy term described by the weighted residual, R'_w (Eqn. (17)). Simulated annealing refinement involves heating the "system" from an initial value of 300 K to 2000-4000 K (it is important to note that the temperature does not represent a physical temperature, but rather a parameter controlling the refinement), performing molecular dynamics and subsequently cooling the system gradually, repeating the dynamics at each cooling step. Although unreasonable from the biological point of view, the advantage of going to high temperatures is that the model can escape local minima in the function being minimized, increasing the radius of convergence of the method with respect to classical least squares.

From a starting temperature of 2000 K, the annealing process was performed in 34 stages with 500 steps of molecular dynamics (0.35 fs) at each stage (a total of 6 ps). The target temperature of the system was decreased in increments of 50 K between stages to a final temperature of 300 K. The system was then minimized by energy restrained conjugate gradient methods for 300 steps. The strain in the vicinity of the T·A base-pair was alleviated as it adopted a non-Watson-Crick hydrogen bonding scheme, and the R-factor dropped to 24%. A refined $|F_o| - |F_c|$ omit map confirmed the T·A geometry (Fig. 32). Spermine molecules were then clearly visible on difference density maps, but were excluded from the refinement until refinement of the DNA alone had completely converged. Restrained conjugate gradient refinement was continued with the gradual incorporation of higher and lower resolution data. The correct orientation of the model around the pseudo two-fold axis of the helix was readily confirmed, as the incorrect orientation refined poorly, yielding unreasonable geometry and a discernibly inferior fit to calculated electron density maps. With the correct orientation, the model fit well into calculated electron density on $2|F_o| - |F_c|$ maps (Fig. 32) while maintaining good geometry. One and a half spermine molecules and seventeen water molecules were located on difference Fourier maps.

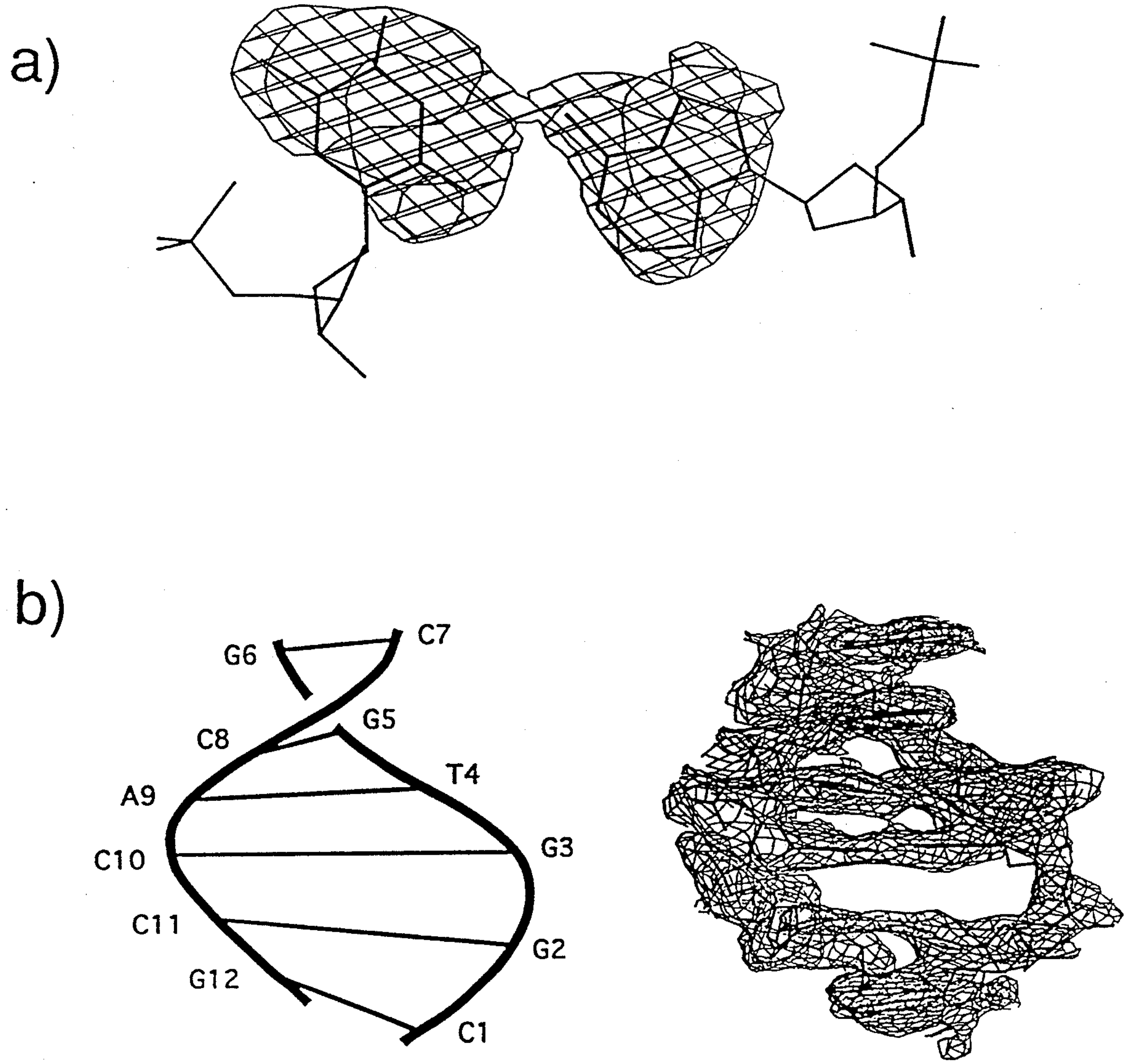


Fig. 32 a) Refined $|F_o| - |F_c|$ omit map contoured at 2.6σ around the T4 thymine and A9 adenine. The map was made by excluding the adenine and thymine atoms from structure factor calculations, refining the remainder of the structure by energy restrained conjugate gradient least squares for 20 cycles, and then calculating the difference map. **b)** A view into the duplex major groove with a $2|F_o| - |F_c|$ map and the corresponding sequence orientation given in the schematic representation on the left. The map is contoured at 1.0σ. The electron density maps in a) and b) were drawn using SETOR[96].

Only those regions of electron density that were approximately spherical and greater than 2.5σ above the mean on difference density maps were considered as possible water molecules. Each potential water molecule was carefully studied on a graphics terminal for reasonable hydration geometry. A large increase in the temperature factor of a possible water molecule or its loss of reasonable hydration geometry resulted in that position being rejected as a solvent site. The spermine molecules were manually fitted into difference density using FRODO[97], and then included in the refinement with restraints on bond lengths and angles. Refined omit maps of the spermine molecules show clear continuous envelopes of electron density around the molecules at the 2.6σ level (Fig. 33). B-factors of all atoms were refined individually. Currently, the structure is refined to an R-factor of 21.8% using 338 8.0 - 2.5 Å data above $\sigma(F_O)$ and 22.1% for 392 data between 20 - 2.5 Å data above $\sigma(F_O)$ (R=21.6% for $F_O>2\sigma(F_O)$). As a further check of the structure, the model was refined against the merged data set (**A** & **B**, described in Section 2.32) yielding an R-factor of 23.2% with no perceptible differences in the structure. The average r.m.s. deviations from ideal bond lengths and angles in the final structure are 0.029 Å and 5.2°, respectively. The positional errors of the atoms in the final structure determined from a Luzzati plot [98] are less than 0.25Å.

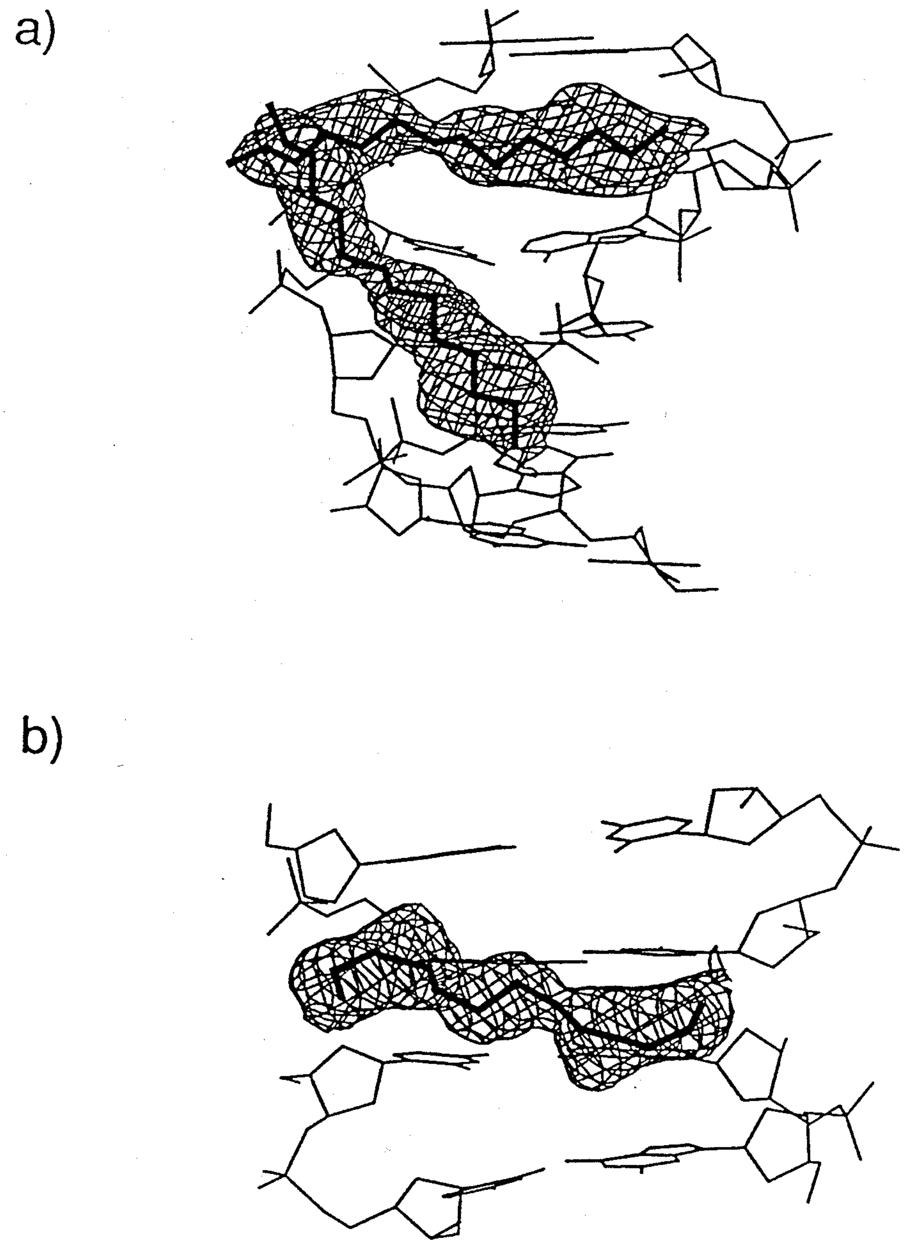


Fig. 33 $|F_o|$ - $|F_c|$ omit maps contoured at 2.6σ around **a)** the first spermine (spermine A, in text) molecule in the asymmetric unit and one of its two-fold symmetry related mates, and **b)** the second spermine molecule in the crystallographic asymmetric unit (spermine B, in text). Spermine B lies on a crystallographic two-fold axis which is approximately perpendicular to the plane of the paper. Both a) and b) were drawn using SETOR[96].

Chapter III

Results and Discussion

3.0 Nomenclature

In the lexicon of solution studies, "opening" is a general term referring to the disruption of a base-pair's internal H-bonds and exposure of the inter-base-pair region to solvent[54,99-101] without regard to any description of the disruption. However, in crystallographic nomenclature for nucleic acids[41,102] the term "opening" has a specific meaning and is distinct from any other term that describes distortion from normal Watson-Crick geometry, *i.e.* "shearing". "Opening" refers to rotation of the bases in their molecular planes relative to each other resulting in exposure of the inter-base-pair region to the major (+opening) or to the minor (-opening) grooves. "Shearing" refers to a lateral shift of the bases (within a base-pair) with respect to each other perpendicular to the helix axis but in opposite directions toward the major and minor grooves (see Figs. 15f, g). To avoid the potential source of confusion in the use of the term "opening", it will be italicized when used strictly in the crystallographic sense.

3.1 The Structure of d(CGGTGG)/d(CCACCG)

Using the techniques described in the previous section, the structure of DHEX was determined at 2.5 Å resolution. The overall helix structure resembles that found for B-DNA. The sugar conformations are generally C2'-*endo*. The minor groove is narrow (5.5 Å, measured by averaging P-P distances across the minor groove and subtracting 5.8 Å to account for the van der Waals radii of the phosphates) and 7 Å deep, while the major groove is 12 Å wide and 9 Å deep.

The average rise/base pair is 3.47 Å, and the average helical twist/base pair is 37.6°, corresponding to 10.2 base pairs/turn of the helix.

3.11 Helix Morphology

Graphs of some of the helical parameters are shown in Fig. 34. While the global structural parameters are unexceptional, striking anomalies do exist in the vicinity of the T4·A9 base pair. The distortion at T4·A9 arises as a result of over-twisting in the duplex at this point, relative to G5·C8. Instead of interacting in a Watson-Crick sense (with T4(O4) hydrogen bonding with A9(N6) and T4(N3) hydrogen bonding with A9(N1), as shown in Fig. 14), only one hydrogen bond is formed between T4(O2) and A9(N6) (Fig. 35). The 46° helical twist, relative to the 36° twist of canonical B-DNA, causes a shortening of the C1'---C1' distance across the T4·A9 base-pair of almost 1 Å and decreases the T4/G5 base step rise from the ideal 3.4 Å to 2.9 Å. Without compensatory structural adjustments these contractions would lead to severe steric clashes between the hydrogen bond donor and acceptor atoms of A4 and T9, and to a series of close contacts between base atoms across the T4/G5 step. However, these highly unfavorable steric interactions are avoided to a large extent by a 20° buckle in T4·A9 towards G3·C10 and, even more dramatically, by a 4 Å shear, in which thymine is rotated by 30° into the major groove, while adenine is rotated by 12° in the opposite direction, into the minor groove. With the exception of the positive shear, the helical parameters in the T4/G5 step adopt the combination of low rise, negative roll and positive cup (across a base-pair step, B1/B2, cup = buckle(B1) - buckle(B2)) typically observed in over-twisted base steps[103]. The results of this work are consistent with those of a recent solution study involving much longer strands of DNA[104]; there it had been deduced that torsional stress, *i.e.* over-twisting, favors base unpairing.

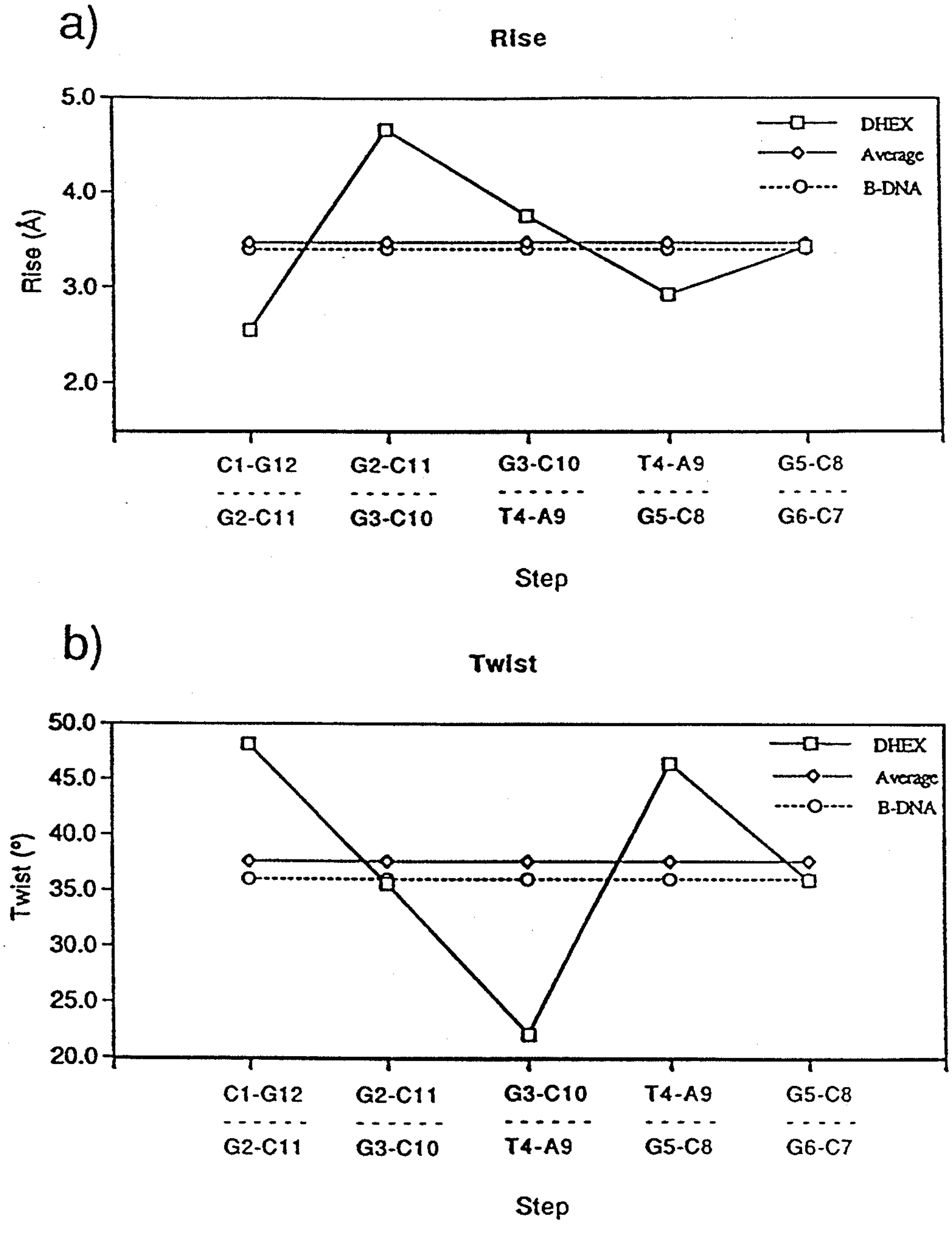


Fig. 34 Plots showing the variations in **a)** rise (Å), and **b)** helical twist (°) across the helix. The average values of the rise and twist, as well as the corresponding values of the parameters in ideal B-DNA are also plotted. The correlation between twist and rise (*i.e.* high twist with low rise and low twist with high rise) is evident when comparing corresponding values of twist and rise in a) and b).

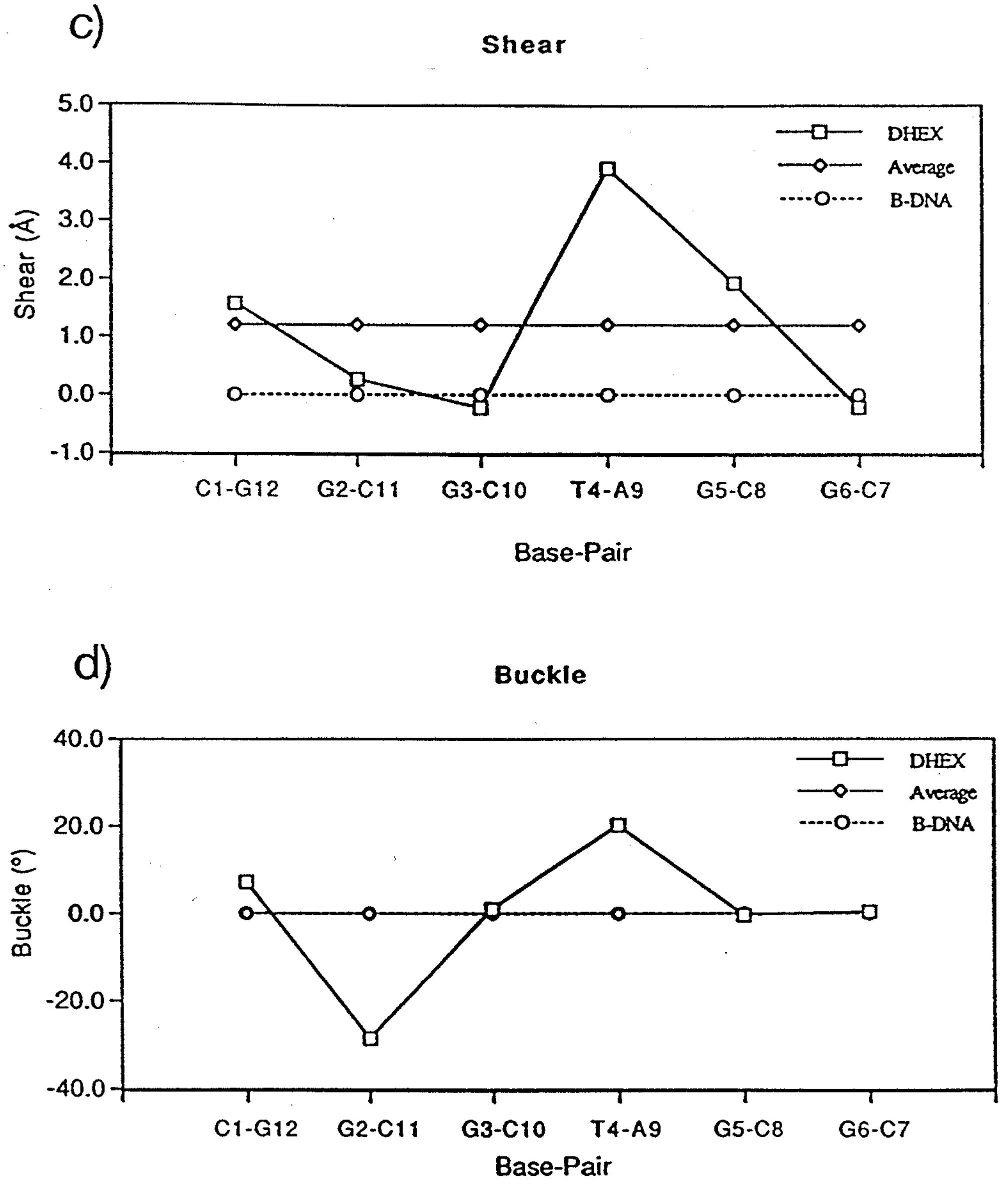


Fig. 34 Plots showing the variations in **c)** shear (Å), and **d)** buckle (°) across the helix. The average values of the shear and buckle, as well as the corresponding values of the parameters in ideal B-DNA are also plotted. The large shear in the T4•A9 base-pair is evident in c). The correlation between buckle, rise and twist (*i.e.* bases in over-twisted steps tend to buckle away from the over-twisted step (positive "cup") to avoid steric clashes across the step, which arise because over-twisting decreases the vertical rise across the step) is evident when comparing corresponding values of twist, rise and buckle in a), b) and d).

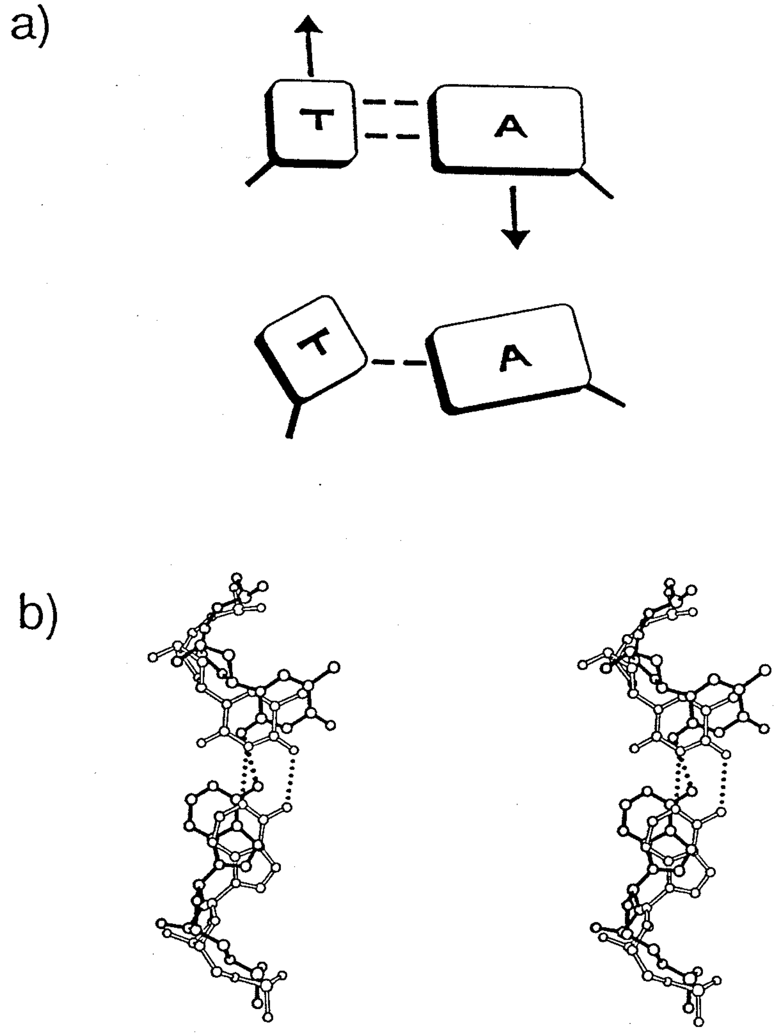


Fig. 35 a) Cartoon comparing the T•A base-pairing scheme observed in this structure with that of ideal B-DNA. TOP: A Watson Crick T•A base-pair, with hydrogen bonds shown as dotted lines. Over-twisting of the helix in this structure causes steric clashes between adenine and thymine, which are relieved by shearing the bases away from each other (in the directions shown by the arrows). BOTTOM: The result of the shearing has thymine rotated out into the major groove, with adenine moved by a lesser extent into the minor groove. A different hydrogen bonding scheme is adopted by the sheared base-pair (hydrogen bond shown as dashed line), which forms only a single hydrogen bond between the primary amine of the adenine (N6) and a keto-oxygen of thymine (O2). **b)** ORTEP[105] stereo diagram of the ideal starting model (hollow bonds) in the region of the T•A base-pair with the experimentally observed conformation superposed (solid bonds). Hydrogen bonds are shown as dotted lines.

The shearing of the A•T base-pair globally alters the structure of the helix, by creating 2.0 Å and 1.3 Å "bumps" in the surfaces of the major and minor grooves, respectively (Fig. 36). Such bumps may play a key role in protein recognition, as discussed in a later section.

