Crystallographic and Synthetic Studies of Several Biologically Relevant Compounds

by

Connie K. Cho

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

Master of Science

Department of Chemistry
University of Manitoa
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 0-612-13029-0

Canada

Name Connie K. Cho

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.

Biochemistry
SUBJECT TERM

0487 U·M·I
SUBJECT CODE

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

CRYSTALLOGRAPHIC AND SYNTHETIC STUDIES OF SEVERAL BIOLOGICALLY RELEVANT COMPOUNDS

BY

CONNIE K. CHO

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

MASTER OF SCIENCE

Abstract

The crystal structures of two compounds were determined, and structural motifs potentially significant in compound activity were enumerated. Unit cell dimensions were determined for a macromolecular crystal, *viz.* a DNA octamer. The synthesis of a pyranosyl-based nucleoside analogue, while successful, was hampered by several significant side reactions.

The first crystal structure described, of 14β-hydroxyprogesterone, is of a mammalian-derived, semi-synthetic hormone which induces muscle contraction in isolated heart muscle. The crystals are monoclinic, $P2_1$, with cell dimensions a = 11.831(3), b = 8.096(2), c = 18.696(6) Å, β = 91.38(2)°, V = 1790.3(8) Å^3, Z = 4. The features observed in the crystal structure cannot fully account for the title compound's activity, although some features ubiquitous to cardiac glycosides are present.

2′-(Deoxy-β-D-*arabino*-hexopyranosyl)thymine, a 2′-deoxyhexopyranose analogue of thymine with known and potential biological properties, was synthesized by two methods and partially purified. Neither synthetic method proved particularly effective, and the compound was sensitive to dethymidination in the presence of heat and methanol. The nature of the

solvent used was critical for successful synthesis, but ironically the most successful solvent may also have been responsible for enhancing the production of side reactions.

The high resolution crystal structure of spermine phosphate hexahydrate was determined to obtain structural details which could not be studied in a previous study. The crystal was monoclinic, $P2_1/c$, a=6.898(2), b=23.217(4), c=7.934(1) Å, β=113.47(1)°, V=1165.47(1) Å^3. Of particular interest was the nature of hydration of the molecule. The inorganic and spermine molecules established discrete layers; the water molecules arranged themselves in a zigzag fashion, perhaps imitating what occurs during spermine-DNA interaction.

The DNA sequence studied was $^{5'}$GATTCCAC$^{3'}$/$^{5'}$GTGGAATC^{3}, which is located in the exterior operator of the E. coli *gal* operon. The sequence $^{5'}$GTG$^{3'}$/$^{5'}$CAC^{3} is of interest because of its unusually frequent appearance in sites which are known to be involved in recognition. No information about the structure was yet obtained due to poor crystal quality, although the likely space group of $P4_n$ and similarity to known octamer structures suggested the DNA might be A type.

Acknowledgements

I would like to express my gratitude to Professor A. S. Secco, for his patience, guidance, and enthusiasm as a supervisor throughout my graduate study.

I would also like to thank Professor L. Diop for all his help and expertise with the nucleoside analogue project, Dr. K. Sadana for his practical advice on nucleoside synthesis, and Dr. J. Templeton for providing the crystal of 14β-hydroxyprogesterone.

Thanks also to Kathy Rzeszutek, Nancy Rosenthal, and Jon Havelock, for their assistance with purification and synthesis of materials; to Angela Toms, for her excellent work on the spermine structure; and to Les Tari, for all his valuable aid and information.

List of Figures

Figure **page**

List of Tables

Table **page**

Table of Contents

I. General Introduction

I. Introduction

The study of biologically relevant compounds by crystallography is an increasingly popular field. While most biological molecules and assemblies do normally exist in solution, very frequently they experience situations that may be described by the crystal state. One example is a situation in which a pharmaceutical molecule may selectively bind to a receptor site; the molecular conformation may be largely fixed in the interaction, much as it might be in the crystal. Another example concerns the interaction between two macromolecules, such as an enzyme and DNA. When the two macromolecules lock together in a complex, movement is minimal in the area of the interaction, and X-ray crystallography is a very plausible study.

The practical development of useful drugs has been enhanced by the use of rational drug design, in which the identification of certain structural motifs, integral to the compound's ability to interact with its target, permits concentration of attention on the most productive areas. The x-ray structure of an active compound may reveal the conformational detail in high resolution, since many drugs tend to be small molecules. Information may also often be gleaned, however, from any observed disorder. The disorder may indicate that while a functional group has a marked preference for one orientation,

it may also occur in significant quantity in some other, or it may indicate that a group is highly labile under even extreme conditions, capable of adopting many energetically similar conformations. Such interpretations of the data may be important in identifying a key feature in activity; the identification may have further practical ramifications in, for example, potential commercial production of a drug.

DNA crystallography is not as advanced as small molecule crystallography, or even protein crystallography. This is partially due to the difficulty in the past in obtaining sufficient quantities of a desired oligonucleotide, although with the advent of automated synthesis DNA is much more readily available. Native DNA, as proteins are, is a single long molecule, but unlike proteins only a selected region fewer than 15 base pairs in length may be practically synthesized and studied at a time.

Diffraction quality DNA crystals are difficult to obtain due to the exacting conditions which are required for optimal growth. Furthermore, crystallization conditions for any oligomer of a given sequence are not likely to be identical to those of another oligomer, even an oligomer with a very similar sequence; much of the total effort involved is consumed in the initial screening step. Internally disordered crystals, which diffract poorly and yield little fine structural detail,

are not uncommon, and even exceptional crystals do not diffract to the extent that small molecule crystals do. DNA crystals are also fragile, susceptible to heat and low humidity, in a manner than small molecule crystals are not.

The boundaries between small molecule and macromolecule study are often blurred, since the structure of the native DNA oligomer itself is not the only feature of interest. There are also co-crystallizing small molecules, such as spermine, whose structural effect on the macromolecule may be noteworthy. High resolution details are lost for the small molecule, but information on conformation, for example, may be obtained for an interaction between the two molecules. In addition, unusual small molecules, such as nucleoside analogues, may be incorporated into the DNA chain, Comparisons of the analogue structures to the native, in both small molecule studies and macromolecular studies, are useful in quantifying key features potentially involved in biological processes.

II. 14-Hydroxy-14β-pregn-4-ene-3,20-dione

(14β-hydroxyprogesterone)

II.1 Introduction

Heart disease is a major cause of illness today, and a common form of treatment is the use of cardiac glycosides (CGs), steroidal compounds which enhance heart muscle contraction by inhibiting Na^+, K^+-ATPase and the sodium pump (1). A major CG in use is digitalis, which is derived from the foxglove plant. Although useful, digitalis and related molecules such as digitoxigenin, a CG derived from a component of digitalis, are foreign to mammalian systems, and some incompatibilities are likely. A more congenial choice would be the compound which is the putative natural target molecule for the digitalis receptor, and many steroid derivatives have been assayed in order to categorize this endogenous digitalis-like hormone (2-7). Although known mammalian steroids are planar and have no CG activity, the steroid nature which is ubiquitous to the CG skeleton suggests that the endogenous hormone is likely to share the same general structure. Mammalian steroids may be enzymatically modified *in vivo* to approximate the characteristic globular CG structure, in which *cis* A/B and C/D ring junctions cause the A and D rings to bend out of the plane, almost perpendicular to the B and C rings (1). The crystal structures of a number of compounds which show CG-like activity have been compared and it is apparent that a close relationship

between the activity of the compounds and several structural features exists (8). 14β-hydroxyprogesterone is a semi-synthetic analog of hydroxyprogesterone, and the first mammalian-derived analog which induces CG-like activity in isolated heart muscle (9). However, the effect is weak, and the crystal structure of 14β-hydroxyprogesterone was determined in an effort to provide some explanation for its activity and to furnish information on characteristics useful in identifying more effective endogenous digitalis-like hormones.

II.2 Experimental

Crystals of 14β-hydroxyprogesterone were supplied by Dr. Templeton of the Department of Pharmacy. A colorless, prismatic crystal of dimensions 0.4 X 0.2 X 0.2 mm was selected and mounted on a glass fibre with epoxy resin. Data were collected on a Nicolet R3*m* diffractometer equipped with a Mo X-ray tube.

Crystal Data. $C_{21}H_{30}O_3$, M_r = 330.47, monoclinic, $P2_1$, a = 11.831(3), b = 8.096(2), c = 18.696(6) Å, β = 91.38(2)°, V = 1790.3(8) Å^3, Z = 4, D_m(floatation) = 1.225 Mg m^{-3}, D_c = 1.226 Mg m^{-3}, λ(Mo Kα) = 0.71069 Å, μ = 0.86 cm^{-1}, F(000) = 720, T = 294 K, R = 0.036 for 1588 reflections with $I \geq 3\sigma(I)$.

Unit cell parameters were determined by least-squares

refinement of 25 reflections chosen from a rotation photo. Axial photos suggested that the space group was monoclinic.

Nineteen hundred and fourteen reflections from the unique quadrant ($-12 \leq h \leq 12$, $0 \leq k \leq 8$, $0 \leq l \leq 18$) were collected by an $\omega/2\theta$ scan mode. The crystal quality limited the 2θ range to 2 - 40°. The data were corrected for Lorentz-polarization effects, resulting in 1838 unique reflections. Three standard reflections monitored during data collection showed no significant intensity variation after 48 h.

II.3 Structure Solution and Refinement

The structure was solved by the direct methods program *MULTAN87* (10) by increasing the number of E's used to 600 from the default 350 and using a randomly oriented fragment. Two independent molecules in the asymmetric unit were revealed. All non-hydrogen atoms appeared in the solution and were initially assigned as carbon atoms, then refined isotropically with unit weights by full-matrix least-squares refinement using the 1588 reflections with $I \geq 3\sigma(I)$. The refinement converged at $R = 12\%$, with some unacceptable geometry, where

$$R = \Sigma \, | \, |F_o| - k|F_c| \, | \, / \Sigma |F_o| \ .$$

Correct assignment of the atom types at this juncture

did not improve the refinement, so refinement was repeated with the correct atom assignment from the beginning, whereupon the *R* factor dropped and the geometries improved. Further refinement with anisotropic temperature factors and subsequent difference Fourier maps revealed positions for the hydrogen atoms; however, attempts to refine their coordinates failed. All hydrogens except those on the hydroxyl groups were placed in calculated positions for the final two cycles without refinement. The two hydroxyl hydrogens were located on a difference Fourier map, and the observed coordinates were inserted without refinement. All hydrogens were assigned isotropic temperature factors twice those of their attached atoms.

After the final cycle of refinement, $R = 0.047$ and $wR = 0.044$ for all 1838 reflections while $R = 0.036$ and $wR = 0.044$ for the 1588 observed reflections used, where

$$wR = [\Sigma w(|F_o|-k|F_c|)^2 / \Sigma w|F_o|^2]^{1/2}.$$

The use of sigma weights did not yield the best refinement, so a modified scheme based on observed intensities was used:

$$w = 1 \text{ for } F < 22.8, \quad w = 22.8/F \text{ for } F \geq 22.8.$$

The goodness-of-fit (*GOF*) = 0.623, where

$$GOF = [\Sigma w(|F_o|-|F_c|)^2/n-m]^{1/2}$$

n= number of observations, m= number of parameters.

The maximum shift/error (Δ/σ) = 0.049; the isotropic extinction coefficient g = 2.1(6) X 10^4 (11). The final difference map was essentially featureless, with maximum and minimum residuals of 0.171 and -0.156 $e\text{Å}^{-3}$, respectively. The refinement results of the other enantiomer were not significantly different. Atomic scattering factors for non-hydrogen atoms were from (12) and from (13) for H atoms. All computation was carried out on the University of Manitoba Computer Services Department's Amdahl 5870 mainframe computer using locally written programs for processing and modified versions of structure solution, refinement and calculations programs (14-16). Final atomic coordinates and thermal factors of non-hydrogen atoms and hydroxyl hydrogens are given in Table 1. The two molecules in the asymmetric unit are shown in Figure 1, and entire unit cell contents are shown in Figure 2.