Conformational features in the vicinity of C1·G12 and G2·C11 are unusual, but this is not totally unexpected, as terminal base pairs are subject to end-effects and have a tendency to fray if not stabilized by adjacent molecules. The large helical twist in the C1/G2 step of 48°, a partial *opening* (25°) into the minor groove and a 1.6 Å shear of the terminal C1·G12 base-pair seems to facilitate hydrogen bonding between C1(N4) and a phosphate oxygen belonging to nucleotide C8 of a symmetry related duplex. The high twist in the C1/G2 step is accompanied by a 29° buckle in G2·C11 towards G3·C10. The results of high twist across the C1/G2 and T4/G5 steps in the context of the global structure of the duplex are shown in Fig. 36.

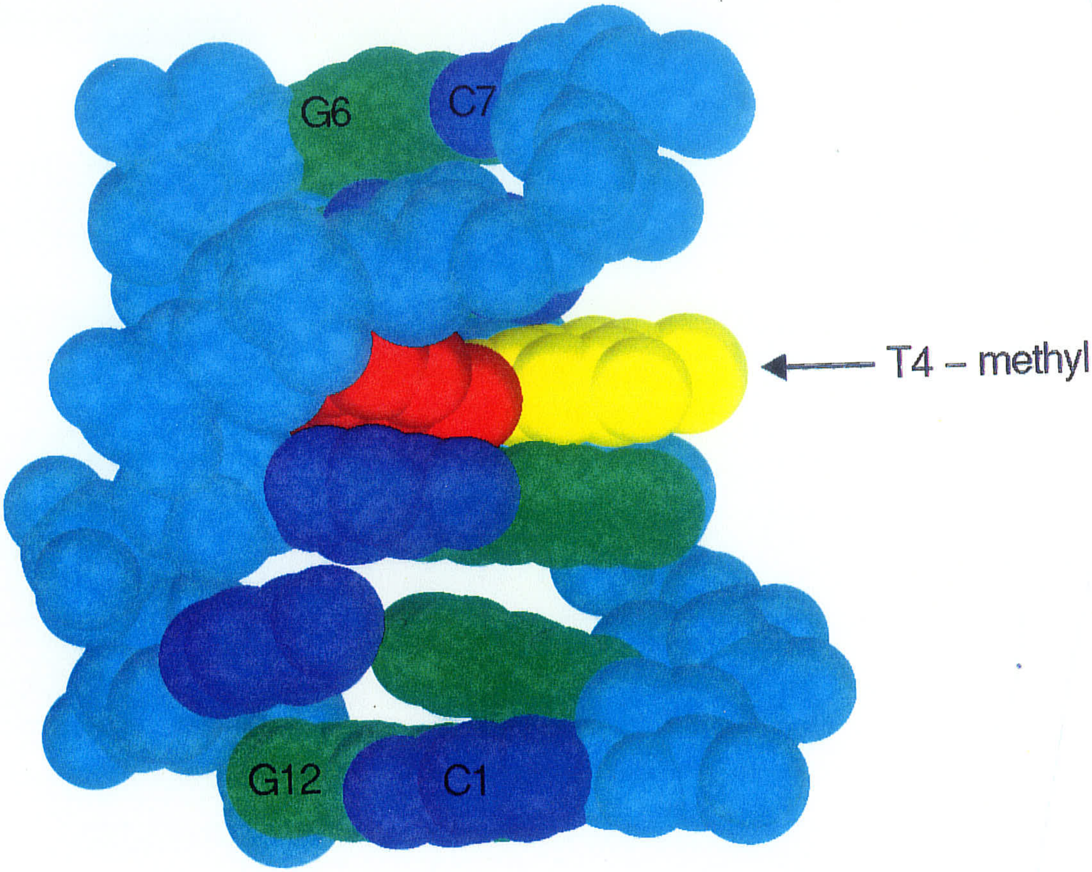


Fig. 36 Space-filling representation of the hexamer duplex looking into the major groove. The sugar-phosphate backbone in shown in blue. The buckle in the G2•C11 base-pair results from helical over-twisting in the C1/G2 step. Helical over-twisting also causes thymine (yellow) to bulge out by 2 Å into the major groove, exposing the imino proton to solvent. The figure was drawn using SETOR[78].

3.12 Sugar-Phosphate Backbone

In the crystallographic refinement, care was taken to impose only negligible restraints on the sugar-backbone torsion angles, to avoid biasing the torsion angles towards the restraint values in the final structure. The deoxyribose sugars of DHEX populate the C2'-*endo* region of the pseudorotational cycle (see Fig. 9), with the exception of the sugar on the terminal C7 nucleotide, which adopts a C2'-*exo* conformation (Table 7).

Nucleotide	**Phase**	**Amplitude**	**Pucker**
C1	128.7	49.0	C1'-*exo*
G2	221.9	33.6	C4'-*endo*
G3	185.4	50.2	C3'-*exo*
T4	153.5	45.5	C2'-*endo*
G5	176.5	53.5	C2'-*endo*
G6	101.7	41.9	O1'-*endo*
C7	334.2	18.7	C2'-*exo*
C8	159.5	44.1	C2'-*endo*
A9	177.4	45.1	C2'-*endo*
C10	115.0	58.2	C1'-*exo*
C11	168.5	27.5	C2'-*endo*
G12	124.8	50.1	C1'-*exo*

Table 7 Sugar puckering parameters for DHEX sugars. The puckering phases and amplitudes were determined using Eqns. (1) and (2).

As evident from Table 7, the largest deviations from C2'-*endo* puckering occur in terminal nucleotides. The ends of oligonucleotides are not stabilized by 3'-phosphodiester linkages, and are subject to fraying. Consequently, terminal nucleotides tend to exhibit a higher degree of thermal motion than the interior nucleotides of the helix. Thus, the sugar-phosphate torsion angles possess more conformational freedom than the same parameters in interior nucleotides. A second consequence of the high thermal motion is that the electron density around terminal nucleotides is less well resolved, making the precise determination of sugar puckering modes more difficult.

The exocyclic torsion angles are distributed fairly evenly around the values observed in canonical B-DNA (Fig. 37), and all glycosidic torsion angles, χ, are in the usual anti-conformation. However, unusual phosphate-backbone conformations exist in nucleotides G2, G5 and C10, which are correlated with the anomalies in helical parameters observed in DHEX. In each of the three nucleotides, the torsion angle α adopts a *+sc* to *+ac* conformation instead of the *-sc* conformation normally observed in B-DNA (Fig. 37). As described in Section 1.224, the *-sc* conformation of α is required for the formation of a right-handed double-helix, if it is assumed that the *usual* conformations are adopted by the other exocyclic backbone torsion angles (Table 4). Therefore, a *+sc* conformation around α is only possible in a right-handed double-helix if compensatory adjustments in other exocyclic torsion angles occur[106]. In DHEX, two distinct families of backbone conformers were observed which maintained the integrity of the helix in the presence of a *+sc* conformation of α; they have been designated the $B_{\alpha^+\gamma^+}$ and $B_{\alpha^+\beta^-}$ conformations, where B refers to the DNA type and the subscripts refer to specific torsion angles superscripted with the sense of rotation required to acheive the conformation from the canonical form of the DNA. The $B_{\alpha^+\gamma^+}$ conformation (Fig. 38) is attained *via* a "crankshaft" rotation around the O5' and C5' atoms of a nucleotide around their bond central point, yielding final conformations with α, 34° (G2) and 105° (G5), and γ, 320° (G2) and 160° (G5). In the $B_{\alpha^+}\gamma^+$ conformation the O5' elbow points away from the helix rather than over the deoxyribose sugar as in ideal B-DNA.

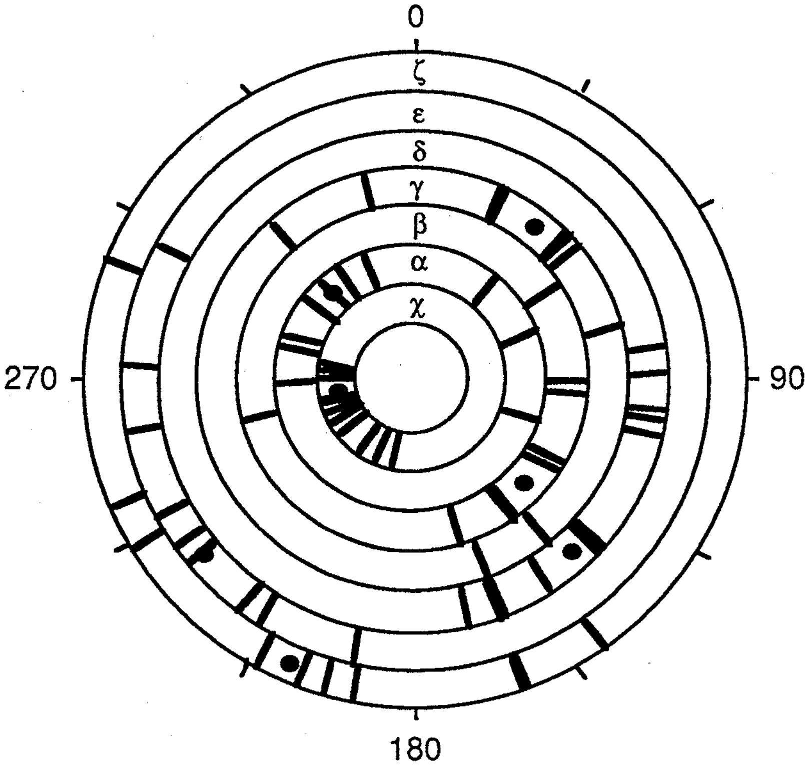


Fig. 37 A conformational ring, showing the distribution of sugar-phosphate backbone and glycosidic torsion angles (°) in DHEX. Black circles represent the values of the respective torsion angles in canonical B-DNA.

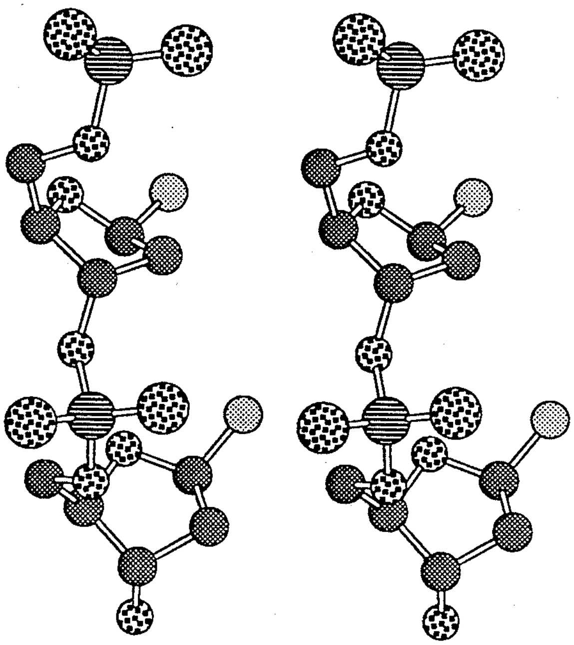


Fig. 38a A stereo view of a dinucleotide section of canonical B-DNA. Oxygen is stippled, carbon is dark gray, phosphorous is banded and nitrogen (representing N1 or N9, the remaining base atoms are omitted for purposes of clarity) is shown in light gray. In this figure the torsion angles α and γ in the *-sc* and *+sc* conformations, respectively.

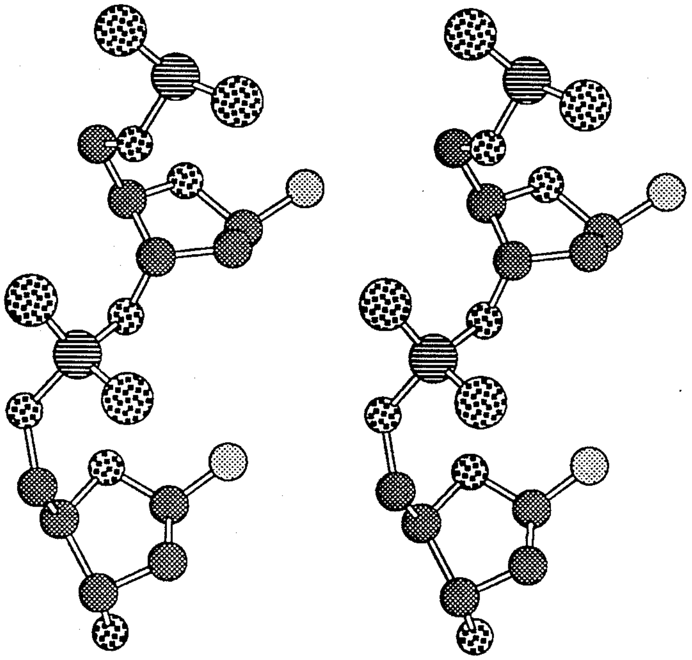


Fig. 38b A stereo view of the T4/G5 base step in DHEX, which adopts a $B\alpha^+\gamma^+$ conformation. T4 is the upper nucleotide. The shading scheme is the same as in Fig. 38a. A crankshaft rotation around O5' and C5' of nucleotide G5 invert α with respect to its value in canonical B-DNA, so that it adopts a *+ac* conformation. A correlated change in γ occurs, with its conformation changing from *+sc* to *+ap*. It is also important to note that the $B\alpha^+\gamma^+$ conformation shortens the intrastrand phosphate-phosphate distance between the $B\alpha^+\gamma^+$ nucleotide and the nucleotide on the 5'-side of it. The P-P distance across the T4/G5 step (shown in the figure) is shortened from its value of 7.0 Å in ideal B-DNA to 5.7 Å.

The $B_{\alpha^+\beta^-}$ conformation (Fig. 39) found in nucleotide C10, is distinct from the $B_{\alpha^+\gamma^+}$ conformation both in terms of the mechanism by which it must be attained and in its effect on the surrounding helical parameters. The $B_{\alpha^+\beta^-}$ conformation is obtained from canonical B-DNA by a crankshaft rotation of P and O5' around the central point of the bond between the atoms. The rotation results in α and β adopting *+sc* conformations. The torsion angle γ is unaffected, and remains in a*+sc* conformation. In both the $B_{\alpha^+\gamma^+}$ and $B_{\alpha^+\beta^-}$ conformations, the sugar puckers are unaffected and remain in the C2'-*endo* family.

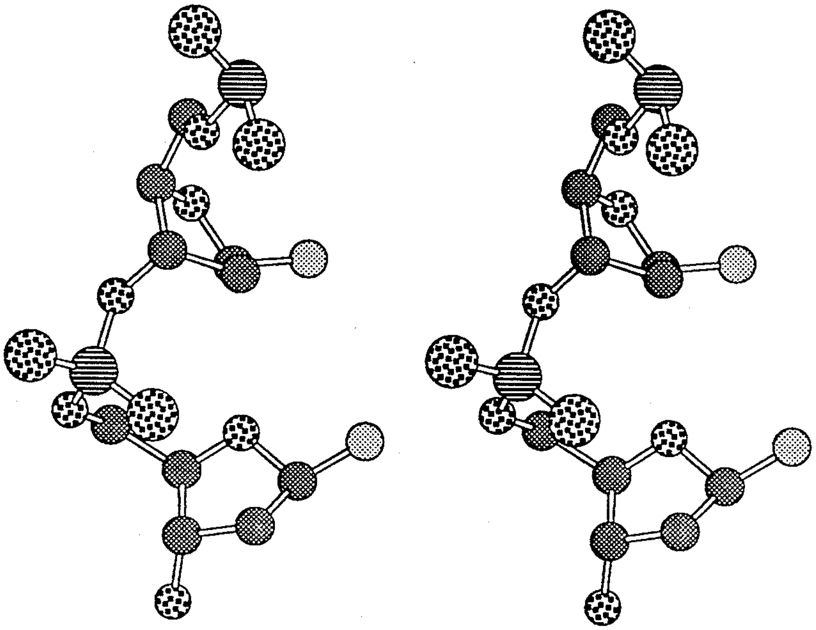


Fig. 39 A stereo view of the A9/C10 base step in DHEX, which adopts a $B_{\alpha^+\beta^-}$ conformation. A9 is the upper nucleotide. The shading scheme is the same as in Fig. 38a. A crankshaft rotation of the P and O5' atoms around their bond central point, inverts α with respect to its value in canonical B-DNA, so that it adopts a *+sc* conformation. A correlated change in β occurs, with its conformation changing from *+ap* to *+sc*. The intrastrand phosphate-phosphate distance is shortened between C10 and C11 (not shown) to 5.8 Å.

In canonical B-DNA, the transition of the phosphate backbone to the $B\alpha^+\gamma^+$ or $B\alpha^+\beta^-$ conformation has important structural consequences: **i)** The intrastrand phosphate-phosphate distance is shortened — in the presence of a $B\alpha^+\gamma^+$ conformation, between the $B\alpha^+\gamma^+$ nucleotide and the adjacent phosphate on the 5' side, and in the presence of a $B\alpha^+\beta^-$ conformation, between the the $B\alpha^+\beta^-$ nucleotide and the adjacent phosphate on the 3' side; **ii)** in the presence of the $B\alpha^+\gamma^+$ conformation, the base of the $B\alpha^+\gamma^+$ nucleotide becomes over-twisted with respect to the adjacent base on the 5' side, and in the presence of the $B\alpha^+\beta^-$ conformation, the base of the $B\alpha^+\beta^-$ nucleotide becomes under-twisted with respect to the adjacent base on the 5' side (Figs. 40, 41).

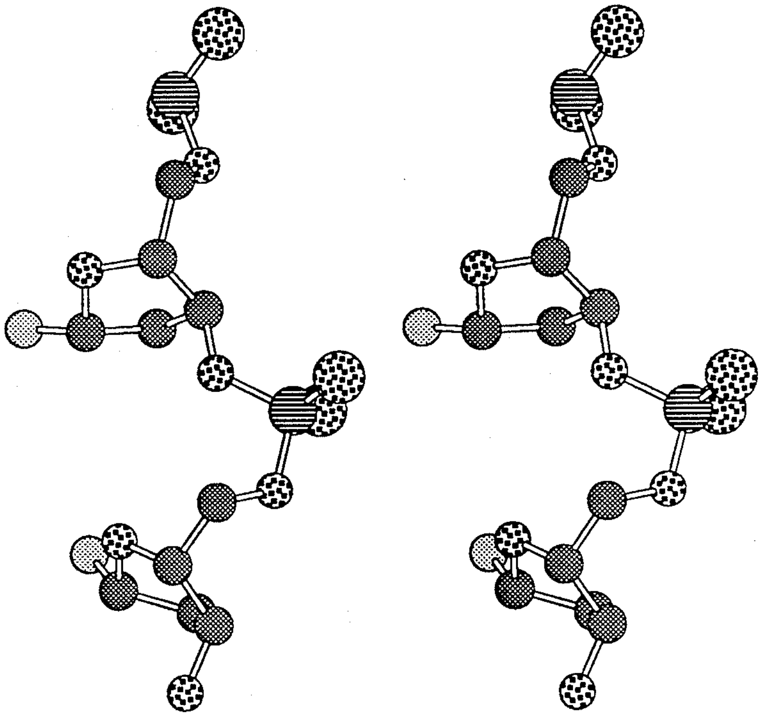


Fig. 40a A stereo view of a dinucleotide section of canonical B-DNA with the same shading scheme as Fig. 38a. This view shows the relative disposition of the sugars and base atoms in a base-step with the ideal 36° twist angle.

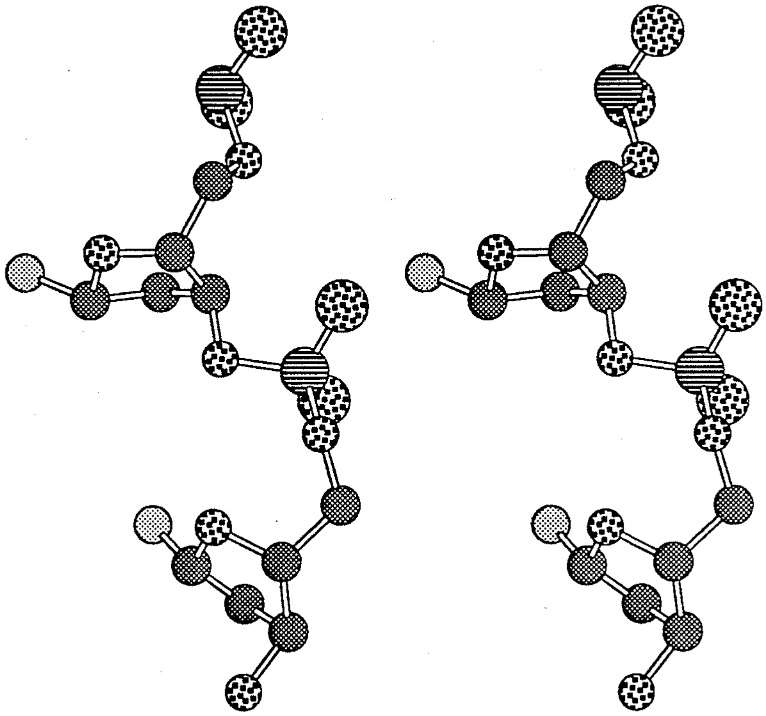


Fig. 40b A stereo view of the T4/G5 base step in DHEX, which adopts a $B\alpha^{+}\gamma^{+}$ conformation. T4 is the upper nucleotide. The shading scheme is the same as in Fig. 38a. The $B\alpha^{+}\gamma^{+}$ conformation in nucleotide G5 induces a high twist angle in the T4/G5 base-step of 46° by displacing T4 (which contains the sheared thymine) towards the major groove (which faces the left side of the diagram).

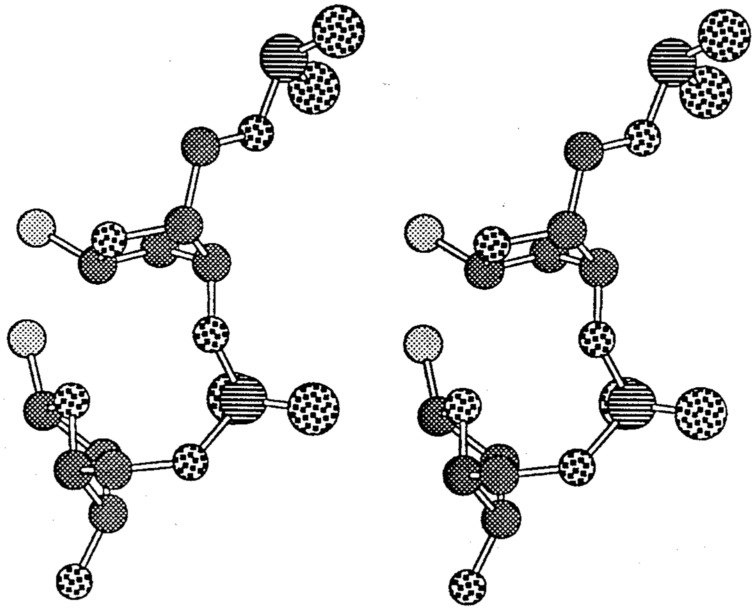


Fig. 41 A stereo view of the A9/C10 base step in DHEX, which adopts a $B\alpha^{+}\beta^{-}$ conformation. A9 is the upper nucleotide. The shading scheme is the same as in Fig. 38a. This view shows how the $B\alpha^{+}\beta^{-}$ conformation in nucleotide C10 reduces the helical twist in the A9/C10 base-step to 22° by displacing A9 towards the minor groove (which faces the right side of the diagram).

From a biological standpoint, the $B_{\alpha^+\gamma^+}$ and $B_{\alpha^+\beta^-}$ conformations (herein referred to as the α^+ conformations) are important because they provide an oligonucleotide with a means of accommodating mechanical stress (*i.e.* over twisting). They are also mechanisms through which the helix can adapt to localized regions of low humidity. When DNA is transferred from high humidity to low humidity (see Section 1.234), the DNA undergoes a B-to-A transition, which involves a decrease in the intrastrand phosphate distances (so that less water is required to hydrate the phosphate backbone) *via* a global change in sugar puckers from the C2'-*endo* to the C3'-*endo* conformation. The α^+ conformations can be utilized in B-DNA to effect an identical contraction across the phosphates in the backbone while maintaining the C2'-*endo* sugar pucker. Localized deficiencies in hydration could be caused by the presence of cations, or, by close contact with another macromolecule (*i.e.* a protein searching for its target site). Alternatively, if the α^+ conformations are stress induced, water is "squeezed" away from the backbone to accommodate the concomitant collapse of the intrastrand phosphates. Since adjacent phosphates in the vicinity of the α^+ nucleotides are less hydrated, the anionic phosphates are less shielded, and can thus experience stronger electrostatic interactions with cationic amino acids, spermine, *etc.* In addition, contraction of the phosphates creates a region of high charge density along the backbone. Molecular electrostatic field calculations on DNA models[107] show that the contraction of intrastrand phosphates caused by a B-to-A transition increases the surface maximal field at the DNA backbone by almost 20%—the calculations do not account for the effects of phosphate hydration, so that the differences in electrostatic fields between A- and B-DNA in solution are probably even more dramatic. In summary, a "twist induced" dehydration of the backbone provides a mechanism for transmitting information about base-pair geometry through the backbone.

Since the α^+ conformations are associated with proximal modulations in the dielectric character of the surrounding medium, increases in the electrostatic field along the backbone and a corresponding lowering of the electrostatic potential, they provide additional features which can act in conjunction with altered base-pair geometry. Whether the α^+ conformations are induced by hydration deficiencies, or are a means of compensating for mechanical stress, or both, they may play a significant role in the protein recognition process.

3.13 Thermal Parameters

In DHEX, average isotropic temperature factors range from 7 to 39 Å^2 for phosphate groups, from 7 to 38 Å^2 for ribose units, and from 3 to 26 Å^2 for bases with overall mean values of 21, 18 and 9 Å^2, respectively. As is commonly observed in macromolecular structures, the B values of a few of the atoms refined to a pre-selected minimum (2.0 Å^2), a physically meaningless situation caused by phasing errors in the structure factors that is observed frequently in macromolecular crystal structures. The phasing errors probably arise from the inability to correctly account for the scattering by disordered bulk solvent in the crystal. The B values increase from the inside to the periphery of the double helix, reflecting the increased flexibility of the helix sugar-phosphate backbone and/or an overall rotational motion of the molecule. The terminal nucleotide C7 displays the highest degree of thermal motion, with average base and sugar B values of 26 and 38 Å^2, respectively. The high thermal motion in C7 is probably the result of fraying of the terminal G6•C7 base-pair, combined with the fact that C7 resides in a relatively "open" region of the crystal lattice. A more detailed depiction of the thermal motion trends in DHEX is displayed in Fig. 42. Average thermal parameters for the two spermine molecules, spermine

A and spermine B, are 20 and 42 Å^2, respectively. The average thermal parameter for solvent is 13 Å^2.

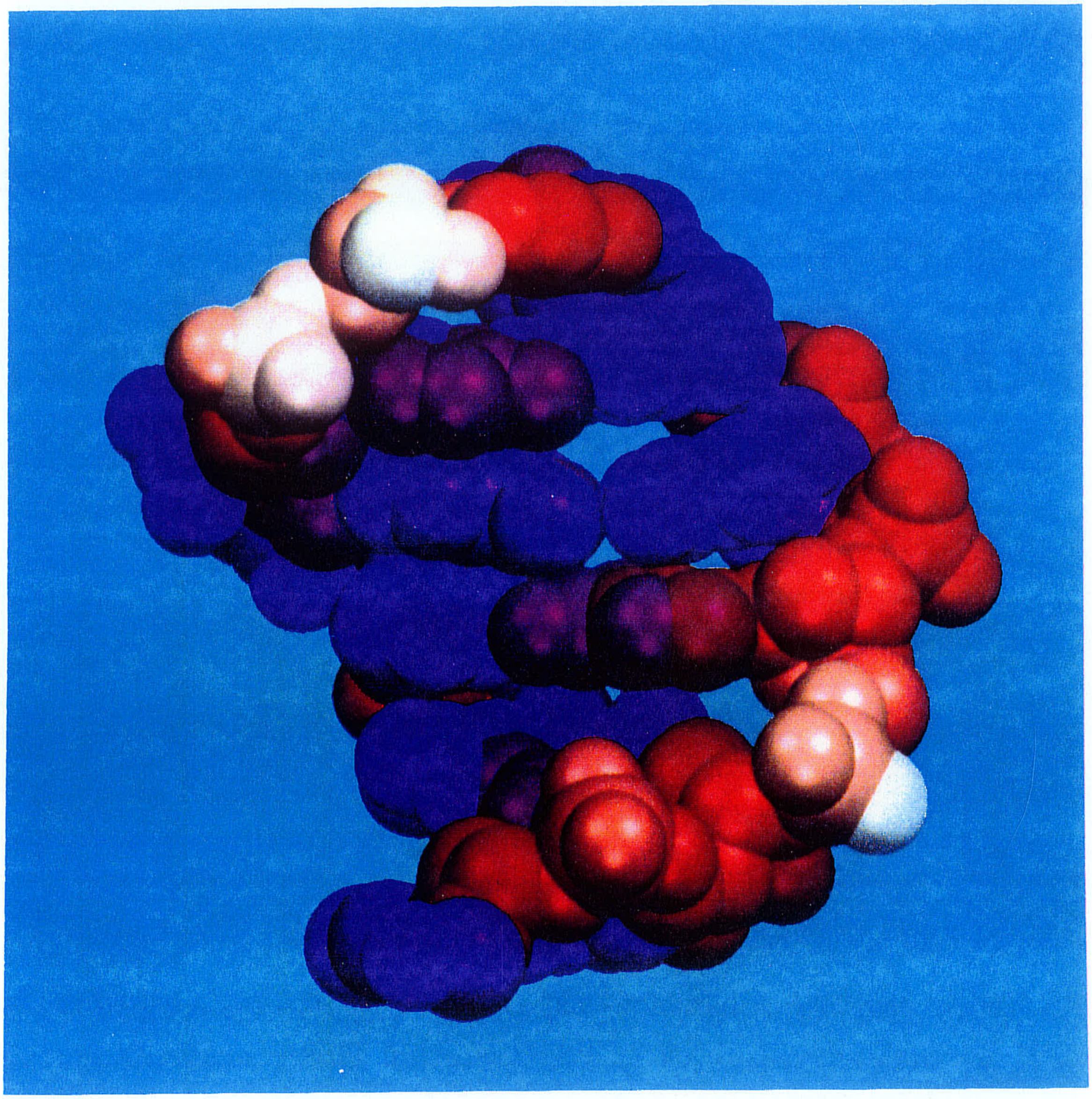


Fig. 42 A space-filling rendition of DHEX, with nucleotide C7 shown on the upper left, and nucleotide C1 shown on the bottom of the diagram, facing the viewer. The atoms are colored according to their thermal parameters. Dark blue represents small (< 8 Å^2) thermal factors and white represents large (> 40 Å^2) thermal factors, with colors ranging from light blue through red, orange and pink as thermal parameter values increase. This Figure was drawn using SETOR[96].

3.14 Crystal Packing

This crystal structure exhibits a packing arrangement which, until now, had not been observed in crystals of native B-DNA. Helices stack end-on-end to form continuous wire-like strands, which are kinked by 35° at every second helix. Adjacent strands criss-cross, and wind around 12 Å solvent channels parallel to the crystallographic *c* axis (Fig. 43), much like wires in a twisted cable having a central core. Non-stacked, symmetry related DNA molecules form direct and water mediated backbone-backbone, and backbone-major groove contacts. Direct (less than 3.4 Å) contacts between potential hydrogen bonding atom pairs exist between a phosphate oxygen of nucleotide C8 and C1(N4) of a symmetry related helix, with another between terminal O3' atoms of nucleotide G12 in related molecules. The sugar-phosphate backbones of crossed helices interlock, forming molecular contacts between the phosphates of nucleotides A9 and C10 on adjoining duplexes. The P(9) - P(10) distance of 4.3 Å is similar to the close approach of phosphates observed in the decamer d(CGATTAATCG)[103] where adjacent symmetry related molecules meet at the backbones of strands running in opposite directions in a fashion similar, but not identical to the interlocking arrangement observed in DHEX crystals. A more detailed analysis of the geometry in this region is limited by the resolution of the data since, at 2.5 Å resolution the electron density appears continuous in the immediate region of the interlocking phosphates.

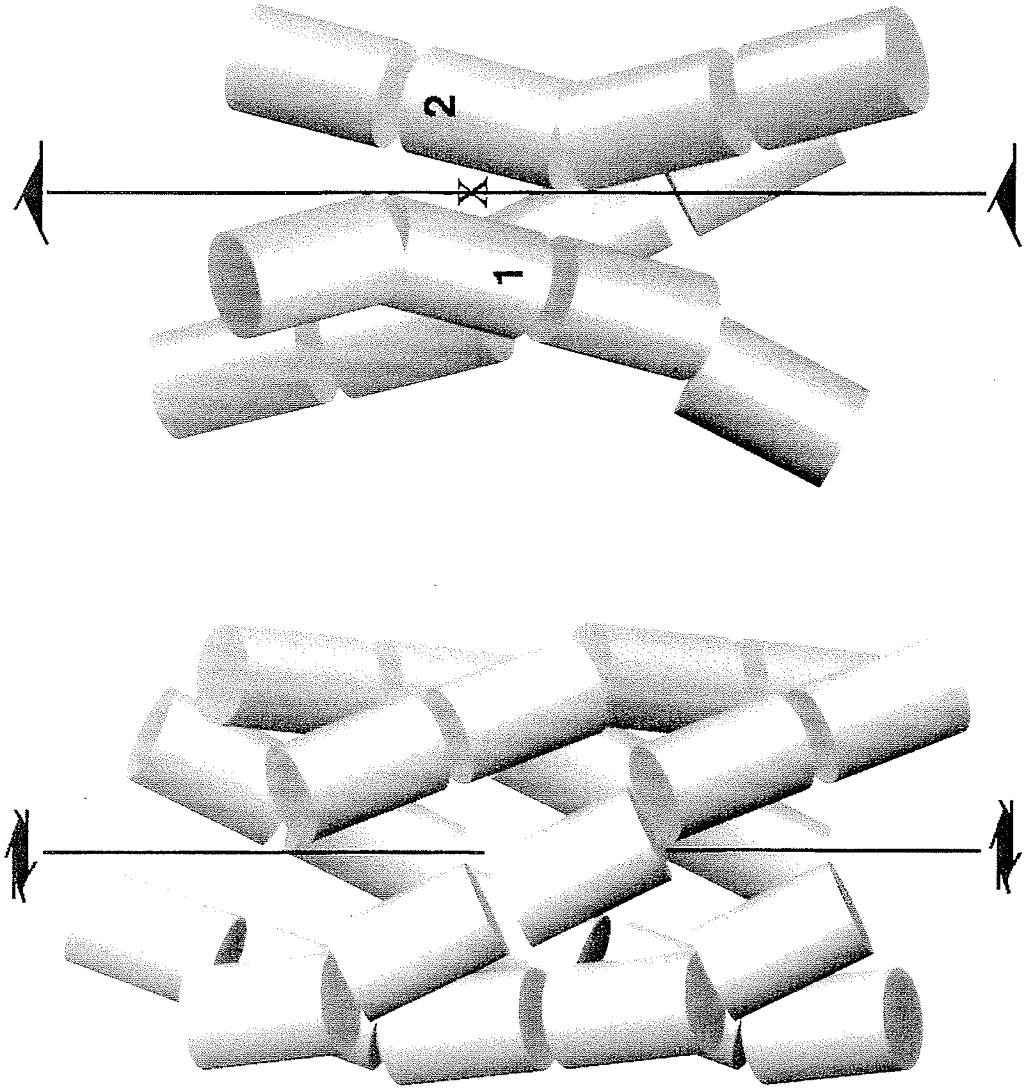


Fig. 43 Continuous wire-like strands of hexamers represented as cylinders around the 6_1 (bottom) and 3_1 (top) axes. A crystallographic two-fold relating molecules 1 and 2 (same molecules shown in Fig. 44) is indicated by an X on the 3_1 axis. This figure was drawn with SETOR[78].

3.15 Spermine-DNA Interactions

The 1.5 spermine molecules in the crystallographic asymmetric unit stabilize the observed crystal packing through an extensive network of solvent mediated DNA-spermine and spermine-spermine lattice contacts. The positively charged primary amino termini of each crystallographically distinct spermine molecule experiences water mediated interactions with symmetry related spermine molecules in the crystal. In addition, the primary amino group at one end of the first spermine (spermine *A*) stitches together two symmetry related DNA helices through water-bridged interactions with backbone oxygen atoms, while the primary amino group at the opposite end of the molecule sits in the major groove of a third DNA molecule forming water mediated contacts with base atoms in nucleotides C10 and G3 (Fig. 44). A direct contact between spermine and DNA was observed, between O5' of nucleotide C1 and one of the secondary amines of spermine *A*. The second spermine molecule (spermine *B*), which is located on a crystallographic two-fold axis, splices stacked DNA molecules by spanning their major grooves and forming water mediated contacts with symmetry related base atoms of nucleotide C7 and adjoining backbone oxygens. Fig. 45 shows some of the specific contacts made between spermine *B* and DNA. The finding of spermine in this structure is the first conclusive evidence for the polyamine complexed with native B-DNA in a crystal structure.

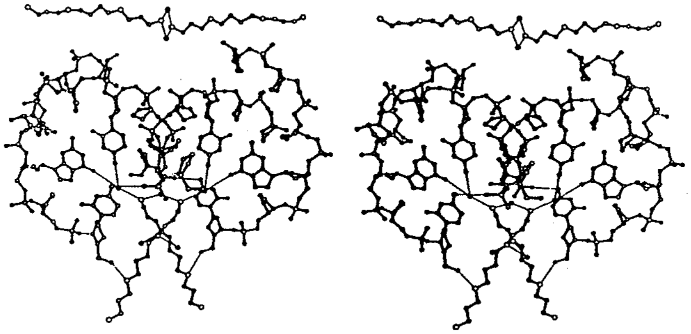


Fig. 44 ORTEP[105] stereo diagram showing a detailed view of some of the intermolecular contacts involving spermine A and crossed helices in the crystal. For clarity, only the sugar-phosphate backbone atoms of the DNA molecules are shown, along with the bases involved in hydrogen bonding interactions. Spermine molecules are drawn with solid bonds, and DNA with hollow bonds. Thin lines depict potential hydrogen bonds, and water molecules are drawn as darkened spheres. The C10-G3 base-pair forms a water mediated contact with the primary amino terminus of one spermine, while one of the secondary amines of the symmetry related spermine shown makes a direct contact with O5' of nucleotide C1 on the same DNA molecule. Simultaneously, these spermine molecules bridge the minor grooves of other symmetry related helices (symmetry related spermine molecules which display this are shown at the top of the figure). Methylene carbon atoms of the crossed spermine molecules form direct hydrophobic contacts. The phosphate backbones of the helices shown interlock, and form direct major groove-backbone hydrogen bonding contacts, between N4 atoms of nucleotide C1 and phosphate oxygens of nucleotide C8.

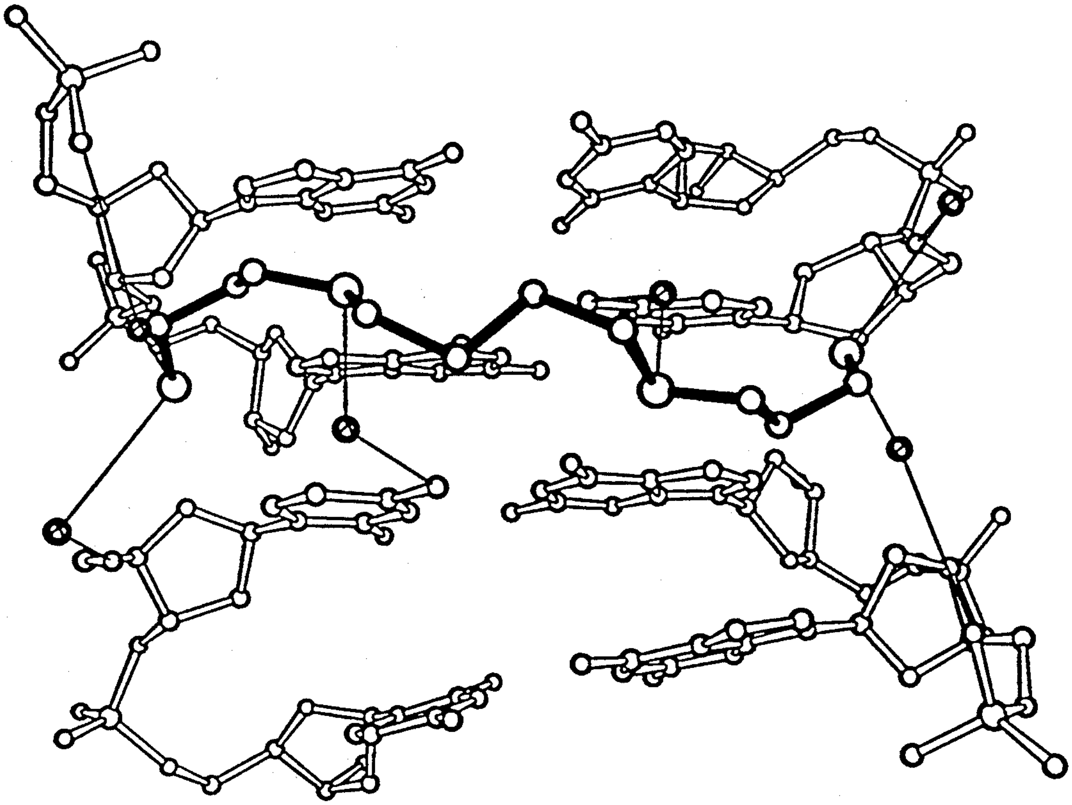


Fig. 45 ORTEP[105] plot showing a detailed view of contacts made by spermine B and stacked DNA molecules in the structure. Spermine is drawn with dark bonds, and DNA with hollow bonds. Water molecules are drawn as cross-hatched spheres, and possible hydrogen bonds are represented by thin lines. Spermine B forms water mediated contacts with the two-fold related primary amines of nucleotide C7, and with adjoining backbone oxygens.

3.2 Structure and Biological Implications — DNA

3.21 The Influence of Crystal Packing

The non-Watson-Crick hydrogen bonding scheme adopted by the sheared T·A base-pair in this structure has not been observed in the GTG regions of any other native B-DNA crystal structures bearing the trimer[108-110]. Thus, it may be argued that the helical distortions observed in the GTG region of this structure are artifacts of crystal packing forces. The explanation for the T·A shear would be that the molecule has yielded to the stress of packing forces by distorting at its weakest point (Fig. 46). Whether the stress is induced by contact with surrounding molecules in the crystal or by contact with a non-specifically bound protein molecule seeking its cognate sequence, the distortion would be expected in the same region, *i.e.* at T·A in the GTG trimer.

The novel packing arrangement observed in this structure, with helices locked together *via* sequence specific backbone-backbone (at the 5'- and 3'-phosphates of A9) and backbone-major groove interactions, provides an analogy to the interactions found in repressor-operator complexes[46]. Docking experiments[111] using an ideal B-DNA model in place of the observed structure, show that canonical B-DNA lacks the proper three-dimensional shape to achieve the close inter-helix contact and electrostatic complementarity realized by the symmetry related helices in this structure. In the context of protein-DNA recognition, this suggests that the inherent conformational flexibility of the GTG trimer, and this sequence as a whole, provide it with the ability to mediate sequence specific interactions between Gal repressor and *gal* O_I.

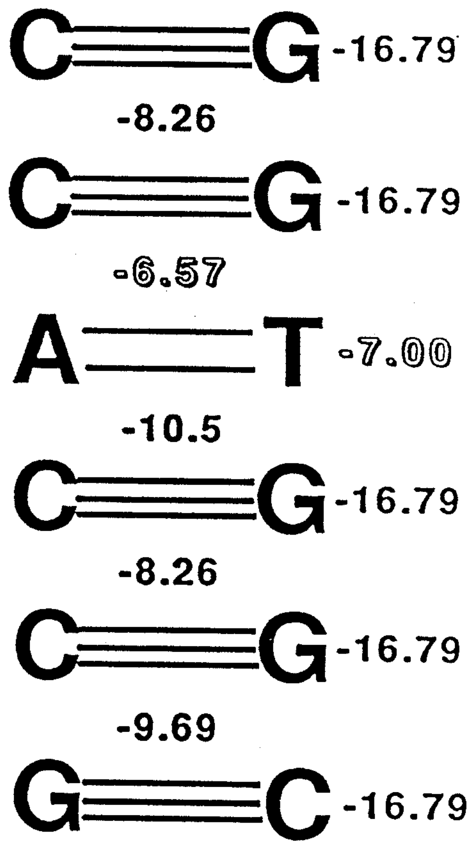


Fig. 46 A ladder diagram showing the variations in stacking energies (between base-pairs) and hydrogen bonding energies (to the right) with sequence for the DHEX structure. All energies shown were derived from quantum mechanical calculations using a canonical B-DNA model (from Refs. 36 and 39). The highlighted numbers represent the stacking energy across the T4/G5 step (which is over-twisted in DHEX), and the hydrogen bonding energy of the A•T base-pair (which is sheared); both energies are at the global minimum in the structure. It is evident from this diagram that DHEX deformed at its weakest point.

3.22 Base-pair Opening

An alternative and more compelling explanation for the distortion at GTG is that base-pair shearing at T·A is an intrinsic property of the trimer that can result in protein recognizable features on the DNA. Solution results of base-pair opening at T·A in the GTG trimer[47,48,52,54,112] suggest that the observed distortion originates in solution leading to the conclusion that the crystallization process, aided by spermine, has captured (rather than induced) a pre-recognition or transition-state-like conformation characterized by base-pair opening at T·A.

Base-pair opening in solution has been well documented[54,100,101,113-117]. but the nature of the open state has remained undefined. While it is possible that the state is not structurally unique, the present analysis provides one description which is consistent with the solution results. Base-pair shear, as observed here, results in an open state that exposes the T4 imino proton to solvent. Shearing of T·A would explain the unusually high rate of thymine imino proton exchange with solvent in the GTG region of the *lac* operator observed well below the melting temperature of the operator duplex[47]. Base-pair shear could be the "specific feature" inherent to GTG that Leroy *et al.*[54] have suggested may account for the anomalously short T·A lifetime in solution in the sequence d(GTGCG)/d(CGCAC). Shearing may also occur in platinated double-stranded oligonucleotides containing GTG, where, with NMR[112] and chemical probes[52], thymine has been shown to bulge out and no longer pair with the complementary adenine.

The distortion at GTG in the present structure affects three of the thymine substituents, *viz.* the N(3)-imino H, C(5)-methyl, and O4, which may provide specific recognizable features to a searching protein. The most novel feature of the base-pair opening is exposure of the imino H to the major groove presenting a hydrophilic interaction site, which may be a potential recognition contact. In normal Watson-Crick geometry this H is inaccessible to either protein or solvent. The opening also accentuates the protrusion of the C(5)-methyl group into the major groove. There is ample experimental evidence for the significant role of the thymine methyl in determining the sequence specificity of many DNA binding proteins that recognize major groove features[118]. A molecular dynamics perturbation thermodynamics study[119] indicates that protein-driven desolvation of this hydrophobic group makes a substantial energetic contribution to the specificity of protein-DNA interactions and that the extent of this contribution increases with the solvent accessible area of the methyl group. Since shearing of the T·A base-pair in the present structure increases the solvent accessible region around the methyl group, an enhancement in the fidelity of protein recognition would be expected. The shearing also results in greater exposure of O4 to potential H-bonding interactions with solvent or protein in the major groove than in the Watson-Crick base-paired T·A, providing an additional recognition contact.

Regardless of its origin in this fragment of *gal* O_I, the helical distortion at the GTG trimer could act as a molecular detent (an image suggested by Lu *et al.*,[47]) in the rod-like track the DNA presents to the searching Gal repressor, thereby retarding the movement of the protein sufficiently to allow a more thorough examination of the flanking sequences. Evidence for GTG involvement in Gal repressor recognition of its cognate DNA sequence is found in ethylation interference studies where it has been shown that ethylating the two phosphates within the GTG trimer (in every case for both the internal and external operators, *i.e.* O_I and O_E) inhibits or reduces repressor binding[120]. One explanation for this observation is that ethylation prevents a conformational contortion at these phosphates which in the native DNA is favorable to protein interaction. Furthermore, there is evidence that conformational flexibility is associated with GTG in the *gal* operon. It has been demonstrated by electrophoretic analysis of repressor binding to *gal* O_E and O_I that the DNA bends at or near the regions containing the GTG trimers[121]. Overall, in the *gal* system, GTG appears functionally important in various stages of the recognition process.

3.23 Structure and Biological Implications — Spermine-DNA

While the results of this study confirm that spermine-DNA binding is polymorphic in nature, the specificity of polyamine-DNA binding is still open to considerable debate. Binding may occur through charge-charge interactions in a dynamic, non-specific manner, driven by the release of bound ions[69,75-77]. Alternatively, a static association may occur where polyamines utilize sequence dependent structural properties of DNA in the binding process, employing direct or water mediated hydrogen bonding interactions with DNA bases and phosphate oxygens, and van der Waals interactions with hydrophobic regions of DNA[79-84]. In this structure spermine binding to DNA embodies aspects of both the dynamic counterion condensation and static-specific models, with the crystallographically unique spermine molecules binding with specificity to different regions of the duplex. However, the results also suggest that an additional feature, not previously considered, exists in DNA potentially conferring specificity on the binding of spermine *viz.* electrostatic field variations in the phosphate backbone which arise from conformational modulations in the phosphate backbone.

There is strong evidence that the complexation of spermine with DNA in the present crystal structure occurs with a high degree of specificity, utilizing differences in the electrostatic properties of the phosphate backbone. Close contacts (< 4 Å, after accounting for the van der Waals radii of the amino groups and phosphates) between the cationic centers of the spermines and the backbone occur predominately in the vicinity of the α^+ phosphates which, as noted above, are regions of the DNA with increased electrostatic potential. The results of this study are corroborated by those from a theoretical study of spermine-DNA complexes by Zakrewska and Pullman[82] in which models of spermine bound to the DNA backbone were used. Zakrzewska's study indicated that spermine complexes with A-DNA are over 1.5 times stronger than the corresponding complexes of B-DNA with spermine. Since the phosphate backbone in the vicinity of the α^+ phosphates is "A-like", it is reasonable to assume that spermine should associate more stongly with these sites. Also, symmetry related spermine *A* molecules span the minor groove across phosphates adjacent to the sheared T·A base pair (Fig. 47) and this interphosphate distance, P(5) to P(10), is 1.3 Å shorter than in ideal B-DNA suggesting that spermine stabilizes this altered conformation.

In light of the evidence that spermine binding in this structure displays sequence specificity, it may be argued that spermine binding occurred in solution, resulting in a complex whose lifetime was sufficiently long so as to furnish the crystallization process with the necessary building blocks (much like the manner in which DNA-drug complexes are considered to form). The observation that the slight curvature in spermine *A* follows the contour of a DNA duplex suggests that spermine is associated with that particular duplex. Furthermore, the observation that spermine spans the minor groove and forms close contacts with α^+ phosphates suggests that spermine stabilizes the altered backbone and base conformations. The appealing aspects of the above description are that it offers a rationale for the crystallization process and, in consideration of the following arguments, has implications for protein recognition.