II.4 Discussion

The refinement posed an unusual problem in that the true structure had to be assigned from the beginning. Usually, it is possible to refine an organic structure with all atoms initially assigned as carbons; the thermal factor reflects the difference in electron density between the assigned

Table 1

Final positional parameters and isotropic thermal parameters for 14β-hydroxyprogesterone, with estimated standard deviations in parentheses

Atom	x	y	z	U_{eq}/U_{iso} †
C(1)	0.4620(5)	0.1009	0.8566(4)	6.2
C(2)	0.4917(6)	0.2257(16)	0.9148(4)	5.7
C(3)	0.4186(7)	0.3726(15)	0.9092(4)	5.4
O(3)	0.4517(6)	0.5111(14)	0.9249(4)	9.6
C(4)	0.2990(6)	0.3417(14)	0.8892(4)	5.0
C(5)	0.2614(5)	0.1941(14)	0.8662(3)	3.9
C(6)	0.1375(5)	0.1677(13)	0.8523(4)	4.8
C(7)	0.1154(5)	0.1023(14)	0.7767(4)	4.7
C(8)	0.1831(5)	-0.0541(13)	0.7624(3)	3.4
C(9)	0.3092(5)	-0.0194(14)	0.7771(3)	4.0
C(10)	0.3367(5)	0.0479(14)	0.8533(3)	4.0
C(11)	0.3797(6)	-0.1702(15)	0.7576(4)	5.8
C(12)	0.3566(5)	-0.2284(15)	0.6807(4)	5.9
C(13)	0.2314(5)	-0.2732(14)	0.6667(3)	4.1

Table 1 (continued)

C(14)	0.1564(5)	-0.1253(13)	0.6877(3)	3.8
O(14)	0.0393(3)	-0.1734(12)	0.6897(2)	4.8
C(15)	0.1707(6)	-0.0032(14)	0.6260(4)	4.9
C(16)	0.1868(7)	-0.1086(15)	0.5590(4)	6.5
C(17)	0.2136(5)	-0.2863(15)	0.5838(4)	4.8
C(18)	0.2034(7)	-0.4342(14)	0.7361(4)	5.8
C(19)	0.3160(6)	-0.0819(14)	0.9117(4)	5.7
C(20)	0.1229(7)	-0.4071(15)	0.5601(4)	5.5
O(20)	0.0232(5)	-0.3736(13)	0.5683(3)	7.4
C(21)	0.1555(7)	-0.5657(16)	0.5272(5)	7.9
C(1)'	0.6483(7)	-0.8018(14)	0.6896(4)	5.8
C(2)'	0.5898(7)	-0.8759(17)	0.6241(4)	7.1
C(3)'	0.5193(7)	-0.7500(16)	0.5867(4)	6.5
O(3)'	0.4265(5)	-0.7804(14)	0.5593(3)	10.4
C(4)'	0.5703(6)	-0.5863(14)	0.5808(4)	5.2
C(5)'	0.6653(5)	-0.5416(14)	0.6168(3)	4.1

Table 1 (continued)

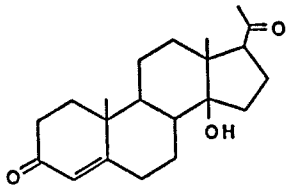
C(6)'	0.7182(6)	-0.3747(14)	0.6053(3)	4.9
C(7)'	0.7273(5)	-0.2803(13)	0.6753(3)	4.1
C(8)'	0.7903(5)	-0.3770(13)	0.7337(3)	3.2
C(9)'	0.7381(5)	-0.5471(13)	0.7433(3)	3.6
C(10)'	0.7227(5)	-0.6503(13)	0.6729(3)	3.9
C(11)'	0.8026(6)	-0.6418(13)	0.8029(3)	5.0
C(12)'	0.8030(6)	-0.5437(14)	0.8726(3)	4.6
C(13)'	0.8568(5)	-0.3727(13)	0.8665(3)	3.3
C(14)'	0.7995(5)	-0.2792(13)	0.8034(3)	2.9
O(14)'	0.8603(3)	-0.1280(12)	0.7904(2)	3.9
C(15)'	0.6857(5)	-0.2219(14)	0.8326(3)	4.3
C(16)'	0.6968(5)	-0.2265(14)	0.9152(3)	4.7
C(17)'	0.8210(5)	-0.2706(13)	0.9330(3)	3.9
C(18)'	0.9845(5)	-0.3887(14)	0.8592(3)	4.9
C(19)'	0.8371(6)	-0.7105(14)	0.6438(4)	5.4
C(20)'	0.8958(6)	-0.1252(14)	0.9524(3)	4.5

Table 1 (continued)

O(20)'	0.9902(5)	-0.1494(12)	0.9777(3)	7.5
C(21)'	0.8566(7)	0.0493(14)	0.9438(4)	5.8
H(O14)	0.0012	-0.0256	0.0646	9.5 *
H(O14)'	0.0936	-0.0154	0.0762	7.9 *

* unrefined coordinates and thermal parameters

† U_{eq}= 1/3 (trace of the diagonalized anisotropic temperature factor matrix)



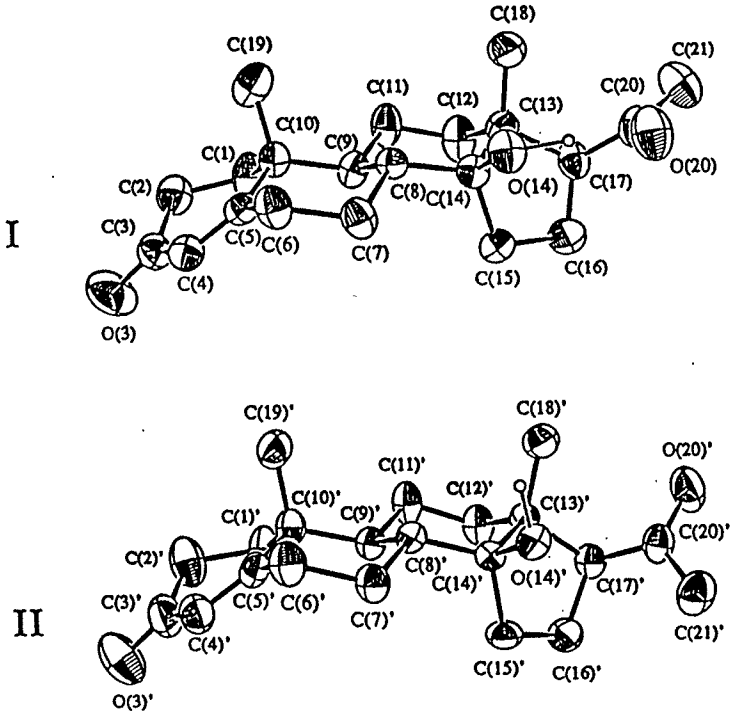


Figure 1.

Schematic drawing and *ORTEP* plots of 14β-hydroxyprogesterone (ellipsoids at 50% probability level). Hydrogen atoms other than hydroxyl groups were omitted for clarity

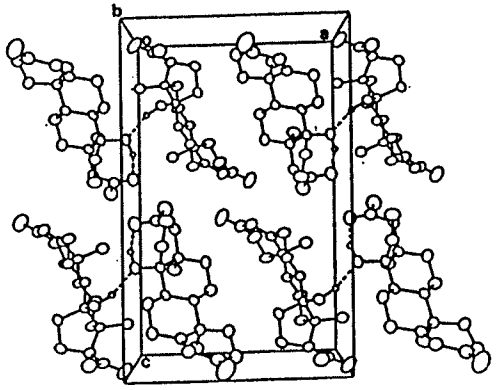


Figure 2.

Packing diagram of unit cell viewed down *b* axis (into page), showing hydrogen bonding.

carbon atom and the actual atom. However, in this situation, setting the oxygen atoms as carbons was sufficient to cause the refinement to converge to a false solution or local minimum. It was necessary to assign all atoms correctly from the beginning and then proceed slowly with the refinement, by introducing a few parameters at each refinement cycle.

The goodness-of-fit (*GOF*) value, also *the standard deviation of an observation of unit weight*, is smaller than desirable. Ideally, *GOF* = 1.0, but in practice it can be higher (17). A *GOF* > 1.0 generally arises from inaccuracies in the model, which may not account for features such as electron pairs, and to non-random errors such as extinction. The weighting scheme used, *w*, is very similar to a unit weight scheme in distribution, scaling down only the very intense reflections. A *GOF* of 0.623 indicates that the weighting scheme used artificially underweighted some reflections. However, since the data are reasonably good, the refinement results are not noticeably dependent upon the weighting scheme (17); the final structure is reasonable, and the final difference Fourier map generally smooth.

The structure solution clearly shows that 14β-hydroxyprogesterone adopts two conformations, I and II (Figure 1), both of which resemble the crescent shape of a typical CG more than the essentially planar steroid skeleton of progesterone

(18,19). This crescent shape was expected in view of the compound's CG-like effects on cardiac tissue (9). Structural elements required for the drug to bind to the receptor have been well characterized (20) and primarily involve the A and D rings, particularly the functional groups attached to these rings. Concerted motions appear to be needed, as appropriate parameters in one region alone are not sufficient to account for all activity; both I and II in the crystalline state exhibit the conformational requirements, but both also have regions that do not conform. These structural deviations from the canonical CG structure may be offsetting the similarities of the more well-known structural features, which may provide a partial explanation for the compound's relatively weak affinity (about one-tenth that of ouabagenin) for the digitalis receptor, Na^+, K^+-ATPase, as inferred from inhibition of [^{3}H]ouabain binding to cardiac tissue (9).

In Figure 3, both structures of 14β-hydroxyprogesterone are superimposed on the crystal structure of digitoxigenin (21), which is used as a standard for comparison. A least squares fit of the B and C rings to the corresponding atoms in digitoxigenin produced on the PC program ALCHEMYII (22) shows good correlation for both molecules, with an r.m.s deviation of 0.029 Å for the C(5)...C(14) atoms, slightly less than the 0.055 Å for the C(5)'...C(14)' atoms. Bond lengths and angles are generally normal in I and II; the

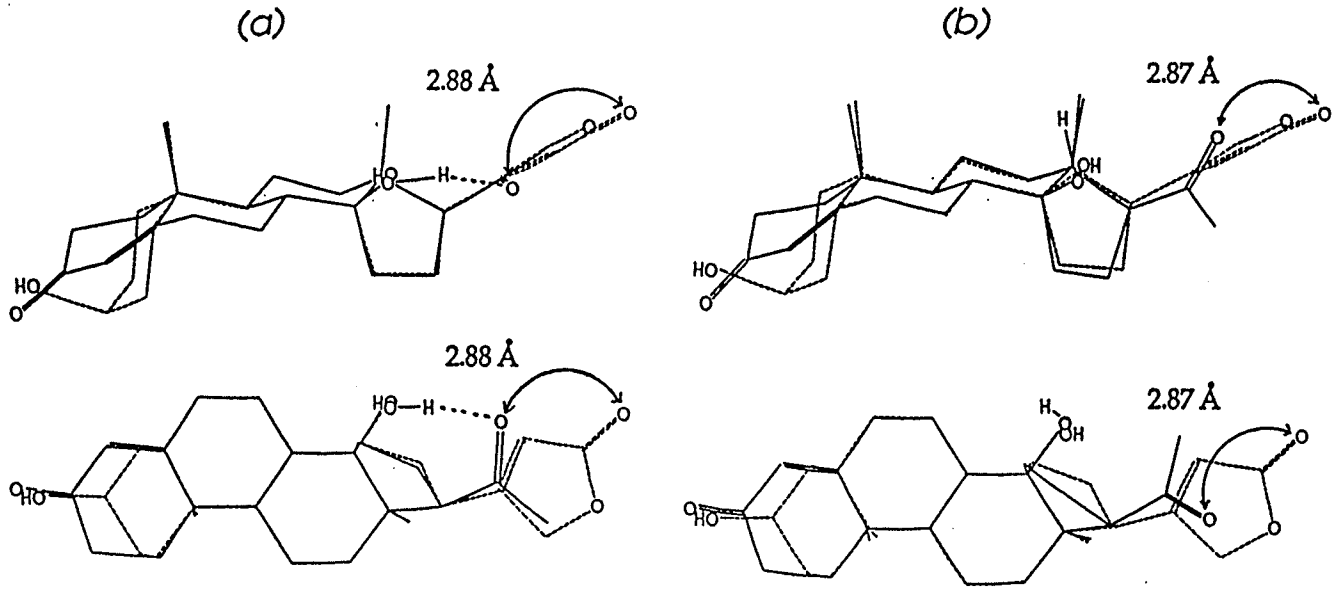


Figure 3.Orthogonal projections showing superposition of B and C ring atoms of digitoxigenin and (a) I and (b) II of 14β-hydroxyprogesterone (least squares fit through atoms C(5)..C(14)).Solid line: 14β-hydroxyprogesterone. Dashed line: digitoxigenin