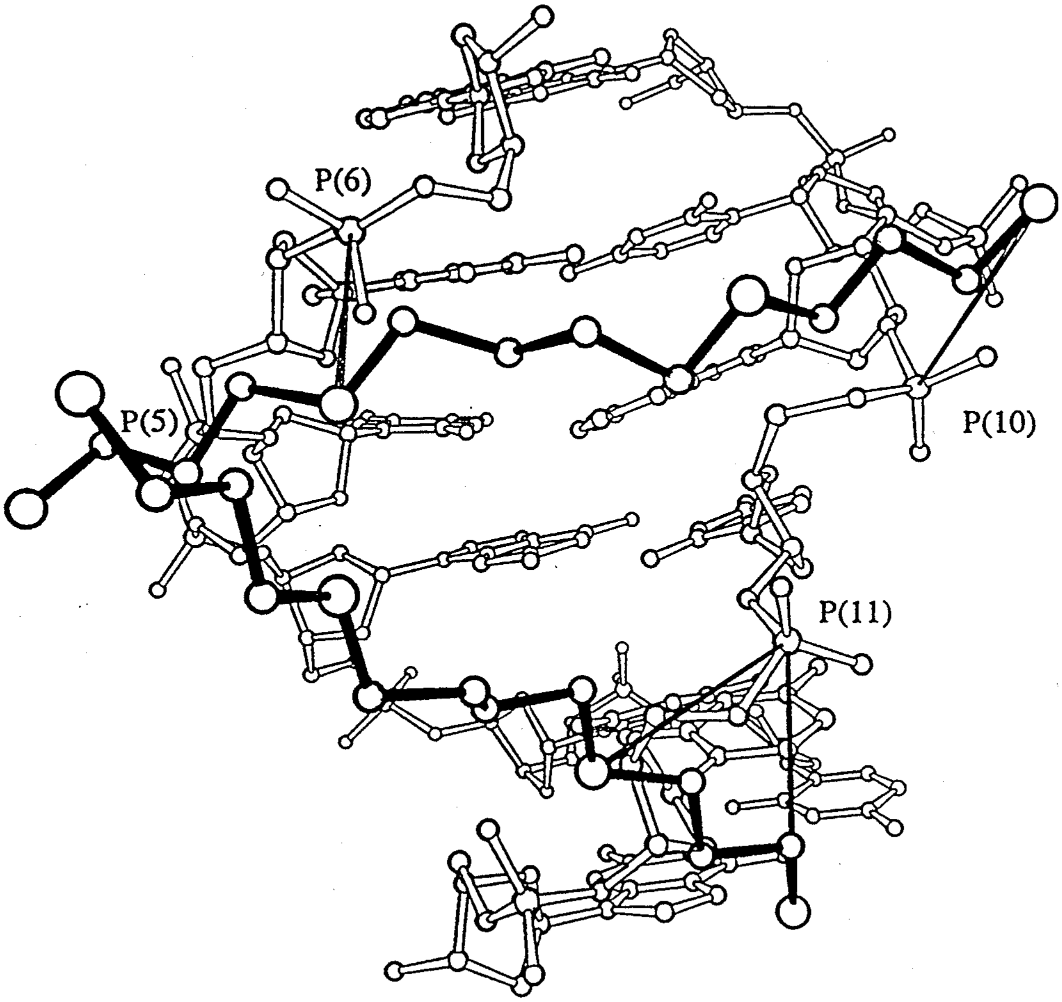


Fig. 47 Symmetry related spermine A (drawn with dark bonds) molecules span the minor groove in the vicinity of the α^+ phosphates and seem to stabilize the sheared T•A base-pair. Thin lines are drawn to show close contacts (< 4 Å, after subtracting the van der Waals radii of the phosphates and amino groups) between anionic phosphates and cationic amino groups. This figure was drawn with ORTEP[105].

Spermine may help aggregate DNA by stabilizing the stacking of duplex strands in solution. In the present hexamer duplex the most energetically favorable of the six end-to-end stacking possibilities is predicted to be 3'-G6 over 5'-C7[122] suggesting the association into dimers of duplexes. Spermine (*B*) can then bind to amenable base and backbone contacts at the stacking junction conferring additional stability on the dimer in solution. In this description, a dimer of hexamer duplexes (resembling a self-complementary dodecamer), the most prevalent strand length observed in B-DNA crystal structures) with three bound spermine molecules (an *A* across each of the T·A base-pairs and a *B* at the dimer junction) collectively constitutes the "crystal packing unit". The role of spermine *B*, inferred from inter-molecular contacts, is to stabilize the dimer of duplexes as spermine *B* makes only one DNA contact in the crystal other than with the two stacked duplexes. Spermine *A* molecules make a number of contacts with several duplexes and other symmetry related spermine molecules; they can be regarded as molecular handles which facilitate the aggregation of DNA molecules and the crystallization process by providing an asymmetric surface that incoming dimers from solution latch onto through direct and water mediated interactions.

The flexibility of DNA in solution and its susceptibility to transient opening at T·A base pairs[54,100,101,113-117] can provide spermine with the necessary sequence specific cues for binding, through the disposition of the phosphates on the backbone and hydration patterns along the duplex. More specifically, the over-twisting across the T4/G5 base step causes shearing of the T·A base pair and the corresponding adoption of α^+ conformations in the phosphate backbone at nucleotides G5 and C10. The contraction of the backbone in the vicinity of G5 and C10 accompanying the conformational adjustments facilitates stronger electrostatic interactions with cationic spermine molecules that are non-specifically bound in the vicinity. The spermine molecules respond by associating with the backbone near G5 and C10, tethering phosphates across the minor groove and, thereby, locking the sheared T·A out of its classical Watson-Crick arrangement. Spermine stabilizes the sheared base-pair by "pinching" together the minor groove across the G5 and C10 phosphates, and additionally by competing with the phosphates around nucleotides G5 and C10 for waters of hydration, thus stabilizing the α^+ conformations. The implication of this spermine-DNA association in solution for protein recognition is that in control regions, where the sequence GTG/CAC is a common element, spermine has the ability to stabilize a sheared T·A base pair. Consequently, the thymine methyl, O4 and imino H substituents protrude into the major groove and can act as signals to a non-specifically bound protein searching in proximity to its target site.

Whether spermine provides the driving force for the observed deviations from canonical B-DNA through its binding alone, or alternatively, by "capturing" a transient distortion in the DNA, cannot be ascertained conclusively from this study. The focus here has been on the latter description which is consistent with the finding that polyamine binding to DNA is sensitive to secondary structure[83,123-126]. With this crystal structure as a model, one can envision spermine increasing the fidelity of type II restriction endonucleases in the following ways; **i)** by associating with specific sequences of DNA, and directly participating in the protein-DNA interaction process, by providing an altered set of hydrophobic/electrostatic recognition contacts, and **ii)** by stabilizing structural deformations in DNA target sequences which would facilitate recognition by the restriction enzyme. It is feasible that spermine accelerates target sequence location by either one or a combination of these factors *in vivo*.

In conclusion, molecules in the crystal structure of the d(CGGTGG)/d(CCACCG) fragment of the *gal* operon display a distorted B-DNA conformation best described as a pre-recognition or transition-state-like conformation which spermine appears to stabilize. The unusual structural features observed in the GTG region provide supporting evidence for the hypothesis that the GTG trimer plays a "signaling role" in the overall regulatory process.

Chapter IV

The Structures and Properties of a Nucleoside and a Nucleotide Dimer Containing Modified Sugars

4.1 The Structure of the Pyranosyl Nucleoside 1-(2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranosyl)thymine and a Comparison with AZT

As part of an ongoing series of studies of modified nucleotides, 1-(2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranosyl)thymine (1) has been synthesized. The compound was first synthesized by Yamazaki, Matsudo, Sugiyama, Seto and Yamaoka[127] to demonstrate a new method of preparing 2',3'-unsaturated pyranosyl nucleosides. Several hexeno- and hexano-pyranosyl nucleosides, like their pento-furanosyl analogs, have long been known to exhibit significant biological activity. Only recently have the pyranosyl compounds been investigated as potential anti-Human Immunodeficiency Virus (HIV) drugs[128]. Presently, the most effective nucleoside drug used against acquired immunodeficiency syndrome (AIDS) is 3'-azido-3'-deoxythymidine (AZT), the structural characteristics of which have been well documented[129-131] and their relationship to its activity inferred[131-132]. However, the adverse side effects of AZT have necessitated the search for drugs with an equal or greater efficacy but lower toxicity. The difficulty in designing nucleosides as therapeutic agents for AIDS patients stems from the lack of detailed structural knowledge of the active sites of the cellular kinases responsible for phosphorylation of the nucleoside analogs. Also, detailed information about the active site of the viral reverse transcriptase which incorporates the resulting nucleotide into the growing DNA strand is still unavailable (The 2.9 Å structure of HIV reverse transcriptase complexed with Nevirapine has recently been published[133], however, only the $C\alpha$ backbone coordinates (except in a few regions) were used in the refinement and some inaccuracies in sequence assignment may exist. It is also reported that the Nevirapine probably alters the conformation of the nucleotide binding site). Therefore, using AZT as the best available model, it may be possible to effectively screen the pyranosyl compounds for anti-HIV activity by comparing the structural parameters of these strategically modified nucleosides with those of AZT. In this context, the present crystallographic study has been undertaken in an attempt to investigate the conformational features of (1) for comparison with those of AZT, as a predictive tool for its activity.

4.2 Experimental Procedures

4.21 Synthesis and Crystallization

Compound (1) was synthesized by condensation of permethylsilyated thymine with 3,4,6-tri-*O*-acetyl-D-glucal (Sigma) (Fig. 47), using a method analogous to that devised by Kondo, Nakai and Goto[134] for uracil.

4.211 Reagents

All reactions, up to and including the coupling of the silyated base to the sugar, had to be executed under completely anhydrous conditions. Solvents and reagents were dried as follows:

1) Benzene -- Anhydrous benzene (1 L, Aldrich) was decanted into a flask containing $CaCl_2$ (5.0 g), sealed, and allowed to sit for 3 hrs. The benzene was then filtered under vacuum, and refluxed over 1 g of finely divided sodium for 6 hrs. The benzene was then distilled into a flame dried flask and sealed.

2) Triethylamine (TEA) -- Reagent grade TEA (500 mL, Aldrich) was refluxed for 6 hrs. over finely divided CaH_2 (2 g) and distilled into a flame dried flask and sealed.

3) Chlorotrimethylsilane (TMSCl) (Aldrich) -- Prepared using the same method as for TEA.

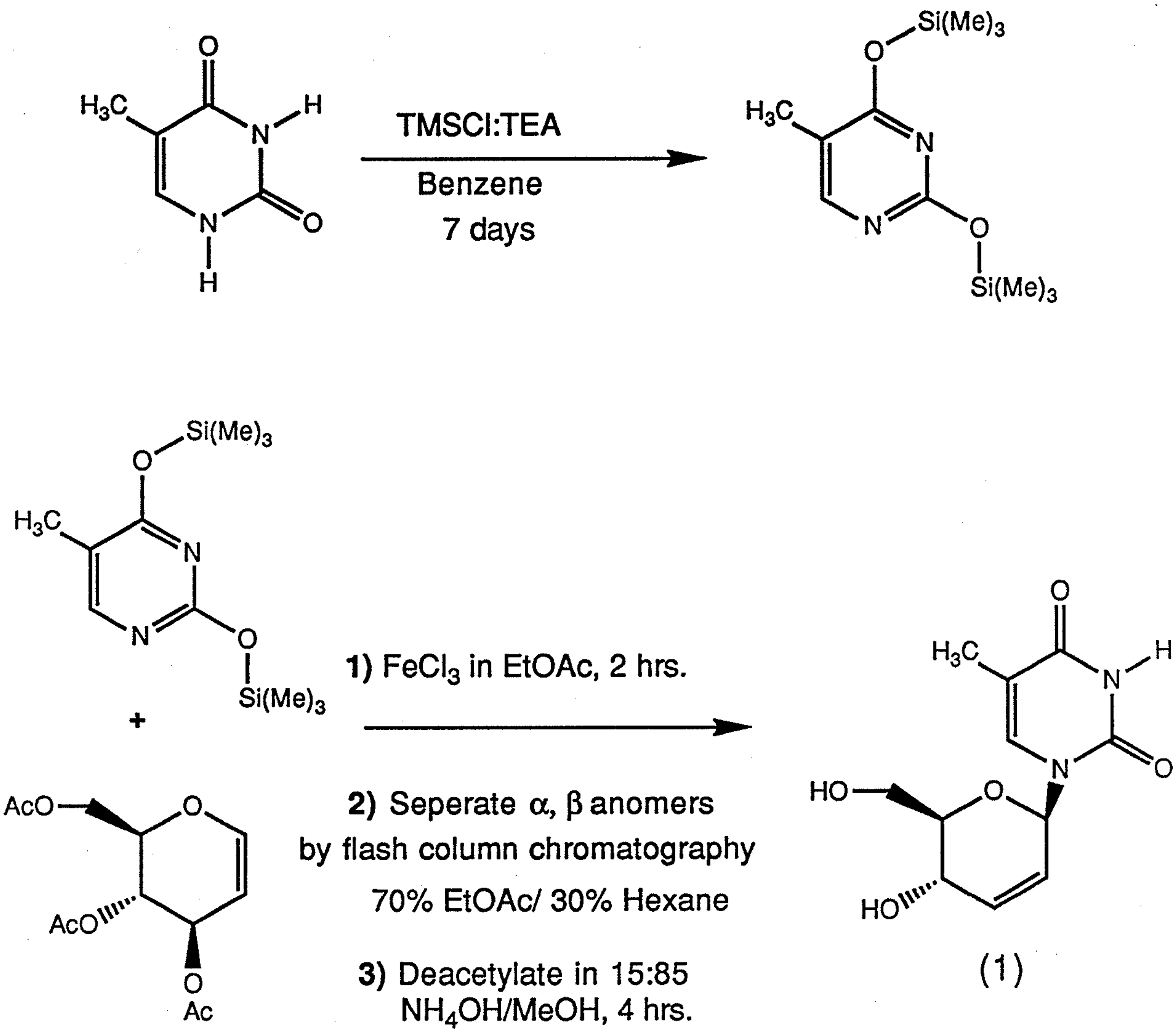


Fig. 48 Reaction scheme for the synthesis of 1-(2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranosyl)thymine (1).

^{1}H NMR spectra were obtained at 300.1 MHz (Bruker AM300). Signals are quoted as δ ppm downfield from internal tetramethylsilane unless otherwise stated. Positive ion FAB mass spectra (glycerol matrix, 8-keV xenon atoms) were obtained on a VG 7070E-HF spectrometer.

4.212 Synthetic Methods

1.0 g (7.9 mmol) of thymine (Sigma) was dried by evaporation of dry benzene (10 mL, 3 times) then suspended in 50 mL of dry benzene. 6.0 mL (48 mmol) of TMSCl and 3.0 mL (20 mmol) of TEA were added to the mixture, which was then sealed under a dry nitrogen atmosphere (using a teflon stopper) and stirred for seven days. The solution adopted a yellowish hue after about 4 days and turned brown after 6 days. After the seven days, the reaction mixture was filtered under positive pressure (using dry nitrogen) through glass wool and the solvent evaporated to a thick brown oil which was used in subsequent steps without further purification. 20 mL of anhydrous ethyl acetate were added to the flask containing the crude bis(trimethylsilyl)thymine, along with 2.6 g (10 mmol) of tri-O-acetyl-D-glucal and 2.6 g (16 mmol) of ferric chloride, and the solution was sealed and stirred for 1 hr. After 1 hr., the reaction was deemed complete (by thin layer chromatography in 70:30 (v/v) ethyl acetate:hexane). The reaction mixture was then washed with 70 mL of saturated sodium bicarbonate solution to remove some of the iron salts, and filtered under vacuum. The ethyl acetate layer was then evaporated to a thick, black oil on a rotary evaporator. The resulting solid was dissolved in 4 mL of ethyl acetate, and loaded on a 50 mL sintered glass funnel filled with silica gel (Aldrich, 60 grade 230-400 mesh) and eluted with a large excess of ethyl acetate. This successfully removed almost all traces of the darker colored material. Solvent was removed on a rotary evaporator, leaving a thick slightly yellow oil. The crude product was purified on silica gel by flash column chromatography, using 70:30 (v/v) ethyl acetate hexane, yielding 542 mg (1.6 mmol, 20%) of 1-(2',3'-dideoxy-4',6'-O-acetyl-β-D-*erythro*-hex-2'-enopyranosyl)thymine as a glassy solid.

^{1}H NMR ($CDCl_3$) δ 8.07 (1 H, br s, NH), 6.97 (1 H, s, 6-H), 6.53 (1 H, s, 1'-H), 6.17 (1 H, m, $J_{2',3'}$ 10.2 Hz, 2'-H), 5.76 (1 H, m, $J_{2',3'}$ 10.2 Hz, 3'-H), 5.39 (1 H, m, 4'-H), 4.22 (2 H, d, 6'-H), 4.01 (1 H, m, 5'-H), 2.12 (3 H, s, OAc), 2.09 (3 H, s, OAc), 1.93 (3 H, s, Me)

MS (positive ion FAB, xenon atoms) *m/z* 339 (25, $[M + H]^+$), 213 (70, $[M - \text{thymine}]^+$), 127 (43, $[\text{thymine} + H]^+$)

518 mg (1.5 mmol) of 1-(2',3'-dideoxy-4',6'-O-acetyl-β-D-*erythro*-hex-2'-enopyranosyl)thymine were dissolved in 40 mL of 15:85 NH_4OH:MeOH, and allowed to stir for four hours in a sealed flask. The reaction was monitored by TLC (ethyl acetate). After the reaction was complete, the solvent was removed, yielding 305 mg (1.2 mmol, 80 %) of compound (1) as a clear, glassy solid.

MS (positive ion FAB) *m/z* 255 (19, $[M + H]^+$), 129 (42, $[M - \text{thymine}]^+$), 127 (100, $[\text{thymine} + H]^+$)

Crystals of (1) were obtained by slow evaporation of a methanol/ethyl acetate solution of the compound.

4.22 Data Collection

The unit cell and diffraction symmetry, as well as basic data collection parameters, were determined using methods analogous to those described for the DHEX data collection in Chapter II.

The crystal data are summarized as follows:

Unit cell; Monoclinic, $a = 7.966(3)$, $b = 18.004(3)$, $c = 8.346$ (3) Å, V = 1196.6 Å^3, Z (no. molecules per cell) = 4, Space group $P2_1$ (from systematic absences and structure analysis), Calculated density = 1.411 Mg•m^{-3}, μ = 9.13 cm^{-1}, $F(000)$ = 536.

Data were collected at room temperature using a colorless crystal of dimensions 0.30 x 0.25 x 0.40 mm on a Rigaku AFC6S diffractometer equipped with an evacuated beam collimator and detector tunnel. Graphite monochromated Cu$K\alpha$ radiation (λ = 1.54178 Å) was employed with the intensities measured 400 mm from the crystal. Measured data to 2θ = 106° included reflections h (0-8), k (0-18), l (-8-8). Scans were of the $\omega - 2\theta$ type, where the ω scan speed was 32° min.$^{-1}$. The scan width in ω was (1.57 + 0.30 tanθ)° and weak reflections were rescanned up to four times and the counts accumulated. Of the 1582 reflections measured, 1460 were considered unique. For averaged data, R_{sym} = 0.035. 1285 reflections (88%) were classified as observed having $I \geq 3\sigma(I)$. Three standard reflections monitored throughout data collection showed less than 1% change in their intensities over the period of the experiment indicating no decay correction was required. Lorentz-polarization corrections were applied and an empirical absorption correction was included in the data processing step (transmission factors, 0.875 - 1.000, based on *psi*-scans of three reflections).

4.23 Structure Solution and Refinement

4.231 Direct Methods

The structure of (1) was solved using direct methods. The term direct methods refers to an *ab initio* method of solving crystal structures by deriving the structure factor phases directly from the observed amplitudes through mathematical relationships. For the statistical relationships relating phases to amplitudes to be valid, two assumptions about the electron density in the model must be made:

1) The electron density is everywhere positive, (*i. e.* $\rho(r) \geq 0$);

2) The electron density is not continuous, but is composed of discrete atoms (*i.e.* Dirac delta functions).

To meet condition 2), the structure factors $F_o(\mathbf{h})$ are normalized, so that they are placed on an absolute scale, and also, so that they represent scattered waves from discrete point atoms with no thermal motion. The expression for the normalized structure factors, $E(\mathbf{h})$, is;

$$|E(\mathbf{h})|^2 = k^2 \frac{|F_e(\mathbf{h})|^2}{\varepsilon \sum_j f_j^2} \qquad (24)$$

k is a scaling factor required to convert the relative $F_e(\mathbf{h})$ values, derived from the intensities, and modified to remove the effects of atomic thermal motion which are inherent in $F_o(\mathbf{h})$ values, to absolute quantities based on the scattering material in the unit cell, and ε is an integer inserted to account for symmetry inherent in particular zones depending on the space group. Using E values, condition (2) (electron density is composed of discrete atoms) can be used to derive the relation upon which direct methods is based. For a structure composed of well resolved and almost equal atoms (for simplicity a centrosymmetric electron density distribution will be considered), the functions $\rho(r)$ and $\rho^2(r)$ are almost identical and show maxima at the same positions (if the electron density is considered to consist of Dirac delta functions, $\rho(r)$ and $\rho^2(r)$ are identical). The Fourier transform of $\rho(r)$ is $(1/V)E(\mathbf{h})$, and for the case of all equal atoms:

$$E(\mathbf{h}) = Nf\sum_{j=1}^{N}\cos(2\pi i\mathbf{h}\bullet r_j). \qquad (25)$$

This expression is analogous to Eqn. (7), with the condition of centrosymmetry reducing the exponential term to a cosine function. As is evident from Eqn. (25), the phase term can only adopt two values; 0 and π (or '+' and '-'). The structure factor of $\rho^2(r)$ is defined similarly;

$$G(\mathbf{h}) = Ng\sum_{j=1}^{N}\cos(2\pi i\mathbf{b}\bullet r_j) \qquad (26)$$

where g is the scattering factor of the "squared" atom.

The Fourier transform of $\rho^2(r)$ is $(1/V)G(\mathbf{h})$ and, because of the convolution theorem (Eqn. (6)), it will correspond to the convolution product $(1/V)E(\mathbf{h})\cap(1/V)E(\mathbf{h})$.

Since E(**h**) is only non-zero at reciprocal lattice points, the convolution integral (Eqn. (6)) can be replaced by a summation;

$$G(\mathbf{h}) = \frac{1}{V}\sum_{\mathbf{k}} E_{\mathbf{k}}E_{\mathbf{h-k}}. \qquad (27)$$

From the ratio of Eqn. (25) and Eqn. (26);

$$E(\mathbf{h}) = (f/g)G(\mathbf{h}) = T \bullet G(\mathbf{h}), \qquad (28)$$

and Eqn. (27) becomes;

$$E(\mathbf{h}) = \frac{T}{V}\sum_{\mathbf{k}} E_{\mathbf{k}}E_{\mathbf{h-k}}, \qquad (29)$$

which is called Sayre's equation[135]. For reflections **h** with |E(**h**)| being sufficiently large, it is likely that the sum on the right side of Eqn. (29) will contain more products E(**k**) x E(**h-k**) having the same sign as E(**h**) itself, than terms of opposite sign. Otherwise Eqn. (29) could not hold. This is especially true for those terms for which |E(**k**)| *and* |E(**h-k**)| are large, since they are the major contributors to the sum. Therefore, there exists a more than 50% probability that for large E values,

$$s(\mathbf{h}) = s(\mathbf{k}) \cdot s(\mathbf{h-k}) \qquad (30)$$

where s(**h**) denotes the phase or "sign" ('+' or '-') of E(**h**). This equation remains valid if on the left side **h** is replaced by -**h** (since s(**h**) = s(-**h**)).

Setting $-\mathbf{h} = \mathbf{h}_1$, $\mathbf{k}=\mathbf{h}_2$ and $\mathbf{h}-\mathbf{k} = \mathbf{h}_3$, Eqn. (30) becomes;

$$s(\mathbf{h}_1) = s(\mathbf{h}_2)\cdot s(\mathbf{h}_3),$$

or

$$s(\mathbf{h}_1)\cdot s(\mathbf{h}_2)\cdot s(\mathbf{h}_3) = 1 \tag{31}$$

if the three reflections $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$ satisfy the equation

$$\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0. \tag{32}$$

Reflection triples for which Eqn. (32) holds are said to be related by a Σ_2 relation, which holds in both centric and acentric structures. If a large number of signs is known, it can happen that an unknown reflection is contained in more than one Σ_2-relation. If this is the case, Eqn (30) becomes;

$$s(\mathbf{h}) = \sum_{\mathbf{k}} s(\mathbf{k})\bullet s(\mathbf{h}-\mathbf{k}) \tag{33}$$

and is a better approximation of Sayre's equation. The probability for the sign of b to be "+" is given by (for an equal atom structure);

$$P_{+} = 1/2 + (1/2)\tanh\left[N^{-1/2}|E(\mathbf{h})|\sum_{\mathbf{k}} E(\mathbf{k})\bullet E(\mathbf{h}-\mathbf{k})\right]. \tag{34}$$

the probability of E(**h**) being negative is simply $P_- = 1 - P_+$.

For phase determination in acentric structures (where phases can adopt any value between 0 and 2π) the expanded form of Eqn. (29) for acentric structures can be separated into its real and imaginary parts, yielding;

$$|E(\mathbf{h})|\sin\varphi(\mathbf{h}) = \frac{T}{V}\sum_{\mathbf{k}} |E(\mathbf{k})\bullet E(\mathbf{h}-\mathbf{k})|\sin[\varphi(\mathbf{k})+\varphi(\mathbf{h}-\mathbf{k})]$$

and

$$|E(\mathbf{h})|\cos\varphi(\mathbf{h}) = \frac{T}{V}\sum_{\mathbf{k}} |E(\mathbf{k})\bullet E(\mathbf{h}-\mathbf{k})|\cos[\varphi(\mathbf{k})+\varphi(\mathbf{h}-\mathbf{k})]$$

which by division yield,

$$\tan\varphi(\mathbf{h}) = \frac{|E(\mathbf{h})|\sin\varphi(\mathbf{h}) = \sum_{\mathbf{k}} |E(\mathbf{k})\bullet E(\mathbf{h}-\mathbf{k})|\sin[\varphi(\mathbf{k})+\varphi(\mathbf{h}-\mathbf{k})]}{|E(\mathbf{h})|\cos\varphi(\mathbf{h}) = \sum_{\mathbf{k}} |E(\mathbf{k})\bullet E(\mathbf{h}-\mathbf{k})|\cos[\varphi(\mathbf{k})+\varphi(\mathbf{h}-\mathbf{k})]} \qquad (35)$$

Expression (35) is called the tangent formula^{136}, and is the key formula for phase determination in acentric structures.

To solve a structure by direct methods, a list of all reflections related by Σ_2-relations must be developed, and then using a small set of starting phases, phases must be assigned to the remaining reflections related by the Σ_2-relations. Sign indicators of greater than a probability of 95% are usually taken as correct.

As already described, a small starting set of phases must be known in advance to initiate the phase determining process. Three and sometimes fewer reflections are assigned arbitrary phases at the start. These reflections are chosen according to specific space group dependent rules to be origin determining. The origin defining reflections are chosen so that they enter into as many Σ_2-relations as possible. Other phases may have to be determined (prior to this point) with a high degree of probability by Σ_1-relations, which are special cases of the more general Σ_2-expression and depend only on the magnitude of the normalized structure factors involved. In the case of non-centrosymmetric space groups an additional phase must be specified in order to define the positive sense of the coordinate system (*i.e.* fixing the enantiomorph). Once this is accomplished the phasing of the remaining bulk of reflections is initiated.

Once all of the Σ_2-relations have been exhausted using known and probability determined phases there will likely remain some reflections to which phases have not been assigned. Using the MITHRIL[137] software package, the unknown phases are permuted with values of 0 and π in the centrosymmetric case, and $\pm\pi/4$ and $\pm3\pi/4$ in the non-centrosymmetric case, thereby generating sets of phases corresponding to each permutation. In acentric structures, initial and subsequently determined approximate phases are refined in an iterative manner using the tangent formula (Eqn. 35). Each set of phases determined for a given permutation of the variable phases is assessed according to its internal consistency, the validity of the structural model the E values produce upon Fourier transformation, and the values of various statistical "figures of merit". Generally, a structurally reasonable model with the best overall figures of merit constitutes the correct solution, which can subsequently be improved by least-squares refinement.

4.232 Solution and Refinement

The structure of (1) was solved by direct methods. E-values were sorted based on their intensities, and Σ_2-relations were generated for the reflections with the 244 largest E-values. This yielded a total of 3500 Σ_2-relations. Sixty-five phase sets were generated corresponding to the different possible combinations of the variable reflections. The best 36 phase sets generated (*i.e.* the 36 phase sets with the highest figures of merit), were subjected to tangent formula refinement. After the refinement, the phase set with the highest combined figure of merit yielded an E-map (an electron density map calculated using E values) showing the complete heavy atom framework of two independent molecules in the asymmetric unit. The non-H atoms were refined to convergence and subsequent difference Fourier maps revealed 21 of the hydrogen atoms. Full-matrix least-squares refinement continued to convergence with all non-H atoms assigned anisotropic thermal parameters and available H's, isotropic temperature factors. At this stage, a difference map indicated positions for four of the six methyl H's, but attempts to refine them were unsuccessful. However, these positions were used in orienting all six methyl hydrogens in ideal positions. The H of O44' A was difficult to locate. Difference maps showed two very small peaks, both in chemically reasonable positions to be the hydrogen of O44' A. When each peak position was refined individually as the hydrogen, the refinement tended to shift the hydrogen toward the other peak coordinates. This difficulty in refining the H position, in combination with the development of a large temperature factor associated with the atom, suggested the possibility that the hydrogen was disordered. Therefore, the H6A occupancy was assigned 0.5 at each of the two observed peak positions on the difference map and refined along with all other atoms. Their coordinates and isotropic temperature factors were allowed to vary in subsequent refinement cycles. No attempt to determine the occupancies of the disordered hydrogen was made. From this point, convergence of the entire structure was reached within three cycles and yielded R = 0.0322 and wR =

0.0413 for 1285 observed data. The maximum parameter shift was 0.16σ on the final cycle of refinement and the standard deviation in an observation of unit weight was 1.831 for 417 variables. The function minimized was Eqn. (17), where $w = 4F_o^2/\sigma^2(F_o^2)$. A weighting analysis confirmed the suitability of the chosen weights by showing uniform average values of Eqn. (17) over ranges of index classes, F_o, and $\sin\theta/\lambda$. Following convergence, the peaks on the difference Fourier map had densities corresponding to $\pm 0.16 e\text{Å}^{-3}$ randomly scattered throughout the unit cell. When the complete data set was used R = 0.0387 and wR = 0.0486. The enantiomer refined was chosen on the basis of the chirality of the starting sugar (*vide supra*). Scattering factors and anomalous dispersion corrections for non-hydrogen atoms were from Ref. 138 and from Ref. 139 for the scattering factors of hydrogen. Final atomic positional parameters and equivalent isotropic thermal parameters are given in Table 8. The two molecules in the asymmetric unit are displayed in Fig. 49 along with the atom labels.

Table 8 Final fractional positional parameters and equivalent isotropic thermal parameters (Å^2) with estimated standard deviations in parentheses.

	x	y	z	B_{eq}^*
O2A	0.4172(5)	0.8913	-0.1269(5)	5.9
O4A	0.4109(5)	0.7177(3)	0.2637(5)	6.4
O5′A	0.7029(4)	1.0151(3)	0.0974(4)	4.5
O44′A	0.6890(6)	1.2131(3)	0.1910(5)	5.4
O55′A	1.0093(4)	1.0647(3)	0.2253(5)	5.6
N1A	0.4987(5)	0.9260(3)	0.1254(5)	4.6
N3A	0.4175(6)	0.8061(4)	0.0740(6)	4.9
C1′A	0.5295(7)	1.0030(4)	0.0760(8)	5.1
C2A	0.4426(6)	0.8753(4)	0.0137(7)	4.6
C2′A	0.4242(7)	1.0572(5)	0.1637(8)	6.0
C3′A	0.4885(8)	1.1176(4)	0.2303(8)	5.7
C4A	0.4412(7)	0.7823(4)	0.2297(7)	4.8
C4′A	0.6694(7)	1.1350(4)	0.2195(6)	4.5
C5A	0.5033(6)	0.8379(4)	0.3388(6)	4.7
C5′A	0.7474(6)	1.0925(4)	0.0847(6)	4.1
C6A	0.5277(7)	0.9063(4)	0.2834(7)	4.9
C7A	0.5378(8)	0.8168(4)	0.5101(8)	6.1
C55′A	0.9347(7)	1.0959(4)	0.0874(7)	5.2
O2B	0.0621(6)	0.7493(3)	0.5914(5)	6.5
O4B	0.0613(6)	0.9131(4)	0.1812(6)	7.0
O5′B	0.0367(4)	0.5965(3)	0.3438(4)	4.2
O44′B	-0.1744(6)	0.4184(3)	0.3568(5)	5.9
O55′B	0.3196(5)	0.5150(4)	0.4040(6)	7.4
N1B	-0.0670(5)	0.7148(4)	0.3567(5)	4.5
N3B	0.0531(6)	0.8313(3)	0.3848(6)	5.1

C1′B	-0.0919(6)	0.6385(4)	0.4140(6)	4.1
C2B	0.0206(7)	0.7637(4)	0.4544(7)	5.1
C2′B	-0.2608(7)	0.6096(4)	0.3739(6)	4.5
C3′B	-0.2825(7)	0.5419(4)	0.3208(7)	4.7
C4B	0.0194(7)	0.8515(4)	0.2278(8)	5.1
C4′B	-0.1391(6)	0.4901(4)	0.2945(7)	4.2
C5B	-0.0717(7)	0.7976(4)	0.1306(7)	4.9
C5′B	0.0174(6)	0.5191(4)	0.3780(7)	4.4
C6B	-0.1058(7)	0.7334(4)	0.2001(7)	4.7
C7B	-0.1125(9)	0.8161(5)	-0.0391(8)	7.1
C55′B	0.1783(7)	0.4835(5)	0.3224(8)	5.4

$$^{*}B_{eq} = \frac{8\pi^2}{3} \sum_{i=1}^{3} \sum_{j=1}^{3} U_{ij} a_i^{*} a_j^{*} \vec{a}_i \cdot \vec{a}_j$$

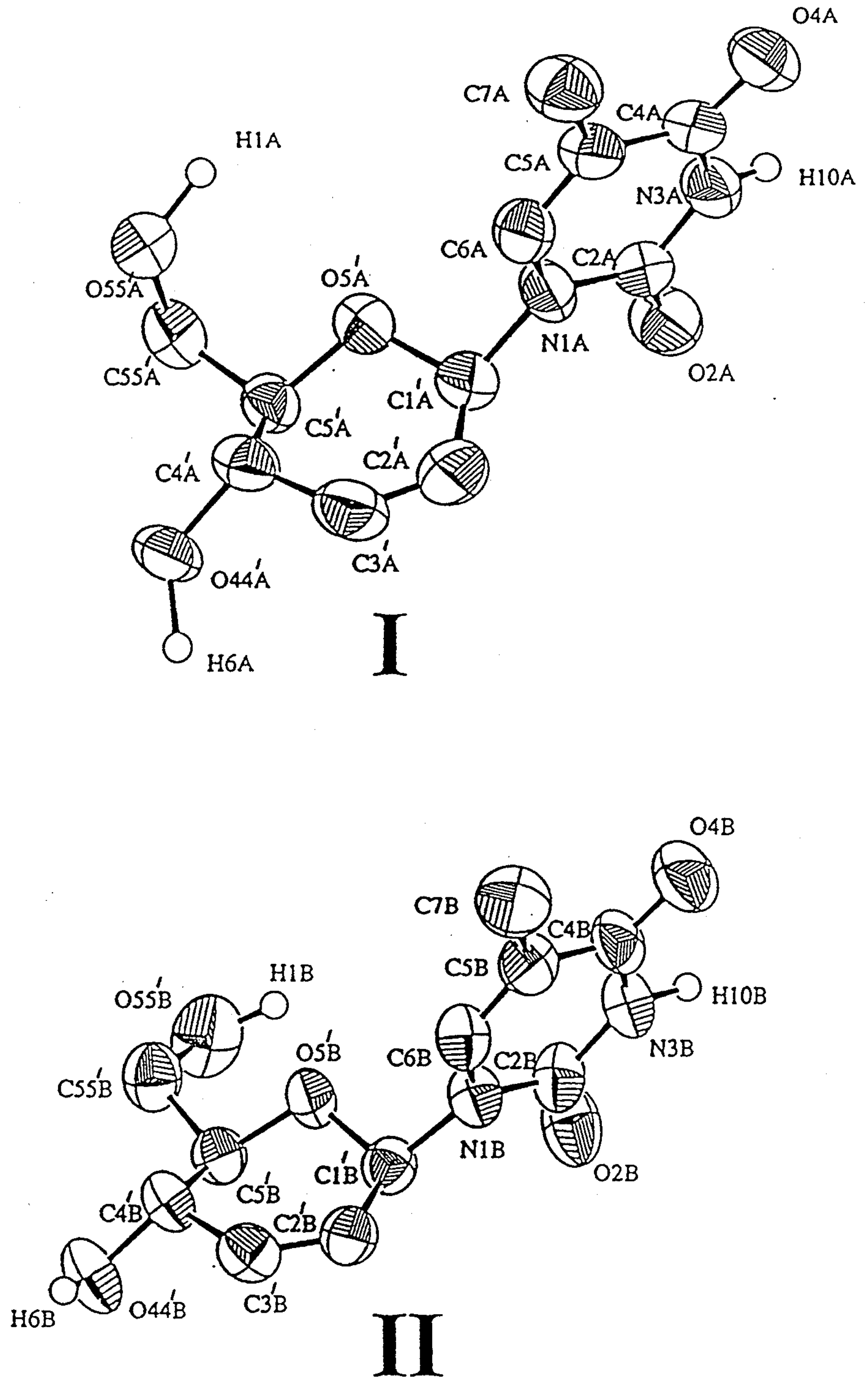


Fig. 49 An ORTEP[105] plot of the two molecules (I & II) in the asymmetric unit with 50% probability anisotropic thermal ellipsoids for the non-H atoms (nominal isotropic for H's) and atomic numbering scheme. Some hydrogens have been omitted for clarity.

4.3 Results and Discussion

4.31 Molecular Geometry

Bond lengths and bond angles are given in Table 9. The two independent molecules in the asymmetric unit possess similar molecular dimensions, with bond lengths and angles that are within the normal range[140]. The two molecules differ most notably in their N-glycosidic torsion angles, χ (O5'-C1'-N1-C6) which are 64.5(6)° (*+sc*) in **I** and 72.7(6)° (*+sc*) in **II**, as well as in the values of their torsion angles γ (C4'-C5'-C55'-O55'), which are 62.0(7)° (*+sc*) in **I**, and 178.6(5)° (*+ap*) in **II**. The value of γ adopted in **I** represents a more sterically hindered (close O55'A• • •C4'A contact), thus less energetically favorable, conformation than the value found in **II**, the attainment of which may be ascribed to the fact that the O55'A hydrogen, H1A, is involved in an intermolecular hydrogen bond (Table 10). The γ torsion angle in **II** is stabilized by an intramolecular hydrogen bond between H1B and O5'B (Table 3). The O5'-C1' bond differs in length from the O5'-C5' bond in **I**, with the O5'-C5' bond length being shorter by 0.04(1) Å. An analogous trend is observed in other pyranosyl nucleosides, *e.g.* amicetin[141], cytosamine triacetate[142] and 1-(2',3'-dideoxy-β-D-erythro-hexopyranosyl)thymine[143]. However, the same is not true in **II**, where the O5'-C1' and O5'-C5' bond lengths are not significantly different (at $p > 0.05$). The C1'-N1 glycosyl bond distances for **I** and **II** are identical within experimental error, with values of 1.468(8) Å, and 1.470(7) Å, respectively. The dimensions of the thymine moiety in I and II are in good agreement with those of 1-methyl thymine[144], thymidine[145] and 1-(2-hydroxyethyl)thymine[146]. The thymine rings in **I** and **II** are essentially planar, with average deviations of

±0.005 Å and ±0.013 Å from their respective mean-planes through N1, C2, N3 C4, C5 and C6.

Table 9 Bond lengths (Å) and angles (°) for non-hydrogen atoms.

O2A	C2A	1.221(6)	O2B	C2B	1.211(6)
O4A	C4A	1.221(7)	O4B	C4B	1.223(7)
O5'A	C1'A	1.405(6)	O5'B	C1'B	1.411(6)
O5'A	C5'A	1.444(6)	O5'B	C5'B	1.431(6)
O44'A	C4'A	1.436(6)	O44'B	C4'B	1.423(7)
O55'A	C55'A	1.401(7)	O55'B	C55'B	1.421(7)
N1A	C1'A	1.468(8)	N1B	C1'B	1.470(7)
N1A	C2A	1.373(7)	N1B	C2B	1.379(7)
N1A	C6A	1.380(7)	N1B	C6B	1.378(7)
N3A	C2A	1.360(7)	N3B	C2B	1.376(8)
N3A	C4A	1.377(7)	N3B	C4B	1.381(7)
C1'A	C2'A	1.490(9)	C1'B	C2'B	1.475(8)
C2'A	C3'A	1.320(9)	C2'B	C3'B	1.306(8)
C3'A	C4'A	1.480(8)	C3'B	C4'B	1.494(8)
C4A	C5A	1.435(9)	C4B	C5B	1.449(8)
C4'A	C5'A	1.505(7)	C4'B	C5'B	1.508(7)
C5A	C6A	1.330(8)	C5B	C6B	1.324(8)
C5A	C7A	1.499(8)	C5B	C7B	1.484(9)
C5'A	C55'A	1.493(8)	C5'B	C55'B	1.516(8)

C1'A	O5'A	C5'A	112.5(4)	C1'B	O5'B	C5'B	110.9(4)
C1'A	N1A	C2A	119.4(5)	C1'B	N1B	C2B	118.3(5)
C1'A	N1A	C6A	119.1(5)	C1'B	N1B	C6B	120.5(5)
C2A	N1A	C6A	121.5(5)	C2B	N1B	C6B	120.4(5)
C2A	N3A	C4A	128.0(5)	C2B	N3B	C4B	126.9(5)
O5'A	C1'A	N1A	106.3(4)	O5'B	C1'B	N1B	105.2(4)
O5'A	C1'A	C2'A	113.5(5)	O5'B	C1'B	C2'B	112.5(5)

N1A	C1’A	C2’A	112.4(5)	N1B	C1’B	C2’B	112.7(5)
O2A	C2A	N1A	122.7(5)	O2B	C2B	N1B	123.1(6)
O2A	C2A	N3A	123.3(5)	O2B	C2B	N3B	122.5(6)
N1A	C2A	N3A	114.0(5)	N1B	C2B	N3B	114.3(5)
C1’A	C2’A	C3’A	122.0(6)	C1’B	C2’B	C3’B	121.3(6)
C2’A	C3’A	C4’A	121.2(6)	C2’B	C3’B	C4’B	122.4(6)
O4A	C4A	N3A	119.5(5)	O4B	C4B	N3B	119.5(6)
O4A	C4A	C5A	125.7(6)	O4B	C4B	C5B	124.4(6)
N3A	C4A	C5A	114.8(5)	N3B	C4B	C5B	116.0(5)
O44’A	C4’A	C3’A	109.1(5)	O44’B	C4’B	C3’B	110.8(5)
O44’A	C4’A	C5’A	109.0(4)	O44’B	C4’B	C5’B	108.3(4)
C3’A	C4’A	C5’A	110.9(5)	C3’B	C4’B	C5’B	110.1(5)
C4A	C5A	C6A	118.4(5)	C4B	C5B	C6B	116.4(5)
C4A	C5A	C7A	118.8(5)	C4B	C5B	C7B	118.7(6)
C6A	C5A	C7A	122.8(6)	C6B	C5B	C7B	124.8(6)
O5’A	C5’A	C4’A	109.2(4)	O5’B	C5’B	C4’B	109.6(4)
O5’A	C5’A	C55’A	106.6(5)	O5’B	C5’B	C55’B	104.8(4)
C4’A	C5’A	C55’A	113.4(5)	C4’B	C5’B	C55’B	114.1(5)
N1A	C6A	C5A	123.2(6)	N1B	C6B	C5B	125.7(6)
O55’A	C55’A	C5’A	113.7(4)	O55’B	C55’B	C5’B	110.5(5)

			Distances (Å)		Angles (°)
D	*A*	*A at*	*D*···*A*	H···*A*	*D*–H···*A*
O55'A–H1A	···O4B	*1+x, y, z*	2.79(7)	1.6(1)	170(7)
O44'A–H6A	···O2A	*1-x,1/2+y,-z*	3.35(7)	2.5(3)	124(15)
O44'A–H6A'	···O2B	*1-x,1/2+y,1-z*	2.74(6)	1.7(1)	165(10)
N3A–H10A	···O44'A	*1-x,-1/2+y,-z*	2.89(6)	2.07(6)	165(5)
O44'B–H6B	···O2A	*-x,-1/2+y,-z*	2.74(6)	1.90(7)	167(7)
N3B–H10B	···O44'B	*-x,1/2+y,1-z*	2.82(6)	1.91(8)	154(6)
O55'B–H1B	···O5'B	*x,y,z*	2.726(6)	2.30(7)	104(5)

Table 10 Distances and angles for hydrogen bonds.

Puckering parameters for the pyrenose rings were found (using the method of Cremer and Pople[147]) to be $\phi = 80.2°$, $\theta = 129.1°$, Q = 0.485 Å in **I** and $\phi = 88.6°$, $\theta = 130.9°$, Q = 0.494 Å in **II**. These parameters are structurally manifested in both molecules as slightly distorted half-chair conformations (distorted towards twist boats), with O5' *endo* and C5' *exo*. The three substituents on the pyrenose rings adopt the less sterically hindered equatorial positions.

Molecules in the crystal form an extensive network of intermolecular hydrogen bonds (Table 10). A packing diagram is given in Fig. 50, in which the disordered hydrogen bonded to O44'A is shown at only one position, *viz.* H6A'. It is interesting to note (although not shown in Fig. 50). that in both positions

H6A and H6A', the disordered hydrogen is involved in hydrogen bonding interactions (Table 10). The ability to bond in both environments provides an explanation for the observed disorder in H6A. Despite the apparent overlap of the thymine rings in the asymmetric unit, their interplanar spacing of 3.65 Å and their 9.8° dihedral angle indicate that π-π stacking interactions play a less important role in the crystal packing.

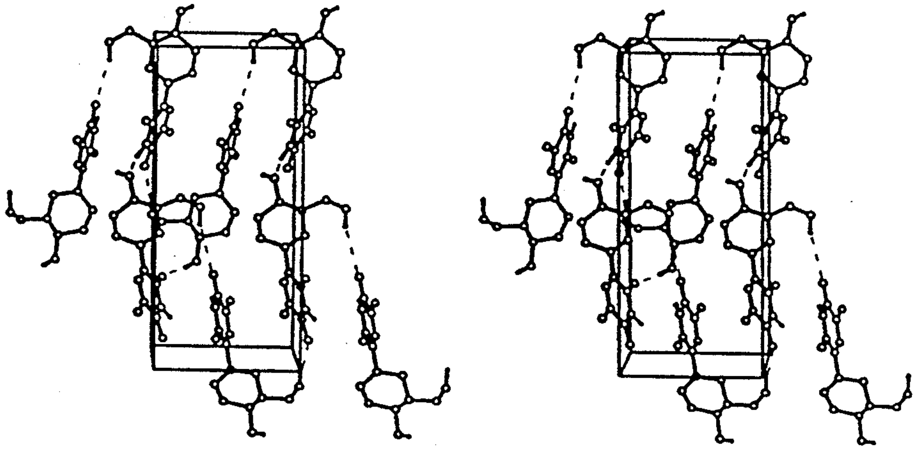
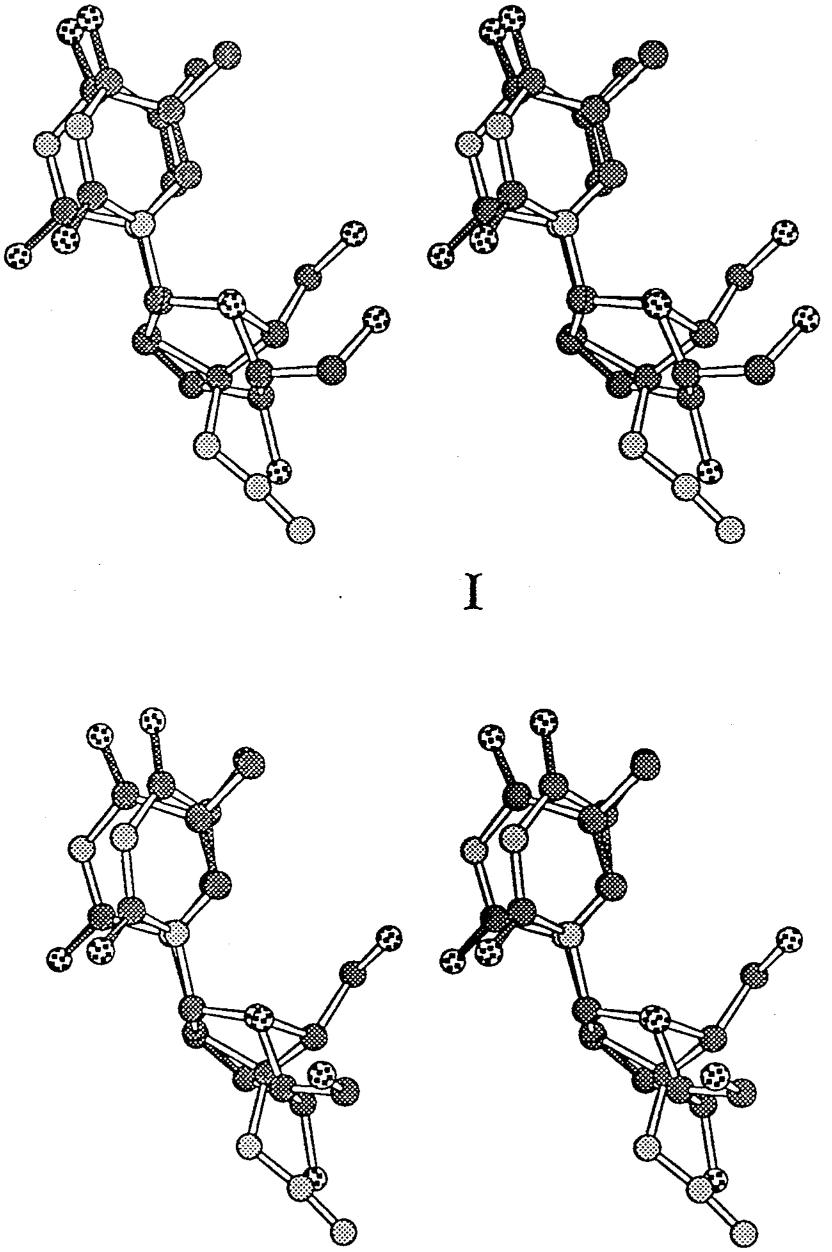


Fig. 50 Stereoscopic view prepared using PLUTO[148] of molecular packing in the unit cell. Some of the hydrogen bonds are indicated by dotted lines. The directions of axes are x(-*a* to the left), y (+*b* at the top) and z (+*c* into the page).

4.32 Biological Implications

Currently, AZT has the highest therapeutic index of the nucleoside analogs that are known to inhibit the replication of HIV-I *in vivo.* AZT has a well defined mode of action. It becomes activated, like natural nucleosides, through phosphorylation by cellular kinases, and is subsequently combined with the HIV's reverse transcriptase resulting in either blockage of the active site as a competitive inhibitor of natural nucleotides, or incorporation into viral DNA leading to premature chain termination. The structural features of AZT, implicated as crucial in recognition by phosphorylating enzymes are two-fold. Firstly, the relative position of the electron-rich terminus of the azido group with respect to the thymine moiety is deemed important[130,131] from the perspective of hydrogen bonding and potential electrostatic interactions within the active site. Secondly, it is important that the rotation of the base about the N-glycosyl bond be energetically feasible, so that the potential for hydrogen bonding interactions between the base and the cognate enzyme can be fully realized, while maintaining the integrity of the interaction between the terminal azido nitrogen and the enzyme[131].

It may be suggested that any active analog of AZT must display similar structural features. The structure of (1), in both forms **I** and **II**, has been compared to the structure of the purported active conformer of AZT[130,131], as illustrated in Fig. 51. The relative position of the O44' hydroxyl (the azido counterpart) and the thymine moiety in (1) is described by the intramolecular distances listed in Table 11; analogous distances in AZT are included in the table for comparison.



I

II

Fig. 51 Stereoviews of the crystal structure conformations of the two independent molecules of the title compound shown with N1, C1' and C2' superimposed on N1, C1' and C2' of the purported active conformer of AZT[131] (the base belonging to AZT can be readily identified, because in AZT the torsion angle χ is 3.5° (+*sp*)). Atom shading scheme is; O, stippled, C, dark gray and N, light gray. Double bonds are drawn in dark gray. Molecule **I** (top). Molecule **II** (bottom).