Table 2

Bond lengths (Å) and angles (°) of 14β-hydroxyprogesterone, with estimated standard deviations in parentheses

bond	length	bond	length
C(1)-C(2)	1.519(10)	C(1)'-C(2)'	1.517(10)
C(1)-C(10)	1.543(9)	C(1)'-C(10)'	1.546(10)
C(2)-C(3)	1.473(12)	C(2)'-C(3)'	1.481(12)
C(3)-O(3)	1.220(10)	C(3)'-O(3)'	1.226(9)
C(3)-C(4)	1.476(11)	C(3)'-C(4)'	1.461(11)
C(4)-C(5)	1.342(10)	C(4)'-C(5)'	1.345(9)
C(5)-C(6)	1.499(9)	C(5)'-C(6)'	1.506(10)
C(5)-C(10)	1.504(10)	C(5)'-C(10)'	1.517(9)
C(6)-C(7)	1.527(9)	C(6)'-C(7)'	1.519(10)
C(7)-C(8)	1.525(9)	C(7)'-C(8)'	1.524(9)
C(8)-C(9)	1.537(8)	C(8)'-C(9)'	1.522(9)
C(9)-C(10)	1.553(9)	C(9)'-C(10)'	1.566(9)
C(9)-C(11)	1.527(10)	C(9)'-C(11)'	1.540(9)
C(10)-C(19)	1.539(10)	C(10)'-C(19)'	1.549(10)
C(11)-C(12)	1.530(10)	C(11)'-C(12)'	1.526(10)
C(12)-C(13)	1.541(9)	C(12)'-C(13)'	1.529(10)
C(13)-C(14)	1.547(9)	C(13)'-C(14)'	1.544(8)

Table 2 (continued)

C(13)-C(17)	1.564(9)	C(13)'-C(17)'	1.560(9)
C(13)-C(18)	1.537(10)	C(13)'-C(18)'	1.526(8)
C(14)-O(14)	1.440(7)	C(14)'-O(14)'	1.444(8)
C(14)-C(15)	1.532(9)	C(14)'-C(15)'	1.537(8)
C(15)-C(16)	1.531(11)	C(15)'-C(16)'	1.548(9)
C(16)-C(17)	1.542(12)	C(16)'-C(17)'	1.541(9)
C(17)-C(20)	1.510(10)	C(17)'-C(20)'	1.512(10)
C(20)-O(20)	1.223(8)	C(20)'-O(20)'	1.219(8)
C(20)-C(21)	1.480(12)	C(20)'-C(21)'	1.494(11)
H(14)-H(O14)	1.102*	H(14)'-H(O14)'	1.074*

atoms	angle	atoms	angle
C(2)-C(1)-C(10)	114.7(6)	C(2)'-C(1)'-C(10)'	113.6(6)
C(1)-C(2)-C(3)	111.2(6)	C(1)'-C(2)'-C(3)'	110.5(8)
C(2)-C(3)-O(3)	122.8(7)	C(2)'-C(3)'-O(3)'	123.2(9)
C(2)-C(3)-C(4)	116.0(7)	C(2)'-C(3)'-C(4)'	115.5(7)
O(3)-C(3)-C(4)	121.1(8)	O(3)'-C(3)'-C(4)'	121.2(9)
C(3)-C(4)-C(5)	122.7(7)	C(3)'-C(4)'-C(5)'	123.2(7)

Table 2 (continued)

C(4)-C(5)-C(6)	119.6(6)	C(4)'-C(5)'-C(6)'	120.9(6)
C(4)-C(5)-C(10)	124.0(6)	C(4)'-C(5)'-C(10)'	123.0(7)
C(6)-C(5)-C(10)	116.1(6)	C(6)'-C(5)'-C(10)'	116.0(5)
C(5)-C(6)-C(7)	110.8(5)	C(5)'-C(6)'-C(7)'	110.4(5)
C(6)-C(7)-C(8)	111.7(6)	C(6)'-C(7)'-C(8)'	112.5(6)
C(7)-C(8)-C(9)	109.2(5)	C(7)'-C(8)'-C(9)'	110.9(5)
C(8)-C(9)-C(10)	114.1(5)	C(8)'-C(9)'-C(10)'	115.0(5)
C(8)-C(9)-C(11)	113.3(5)	C(8)'-C(9)'-C(11)'	109.9(5)
C(10)-C(9)-C(11)	113.3(5)	C(10)'-C(9)'-C(11)'	112.9(6)
C(1)-C(10)-C(5)	110.3(6)	C(1)'-C(10)'-C(5)'	110.6(5)
C(1)-C(10)-C(9)	108.3(5)	C(1)'-C(10)'-C(9)'	107.9(5)
C(1)-C(10)-C(19)	109.3(6)	C(1)'-C(10)'-C(19)'	109.1(6)
C(5)-C(10)-C(9)	108.2(5)	C(5)'-C(10)'-C(9)'	108.2(5)
C(5)-C(10)-C(19)	108.6(5)	C(5)'-C(10)'-C(19)'	108.7(5)
C(9)-C(10)-C(19)	112.2(6)	C(9)'-C(10)'-C(19)'	112.2(5)
C(9)-C(11)-C(12)	112.6(6)	C(9)'-C(11)'-C(12)'	110.5(6)
C(11)-C(12)-C(13)	112.4(6)	C(11)'-C(12)'-C(13)'	113.6(5)
C(12)-C(13)-C(14)	109.2(6)	C(12)'-C(13)'-C(14)'	109.0(5)

Table 2 (continued)

C(12)-C(13)-C(17)	106.9(5)	C(12)'-C(13)'-C(17)'	107.4(5)
C(12)-C(13)-C(18)	109.6(6)	C(12)'-C(13)'-C(18)'	110.1(6)
C(14)-C(13)-C(17)	103.9(6)	C(14)'-C(13)'-C(17)'	103.1(5)
C(14)-C(13)-C(18)	113.8(5)	C(14)'-C(13)'-C(18)'	113.0(5)
C(17)-C(13)-C(18)	113.1(6)	C(17)'-C(13)'-C(18)'	113.8(5)
C(13)-C(14)-O(14)	110.8(5)	C(13)'-C(14)'-O(14)'	109.4(5)
C(13)-C(14)-C(15)	103.5(6)	C(13)'-C(14)'-C(15)'	104.4(5)
O(14)-C(14)-C(15)	108.6(5)	O(14)'-C(14)'-C(15)'	104.3(5)
C(14)-C(15)-C(16)	105.9(6)	C(14)'-C(15)'-C(16)'	107.1(5)
C(15)-C(16)-C(17)	107.7(6)	C(15)'-C(16)'-C(17)'	106.1(5)
C(13)-C(17)-C(16)	104.8(6)	C(13)'-C(17)'-C(16)'	103.1(5)
C(13)-C(17)-C(20)	114.6(6)	C(13)'-C(17)'-C(20)'	115.8(5)
C(16)-C(17)-C(20)	112.1(6)	C(16)'-C(17)'-C(20)'	114.9(6)
C(17)-C(20)-O(20)	120.0(7)	C(17)'-C(20)'-O(20)'	119.6(7)
C(17)-C(20)-C(21)	119.6(7)	C(17)'-C(20)'-C(21)'	122.1(6)
O(20)-C(20)-C(21)	120.4(8)	O(20)'-C(20)'-C(21)'	118.3(7)
C(14)-O(14)-H(O14)	113.9*	C(14)'-O(14)'-H(O14)'	109.9*

* calculated from unrefined hydrogen coordinates

largest deviations occurring between the two structures are found in the B ring (Table 2). All B and C rings are in the chair conformation.

The A/B ring junctions are quasi-*trans* in both I and II, as in other progestins, and have similar endocyclic angles. Ring characteristics are described by asymmetry parameters, which reflect the conformational tendencies of the ring. The asymmetry parameters are defined as

$$\Delta C_s = [{}^{m}\Sigma_{i-1} (\varphi_i + \varphi_i')^2/m]^{1/2}$$

$$\Delta C_2 = [{}^{m}\Sigma_{i-1} (\varphi_i - \varphi_i')^2/m]^{1/2}$$

where φ_i and φ_i' are the symmetry related torsion angles and m is the number of individual comparisons (23). Asymmetry parameters of the A ring are: $\Delta C_s(1) = 20.2°$, $\Delta C_2(1,2) = 0.9°$; $\Delta C_s(1)' = 24.1°$, $\Delta C_2(1,2)' = 2.6°$, indicating that both A rings tend toward the $1\alpha,2\beta$-half-chair instead of the 1α-sofa. Both rings are in the normal conformation, with the 2β hydrogen axial. The C(2)-C(3) and C(2)'-C(3)' bonds, 1.47(1) and 1.48(1) Å respectively, are close to the observed average C-C bond length of 1.49 Å. The torsion angles O(3)-C(3)-C(4)-C(5) of 174.6° and O(3)'-C(3)'-C(4)'-C(5)' of 171.9° (Table 3) indicate less than perfect conjugation. The values observed in the title compound lie at the extreme end of the range of previously observed deviations from planarity, generally small,

Table 3

Selected torsion angles (°) of 14β-hydroxyprogesterone, with estimated standard deviations in parentheses

O(3)-C(3)-C(4)-C(5)	174.6(7)
C(17)-C(13)-C(14)-C(15)	-37.9(6)
C(15)-C(16)-C(17)-C(13)	-8.8(8)
C(14)-C(15)-C(16)-C(17)	-14.8(8)
C(14)-C(13)-C(17)-C(16)	28.8(7)
C(13)-C(14)-C(15)-C(16)	32.7(6)
C(13)-C(17)-C(20)-O(20)	72.4(9)
C(13)-C(17)-C(20)-C(21)	-107.5(8)
C(16)-C(17)-C(20)-O(20)	-46.8(9)
C(16)-C(17)-C(20)-C(21)	133.4(8)

Table 3 (continued)

O(3)'-C(3)'-C(4)'-C(5)'	171.9(7)
C(17)'-C(13)'-C(14)'-C(15)'	-36.7(6)
C(15)'-C(16)'-C(17)'-C(13)'	-27.8(7)
C(14)'-C(15)'-C(16)'-C(17)'	5.3(8)
C(14)'-C(13)'-C(17)'-C(16)'	39.9(6)
C(13)'-C(14)'-C(15)'-C(16)'	19.6(7)
C(13)'-C(17)'-C(20)'-O(20)'	-71.1(8)
C(13)'-C(17)'-C(20)'-C(21)'	111.0(7)
C(16)'-C(17)'-C(20)'-O(20)'	168.8(6)
C(16)'-C(17)'-C(20)'-C(21)'	-9.2(9)

in a number of steroids (23). While neither A ring of I nor of II approaches the chair form with a *cis* A/B junction (as in the hydroxylated A ring of digitoxigenin), both show increased displacement in the α direction, opposite to the orientation of the C(18) and C(19) methyl groups, compared to either form of progesterone. The 3-carbonyl oxygen is directed considerably away from the plane of the ring, O(3)' more than O(3). This out-of-plane bending is also demonstrated, to a lesser degree, by chlormadinone acetate (24), the most potent semisynthetic analog of digitoxigenin discovered so far (1). The protrusion of the 3-carbonyl group may be of importance in determining the compound's biological activity, by bringing it into a more favorable position for hydrogen bonding to the digitalis receptor. It is not clear what is responsible for this pucker of the A ring, or why the two asymmetric molecules should show different degrees of bending. The only intermolecular contact of note for either molecule is the O(14)···H-O(14)' hydrogen bond, which appears to have a significant effect on the D ring, as described later in the chapter.

Long range intramolecular effects, namely those resulting from the nature of functional groups on the other rings, may have repercussions on the conformation of the A ring. 17α-acetoxy substitution, for example, restricts the freedom of the C(17) side chain and is known to induce a 1α-sofa conformation on

the A ring (4), while B ring substitution in the 6α position induces formation of a 1α,2β-half-chair on the A ring (3). 14β-hydroxylation may have similar effects, since the two 14β-OH groups possess different orientations and H-bonding, and also show different overall downward displacement; that of II is greater than that of I, both as a result of increased A ring pucker and out-of-plane distortion of O(3)'. The actual shape of the ring (*i.e.* sofa vs. chair) may not be significant provided that the functional groups themselves are aligned properly. Despite the radically different conformations of the A rings of 14β-hydroxyprogesterone and digitoxigenin, O(3) and O(3)' are spatially positioned very similarly to O(1) of III in the superposition discussed above, with separations between oxygen atoms of 0.62 Å and 0.73 Å for $O(3)\cdots O(1)_{III}$ and $O(3)'\cdots O(1)_{III}$, respectively. However, the orientations of the particular groups are quite different, which may limit binding to highly specific enzymes such as Na^+, K^+-ATPase and thus may be partially responsible for 14β-hydroxyprogesterone's low affinity.

The conformational differences between I and II may be due to the different hydrogen bonding experienced by the two molecules. O(14)(x,y,z) of conformer I is involved in intramolecular hydrogen bonding to O(20) and intermolecular hydrogen bonding to O(14)'(x-1,y,z) of conformer II, whereas O(14)' experiences only the intermolecular hydrogen bond

(Figure 3). H(O14)···O(20) and O(14)···H(O14)′ have distances of 1.74 Å and 1.83 Å, respectively. The O(14)···H(O14)′-O(14)′ angle is 161.5° while the O(14)-H(O14)···O(20) angle is 158.1°.

Although angles involving C(14) of the C/D ring junction differ by up to 4°, it may be significant that C(14)-O(14) and C(14)′-O(14)′ are similarly oriented with respect to each other and with respect to C(14)-O(2) of digitoxigenin. The C/D ring junctions in I and II are fully *cis*, as in digitoxigenin, illustrating the observed importance of 14β-hydroxyl substitution in CG activity (1). C/D ring junction notwithstanding, the D ring is very different between I and II, as demonstrated by the appropriate asymmetry parameters: $\Delta C_s(14) = 10.6°$, $\Delta C_2(16) = 5.1°$; $\Delta C_s(13)' = 6.2°$, $\Delta C_2(16)' = 27.5°$. The pseudorotation parameter Δ (25), which indicates overall asymmetry of the ring, has values Δ(I)= -9.7°, Δ(II)= 50.6°. The D ring of I exists primarily in the half-chair conformation with a slight tendency towards the 14-envelope; as shown in Figure 3 the D ring of I is superimposed with almost perfect registry onto the D ring of digitoxigenin. However, in the D ring of II, the 13-envelope conformation is pronounced. The torsion angles C(14)-C(15)-C(16)-C(17) and C(14)′-C(15)′-C(16)′-C(17)′, respectively -14.8(8)° and 5.3(8)°, differ most in that they switch from *-syn* to *+syn* between the two structures; other corresponding torsion angles in the D rings

also differ considerably but do not demonstrate this inversion from *-syn* to *+syn*. Endocyclic bond angles are similar in the two structures and are generally close to the average values for progestins except at the C/D ring junction– C(14)-C(13)-C(17) and C(14)′-C(13)′-C(17)′ angles of 103.9(6)° and 103.1(5)°, respectively, are 4° larger than average, as are C(17)-C(13)-C(18) at 113.1(6)° and C(17)′-C(13)′-C(18)′ at 113.8(5)°.

Which of the structures, if not both, is the active conformer is not clear. Several possibilities must be considered:

-one conformer is active, the other inert

-one is active, the other has negative effects

-both are active, to varying degrees

While two minimum energy conformations exist in the crystal state, a similar result cannot be concluded from the solution data. ^{1}H NMR studies for the title structure show a single strong peak arising from a proton involved in hydrogen bonding (9). This single hydrogen bond in solution is likely the signal from the intramolecular H-bond of molecule I, which has a long lifetime on the NMR timescale. The intermolecular O(14)···H(O14)′ bond may exist in solution, and probably does, but H(O14)′ likely exchanges too rapidly in solution for the signal to be detected. Another possibility is that the intermolecular H-bond is an artifact of crystal packing

which induces the alternative conformation of the molecule. Static energy calculations reveal that, while I is indeed a lower energy conformation than II, the difference is minimal. All energy calculations were carried out on a Silicon Graphics IRIS 4D70/GT workstation using the program *PCMODEL* (26) with default parameters.

Since there is no bulky 17α substituent, such as an acetoxy group, one would expect the 17β side chain to have almost free rotation, as suggested by an examination of Dreiding models of the title compound. This expectation is not supported by crystallographic data on other unhindered progestins, which reveal that the positionings of the side chains are generally restricted to a common range (23). The inference from the Dreiding model might have been erroneous because Dreiding models assume little spatial occupancy, and interatomic repulsions/attractions are not considered. Energy calculations of rotation about the C(17)-C(20) bond using a rigid rotor approximation (26) on conformer I suggest that the previous crystallographic results are not unreasonable; there appear to exist two considerable energy barriers that discourage free rotation about the C(17)-C(20) bond of I. The energy profile is similar to that seen for 16β-methylprogesterone (27). A large barrier of over 230 kcal mol^{-1} appears to block the counterclockwise (as viewed along C(20)-C(17)) rotation of the side chain of I to the orientation in II; in this

pathway, repulsion occurs between the hydrogens of C(21) and those of C(18) and O(14). The barrier in the clockwise rotation is much smaller, at approximately 35 kcal mol^{-1}. Even this latter energy barrier requires that the pathway between the two observed orientations include a concomitant change in the D ring (although not to the same extent exhibited in II). The actual energy values are not as important as the general trend in energy level fluctuation, since the rigid rotor model is only an approximation of the rotational behavior. Calculations of energy barriers to free rotation in an "unconstrained" molecule also show a similar trend, with values of 8.0 and 4.5 kcal mol^{-1} for the counterclockwise and clockwise rotations of the side chain, respectively. The C(16)-C(17)-C(20)-O(20) torsion angle, τ, of -46.8° (molecule I) is not exceptional, but a τ of 168.8° (molecule II) is very unusual. Calculations with the rigid rotor approximation involved rotating τ through 360° in steps of 8°; the "unconstrained" energies were obtained by fixing τ in 10° increments and minimizing the energy of the remainder of the molecule at each step (*i.e.* allowing the molecule to relax). The unusual conformation of the side chain in II would account for its absence in solution; it would appear that while in solution 14β-hydroxyprogesterone is stabilized by hydrogen bonding of the more common side chain, either crystal packing forces cause the side chain of conformer II

to adopt the uncommon orientation as the intramolecular H-bond is disrupted and the intermolecular H-bond is formed, or the unusual H-bond simply occurs very infrequently. This finding is quite contrary to earlier conclusions regarding crystal packing forces on side chain orientation in 17β pregnanes (28). It may be significant that none of those steroids studied had the 14β-OH group, whose effects appear to be as influential as those attributed to 16β-substitution (28). Other steric effects may also influence stabilization of the C(17)′ side chain. The C(18) methyl group, for example, is 3.35(1) Å from O(20), so it is unlikely that any contact has a significant effect. C(18)′, however, is only 2.94(1) Å from O(20)′, close enough for a possible C-H···O interaction (29). Unfortunately, attempts to refine hydrogen coordinates from the difference maps were unsuccessful, due to an insufficient observation to variable ratio, so actual location of the methyl hydrogens is not known.

The 14β-OH group's influence on the side chain likely plays a dominant role in the activity of the title compound. There has been a close relationship observed between the Na^+, K^+-ATPase binding of a number of digitalis-like hormones and the relative separation of the 20-carbonyl oxygen of the hormone from O(4) of digitoxigenin, even in different mammalian tissue (1,8). Activity also appears to be linked to a preferred orientation of the carbonyl oxygen, which presumably

causes correct alignment with the receptor binding site. In Figure 3 it can be seen that neither C(20)-O(20) nor C(20)'-O(20)' lies in a favorable (*i.e.* parallel) orientation with respect to C(23)-O(4) of digitoxigenin. In fact, what is interesting is that they appear almost symmetrically distributed around $C(23)\text{-}O(4)_{III}$. The relative displacements are also very similar, with an $O(3)\cdots O(4)_{III}$ distance of 2.88 Å and $O(3)'\cdots O(4)_{III}$ of 2.87 Å. This quasi-symmetry may conceivably be exploited by the cognate enzyme, which might interconvert the structures, depending on which conformation is actually responsible for the activity; conversely, as mentioned previously, rotation away from the central orientation may have a depressing effect on enzyme recognition. The displacements are surprisingly short in view of the compound's fairly low measured activity, since the observed trend has been that the shorter the separation relative to digitoxigenin, the greater the inhibition of N^+,K^+-ATPase (1). The unfavorable orientations of the carbonyl groups likely account, in part, for the unexpectedly low affinity, which provokes further question regarding the correlation between separation distance and activity. The relationship outlined to date between observed activity and displacement is a simple one, which assumes that all groups are oriented in a similar fashion (which happened to be valid for the listed compounds). However, as demonstrated by the title compound, this assumption ignores the spatial concerns;

there may well be a direct relationship between the angle offsetting the 17β side chain from the digitoxigenin group and the observed activity. Other structures possessing the title compound's unusual side chain angles would need to be considered, such as those with either the O(20) or O(14) position protected from intramolecular hydrogen bonding, or converted into more complex substituents with additional carbonyl or hydroxyl groups which could also serve as critical functional groups. Any negative effects of poor orientation of the 17β side chain might perhaps be mitigated by increasing the outward displacement (relative to the main body of the steroid) of the keto group, the effect of which has been seen to be increased enzyme inhibition (8).

II.5 Summary

The crystal structure, which shows an intriguing combination of favorable and unfavorable aspects, cannot entirely account for the activity of the title compound. Interesting observations and questions arise from the structural effect of 14β-hydroxyl substitution; while necessary for CG activity, such substitution also induces a unique orientation in the 17β side chain that may inhibit binding. Strong conclusions are not possible, as it is not clear which minimum energy conformation is the biologically active conformation. As a relatively unusual

species in terms of features and activity, 14β-hydroxyprogesterone encourages further work toward characterizing an endogenous digitalis-like hormone, in a variety of similarly substituted progestins.

III. (2′-Deoxy-β-D-*arabino*-hexopyranosyl)thymine

III.1 Introduction

Native deoxyribonucleic acid (DNA) found in eukaryotes is formed of double-helical associations of nucleotide chains. The basic components of the nucleotide are a 2′ dehydroxylated ribose, a phosphate moiety and a pyrimidine or purine base. Ribonucleic acid (RNA) is similar to DNA, except that it generally occurs in single stranded forms, uses a different pyrimidine, and incorporates a ribose molecule. In natural systems there is little variation in the chemical structure of the nucleotides, and most of those modified bases which do occur are found in transfer RNA (30).

Much cellular activity, including cell growth and replication, involves the recognition of the DNA strand(s) by cognate proteins. One area in which recognition plays a tremendous role is in the area of cell proliferation. Under normal conditions, cells are produced at an adequate rate to replace dead and damaged ones, and the production is naturally stopped when necessary; sometimes, however, the usual sequence of genesis and death goes awry, as in cancer, and the cells proliferate unchecked. Cancerous growths are attributable to a loss in the stop signal; various treatments involve providing a new stop signal at some point in the complex chain of cell proliferation.

One field of study addresses the protein/DNA recognition step and the subsequent gene expression. As discussed later in the section on DNA, a protein may interact with a particular DNA region as a result of recognizing a particular sequence, electronic pattern, and other reasons. Interfering with the interaction, whether for positive effect or negative, may be a viable means of affecting gene expression or other cellular activity.