Molecule I	Molecule II	AZT
O2A–O44’A···6.712(6)	O2B–O44’B···6.535(6)	O2–N3’γ···7.373(3)
N3A–O44’A···7.697(7)	N3B–O44’B···7.654(6)	N3–N3’γ···8.836(4)
O4A–O44’A···9.213(6)	O4B–O44’B···9.227(6)	O4–N3’γ···10.625(4)

Table 11 Selected intramolecular distances (Å) for (1) and AZT[131].

Considering the results in Table 11, and the fact that the interatomic distances (Fig. 51) from the terminal azido nitrogen to O44' in (1) are only 1.30(1) Å (**I**) and 1.16(1) Å (**II**), it appears that (1) exhibits enough structural similarity with AZT to possess a similar capacity for binding within the active site of enzymatic phosphorylation. Although a spatial correspondence between (1) and AZT exists, the N-glycosidic torsion angle χ differs by approximately 70° and 79° when comparing AZT with **I** and **II** respectively. To ascertain the importance of the differences in χ, the energy required to rotate the base 360° around the N-glycosyl bond (for both (1) and AZT) was calculated using PCMODEL[149] with default parameters. The base was rotated around χ through 360° in 10° steps, with minimization of the remainder of the molecule at each step. The barrier to rotation was found to be less than 1 kcal/mol higher for (1) than for AZT. Furthermore, the energy/χ-angle profile is virtually identical in both systems. As this small difference in rotational energy would indicate, the relative values in the crystal of the N-glycosidic torsion angle, χ, in AZT and (1) are not important features in this comparison.

The completion of the enzymatic phosphorylation process requires that, at least, the following two conditions be met. Firstly, recognition of and binding to the active site must occur. Secondly, it is critical that once bound, the nucleoside analog presents the oxygen atom to be phosphorylated to the appropriate enzymatic site with positional specificity. Phosphorylation will not likely occur in any nucleoside analog which has its phosphorylation site oriented in too different a manner from that in AZT. Fig. 51 clearly shows that the sites of phosphorylation in AZT (O5') and in the title compound (O55') are oriented quite differently in the superimposed structures. These oxygens, *i. e.* O5' and O55', can only be brought into proximity with each other by having the pyrenose ring undergo the unlikely ring-flip to an alternative conformation with O5'A (or B) *exo* and C5'A (or B) *endo*, which would result in severe steric clashes between the thymine and the C6' hydrogens. The distances between O5' in AZT and O55' in (1) are 2.336(9) Å (**I**) and 4.02(1) A (**II**) in the respective comparisons.

On the basis of the structural comparisons of (1) with AZT, the inference is that the title compound would not effectively inhibit the HIV virus. As corroborating evidence for this conclusion, it is interesting to note, that while this work was in progress, the study cited in Ref. 128 appeared in the literature indicating that the title compound possesses little activity against the HIV-1 virus *in vitro*.

4.4 The Evaluation of 2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranosyl Nucleosides as Potential Antisense Constructs: Synthesis, Biophysical Properties and Enzymatic Stability of 2'-deoxyadenosine-(3'-6')-[1-(2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranosyl)thymine] phosphate

In recent years, the use of antisense constructs as a potential therapeutic tool has received much attention. Antisense oligodeoxynucleotides (ODNs) are short, single stranded oligonucleotide analogs that are able to block the expression of specific genes within cells. By duplexing with specific messenger ribonucleic acid (mRNA) targets, ODNs inhibit the translation of the mRNA into protein at two levels[150,151] (Fig. 52). A major problem with the antisense approach, however, is the rapid enzymatic degradation of natural oligodeoxynucleotides. Therefore, enzymatically stable oligonucleotide analogs that form stable duplexes with their natural complements must be developed. While extensive studies have been carried out on oligonucleotides modified in their bases and phosphate backbones, with phosphorothioate[151] and C-5 propynyl-pyrimidine[152,153] oligonucleotide derivatives showing relatively potent effects, sugar modifications have only received a small amount of attention. However, recent studies on 2'-O'-allyl oligoribonucleotides[154], carbo-oligodeoxynucleotides[155], α oligonucleotides[156], acyclic oligonucleotides[157] and hexopyranose oligodeoxynucleotides[158-160], indicate that sugar modified oligonucleotides may also be effective antisense agents. Theoretical studies[161], and physico-chemical studies[162] on hexopyranose nucleic acids by Eschenmoser have indicated that steric problems preclude the formation of Watson-Crick double helices between single stranded hexose oligonucleotides. However, several recent studies by Augustyns *et al.*[158-160], indicate that double-helical duplex formation between single-stranded hexose/furanose oligonucleotide hybrids and their natural complements can occur, particularly when the hexopyranose sugars are incorporated as the 3'- or 5'- terminal nucleotides of the hybrid chains. While the modified hexose nucleosides studied conferred greater resistance against enzymatic degradation when incorporated at various positions in natural oligonucleotide strands, they suffered from a reduced duplex-forming capacity with the natural complement. In addition to possible steric problems, a second deterrent to duplex formation is the decrease in entropy (due to the loss of conformational freedom) when an oligonucleotide containing flexible hexose sugars binds to its natural complement. Therefore, assuming the presence of identical intermolecular forces in duplex formation, complexation between single strands containing more rigid sugars is more energetically favorable. To minimize the effects of this entropic factor, we have, as a first step, synthesized the DNA dimer 2'-deoxyadenosine-(3'-6')-[1-(2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranosyl) thymine] phosphate (ApT*, compound (4)), which contains the conformationally rigid unsaturated sugar 2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranose, as an antisense construct with potentially increased hybridization capabilities. A 3'-end protected molecule has been selected in this study, since (3'-5')-exonuclease activity is the major cause of the degradation of natural oligonucleotides in serum[150]. Here the synthesis of ApT* is presented, as well as the results of CD studies on the compound, and enzyme degradation results.

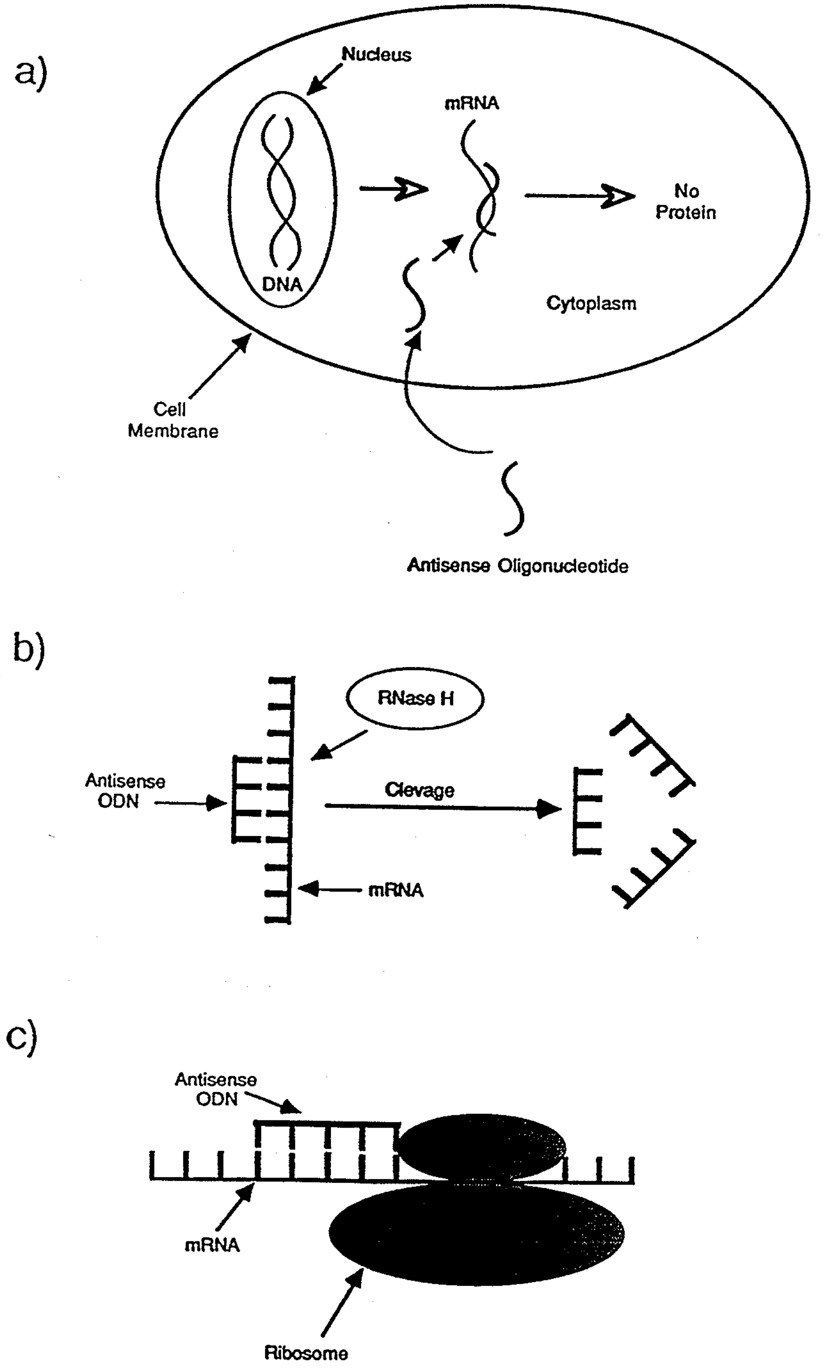


Fig. 52 a) Principle of action of antisense oligonucleotides. Inhibition proceeds *via* two mechanisms, shown in **b)** and **c)**. The mRNA strand of a DNA:RNA hybrid is a substrate for the cellular enzyme RNase H, which selectively hydrolyzes the mRNA phosphodiester backbone (as shown in **b)**). A second mechanism of inhibition is shown in **c)**. The formation of a DNA:RNA hybrid blocks the translation of the mRNA strand.

4.5 Experimental Section

^{1}H NMR spectra were obtained at 300.1 MHz (Bruker AM300). Signals are quoted as δppm downfield from internal tetramethylsilane unless otherwise stated.

Negative ion FAB mass spectra (glycerol matrix, 8-keV xenon atoms) were obtained on a VG 7070E-HF spectrometer.

CD spectra were obtained on a Jasco J-500A spectropolarimeter using a 1.0 mm cell equipped for temperature control with a water bath. ApT* was studied at 4.41 x 10^{-4} M (calculated from UV absorbance measurements) in 0.01 M sodium phosphate buffer (pH 6.8), and 0.1 M NaCl to maintain a constant ionic strength. The CD spectrum of ApT[163], was obtained in the same buffer (with 0.001 M EDTA) at approximately the same concentration.

Final purification of ApT* and analysis of enzyme degradation products were carried by reverse phase HPLC on an analytical C-18 (Phenomenex) column using a Varian 9010 HPLC solvent delivery system equipped with a Varian 9050 UV/vis detector. The flow rate was 1 ml/min. For purification of ApT*, the solvent system used was: A (0.1 M NH_4OAc) and B (MeOH). A linear gradient was used, where A was varied from 100% to 40% over 40 minutes. For the enzyme degradation studies, isocratic elution was performed using a ratio of C (5% MeOH in 0.1 M triethyl-ammonium acetate (pH 7.0)) and D (50% MeOH in 0.1 M triethyl-ammonium acetate (pH 7.0)) of 2:1. The hydrolysis of the dimers was followed by integration of the dimer signal. The decrease in the dimer concentration with time was fitted to a simple exponential function.

Nuclease S1 1 OD^{260nm} of dimer in 270 mL of H_2O was mixed with 30 ml of the following buffer: 0.33 M sodium acetate (pH 4.8), 0.5 M sodium chloride, 0.01 M zinc chloride. Nuclease S1 (4.6 mL, 1068 U) (Sigma) was then added. The mixture was incubated at 37° C. 20 ml aliquots were removed at set times and mixed with 180 mL of the following blocking buffer: 10 mM sodium phosphate (pH 6.8). This mixture was heat inactivated at 95° C for 5 min. and cooled. 10 mL aliquots were analyzed by HPLC.

Snake Venom Phosphodiesterase 1 OD^{260nm} of dimer in 200 mL of H_2O was mixed with 100 mL of the following buffer: 0.3 M Tris.HCl (pH 8.8), 0.3 M sodium chloride, 0.040 mM magnesium chloride. 3 mL (0.1 U) of a solution of snake venom phosphodiesterase (Sigma) in 5 mM Tris.HCl (pH 7.5), 50% glycerol (v/v) was then added. The mixture was incubated at 37° C. Inactivation and analysis was carried out as described for nuclease S1.

Bovine Spleen Phosphodiesterase 1 OD^{260nm} of dimer in 200 mL of H_2O was mixed with 100 ml of the following buffer: 300 mM sodium citrate (pH 6.0). 3 mL (0.1 U) of a solution of bovine spleen phosphodiesterase (Sigma) in 50 mM sodium citrate (pH 6.0), 50% glycerol (v/v) was then added. The mixture was incubated at 37° C. Inactivation and analysis was carried out as described for nuclease S1.

4.51 Synthetic Methods

1-(2',3'-dideoxy-4'-O-acetyl-β-D-*erythro*-hex-2'-enopyranosyl) thymine (2)

254 mg (1.0 mmol) of 1-(2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranosyl) thymine (1) (Fig. 53), was dried by adding and evaporating dry pyridine (3

times) and then dissolved in 5 mL of 50/50 (v/v) anhydrous methylene chloride/pyridine. 373 mg (1.1 mmol) of dimethoxytrityl chloride and 12 mg (0.1 mmol) of 4-(dimethylamino)pyridine were then added to the solution, which was sealed and stirred for 10 hrs. At this point the reaction was complete (as determined by TLC, ether), and 0.19 mL (2.0 mmol) of acetic anhydride were added to the mixture, which was sealed again and stirred for 10 hrs. The reaction was then quenched with 1.0 mL of water, which was added dropwise, while the reaction mixture was immersed in an ice-bath. The reaction mixture was then washed twice with water, and concentrated to an yellowish solid on a rotary evaporator. The solid was dissolved in 3 mL of 80/20 (v/v) acetic acid/H_2O and stirred for 6 hrs. The reaction mixture was then concentrated on a rotary evaporator and ethyl acetate evaporated four times to remove all traces of acetic acid. Finally the product was concentrated to a syrup in ethyl acetate, and purified on silica gel (Aldrich, 60 grade 230-400 mesh) using ether as an eluent by dry column flash chromatography, yielding 170 mg (0.57 mmol, 57% yield) of (2) as a white powder.

^1H NMR ($CDCl_3$) δ_H 8.10 (1 H, br s, NH), 6.99 (1 H, s, 6-H), 6.53 (1 H, s, 1'-H), 6.20 (1 H, m, $J_{2',3'}$ 12.3 Hz, 2'-H), 5.76 (1 H, m, $J_{2',3'}$ 12.3 Hz, 3'-H), 5.48 (1 H, m, 4'-H), 3.82 (2 H, m, 6'-H), 3.70 (1 H, m, 5'-H), 2.13 (3 H, s, OAc), 1.93 (3 H, s, Me).

2'-deoxyadenosine-(3'-6')-[1-(2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranosyl)thymine] phosphate (4)

430 mg (0.65 mmol) of 5'-O-4,4'-dimethoxytrityl-6-N-benzoyl-2'-deoxyadenosine (Cruachem) and 150 mg (.70 mmol) of *p*-chlorophenyl phosphate (Fig. 53) were dried by adding and evaporating dry pyridine (4

times). The resulting foam was dissolved in 5 mL dry pyridine, 610 mg (2.0 mmol) of 2,4,6-triisopropylbenzenesulfonyl chloride (TPSCl) were added, and the mixture sealed and stirred for 4 hrs. At this point, the phosphorylation was complete (determined by TLC, ether, followed by 12:1 CH_2Cl_2: MeOH), and 150 mg (0.5 mmol) of (2) (dried by evaporation of pyridine), an additional 370 mg (0.4 mmol) of TPSCl, and 250 mg (1.5 mmol) of 1H-tetrazole (dried by evaporation of dry acetonitrile) were added, and the mixture sealed and stirred for 24 hrs (monitored by TLC, as above). 3 mL of H_2O were then added dropwise to the reaction mixture, which was placed in an ice-bath. The product was then extracted into methylene chloride, evaporated to a thick syrup, and purified on silica gel (same grade as used above), using dry column flash chromatography. A hexane/ethyl acetate solvent mixture was used as the eluent, changing the hexane content linearly from 100% to 0% through the course of the purification. Fractions were pooled, and solvent was removed, yielding 305 mg of semi-pure 5'-O-4,4'-dimethoxytrityl-6-N-benzoyl-2'-deoxyadenylyl-(3'-6')-[1-(2',3'-dideoxy-4'-acetyl-β-D-*erythro*-hex-2'-enopyranosyl)thymine]*p*-chlorophenyl phosphate (3) (~ 50% yield) as an orange solid.

Detritylation of (3) was carried out by adding 300 mg (0.26 mmol) to a 5 mL solution of 1.0 M zinc bromide in 85/15 (v/v) methylene chloride/isopropanol and stirring for 20 min. Solvent was then removed and the detritylated product dissolved in methylene chloride and washed with a minimum amount of H_2O. The solvent was then removed and the solid washed twice with ether to remove all traces of trityl alcohol. The phosphate group was then deprotected by adding 200 mg (0.2 mmol) of detritylated dimer to a solution of 330 mg (2.0 mmol) o-nitrobenzaldoxime and 0.22 mL (1.76 mmol) 1,1,3,3,N,N, N',N'-tetramethyl guanidinium in 3.0 ml 50/50 (v/v) dry dioxane/dry acetonitrile and stirred for 48 hrs. Solvent was removed, and the product washed 5 times with ether, and dissolved in 10 mL of conc. NH_4OH. The ammonia solution was sealed in a pyrex tube, and heated at 60° C for 8 hrs. Finally, the ammonia was removed by evaporation and the solid residue dissolved in H_2O. Purification of the crude product is described above (see general procedures). The purified product was de-salted on a C18 Sep-Pak (Sigma, using the same procedure as described in Chapter II), yielding 43 mg (17 %) of (4) as a free-acid.

UV (in CD buffer) λ_{max} 263, $\varepsilon = 20400$ l mol^{-1} cm^{-1}

MS (negative ion FAB, xenon atoms) *m/z* 566 (100, $[M - H]^-$), 333 (26, $[6'\text{-}dT^*MP - H]^-$), 330 (24, $[3'\text{-dAMP} - H]^-$), 134 (107, $[\text{Adenine} - H]^-$), 125 (71, $[\text{thymine} - H]^-$)

^1H NMR (D_2O, using HOD-signal as standard (δ=4.74)) δ_H 8.29 (1 H, s, 8-H(A)), 8.26 (1 H, s, 2-H(A)), 7.32 (1 H, s, 6-H(T)), 6.44 (1 H, t, 1'-H(A)), 6.36 (1 H, s, 1'-H(T)), 6.23 (1 H, d, $J_{2',3'}$ 12.0 Hz, 2'-H(T)), 5.69 (1 H, d, $J_{2',3'}$ 12.0 Hz, 3'-H(T)), 4.95 (1 H, m, 3'-H(A)), 4.54 (1 H, m, 4'-H(T)), 4.38 (1 H, m, 4'-H(A)), 4.21 (2 H, m, 6'-H(T)), 3.86 (1 H, m, 5'-H(T)), 3.79 (2 H, m, 5'-H(A)), 2.84-2.67 (2 H, m, 2'-H(A)), 1.64 (3 H, s, Me(T))

4.6 Results and Discussion

4.61 Synthesis of ApT*

Using solution phase phosphotriester chemistry (see Fig. 53) the dimer ApT* (4) was obtained by coupling 5'-O-4,4'-dimethoxytrityl-6-N-benzoyl-2'-deoxyadenosine-3'-*p*-chlorophenyl phosphate with 1-(2',3'-dideoxy-4'-O-acetyl-β-D-*erythro*-hex-2'-enopyranosyl) thymine (2). After deprotection with zinc bromide, oximate reagent and base, and purification by RP-HPLC, the final yield of fully deprotected dimer (4) was 17%. The coupling step was extremely slow, (24 hrs. to reach completion). This was probably due to the conformational rigidity of the unsaturated hexopyranose sugar, which may have hindered the formation of the proper transition state complex. Coupling was initially attempted using H-phosphonate chemistry, but was unsuccessful. H-phosphonate couplings were always dominated by a side reaction, where the activating agents tried, (pivaloyl chloride, TPS-Cl and TPS-tetrazole) reacted preferentially with O6' of (2).

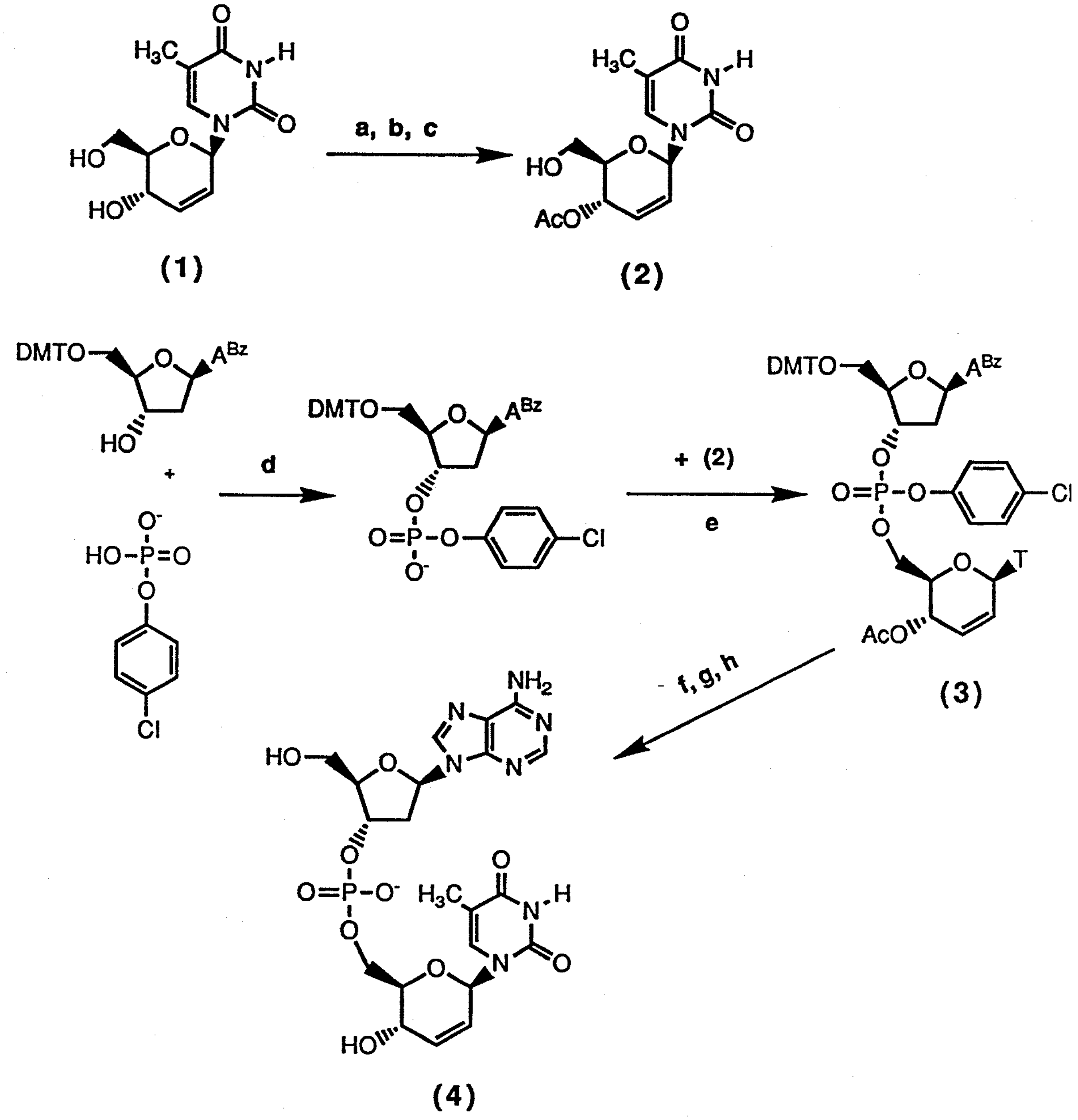


Fig. 53 **a)** 1.1 eq. DMT-Cl, 0.1 eq. DMAP. in pyridine **b)** 2.0 eq. Acetic anhydride. **c)** 80:20 acetic acid:H_2O. **d)** 3.3 eq. TPS-Cl in pyridine. **e)** TPS-tetrazolide (1 eq. TPS-Cl + 3 eq. 1H-tetrazole) in pyridine. **f)** $ZnBr_2$ (1.0 M in 85:15 CH_2Cl_2:Isopropanol). **g)** 10 eq. o-Nitrobenzaldoxime + 8.8 eq. TMG in 50:50 dioxane:acetonitrile. **h)** conc. NH_4OH

4.62 Circular Dichroism Studies

The CD spectra of ApT* at 10° C and 70° C, and the CD spectrum of natural ApT at 10° C (from Ref. 163) are shown in Fig. 54. The virtually identical positioning of λ max., λ min. and the cross-over points for ApT and ApT*, indicate that substitution of a 2'-deoxyribose sugar with 2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranose at the 3'-terminus of ApT has a minimal influence on the global structural parameters. The positive and negative Cotton bands of the ApT* spectra are highly symmetric, and as the temperature is increased to 70°C, the magnitudes of the CD maxima decrease by approximately 60%. Overall, these features (and the signs of the Cotton bands) are consistent with the formation of stacked, B-form minihelices in solution[164,165].