Antisense oligonucleotides are short synthetic strands of DNA whose sequences complement the target sequences that are read by the proteins during transcription (the sense strands). Since a particular sequence of 15 bases or longer (for example, a gene coding site) statistically occurs once in the human genome, antisense oligonucleotides can theoretically be finely tuned to attach to a desired site, and indeed have been observed to do so (31). By binding to a target sequence, an antisense oligonucleotide may either inhibit or permit gene expression, by preventing transcription from occurring. Designed to circumvent the restrictions of identification, antisense oligonucleotides thus have several specific requirements. In order to be effective it is necessary that they: be sufficiently similar to the host DNA so as to escape the degradative nucleases in the extranuclear material; be able to be transported across the nuclear membrane; not accumulate in any particular site or organ; be sufficiently

foreign in nature to avoid digestion by specific intranuclear nucleases and to prolong their effective lifespan. Both unmodified and modified oligonucleotides have been assayed; the properties and mechanisms of antisense oligonucleotides have been reviewed at length (32).

Another means of countering the cell growth observed in conditions such as cancer is a more traditional chemotherapeutic method: incorporation of a modified nucleotide in the chain lengthening step of DNA translation. Chain termination is achieved since the modified nucleoside does not permit further extension. The anti-HIV drug AZT (33) is used in this manner.

Most modifications to nucleosides have occurred on the bases, which are intimately involved in hydrogen bonding to stabilize the DNA duplex (34,35). However, there is also the possibility that modifications to the sugar ring itself may affect recognition. One such family of modified sugars encompasses the 2′-deoxyhexopyranose nucleosides. One member of the family, 2′-(deoxy-β-D-*arabino*-hexopyranosyl)thymine, has been observed to be a strong inhibitor, both *in vivo* and *in vitro*, of uridine-deoxyuridine phosphorylase (36-39). The β-anomer was observed to be particularly effective. This modified nucleoside is a potential antisense oligonucleotide candidate, since it fulfils at least some of the criteria. As a six-carbon sugar, the nucleoside is larger, with either a

more hydrophobic backbone, or more hydrophilic, depending upon the extent of hydroxylation (note that the hydroxylated forms are analogues for RNA as well). Preliminary molecular mechanics studies have suggested that it is possible for the molecule to be incorporated into a helix, and a pyranosyl-based oligonucleotide may form a minihelix (40).

III.2 Nomenclature

The nomenclature of hexopyranose sugars deserves mention. Although based on IUPAC recommendations (41), it should be noted that not all authors follow the recommendations, and a compound may be listed in the literature under a variety of names.

"(2′-deoxy-β-D-*arabino*-hexopyranosyl)thymine" (Figure 4) describes Figure 4 as the sum of several parts. "Hexopyranosyl" indicates the molecule possesses a six carbon sugar moiety, including the primary hydroxyl group; this is in comparison to a "pyranosyl" sugar, as diagrammed in Figure 5.

There is no hydroxyl group on the 2′ position, as there is in the canonical glucose molecule, hence the prefix "2′-deoxy". The "arabino" prefix describes the configuration of the hydroxyl groups as compared to the four-, five-, and

six- carbon sugars; the comparison ignores any atoms without hydroxyl groups.

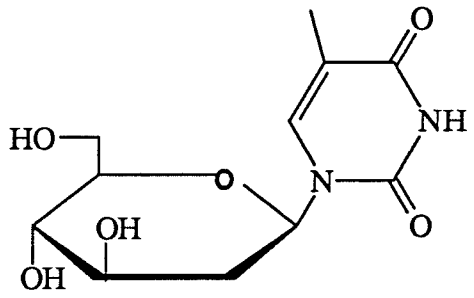
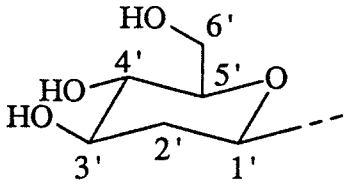
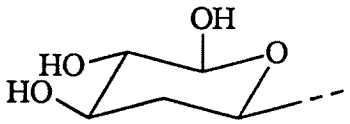


Figure 4.Structure of (2′-deoxy-β-D-*arabino*-hexopyranosyl)thymine



hexopyranosyl



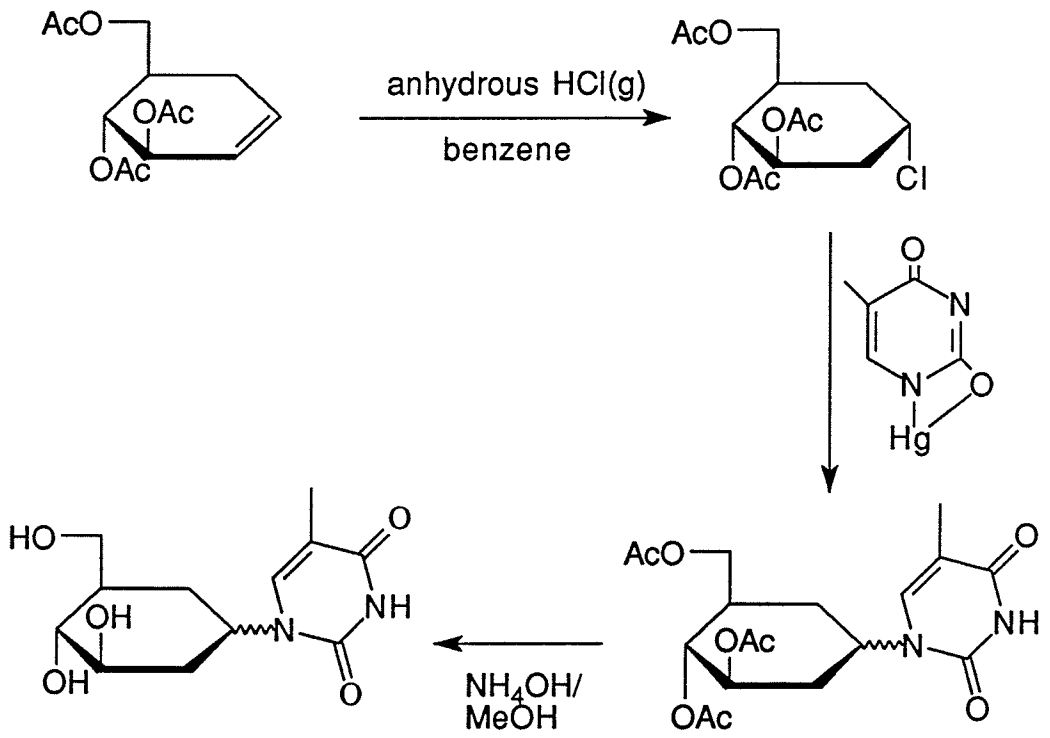
pyranosyl

Figure 5. Structural comparison of hexo-pyranosyl and pyranosyl sugars, showing numbering.

III.3 Experimental

The title compound was synthesized by two different methods: the "mercuri method" used in a previous synthesis of the compound (42), with several modifications to the protocol, and a sugar-phosphorodithioate/thymine coupling using strong Lewis and protic acid catalysts (43-45).

1. "Mercuri" method



Scheme 1. "Mercuri" Method

a) Synthesis of monothyminylmercury: 12.6 g (0.1 mol) thymine in 50 mL acetic anhydride and 1 mL pyridine was refluxed, with stirring, in a heating mantle. The thymine dissolved readily in the warm solution, resulting in a brownish-white, cloudy solution that was refluxed for 15 minutes, during which time a white suspension developed in the brown solution. After cooling, the white crystals of acetylthymine were filtered in a Buchner funnel, and washed several times with cold acetic anhydride. The yield was 14.1 g (85%).

6.37 g (0.02 mol) mercuric acetate was dissolved in 200 mL warm methanol. 3.36 g (0.02 mol) acetylthymine was stirred into the solution, and refluxed for 2.5 hours before being allowed to cool overnight. The white powdery material was suction filtered repeatedly in a Buchner funnel, over the same filter paper (Whatman no. 1), until the filtrate ran clear. The precipitate was washed with a small amount of cold methanol; the precipitate and filter paper were placed on a watch glass and dried at 85°C overnight. The caked white monothyminylmercury was removed from the paper the following day; the yield was 6.1 g (quantitative).

5.4 g (0.02 mol) 3,4,6-tri-O-acetyl-D-glucal was dissolved in 20 mL anhydrous benzene in a 100 mL round bottom flask placed in a pan of ice. The flask was fitted with glass inlet

and outlet tubes set in a rubber stopper. Dry HCl was bubbled through at a fairly rapid rate for 20 min, after which the amount of gas was reduced to a slow rate. The HCl gas was generated by dropwise addition of H_2SO_4 to NaCl; the gas evolved was then bubbled through an H_2SO_4 trap to remove any moisture, before being fed through glass tubing into the reaction vessel. After reaction the benzene was removed under reduced pressure and the brownish syrup coevaporated with 25 mL dry benzene three more times, before finally being redissolved in 10 mL of dry benzene.

2.25g (7 mmol) of monothyminylmercury was dried in a 50 mL flask by adding and then removing 20 mL of N,N-DMF and 10 mL dry toluene by distillation. A drying tube filled with silica gel was fitted to the top of the flask, and Parafilm tape sealed the junctions between glass and rubber. After cooling the solution was rapidly stirred as the benzene-halide solution was added in one portion (exposure to the atmosphere minimized). The solution was stirred for 4 hours, then decanted into a 250 mL separatory funnel. 40 mL chloroform and 70 mL of 15% aqueous KI were added. The yellow aqueous layer was discarded. The chloroform layer was dried with $MgSO_4$ and evaporated to dryness under reduced pressure. The material remaining was then extracted with two 25 mL portions of hot cyclohexane, which were discarded; a final 25 mL portion of hot cyclohexane was added, and removed under reduced pressure.

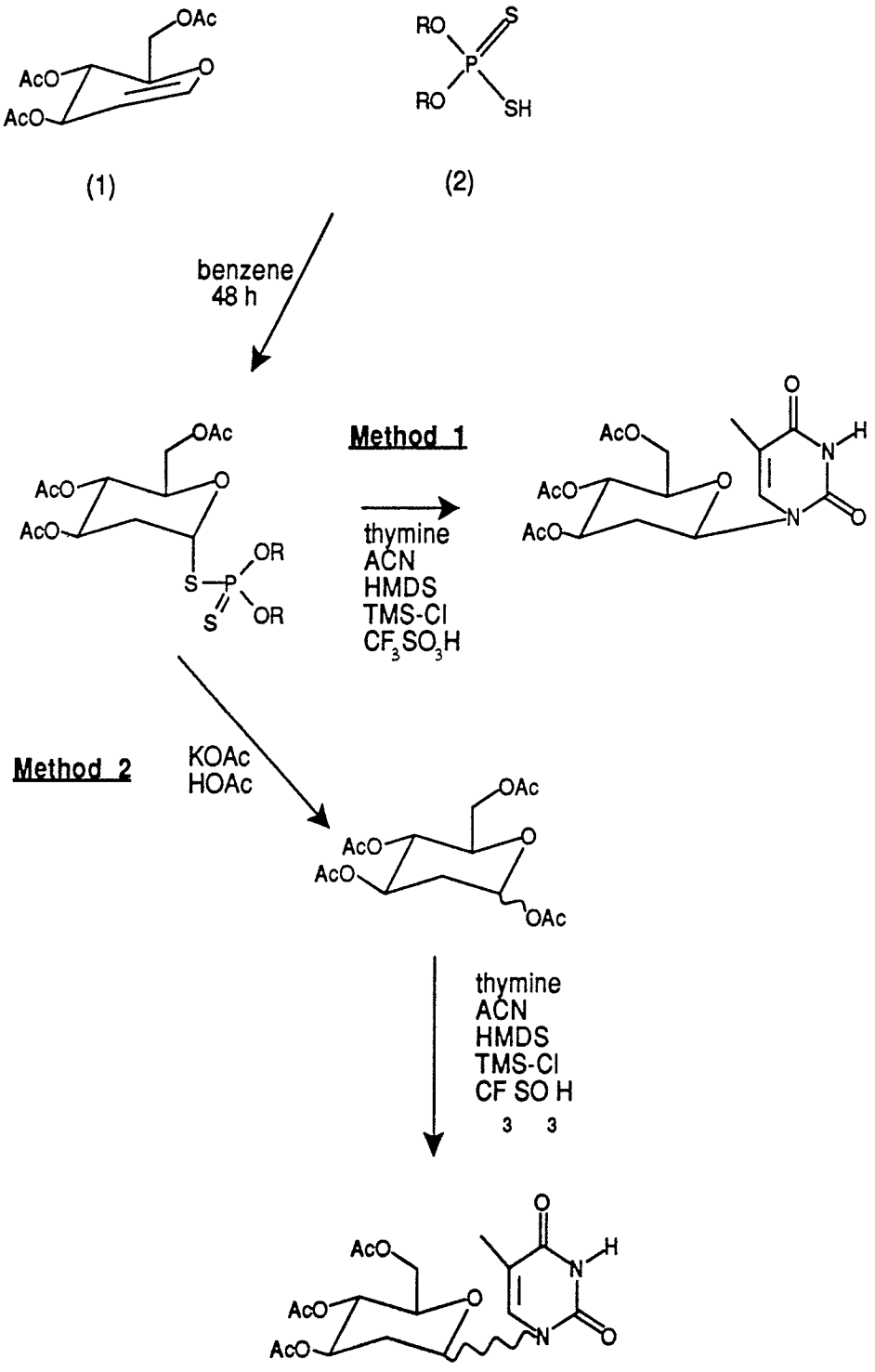
Examination by TLC (ethyl acetate: hexane, 7:3) revealed that the syrupy residue was composed of several species. The yield of the desired β anomer was estimated at < 10%.