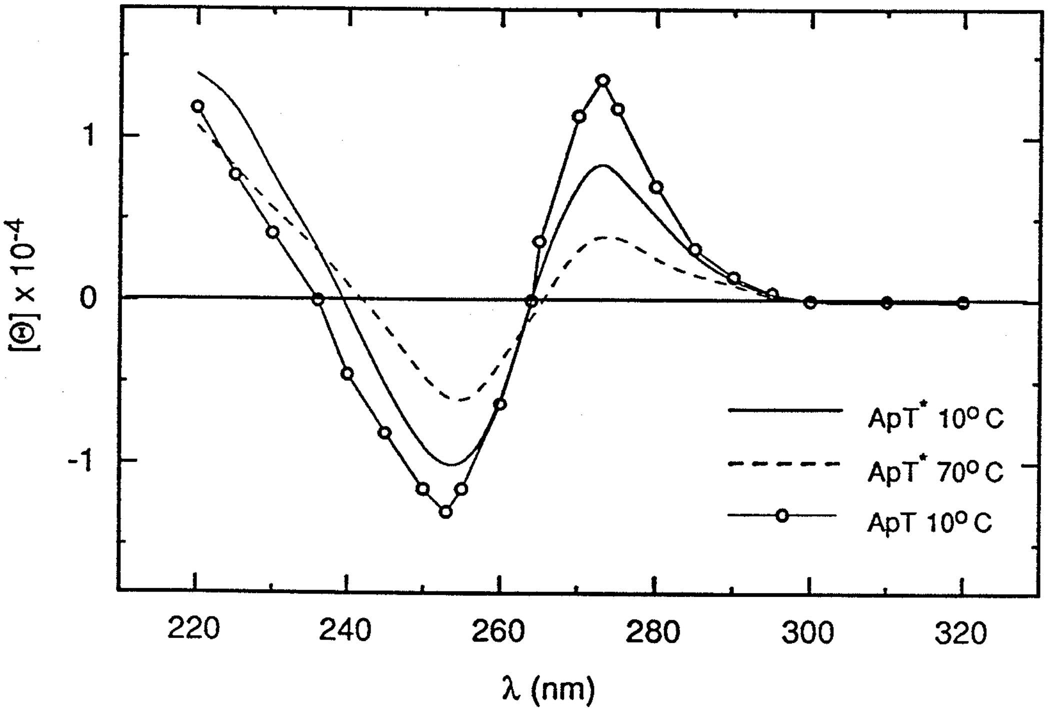


Figure 54 Circular Dichroism (CD) spectra of ApT and ApT* at 10° C, and ApT* at 70° C. The ApT spectrum is from Ref. 163. The vertical axis gives [Θ] in deg M^{-1} cm^{-1}, the molar ellipticity per residue.

4.63 Enzymatic Hydrolysis of ApT* and ApT

As a means to assess the enzymatic stability of ApT*, relative to its natural analog ApT, the following nucleases were used: nuclease S1 (endonuclease), snake venom phosphodiesterase (SV PDE, (3'-5')-exonuclease) and bovine spleen phosphodiesterase (BS PDE, (5'-3')-exonuclease). RP-HPLC was used to monitor the degradation of the dimers, which follows first order kinetics. With each enzyme (Table 12), ApT* showed an increased resistance to degradation relative to ApT.

ApT* was degraded more slowly than ApT by factors of 4, 3 and 2 in nuclease S1, SV PDE and BS PDE, respectively. These results suggest that a 2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranose moiety at the 3'-end of an oligonucleotide can provide significant protection against (3'-5')-exonuclease degradation. However, compared to results obtained from TpT# dimers, where T# refers to a thymine nucleoside containing the fully saturated 2',3'-dideoxy-β-D-*erythro*-hexopyranose sugar[158], the unsaturated sugar moiety seems considerably less effective. Since the spatial distribution of important functional groups on nucleosides containing both the saturated and unsaturated hexopyranose moieties is almost identical (Fig. 55), structural differences are probably not the cause of the increased rates of degradation observed in ApT*. Also, the equal abundance of the 6'- and 3'-phosphate fragment ions in the negative ion FAB mass spectrum (see Section 4.5) indicate that the stability of the 3'-6'-phosphodiester linkage is not compromised by unsaturation of the hexopyranose sugar, ruling it out as a contributing factor. Thus, the more rapid degradation of the title compound is probably the result of sequence and/or entropic effects. Experimental evidence suggests that sequence is important, since under similar experimental conditions, natural ApT is digested almost five times more rapidly than TpT (from Ref. 158). Additionally, entropic factors (see Section 4.4) dictate that the rigidity of the T* relative to T$^\#$ may favor the binding of oligonucleotides containing the rigid sugar to nucleases.

Table 12

	$t_{1/2}$ (min.)		
	nuclease S1	SV PDE	BS PDE
ApT	10.6	2.1	1.6
ApT*	42.1	6.2	3.1

4.7 Summary

The results demonstrate that substitution of a 2'-deoxyribose sugar at the 3'-end of the dinucleotide ApT with 2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranose provides a substantial increase in (3'-5')-exonuclease resistance. ApT* maintains a global structural isomorphism with ApT, suggesting that the hybridization properties of oligonucleotides containing the unsaturated hexose sugar at the 3'-end should not be compromised. Because of favorable entropic factors, the conformational rigidity of the sugar used in this study may, actually, increase the hybridization capabilities of oligonucleotides containing it in positions other than at the 5'- or 3'-ends, relative to oligonucleotides containing more flexible saturated hexopyranose sugars. In future studies, attempts should be made to synthesize longer hybrid oligonucleotide sequences incorporating 2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranose nucleotides at the ends and the interior of the strands, to assess their structural and hybridization properties, as well as their enzymatic stabilities.

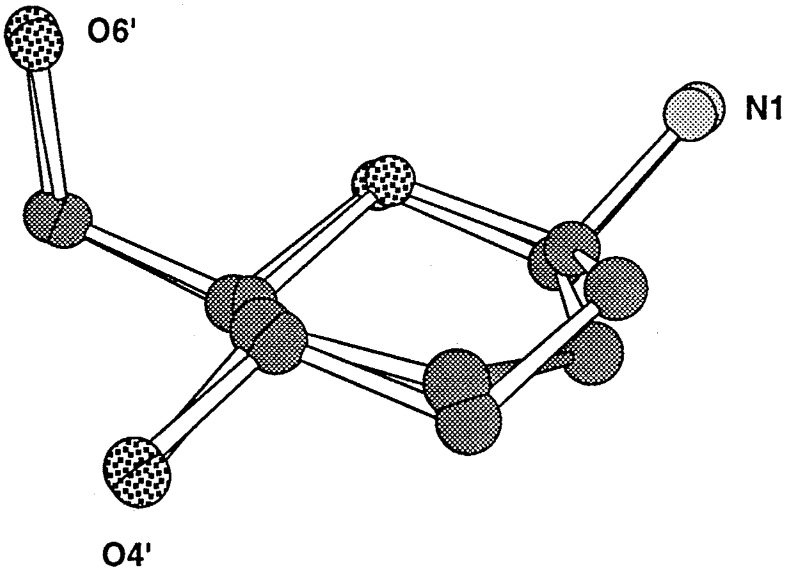
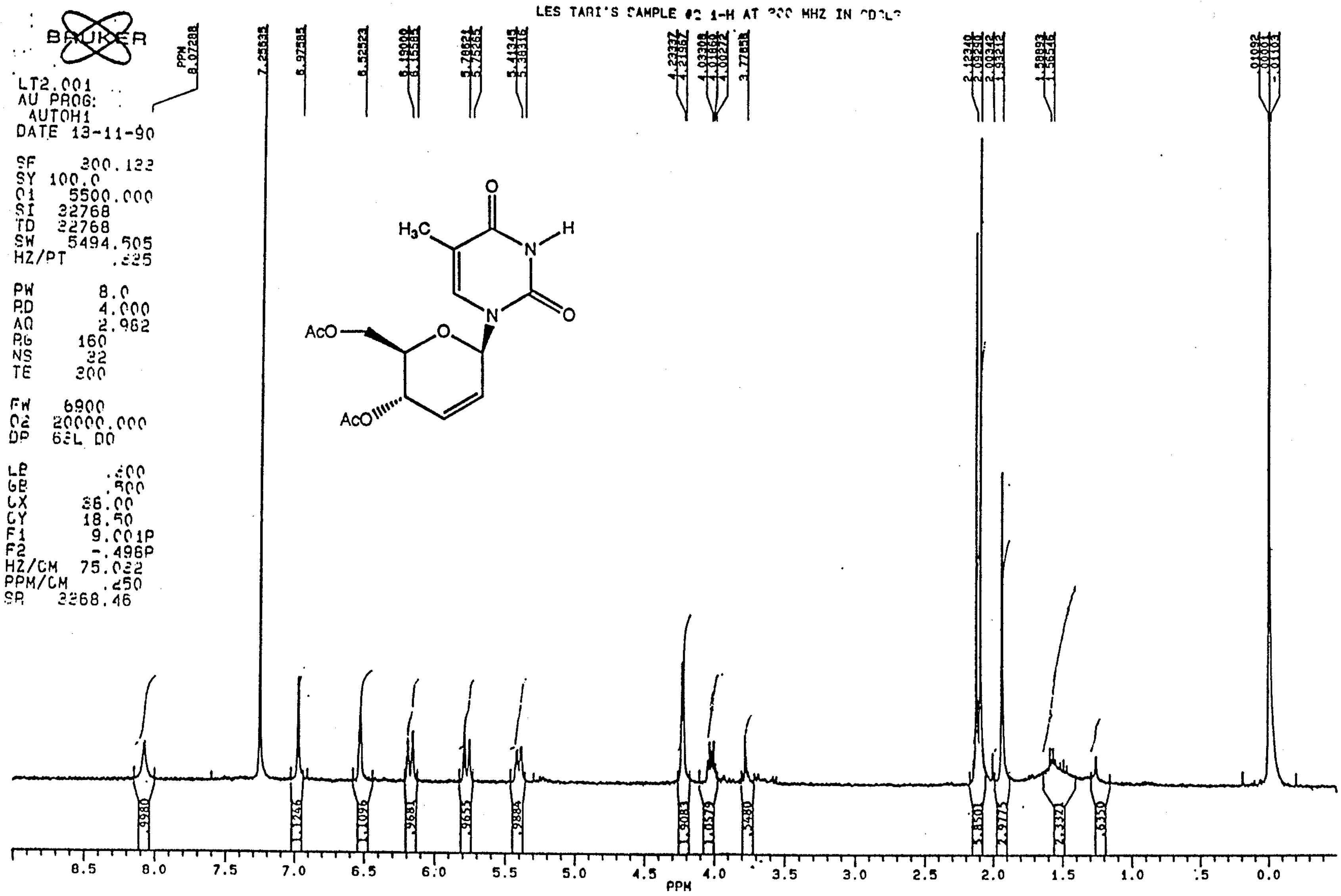
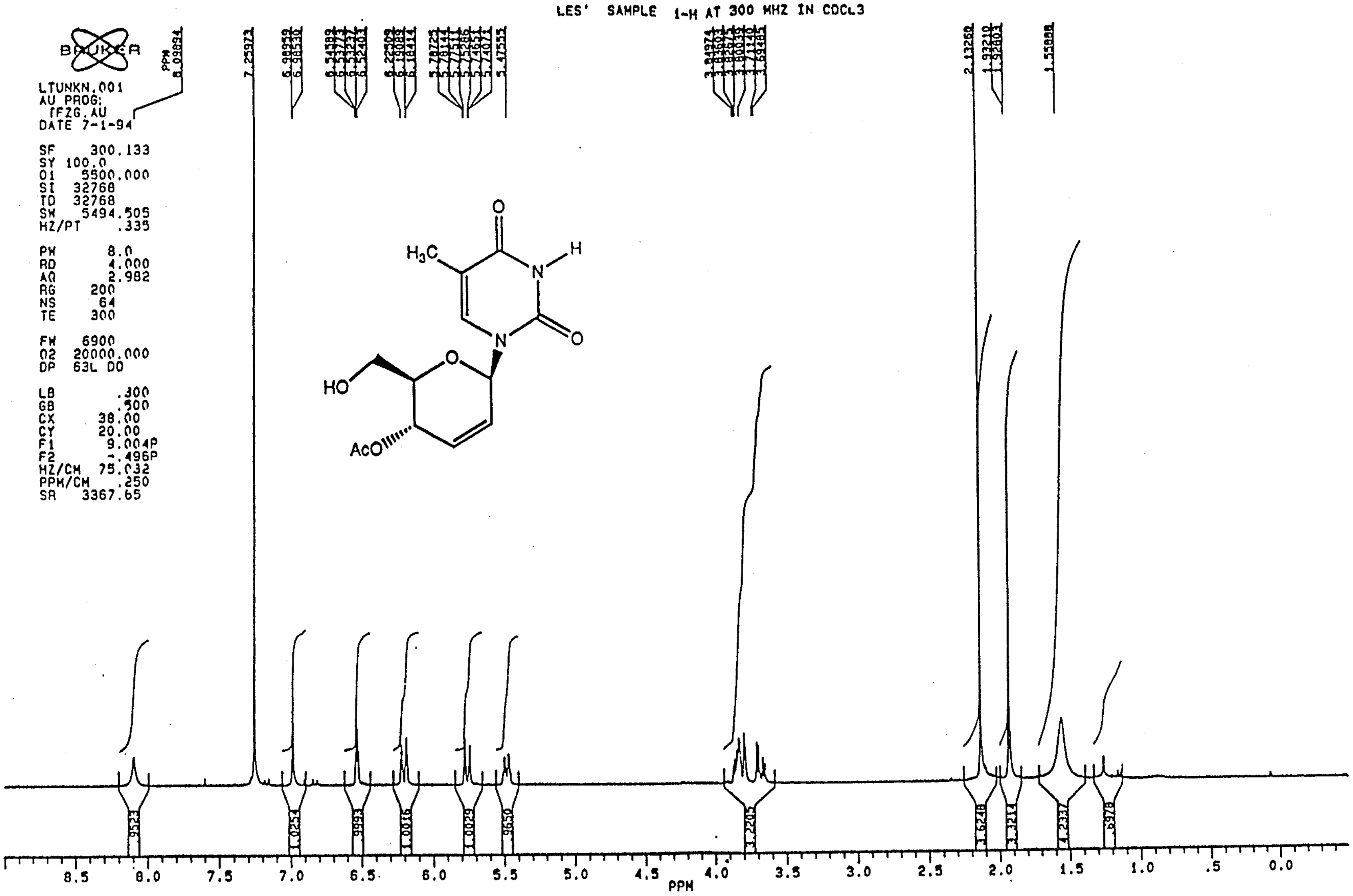


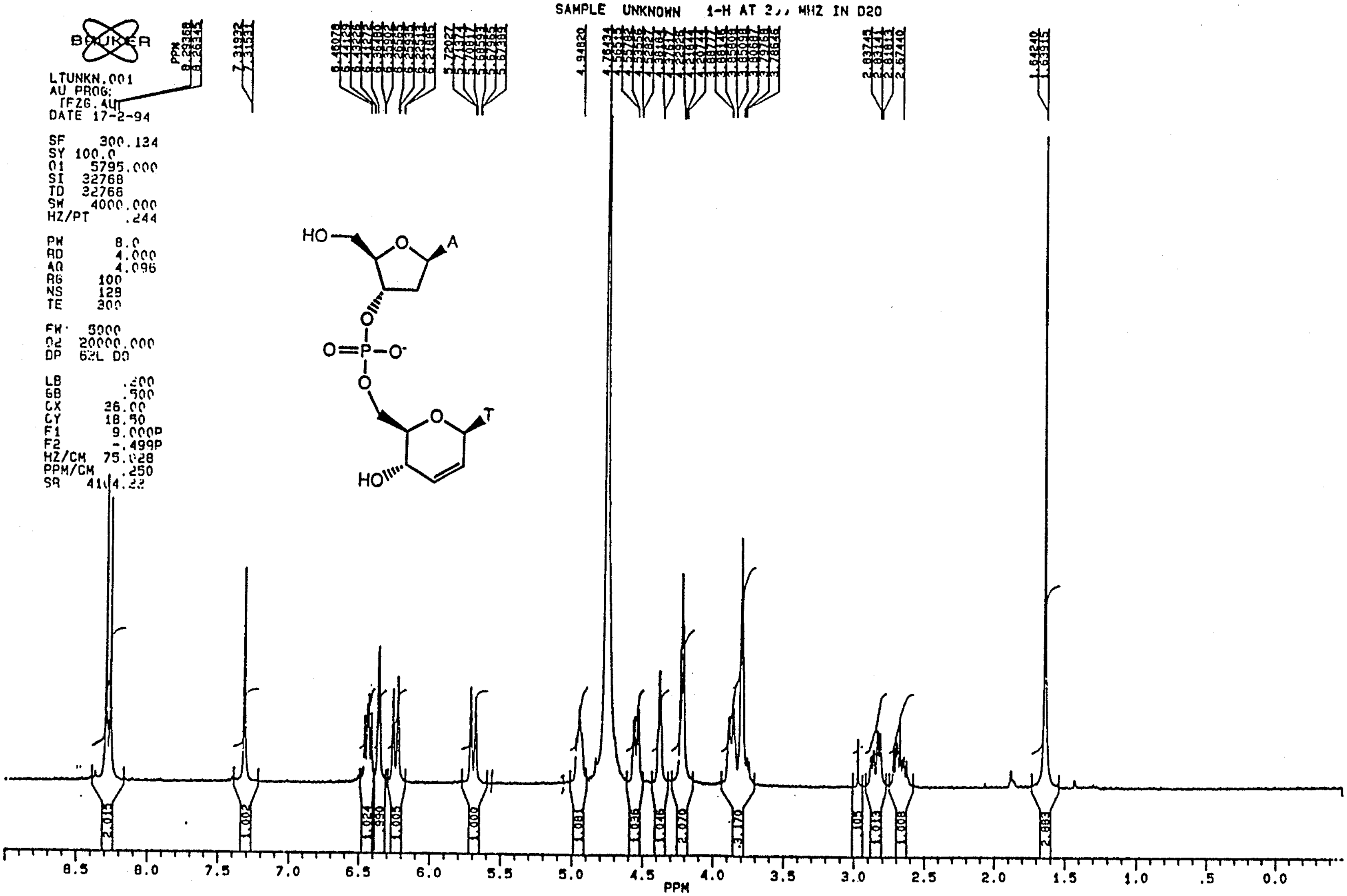
Figure 55 N1, O4' and O6' of 1-(2',3'-dideoxy-β-D-*erythro*-hex-2'-enopyranosyl) thymine (T^*, compound (1), double bond is darkened) superimposed on the same atoms from 1-(2',3'-dideoxy-β-D-*erythro*-hexopyranosyl) thymine ($T^\#$). Except for N1, the remaining base atoms from each nucleoside are not shown. Coordinates for both sugars were obtained from crystal structures. (Coordinates for T^* are from Ref. 166, while those for $T^\#$ are from Ref. 143). The average r.m.s. difference in atomic coordinates between models is 0.09 Å.

Appendix A

NMR Spectra

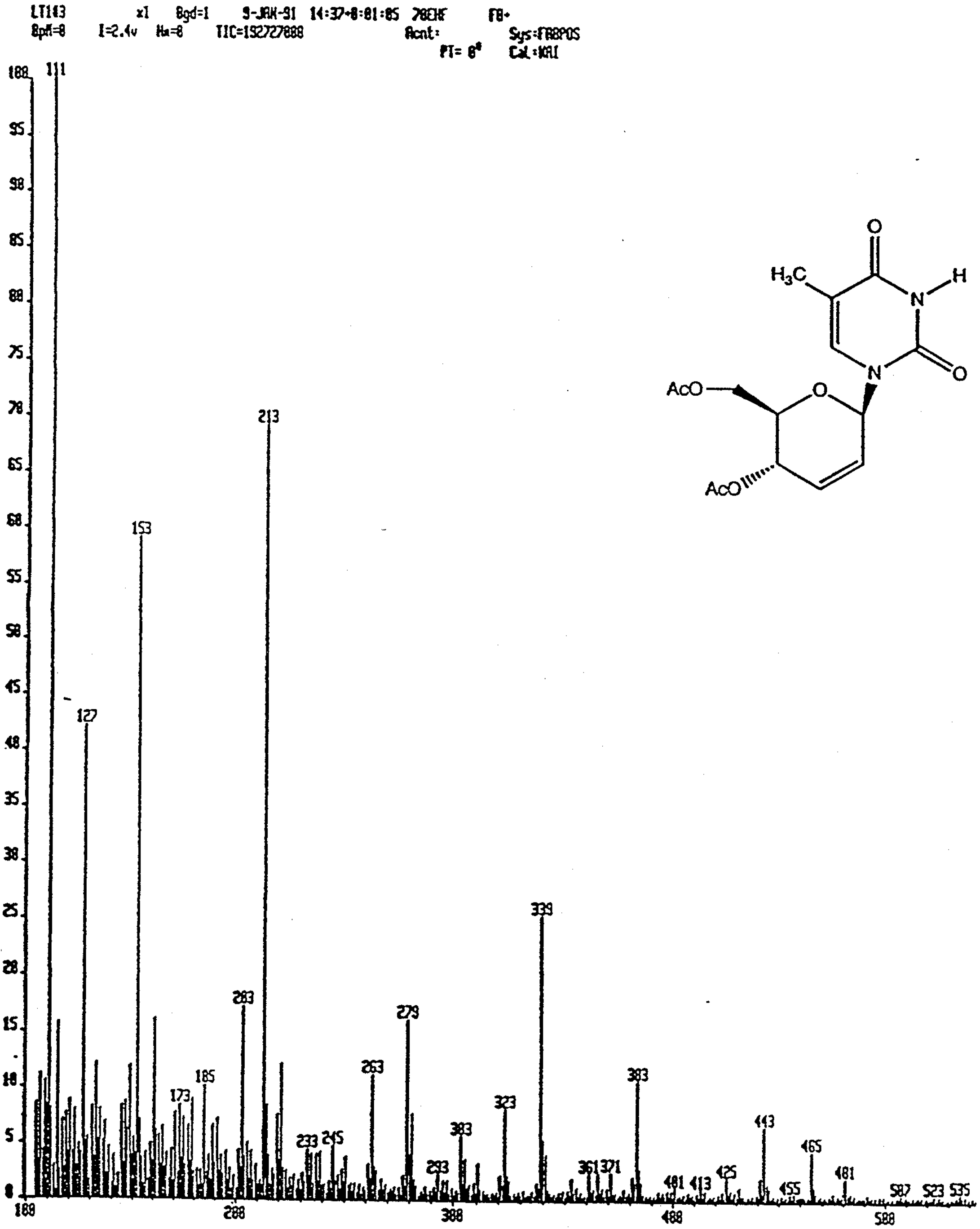

LES TARI'S SAMPLE #2 1-H AT 200 MHZ IN CDCL3
BRUKER
LT2.001
AU PROG:
AUTOH1
DATE 13-11-90
SF 200.122
SY 100.0
O1 5500.000
SI 32768
TD 32768
SW 5494.505
HZ/PT .335
PW 8.0
RD 4.000
AQ 2.982
RG 160
NS 32
TE 300
FW 6900
O2 20000.000
DP 63L DO
LB .200
GB .500
CX 36.00
CY 18.50
F1 9.001P
F2 -.496P
HZ/CM 75.032
PPM/CM .250
SR 2268.46
PPM
8.07288
7.25635
6.97585
6.52523
6.19000
6.15585
5.78821
5.75265
5.41345
5.38316
4.23337
4.21967
4.03308
4.01869
4.00272
3.77858
2.12340
2.09290
2.00342
1.93212
1.58893
1.56546
.01092
.00001
-.01103
O
H₃C
N
H
O
N
O
AcO
O
AcO
.9980
1.1246
1.1096
.9581
.9635
.9884
1.9083
1.0579
.5480
5.8501
2.9775
2.3327
.6350
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
.5
0.0
PPM


LES' SAMPLE 1-H AT 300 MHZ IN CDCL3
BRUKER
LTUNKN.001
AU PROG:
FZG.AU
DATE 7-1-94
SF 300.133
SY 100.0
O1 5500.000
SI 32768
TD 32768
SW 5494.505
HZ/PT .335
PW 8.0
RD 4.000
AQ 2.982
RG 200
NS 64
TE 300
FW 6900
O2 20000.000
DP 63L DO
LB .300
GB .500
CX 38.00
CY 20.00
F1 9.004P
F2 -.496P
HZ/CM 75.032
PPM/CM .250
SR 3367.65
PPM
8.09894
7.25973
6.98959
6.98530
6.54589
6.53773
6.53237
6.52403
6.22309
6.19085
6.18414
5.78725
5.78144
5.77511
5.75285
5.74651
5.74071
5.47555
3.84974
3.83603
3.82675
3.80039
3.71140
3.69485
2.13260
1.93210
1.92803
1.55888
H3C
O
N
H
O
N
HO
O
AcO
.9523
1.0252
.9993
1.0016
1.0029
.9650
3.2205
3.5248
3.3214
4.2337
.6978
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
.5
0.0
PPM

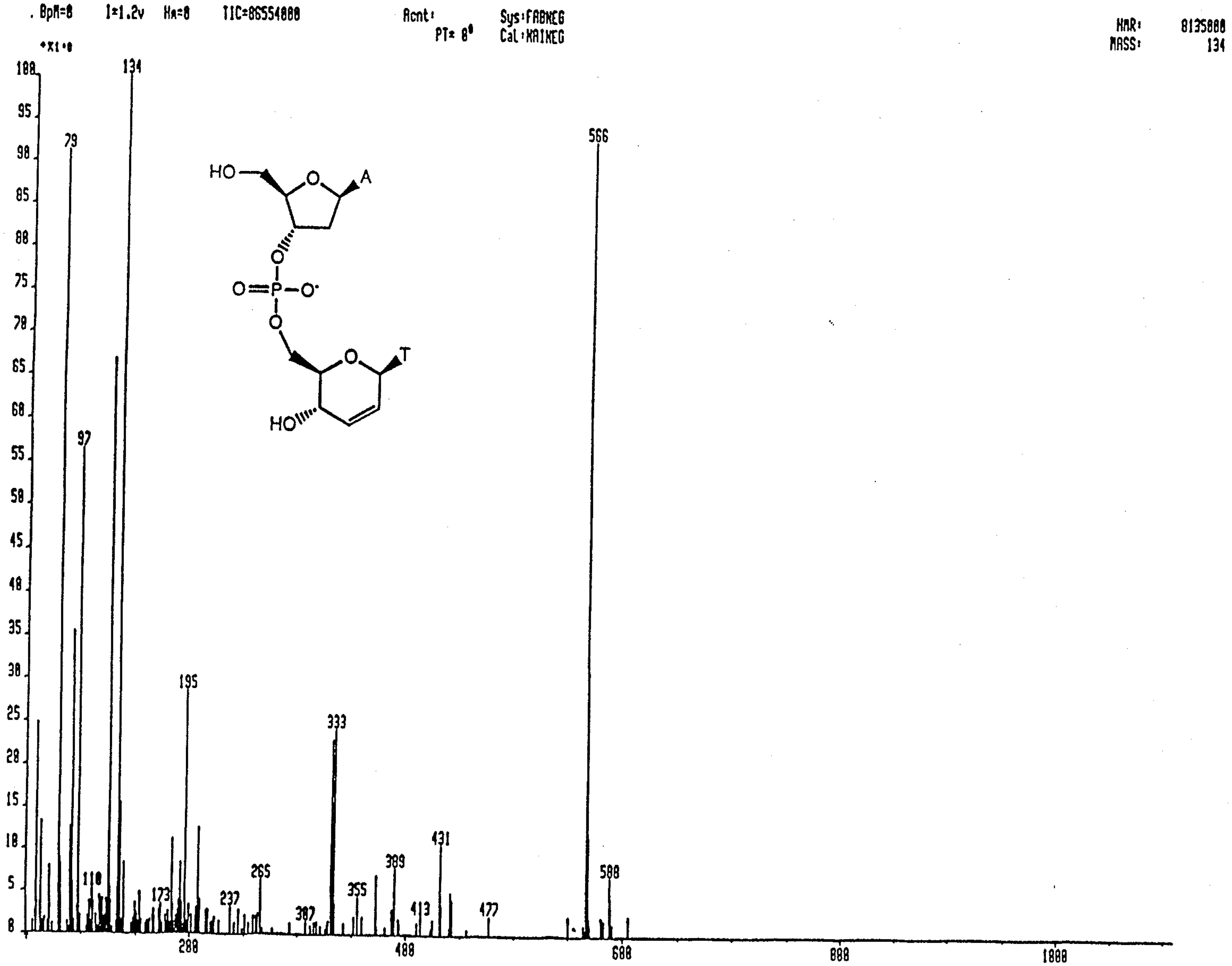

SAMPLE UNKNOWN 1-H AT 3,, MHZ IN D2O
BRUKER
LTUNKN.001
AU PROG:
TFZG.AU
DATE 17-2-94
SF 300.124
SY 100.0
O1 5795.000
SI 32768
TD 32766
SW 4000.000
HZ/PT .244
PW 8.0
RD 4.000
AQ 4.096
RG 100
NS 128
TE 300
FW 5000
O2 20000.000
DP 62L DO
LB .200
GB .500
CX 26.00
CY 18.50
F1 9.000P
F2 -.499P
HZ/CM 75.028
PPM/CM .250
SR 4104.22
PPM
8.29368
8.26345
7.31932
7.31531
6.46078
6.44129
6.43226
6.41277
6.36480
6.35902
6.26563
6.25935
6.22513
6.21885
5.72027
5.71374
5.70817
5.68593
5.67965
5.67389
4.94820
4.78434
4.56515
4.55784
4.53556
4.52827
4.38184
4.37617
4.22928
4.21844
4.20744
3.88777
3.88146
3.85808
3.85098
3.80687
3.79768
3.78646
2.83745
2.83141
2.81813
2.67440
1.64240
1.63915
HO
O
A
O
O=P-O⁻
O
O
T
HO
2.015
1.002
1.024
.990
1.005
1.000
1.081
1.036
1.046
2.070
3.170
.105
1.013
1.008
2.883
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
.5
0.0
PPM

Appendix B

Mass Spectra


LT143 x1 Bgd=1 9-JAN-91 14:37+0:01:05 70EHF FB+
BpM=0 I=2.4v Ha=0 TIC=152727000 Acnt: Sys:FABPOS
PT= 0° Cal:KAI
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
111
127
153
173
185
203
213
233
245
263
279
293
303
323
339
361
371
383
401
413
425
443
455
465
481
507
523
535
100
200
300
400
500
H_3C
O
N
H
N
O
AcO
O
AcO


LT10GLY#165 x1 Bgd=159 2-APR-91 16:40+0:12:57 70EHF FB+
BpM=0 I=3.0v Ha=0 TIC=150300000 Acnt: Sys:FABPOS
PT= 0° Cal:NAI
HMR: 19540000
MASS: 127
*X1.0
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
127
129
185
111
255
159
171
213
239
277
301
317
329
347
369
381
413
425
439
471
485
509
531
553
583
601
634
100
200
300
400
500
600
700
O
H3C
N
H
N
O
HO
O
HO


. BpM=0 I=1.2v Ha=0 TIC=86554000
*X1.0
Acnt:
PT= 0°
Sys:FABNEG
Cal:KAINEG
HMR: 8135000
MASS: 134
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
134
79
566
97
195
333
265
389
431
588
110
173
237
307
355
413
477
200
400
600
800
1000
HO
O
A
O
O
P
O⁻
O
O
T
HO

REFERENCES

1. Watson, J. D. & Crick, F. H. C. *Nature* **171**, 737-738, (1953).

2. Wilson, D. B. & Hogness, D. S. *J. Biol. Chem.* **244**, 2413-2419, (1968).

3. Tokeson, J. P. E., Garges, S. & Adhya, S. *J. Bacteriol.* **173**, 2319-2327, (1991).

4. Weickert, M. J. & Adhya, S. *Mol. Microbiol.* **10**, 245-251, (1993).

5. Makman, R. S. & Sutherland, E. Q. *J. Biol. Chem.* **240**, 1309-1315, (1965).

6. Musso, R. E., DiLaurio, R., Adhya, S. & de Crombrugghe, B. *Cell* **12**, 847-854, (1977)

7. Altona, C. & Sundaralingam, M. *J. Amer. Chem. Soc.* **94**, 8205-8212, (1972).

8. Levitt, M. & Warshel, A. *J. Amer. Chem. Soc.* **100**, 2607-2613, (1978).

9. Viswamitra, M. A., Reddy, B. S., Lin, G. H-Y & Sundaralingam, M. *J. Amer. Chem. Soc.* **93**, 4565-4572, (1971).

10. Tran-Dinh Son, Guschlbauer, W. & Guéron, M. *J. Amer. Chem. Soc.* **94**, 7903-7911, (1972).

11. Guschlbauer, W., Fric, I. & Holy, A. *Eur. J. Biochem.* **31**, 1-13, (1972).

12. Wang, A. H-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. & Rich, A. *Nature* **282**, 680-686, (1979).

13. Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W. & Ratcliff, R. L. *Nature* **283**, 743-745, (1980).

14. Olson, W. K. *Biopolymers* **12**, 1787-1814, (1973).

15. Berthod, H. & Pullman, B. *FEBS Lett.* **33**, 147-150, (1973).

16. Sundaralingam, M. *Jerus. Symp. Quant. Chem. Biochem.* **5**, 417-456, (1973).

17. Yathindra, N. & Sundaralingam, M. *Biopolymers* **12**, 297-314, (1973).

18. Davies, D. B. *Prog. NMR Spectrosc.* **12**, 135-225, (1978).

19. Sundaralingam, M. *Biopolymers* **7**, 821-860, (1969).

20. Sundaralingam, M. *Ann. N. Y. Acad. Sci.* **255**, 3-42, (1976).

21. Saenger, W. *Angew. Chem.* **85**, 680-690, (1973).

22. Saran, A., Pullman, B. & Perahia, D. *Biochim. Biophys. Acta* **287**, 211-231, (1973).

23. Wolfe, S. *Acc. Chem. Res.* **5**, 102-111, (1972).

24. Sussman, J. L., Seeman, N. C., Kim, S.-H. & Berman, H. *J. Mol. Biol.* **66**, 403-421, (1972).

25. Rubin, J., Brennan, T. & Sundaralingam, M. *Biochemistry* **11**, 3112-3128, (1972).

26. Egert, E., Lindner, H.-J., Hillen, W. & Gassen, H. G. *Acta Crystallogr.* B**34**, 2204-2208, (1978).

27. Taylor, R. & Kennard, O. *J. Amer. Chem. Soc.* **104**, 5063-5070, (1982).

28. Lee, C.-H., Ezra, F. S., Kondo, N. S., Sarma, R. H. & Danyluk, S. S. *Biochemistry* **15**, 3627-3639, (1976).

29. Alderfer, J. L. & Ts'o, P. O. P. *Biochemistry* **16**, 2410-2416, (1977).

30. Sasisekharan, V. & Pattabiraman, N. *Nature* **275**, 159-162, (1978).

31. Perahia, D., Pullman, B. & Saran, A. *Biochim. Biophys. Acta* **340**, 299-313, (1974).

32. Tosi, C. & Lipari, G. *Theoret. Chim. Acta* **60**, 41-51, (1981).

33. Tewari, R., Nanda, R. K. & Govil, G. *J. Theor. Biol.* **46**, 229-239, (1974).

34. Govil, G. *Biopolymers* **15**, 2303-2307, (1976).

35. Taylor, R. & Kennard, O. *Acc. Chem. Res.* **17**, 320-326, (1984).

36. Kudritskaya, Z. G. & Danilov, V. I. *J. Theor. Biol.* **59**, 303-318, (1976).

37. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Scheider, B. *Biophys. J.* **63**, 751-759, (1992).

38. Waugh, D. F. *Adv. Protein Chem.* **9**, 326-437, (1954).

39. Ornstein, R. L., Rein, R., Breen, D. L. & MacElroy, R. D. *Biopolymers* **17**, 2341-2360, (1978).

40. Saenger, W. *Principles of Nucleic Acid Structure.*,Springer Verlag, New York, (1984).

41. Dickerson, R. E. *J. Biomol. Str. Dynam.* **6**, 627-634, (1989).

42. Blackburn, G. M. & Gait, M. J. In: *Nucleic Acids in Chemistry and Biology* (Blackburn, G. M. & Gait, M. J. Eds.), pp. 33, Oxford University Press, New York, (1990).

43. Arnott, S. & Huskins, D. *Biochem. Biophys. Res. Commun.* **47**, 1504-1509, (1972).

44. Matthews, B. W. *Nature (London)* **335**, 294-295, (1988).

45. Sarai, A. & Takeda, Y. *Proc. Natl. Acad. Sci. USA* **86**, 6513-6517, (1989).

46. Pabo, C. O. & Sauer, R. T. *Annu. Rev. Biochem.* **61**, 1053-1095, (1992).

47. Lu, P., Cheung, S. & Arndt, K. *J. Biomol. Struct. Dynam.* **1**, 509-521, (1983).

48. Donlan, M. E. & Lu, P. *Nucleic Acids Res.* **20**, 525-532, (1992).

49. Jordan, S. R. & Pabo, C. O. *Science* **242**, 893-907, (1988).

50. Brennan, R. G., Roderick, S. L., Takeda, Y. & Matthews, B. W. *Proc. Natl. Acad. Sci. USA* **87**, 8165-8169, (1990).

51. Schultz, S. C., Shields, G. C. & Steitz, T. A. *Science* **253**, 1001-1007, (1991).

52. Marrot, L. & Leng, M. *Biochemistry* **28**, 1454-1461, (1989).

53. Anin, M., and Leng, M. *Nucl. Acids Res.* **18**, 4395-4400, (1990).

54. Leroy, J.-L., Kochoyan, M., Huynh-Dinh, T. & Gueron, M. *J. Mol. Biol.* **200**, 223-238, (1988).

55. Rastinejad, F., Artz, P. & Lu, P. *J. Mol. Biol.* **233**, 389-399, (1993).

56. Cruse, W. B. T., Salisbury, S. A., Brown, T., Cosstick, R., Eckstein, F. & Kennard, O. *J. Mol. Biol.* **192**, 891-905, (1986).

57. Irani, M. H., Orosz, L. & Adhya, S. *Cell* **32**, 783-788, (1983).

58. Drew, H. R. & Dickerson, R. E. *J. Mol. Biol.* **151**, 535-556, (1981).

59. Igarashi, K., Sakamoto, I., Goto, N., Kashiwagi, K., Honma, R., & Hirose, S. *Arch. Biochem. Biophys.* **219**, 438-443, (1982).

60. Meers, P., Hong, K., Bentz, J. & Papahadjopoulos, D. *Biochemistry* **25**, 3109-3118, (1986).

61. Morris, D. R. In:*Polyamines in Biology and Medicine* (Morris, D. R. & Marton, L. J., Eds.) pp. 223-242, Marcel Dekker, New York, (1981).

62. Tabor, C. W. & Tabor, H. *Ann. Rev. Biochem.* **53**, 749-790, (1984).

63. Feuerstein, B. G., Williams, L. D., Basu, H. S. & Marton, L. J. *J. Cell. Biochem.* **46**, 37-47, (1991).

64. Williams, L. D., Frederick, C. A., Gessner, R. V. & Rich, A. In:*Molecular Conformation and Biological Interactions* (Balaram, P. & Ramaseshan, S., Eds.) pp. 295-309, Indian Academy of Sciences, Bangalore, India, (1991).

65. Tabor, C. W. & Tabor, H. *Annu. Rev. Biochem.* **45**, 285-306, (1976).

66. Sakai, T. T. & Cohen, S. S. *Progr. Nucl. Acids Res. Mol. Biol.* **71**, 15-42, (1976).

67. Mandel, M. *J. Mol. Biol.* **5**, 435-441, (1962).

68. Tabor, H. *Biochemistry* **1**, 496-501, (1962).

69. Bloomfield, V. A. & Wilson, R. W. In:*Polyamines in Biology and Medicine* (Morris, D. R. & Marton, L. J., Eds.) pp. 183-206, Marcel Dekker, New York, (1981).

70. Sen, D. & Crothers, D. *Biochemistry* **25**, 1495-1503, (1986).

71. Thomas, T. & Thomas, T. J. *Biochemistry* **32**, 14068-14074, (1993).

72. Rao, M. V. R., Atreyi, M. & Saxena, S. *Biopolymers* **29**, 1495-1497, (1990).

73. Pinagoud, A., Urbanke, C., Alves, J., Ehbrecht, H., Zabeau, M. & Gualerzi, C. *Biochemistry* **23**, 5697-5703, (1984).

74. Pinagoud, A. *Eur. J. Biochem.* **147**, 105-109, (1985).

75. Thomas, T. J. & Bloomfield, V. A. *Biopolymers* **23**, 1295-1306, (1984).

76. Braunlin, W. H., Strick, T. J. & Record, M. T., Jr. *Biopolymers* **21**, 1301-1314, (1982).

77. Wemmer, D. E., Srivenugopal, K. S., Reid, B. R. & Morris, D. R. *J. Mol. Biol.* **185**, 457-459, (1985).

78. Manning, G. S. *Q. Rev. Biophys.* **11**, 179-246, (1978).

79. Tsuboi, M. *Bull. Chem. Soc. Jpn.* **37**, 1514-1522, (1964).

80. Liquori, A. M., Costantino, L., Crescenzi, V., Elia, V., Giglio, E., Puliti, R., de Santis Savino, M. & Vitagliano, V. *J. Mol. Biol.* **24**, 113-122, (1967).

81. Feuerstein, B. G., Pattabiraman, N. & Marton, L. J. *Proc. Natl. Acad. Sci. USA* **83**, 5948-5952, (1986).

82. Zakrzewska, K. & Pullman, B. *Biopolymers* **25**, 375-392, (1986).

83. Feuerstein, B. G., Basu, H. S. & Marton, L. J. In *Progress in Polyamine Research* (Zappia, V. & Pegg, A. E., Eds.) pp 517-523, Plenum Press, New York, (1988).

84. Williams, L. D., Fredrick, C. A., Ughetto, G. & Rich, A. *Nucleic Acids Res.* **18**, 5533-5541, (1990).

85. Fasman, G. D., ed. *Handbook of Biochemisty and Molecular and Molecular Biology: Nucleic Acids.* 3rd edition. CRC Press, Cleveland, pp 26-27, (1976).

86. Ducruix, A. & Giegé, R., eds. *Crystallization of Nucleic Acids and Proteins A Practical Approach.* Oxford University Press, Oxford, (1992).

87. Taylor, C. A. & Lipson, H., eds. *Optical Transforms.* G. Bell and Sons, London, (1964).

88. Nicolet P3 Diffractometer Data Collection and Operation Manual, pp. 373, (1985).

89. Dickerson, R. E. *Methods Enzymol.* **211**, 67-111, (1992).

90. Brunger, A. *XPLOR Software* The Howard Hughes Medical Institute Yale University, New Haven, CT., (1988).

91. Rossmann, M. G. & Blow, D. M. *Acta Crystallogr.* **15**, 24-31, (1962).

92. Lattman, E. E. *Acta Crystallogr.* **B28**, 1065-1068, (1972).

93. Wilson, A. J. C. *Acta Crystallogr.* **3**, 397-406 (1950).

94. Press, W. H., Flannrey, B. P., Teukolsky, S. A. & Vetterling, W. *Numerical Recipes, The Art of Scientific Computing,* Cambridge University Press, pp. 498-504, (1986).

95. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J, Swaminathan, S. & Karplus, M. *J. Comput. Chem.* **4**,187-217, (1983).

96. Evans, S. V. *J. Mol. Graphics* **11**, 134-138, (1993).

97. Jones, T. A. *J. Appl. Crystallogr.* **11**, 268-272, (1978).

98. Luzzati, V. *Acta Crystallogr.* **5**, 802-810, (1952).

99. Briki, F. & Genest, D. *J. Biomol. Struct. Dynam.* **11**, 43-56, (1993).

100. Gueron, M. Kochoyan, M. & Leroy, J. -L. *Nature* **328**, 89-92, (1987).

101. Hartmann, B., Leng, M. & Ramstein, J. *Biochemistry* **25**, 3073-3077, (1986).

102. Lavery, R. & Sklenar, H. *J. Biomol. Struct. Dynam.* **6**, 655-667, (1989).

103. Quintana, J. R., Grzeskowiak, K., Yanagi, K. & Dickerson, R. E. *J. Mol. Biol.* **225**, 379-395, (1992).

104. Kahn, J. D., Yun, E. & Crothers, D. M. *Nature* **368**, 163-166, (1994).

105. Johnson, C. K. *ORTEP II, Oak Ridge National Laboratory, Report ORNL-5138*; Oak Ridge, Tennessee, (1976).

106. Olson, W. K. *Nucleic Acids Res.* **10**, 777-787, (1982).

107. Pullman, B. *J. Biomol. Str. Dynam.* **1**, 773-793, (1983).

108. Tismit, Y., Westhof, E., Fuchs, R. P. P & Moras, D. *Nature (London)* **341**, 459-462, (1989).

109. Larsen, T. A., Kopka, M. L. & Dickerson, R. E. *Biochemistry* **30**, 4443-4449, (1991).

110. Narayana, N., Ginell, S. L., Russo, I. M. & Berman, H. M. *Biochemistry* **30**, 4449-4450, (1991).

111 Tari, L. W. & Secco, A. S. unpublished results.

112. den Hartog, J. H. J., Altona, C., van del Elst, H., van der Marel, G. A. & Reedijk, J. *Inorg. Chem.* **24**, 983-986, (1985).

113. Leroy, J. -L., Gao, X., Gueron, M. & Patel, D. J. *Biochemistry* **30**, 5653-5661, (1991).

114. Leijon, M. & Graslund, A. *Nucleic Acids Res.* **20**, 5339-5343, (1992).

115. Gralla, J. & Crothers, D. M. *J. Mol. Biol.* **78**, 301-319, (1973).

116. Fawthrop, S. A., Yang, J. -C. & Fraser, J. *Nucleic Acids Res.* **21**, 4860-4866, (1993).

117. Maltseva, T. V., Agback, P. & Cattopadhyaya, J. *Nucleic Acids Res.* **21**, 4246-4252.

118. Goeddel, D. V., Yansula, D. B. & Caruthers, M. H. *Nucleic Acids Res.* **4**, 3038-3055, (1977).

119. Plaxco, K. & Goddard, W. A. *Biochemistry* **33**, 3050-3054, (1994).

120. Majumdar, A. & Adhya, S. *J. Mol. Biol.* **208**, 217-223, (1989).

121. Zwieb, C., Kim, J. & Adhya, S. *Genes Dev.* **3**, 606-611, (1989).

122. Ornstein, R. L., Rein, R., Breen, D. L. & MacElroy, R. D. *Biopolymers* **17**, 2341-2360, (1978).

123. Xiao, L., Swank, R. A. & Matthews, H. R. *Nucleic Acids Res.* **19**, 3701-3708, (1991).

124. Thomas, T. & Kiang, D. T. *Nucleic Acids Res.* **16**, 4705-4720, (1988).

125. Thomas, T. J., Gunnia, U. B. & Thomas, T. *J. Biol. Chem.* **10**, 6137-7141, (1991).

126. Thomas, T. & Thomas, T. J. *Biochemistry* **32**, 14068-14074, (1993).

127. Yamazaki, T., Matsuda, K., Sugiyama, H., Seto, S. & Yamaoka, N. *J. Chem. Soc. Perrkin Trans.I* 1654-1659, (1977).

128. Bessodes, M. Egron, M. -J, Filippi, J. & Antonakis, K. *J. Chem. Soc. Perkin Trans. I* 3035-3039, (1990).

129. Dyer, I., Low, J. N., Tollin, P., Wilson, H. R. & Howie, R. A. *Acta Crystallogr.* **C**44, 767-769, (1988).

130. Camerman, A., Mastropaolo, D. & Camerman, N. *Proc. Natl. Acad. Sci. USA* **84**, 8239-8242, (1987).

131. Birnbaum, G. I., Giziewicz, J., Gabe, E. J., Lin, T. -S. & Prusoff, W. H. *Can. J. Chem.* **65**, 2135-2139, (1987).

132. Van Roey, P., Salerno, J. M., Chu, C. K. & Schinazi, R. F. *Proc. Natl. Acad. Sci. USA* **86**, 3929-3933, (1989).

133. Smerdon, S. J., Jäger, J., Wang, J., Kohlstaedt, L. A., Chirino, A. J., Friedman, J. M., Rice, P. A. & Steitz, T. A. *Proc. Natl. Acad. Sci. USA*, **91**, 3911-3915, (1994).

134. Kondo, T. Nakai, H. & Goto, T *Agri. Biol. Chem.* **35**, 1990-1991, (1971).

135. Sayre, D. *Acta. Crystallogr.* **5**, 60-68, (1952).

136. Hauptman, H. & Karle, J. *Acta Crystallogr.* **9**, 635-648, (1956).

137. Gilmore, G. J. ***MITHRIL*** *J. Appl. Crystallogr.* **17**, 42-50, (1984).

138. Cromer, D. T. & Waber, J. T. In *International Tables for X-ray Crystallography,* Vol. IV. Kynoch Press, Birmingham, England pp. 99-102 and 149, (1974).

139. Stewart, R. F., Davidson, E. R. & Simpson, W. T. *J. Chem. Phys.* **42**, 3175-3185, (1965).

140. Allen, F. H., Kennard, O., Watson, D. G., Brammer, L., Orpen, A. G. & Taylor, R. *J. Chem. Soc. Perkin Trans. II* S1-S19, (1987).

141. Smith, J. L & Sundaralingam, M. *Acta Crystallogr.* **B**37, 1095-1101, (1981).

142. Sygusch, J. Brisse, F. & Hannessian, S. *Acta Crystallogr.* **B**30, 40-47, (1974).

143. De Winter, H. L.. Blaton, N. M., Peeters, O. M., De Ranter, C. J., Van Aerschot, A. & Herdewijn, P. *Acta Crystallogr.* **C**47, 838-842, (1991).

144. Hoogsteen, K. *Acta Crystallogr.* **16**, 28-38, (1963).

145. Young, D. W., Tollin, P. & Wilson, H. R. *Acta Crystallogr.* **B**25, 1423-1430, (1969).

146. Shibata, M. Takenaka, A., Sasada, Y. & Ohki, M. *Acta Crystallogr.* **C**41, 1356-1358, (1985).

147. Cremer, D. & Pople, J. A. *J. Amer. Chem. Soc.* **97**, 1354-1358, (1975).

148. Motherwell, S. & Clegg, W. *Pluto. Program for plotting molecular and crystal structures*, Univ. of Cambridge, England (1978).

149. PCMODEL v. 4.0, Serena Software, Box 3076, Bloomingtom, Indiana, (1990).

150. Uhlmann, E. & Peyman, A. *Chemical Reviews* **90**, 543-584, (1990).

151. Wagner, R. W. *Nature* **372**, 333-335, (1994).

152. Wagner R. W., Matteucci, M. D., Lewis, J. G., Gutierrez, A. J., Moulds, C. & Froehler, B. C. *Science* **260**, 1510-1513, (1993).

153. Fenster, S. D., Wagner, R. W., Froehler, B. C. & Chin, D. J. *Biochemistry* **33**, 8391-8398, (1994).

154. Milligan, J. F., Matteucci, M. D. & Martin, J. C. *J. Med. Chem.* **36**, 1923-1937, (1993).

155. Perbost, M. Lucas, M., Chavis, C., Pompon, A., Baumgartner, H., Rayner, B., Griengl, H. & Imbach, J. -L. *Biochem. Biophys. Res. Com.* **165**, 742-747, (1989).

156. Morvan, F., Rayner, B., Leonetti, J. -P. & Imbach, J. -L. *Nucl. Acids Res.* **16**, 833-847, (1988).

157. Vandendriessche, F., Augustyns, K., Aerschot, A. V., Busson, R., Hoogmartens, J. & Herdewijn, P. *Tetrahedron* **49**, 7223-7238, (1993).

158. Augustyns, K., Van Aerschot, A., Urbanke, C. & Herdewijn, P. *Bull. Soc. Chim. Belg.* **101**, 119-124, (1992).

159. Augustyns, K., Godard, G., Hendrix, C., Van Aerschot, A., Rozenski, J., Saison-Behmoaras, T. & Herdewijn, P. *Nucleic Acids Res.* **21**, 4670-4676, (1993).

160. Augustyns, K., Vandendriessche, F. Van Aerschot, A., Busson, R., Urbanke, C. & Herdewijn, P. *Nucleic Acids Res.* **20**, 4711-4716, (1992).

161. Pitsch, S., Wendeborn, S., Jaun, B. & Eschenmoser, A. *Helvetica Chim. Acta* **76**, 2161-2183, (1993).

162. Eschenmoser, A. *Pure & Appl. Chem.* **65**, 1179-1188, (1993).

163. Buchko, G. W., Hruska, F. E. & Sadana, K. L. *Can. J. Chem.* **68**, 2011-2021, (1990).

164. Cantor, C. R., Warshaw, M. M. & Shapiro, H. *Biopolymers,* **9**, 1059-1077, (1970).

165. Ivanov, V. I, Minchenkova, L. E., Schyolkina, A. K. & Poletayev, A. I. *Biopolymers* **12**, 89-100, (1973).

166. Tari, L. W. & Secco, A. S. *Can. J. Chem.* **90**, 894-899, (1992).