The acetylated nucleoside mixture was deacetylated by treatment overnight with a 15% NH_4OH in methanol. The solvent was removed under reduced pressure.

2. Phosphorodithioate coupling

Several different experiments using this technique were done, in collaboration with Dr. L. Diop of the University of Manitoba (St. Boniface College), to investigate the influence of reagents and solvents. It had been observed that varying the polarity of the solvent had a strong effect on the success of the reaction (44,46) The effects of two highly polar solvents with different properties, acetonitrile and acetic acid, were examined. Two different catalysts, trifluoromethane sulfonic acid (triflic acid) and $SnCl_4$,were also used. Trifluoromethane sulfonic acid is a strong protic acid, whereas $SnCl_4$ is a particularly strong Lewis acid.

0.315 g (2.5 mmol) thymine and 1.075 g (2.5 mmol) 0,0-dimethyl-S-(3,4,6-tri-O-acetyl-2′-deoxy-α-D-*arabino*-pyranosyl)phophorodithioate were dissolved in 25 mL dry acetonitrile. To this solution were added, in order, 0.575 mL (2.75 mmol) hexamethyldisilazane and 0.45g (3 mmol)



Scheme 2. Phosphorodithioate coupling

trifluoromethane sulfonic acid. A balloon of dry argon gas was attached to the flask in place of a drying tube, to exclude moisture from the system, and the solution was refluxed for 24 hours. After the reaction was determined complete by thin layer chromatography (TLC), 33 mL methylene chloride were added, and solution transferred to a separatory funnel. The organic layer was extracted with two 50 mL portions of saturated $NaHCO_3$ solution, then dried over $MgSO_4$, before the solvent was removed under reduced pressure. Examination by TLC (ethyl acetate: hexane, 7:3) revealed that the syrupy residue was composed of several species, similar but not identical to those from the previous synthesis. The yield of the desired β anomer was estimated at < 10%.

The experiment was repeated using $SnCl_4$ in place of the trifluoromethane sulfonic acid. Examination by TLC revealed that the syrupy residue was again composed of several species, but one appeared to be in higher yield than the others. The yield of the desired β anomer was estimated at < 10%.

O,O-Dimethyl-S-(3,4,6-tri-O-acetyl-2′-deoxy-α-D-*arabino*-pyranosyl)phophorodithioate was reacted with 1.5 equivalents of potassium acetate in 50 mL glacial acetic acid, at room temperature. However, since the resulting syrup was composed of numerous species, this avenue of synthesis was not pursued.

III.4 Discussion

The desired product was obtained from both reactions, as determined by mass spectrometric (MS) and nuclear magnetic resonance (NMR) analysis, but it is difficult to conclude that either the "mercuri" method or the phosphorodithioate coupling is appreciably better than the other.

The "mercuri" method preceded the other, and, of the various means attempted, works best at generating the desired (2'-deoxy-β-D-*arabino*-hexopyranosyl)thymine. However, the method does have some notable disadvantages. It is not a particularly stereoselective method (actually designed for the synthesis of anomeric pairs), and many side reactions occur. Although (2'-deoxy-β-D-*arabino*-hexopyranosyl)thymine is a major product, it is still in very poor yield because of the ten or more side products formed, as shown in an analytical high performance liquid chromatography run (Figure 6). The reaction poses some health problems in that it generates mercury salts. Furthermore, the system must be opened to the atmosphere fairly frequently, increasing the chances of accidentally introducing moisture; the coupling of the monothyminylmercury and the glucosyl halide shown in Scheme 1 will not occur if the reaction is not conducted in anhydrous conditions, since H_2O will readily displace the Cl group.

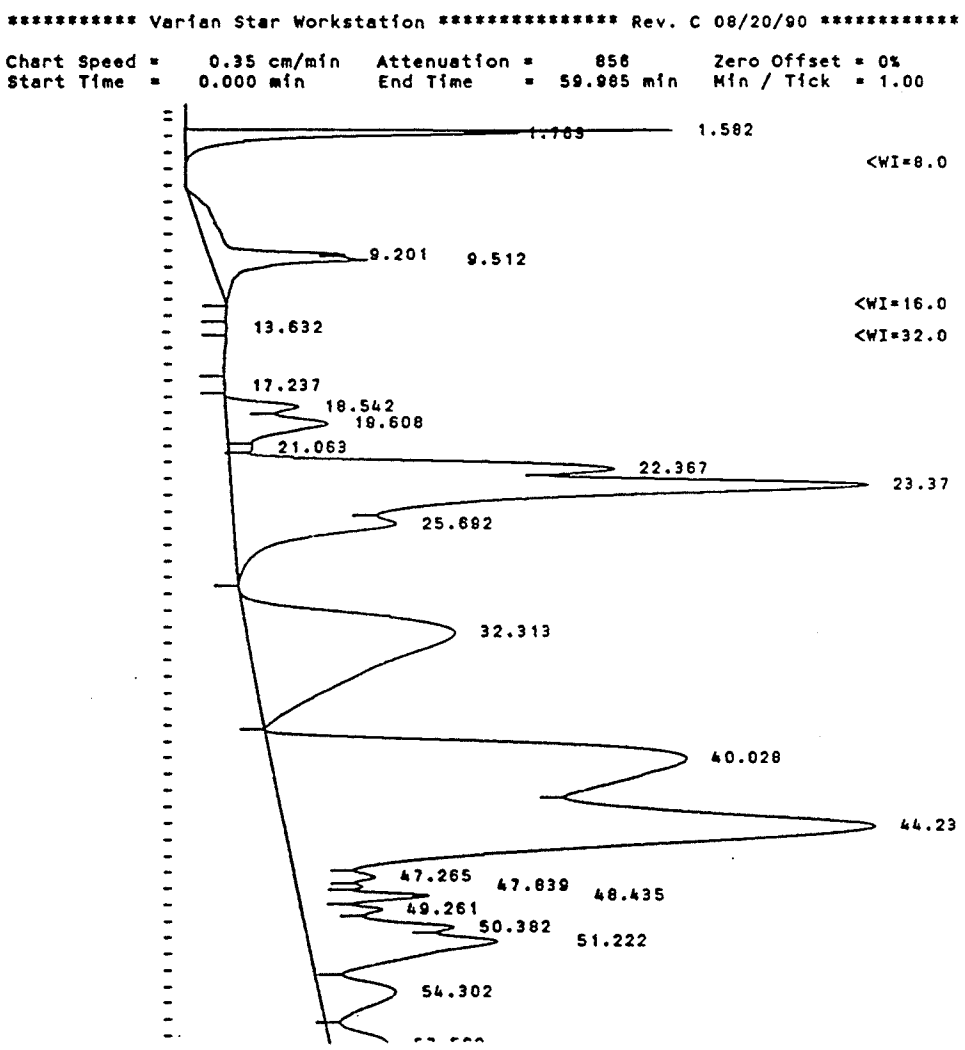


Figure 6. Analytical HPLC chromatogram of crude material from 'mercuri' method.

The advantage of the phophorodithioate method is that it is a "one pot" synthesis; there are fewer steps in which water may inadvertently enter the system and poison the reaction, further reducing an already low yield. The reaction is designed for stereoselective synthesis, is rapid and easily monitored. However, the reaction is surprisingly unspecific in this instance; surprising, in that both $SnCl_4$ and trifluoromethane sulfonic acid have been used quite successfully in synthesizing several nucleosides (44). The better yield of desired compound using trifluoromethane sulfonic acid acid rather than $SnCl_4$ is due to the former's weaker Lewis acid nature. There is a considerable σ-complex formation in the silylated base in the $SnCl_4$ reaction, that causes increased production of unwanted by-products such as N-3 nucleosides (44).

Part of the reason for the low success of the Lewis acid catalyzed method lies in the lack of a 2′ functional group in (2′-deoxy-β-D-*arabino*-hexopyranosyl)thymine. Without a strong directing group, the 1′ position is not particularly electrophilic (hence the use of strong acids). The phosphorodithioate group serves as a labile leaving group and is also intended for cyclization to a sulfur-based analogue of the acyloxonium salt, a general diagram of which is given in Figure 7, which would direct attack to the β side.

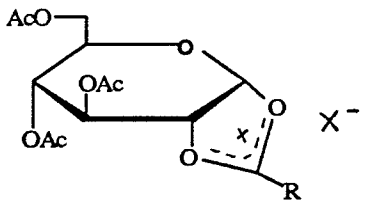


Figure 7. Structure of a protected hexopyranosyl acyloxonium salt.

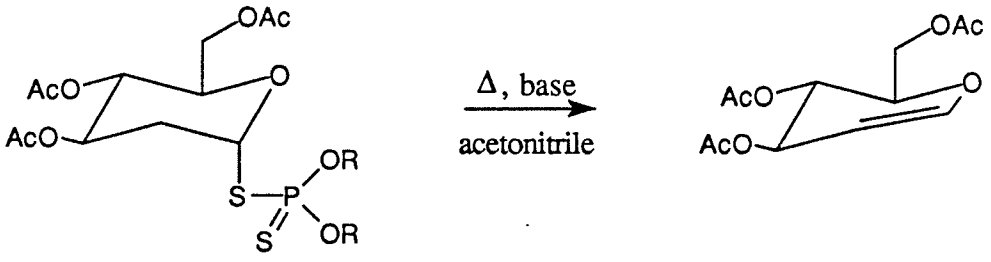


Figure 8. Regeneration of 3,4,6-tri-O-acetylglucal as a side reaction.

The polarity of the solvent plays a considerable role in the success of the reaction. No nucleoside was observed after 24 hours in 1,2-dichloroethane (46). However, acetonitrile, a polar nucleophilic solvent which competes with the silylated base, works well; interestingly, trimethylsilylchloride alone, the usual silylating agent, does not work in this reaction, hence the addition of hexamethyldisilazane. Reaction in acetic acid was attempted to see if an even more polar solvent (albeit protic) would have an effect, but the preliminary step resulted in a confusion of products in too low a yield to be of use in further steps.

Although acetonitrile was necessary in synthesizing the desired compound, the solvent was also likely involved in the concomitant production, in considerable amounts, of both anomers of (2′,3′-dideoxy-β-D-*erythro*-hex-2′-enopyranosyl)-thymine. The latter was in fact the primary reaction product in the reaction using $SnCl_4$ as the catalyst, and a significant secondary product in the reaction with trifluoromethane sulfonic acid. Acetonitrile in the presence of base was observed to have regenerated 3,4,6-tri-O-acetylglucal (43) (Figure 8). 3,4,6-tri-O-acetylglucal was probably produced in considerable quantity in both the $SnCl_4$- and triflic acid-catalyzed reactions, from which (2′,3′-dideoxy-β-D-*erythro*-hex-2′-enopyranosyl)-thymine was subsequently formed. This side production was not seen appreciably in the "mercuri" method.

The title compound was very sensitive to depyrimidination in the presence of heat and methanol. A purified sample (see Appendix A for spectra and assignments), which was dissolved and coevaporated with methanol in a warm water bath, developed an impurity that was discovered by TLC and NMR to be thymine. Elution with methanol on a silica gel column also caused a considerable amount of thymine to be generated.

It should be noted that the (2′-deoxy-β-D-*arabino*-hexopyranosyl)thymine was indirectly identified by TLC, MS, and NMR. Assignment of the NMR peaks and determination of the parent ion could yield a tentative identification of the compound, but not which anomer it was with any great degree of confidence. Deduction from TLC results was used. One interesting observation was the behavior of acetylated (2′-deoxy-β-D-*arabino*-hexopyranosyl)thymine and that of the acetylated (2′,3′-dideoxy-β-D-*erythro*-hex-2′-enopyranosyl)-thymine in an 70:30 ethyl acetate:hexane system on silica gel. Both α and β anomers appeared to run with almost identical R_f values as their counterparts ($R_f \approx 0.4$ for the β anomer, 0.2 for the α anomer) This was not known until the "pure" β collected turned out by NMR to be composed of about 60%:40% *erythro*:*arabino*, for the phophorodithioate reaction; the result from the "mercuri" method was about 40%:60% *erythro*:*arabino*. It was not until the TLC plates were run in ether that the two "clean" spots further split into four. The

α and β anomers of (2′,3′-dideoxy-β-D-*erythro*-hex-2′-enopyranosyl)thymine were available for a comparison run, and it was found that each ran along the plate with a companion spot. The similar behavior led to the tentative conclusion that the two β anomers ran together.

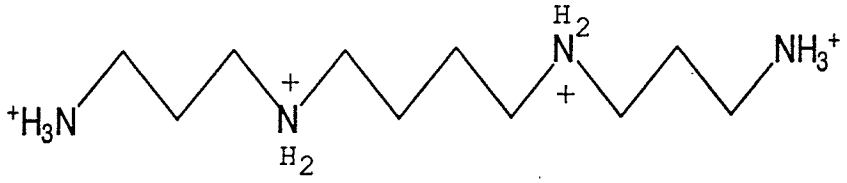
III.5 Summary

The synthesis of (2′-deoxy-β-D-*arabino*-hexopyranosyl)thymine, while successful, suffered from a number of side reactions that dramatically reduced the yield. Acetonitrile was found to be the solvent with the best properties for synthesizing the desired compound, but ironically was also the likely culprit in indirectly producing the major side reaction.

IV. Spermine phosphate hexahydrate

IV.1 Introduction

Spermine is a member of a group of molecules known as polyamines. The group is characterized by the symmetric distribution of amino groups along an aliphatic carbon chain. Several of the shorter chain members of the group are noted for their foul odor.



spermine

$H_2N-(CH_2)_5-NH_2$ cadaverine

Figure 9. Molecular structures of spermine (fully protonated form) and cadaverine.

Spermine and related polyamines occur naturally within cells at millimolar concentrations, especially in mammalian pancreas and prostate. The amounts present have been observed to increase substantially at certain stages of meiosis and have been implicated in DNA regulation during these critical moments of cell growth and proliferation. Since DNA aggregation occurs at these cell stages, spermine may act, in some as yet not fully known manner, as a stabilizing agent to enhance the condensation of the macromolecular strand (47). The molecule's influence upon DNA has been observed in both solution and crystal (48-50). All three of the most common forms of DNA-

B, Z, and A- have demonstrated examples in which spermine is involved in stabilization of the DNA in the crystal (51-53); there have also been observations made that the addition of spermine appeared necessary for the growth of a number of A-type DNA crystals, although none was to be found within the crystal lattice itself (54,55).

A structurally flexible molecule, spermine has been observed in various forms in associations with DNA in crystal (51-53). There are numerous ways in which the molecule may interact with the DNA duplex, of which several have been examined with molecular mechanics (56). One potentially attractive site is the sugar-phosphate backbone, with its numerous negative charges exposed to the solvent in which the spermine is dissolved. Although molecular mechanics studies have indicated that spermine interaction purely along the phosphate backbone is unfavorable, the terminal NH_3^+ groups appear strongly attracted to the phosphate oxygens (56). Preferred mode of binding appears to be within a groove, with the terminal groups hydrogen bonded to phosphates (57,58).

A high resolution examination of the interaction of spermine and phosphate groups is not easily achieved in the macromolecular crystal, so a small molecule analogy, spermine phosphate hexahydrate, is considered. The crystal structure of spermine phosphate hexahydrate was first studied in 1965 (59). However, fine detail could not be determined due to the resolution limits of the data; no hydrogens were refined in the final structure, which had an *R* factor of 12.7%. One feature of interest, which was discussed but not actually studied, was the nature of the hydration of the molecule. Water molecules frequently form fairly well ordered "shells" or "spines" of hydration about a macromolecule; it is of interest to determine whether water molecules would order themselves similarly about the spermine and phosphate groups, perhaps imitating what occurs during spermine-DNA interaction. In addition, a higher resolution structure would provide more precise spermine parameters for use in analyses of macromolecule:spermine complexes.

IV.2 Experimental

The crystals of spermine phosphate hexahydrate were harvested from dips (*i.e.* crystallization droplets) originally intended for the growth of oligomeric DNA crystals. Spermine, DNA, and $MgCl_2$ solutions of various concentrations had been

placed in hanging drop vapor diffusion wells at room temperature (25°C). Phosphate contamination arose from the inadequate desalting of the DNA, which had been purified by ion-exchange chromatography with a sodium phosphate buffer. Discrete crystals of what was ultimately determined to be spermine phosphate hexahydrate grew rapidly within hours in the majority of the wells, reaching a large size within a few days. Two major habits were observed: a long form with poorly defined faces, gently rounded on one side and with very thin edges, and roughly squared-off needles, elongated along one axis, with well-defined faces at all ends. The very rapid, clean growth, in a wide range of conditions, fostered the suspicion that the crystals were in fact not of macromolecular origin, a belief supported by the observation that the crystals showed considerable resistance to the edge of a sharp scalpel. The crystals broke cleanly under force, whereas macromolecular crystals, noted for their fragility, would have parted much more readily, typically yielding a shower of crystallites from the sectioning. Crystal deformation is also potentially a problem when macromolecular crystals are being manipulated, but the robust crystals observed withstood examination easily. Since the identity and solvent content of the crystals were uncertain, a clear, tabular crystal of the poorly defined habit, with dimensions 0.160 X 0.090 X 0.400 mm^3 , was mounted

in a glass capillary with a droplet of mother liquor, the long axis of the crystal colinear to the axis of the capillary. The capillary was then sealed with epoxy resin.

Data were collected on a Rigaku AFC6S sealed tube, automated diffractometer, equipped with a Cu X-ray tube (graphite monochromated Cu Kα radiation), operated at 50 kV and 40 mA.

Crystal data: $C_{10}H_{44}O_{14}N_4P_2$; M_r =506.42, monoclinic, $P2_1/c$, a=6.898(2), b=23.217(4), c=7.934(1) Å, β=113.47(1)°, V=1165.47(1) Å^3, Z=2, D_c= 1.44 g cm^{-3}, λ(Cu Kα)=1.54178 Å, μ(Cu Kα)=11.66 cm^{-1}, T=296 K, F(000)=274 electrons.

Twenty-five reflections, collected by an ω–2θ scan in the range $22.7 \leq 2\theta \leq 40.3°$, were centered individually and used for indexing the unit cell and obtaining the cell constants. The systematic absences

h0l: l≠2n

0k0: k≠2n

were consistent with the space group $P2_1/a$ that had been assigned by the previous investigators. However, $P2_1/c$ was selected by convention. Data were collected in the unique quadrant ($0 \leq h \leq 7$, $0 \leq k \leq 25$, $-8 \leq l \leq 8$) to a maximum 2θ of 100.1°. The scan rate was 16°/min in ω, with 3 rescans of

weak reflections ($I<10.0\sigma(I)$).One-third of the count time was spent counting the background. The scan width was determined by $(1.31 + 0.30\tan\theta)°$. Three strong reflections were selected as standards, and rescanned after every 150 reflections; after collection time they showed little variation, with a mean intensity of 99%, so no decay correction was applied. A total of 981 reflections were measured, of which 857 were unique. All reflections were corrected for Lorentz polarization. The linear absorption coefficient of Cu Kα radiation was not negligible (μ=11.66 cm^{-1}) despite the fairly lightly scattering (*i.e.* low atomic mass) atoms present in the sample, and the grossly uneven shape of the crystal led to a transmission factor range of 0.7-1.0, so an empirical absorption correction was applied.

IV.4 Structure Solution and Refinement

A literature search of unit cell dimensions revealed the identity of the unknown crystal as spermine phosphate hexahydrate. The data were processed using the TEXSAN crystallographic software package (60) running on a VAXstation 3520. Direct methods using the program MITHRIL was attempted first. A model molecule of spermine phosphate hexahydrate defined by the coordinates from the previous study was used to provide a more accurate scattering group in the orientation process. The phosphate group was clearly visible in the E-map

subsequently generated by MITHRIL; the coordinates of the observed group and correct atom assignments could have been input for a least-squares analysis, but the previously published coordinates (adjusted for the change from $P2_1/a$ to $P2_1/c$) were simply refined instead.

The coordinates were entered through the program ATOMED (60), which displayed an E-map of the largest 25 peaks. The phosphate group was clearly visible, and the phosphorus and oxygen atoms were assigned first; thermal factors were assigned the default program values. Several cycles of full-matrix least-squares refinement of the coordinates were run until the ratio of the shifts in the parameters to their errors was less than unity. A difference Fourier map, in which the contribution of the assigned atoms' electron density is removed from the Fourier map, was then generated. After the initial difference Fourier map, several of the carbon and nitrogen atoms were assigned. Refinement and generation of difference Fourier maps were repeated, until all non-hydrogen atoms were correctly assigned.

All non-hydrogen atoms were refined first isotropically, then anisotropically, with unit weights. No restraints were placed on the thermal factors, which were initially assigned the default values. Refinement by full matrix least squares was repeated until the shift/error ratio was less than unity. The weighting scheme was changed to use sigma weights, which more accurately considered the observed intensities of the reflections. The R factor dropped to 12%, which was similar to the final R factor reported in the previous study. The high residual error, which reflected the difference between the true structure, as indicated by the data observed, and the proposed structure, attested that the model accounted for only part of the electron density.

The hydrogen atoms were located on a further difference Fourier map, after examination of the location of the peaks of interest relative to their companion atoms. All but four of the hydrogen atoms were refined isotropically; thermal factors of the remaining four could not be controlled, so the observed positions and average thermal factors were fixed and not refined. The area around two of the water oxygen atoms was of particular interest, since there appeared to be several viable assignments; however, attempts at refining all these peaks at full occupancy failed. Four hydrogen atoms, two each surrounding two of the water oxygen atoms, were then assigned half occupancy; the two on one oxygen, along with

its full occupancy companion, were refined successfully, while those on the remaining water oxygen (two half-occupancy and one full-occupancy hydrogen) were fixed. After the final full cycle of refinement, in which the shift/error ratio was 0.009, the unweighted R factor had dropped to 2.6%, while the weighted R factor had dropped to 3.1%.

The difference Fourier map after the final cycle of refinement was essentially featureless; the highest remaining peak had a density of 0.17 $e\text{Å}^{-3}$, while the deepest trough had -0.18 $e\text{Å}^{-3}$. Final atomic coordinates and thermal factors of non-hydrogen atoms are given in Table 4.

IV.4 Discussion

Spermine's great affinity for the phosphate group could be seen in the high crystal quality and ease of crystallization, despite the heterogeneity of the crystallizing solutions. The ease of crystallization is especially notable when compared with the frustrating attempts to obtain diffraction quality

Table 4

Final positional parameters and equivalent isotropic thermal parameters for non-hydrogen atoms of spermine phosphate hexahydrate with e.s.d.'s in parentheses

atom	x	y	z	B(eq)
C(1)	-0.0114(11)	-0.0310(2)	0.9653(7)	2.6(2)
C(2)	-0.0010(10)	-0.0340(2)	0.7791(7)	2.5(2)
N(1)	-0.0061(7)	-0.0944(2)	0.7127(5)	2.1(2)
C(3)	-0.0090(10)	-0.0949(2)	0.5230(7)	2.6(2)
C(4)	-0.0305(11)	-0.1558(2)	0.4492(7)	2.8(3)
C(5)	-0.0310(10)	-0.1532(2)	0.2568(7)	2.7(3)
N(2)	-0.0648(8)	-0.2108(2)	0.1672(6)	2.3(2)
P	0.3673(2)	-0.19675(5)	-0.0008(2)	2.16(6)
O(1)	0.1339(5)	-0.2111(1)	-0.0692(4)	2.9(1)
O(2)	0.4998(5)	-0.2469(1)	-0.0197(4)	2.6(1)
O(3)	0.4002(5)	-0.1420(1)	-0.0930(4)	2.7(1)
O(4)	0.4495(6)	-0.1802(1)	0.2086(4)	2.9(2)
O(W1)	0.5468(7)	-0.0467(2)	0.1181(6)	4.3(2)
O(W2)	0.4405(6)	-0.0527(1)	0.4175(5)	5.0(2)
O(W3)	0.5957(6)	-0.1470(2)	0.6674(5)	3.5(2)

The equivalent isotropic thermal parameter is given by the expression:

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

spermine crystals (61).

In general, the structure was in good agreement with the previously determined structure. The nature of the crystallizing medium seems not to have had major effect upon the structure (unlike previously, which was conducted in a highly acidic environment, the current crystallization solution was adjusted to a pH of 7). As expected the molecule is centrosymmetric, with the center of the molecule lying on the symmetry element (Figure 10). The anisotropic thermal factors are small, indicating that the atom positions are well-defined in the crystal, and fairly symmetrical in all dimensions, indicating that none of the atoms experience a strongly directing force.

Errors in positional parameters are much smaller than those of the previous study, an order of magnitude for most atoms. The phosphorus coordinates in particular had small errors, which is often the case for heavy atoms. The largest errors occurred in the spermine carbon atoms, which might be due to small positional errors. The larger errors may also have been due to counting statistics, which may have been further improved had there been on the order of 10 rescans of weak reflections instead of the 3 rescans actually chosen.

All atoms except N(2) lie almost in the plane; the mean deviation from the least squares plane (0.0355 Å) is not substantially large. C(1) and C(2) showed considerable disparity

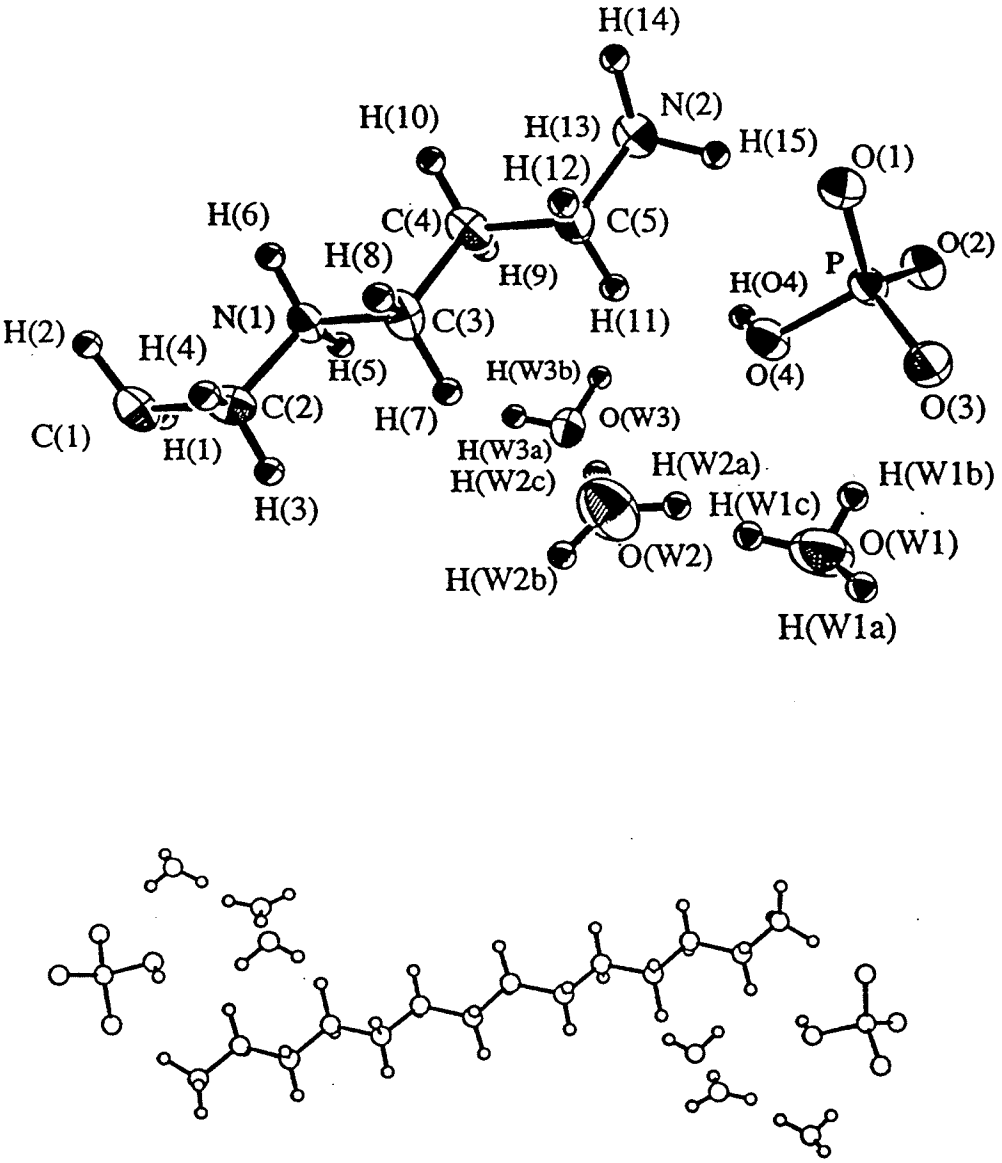


Figure 10. (top) ORTEP plot of the asymmetric unit of spermine phosphate hexahydrate (50% probability ellipsoids); (bottom) the symmetry-expanded structure.

from their counterparts in the previous study (0.03 Å and 0.04 Å, respectively) while the others were fairly consistent. As expected the least-squares analysis showed that the terminal amino groups (identical by symmetry) displayed the greatest displacement from the plane at -0.1719 Å, slightly less than the -0.19 Å previously observed. The results are given in Table 5. The displacement of N(2) is small, as indicated by the small deviations from the plane and by the almost fully staggered torsion angles (Table 6). The spermine molecule is generally extended, with normal C-C and C-N bond lengths (Table 7), and angles (Table 8), comparable to values obtained in the previous study.

Interatomic contacts via the terminal NH_3^+ moiety (both termini, due to the centrosymmetry of the molecule), are numerous; although the amino groups are not greatly displaced from the plane, they are likely able to adjust to a conformation more suitable for binding to another molecule, whereas the central secondary amino group is necessarily constrained in its potential interaction. There are several strong contacts to N(1) as well (Table 9).

Anisotropic thermal parameters of the non-hydrogen atoms, and bond lengths and angles involving hydrogen atoms are

Table 5

Least squares plane analysis of spermine in spermine phophate hexahydrate

Atom	Deviation (Å)	e.s.d. (Å)
C(1)	-0.0761	0.0068
C(2)	0.0043	0.0064
N(1)	0.0413	0.0042
C(3)	0.0349	0.0064
C(4)	-0.0313	0.0067
C(5)	-0.0250	0.0064
N(2)*	-0.1719	
mean deviation from plane:		0.0355

* denotes atom not used in calculation of least squares plane

Table 6

Torsion Angles (°) for spermine phophate hexahydrate with estimated standard deviations in parentheses

Atoms	angle(°)
C(1)*-C(1)-C(2)-N(1)	-175.6(6)
C(1)-C(2)-N(1)-C(3)	176.5(5)
C(2)-N(1)-C(3)-C(4)	-175.8(5)
N(1)-C(3)-C(4)-C(5)	-179.6(5)
C(3)-C(4)-C(5)-N(2)	-176.5(5)

Table 7

Bond lengths (Å) for non-hydrogen atoms of spermine phosphate hexahydrate with estimated standard deviations in parentheses

Bond	length(Å)
C(1)*-C(1)	1.529(9)
C(1)-C(2)	1.510(7)
C(2)-N(1)	1.494(5)
N(1)-C(3)	1.496(6)
C(3)-C(4)	1.515(6)
C(4)-C(5)	1.526(7)
C(5)-N(2)	1.488(6)

Table 8

Bond Angles (°) for spermine phosphate hexahydrate with estimated standard deviations in parentheses

Atoms	angle(°)
C(1)*-C(1)-C(2)	111.0(5)
C(1)-C(2)-N(1)	112.6(4)
C(2)-N(1)-C(3)	110.5(4)
N(1)-C(3)-C(4)	110.9(4)
C(3)-C(4)-C(5)	108.1(4)
C(4)-C(5)-N(2)	112.3(4)
O(1)-P-O(2)	113.0(2)
O(1)-P-O(3)	110.5(2)
O(1)-P-O(4)	108.4(2)
O(2)-P-O(3)	112.3(2)
O(2)-P-O(4)	108.2(2)
O(3)-P-O(4)	104.0(2)

Table 9

Selected intermolecular distances (Å) for non-hydrogen atoms of spermine phosphate hexahydrate, with estimated standard deviations in parentheses

Bond	length(Å)
N(1)-O(3)(x,y,z+1)	2.830(5)
N(1)-O(W3)(x-1,y,z)	2.894(6)
N(1)-O(1)(x,y,z+1)	3.151(5)
N(2)-O(1)(x,y,z)	2.729(5)
N(2)-O(2)(x-1,y,z)	2.894(6)
N(2)-O(1)(x,y-1,z)	2.694(5)
N(2)-O(4)(x,y,z)	3.503(6)
N(2)-O(4)(x-1,y,z)	3.566(6)

Table 10

Bond lengths (Å) for non-hydrogen phosphate atoms in spermine phophate hexahydrate with estimated standard deviations in parentheses

Bond	length (Å)
P-O(1)	1.517(3)
P-O(2)	1.524(3)
P-O(3)	1.527(3)
P-O(4)	1.575(3)

tabulated in Appendix B.

The bond lengths in the phosphate group (Table 10) are very similar to those previously determined. P-O(2) is slightly longer than expected, but it is not clear whether the difference is significant in terms of the degree of double bond character, which may be inferred from the bond length; a longer bond length may indicate that the bond possesses more single bond character, and that the hydrogen atom resides at that position more often than expected. The lone hydrogen on the phosphate group refined satisfactorily (assigned full occupancy as a reasonable default), with a discrete B factor, although its length is somewhat short. Since the O(4) hydrogen is intimately involved in the network of hydrogen bonding, the extent of the hydrogen's presence in that location influences the strength of the interaction with spermine, water molecules, and other phosphate groups.

The previous study presented some speculations on the disordering of the water layer, which could not be determined in that particular case. The proposal consisted of a zigzag arrangement of partially occupied water sites, in which the hydrogens occurred at multiple positions. The presence of disordered hydrogen atoms was supported by the current study, in which three hydrogen atoms were clearly visible on the difference Fourier map around two of the water oxygens, O(W1)

and O(W2). The hydrogen atoms of one water molecule, O(W3), were not disordered and refined at full occupancy, with bond lengths 0.83(4) Å and 0.94(6) Å and an H-O-H angle of 109(5)°. The other two molecules were disordered, and peaks selected from the difference Fourier possessed reasonable geometry. It should be noted that the disordered hydrogen atoms were assigned 50% occupancy as a average default value only to indicate that they are partially occupied; with the data available, it is not possible to refine their B factors so that the relative occupancies can be estimated.

The various molecules aggregate into well defined sheets, forming discrete organic and inorganic layers. As can be seen in Figure 11*a*, the water-phosphate channel is clearly distinct from the closely packed spermine molecules. These channels run through the entire crystal. Such particular ordering is often seen in macromolecular assemblies, in which a shell or spine of hydration is necessary to provide the free energy required for crystallization (62).

The spermine molecules associate with the phosphate groups (particularly via N(2)-H···O(1) hydrogen bonding) and the water molecules from several surrounding unit cells, forging a strongly linked network of hydrogen bonds. The hydrogen bonding is particularly strong within the water:phosphate layer. A view down the *a* axis, in which the

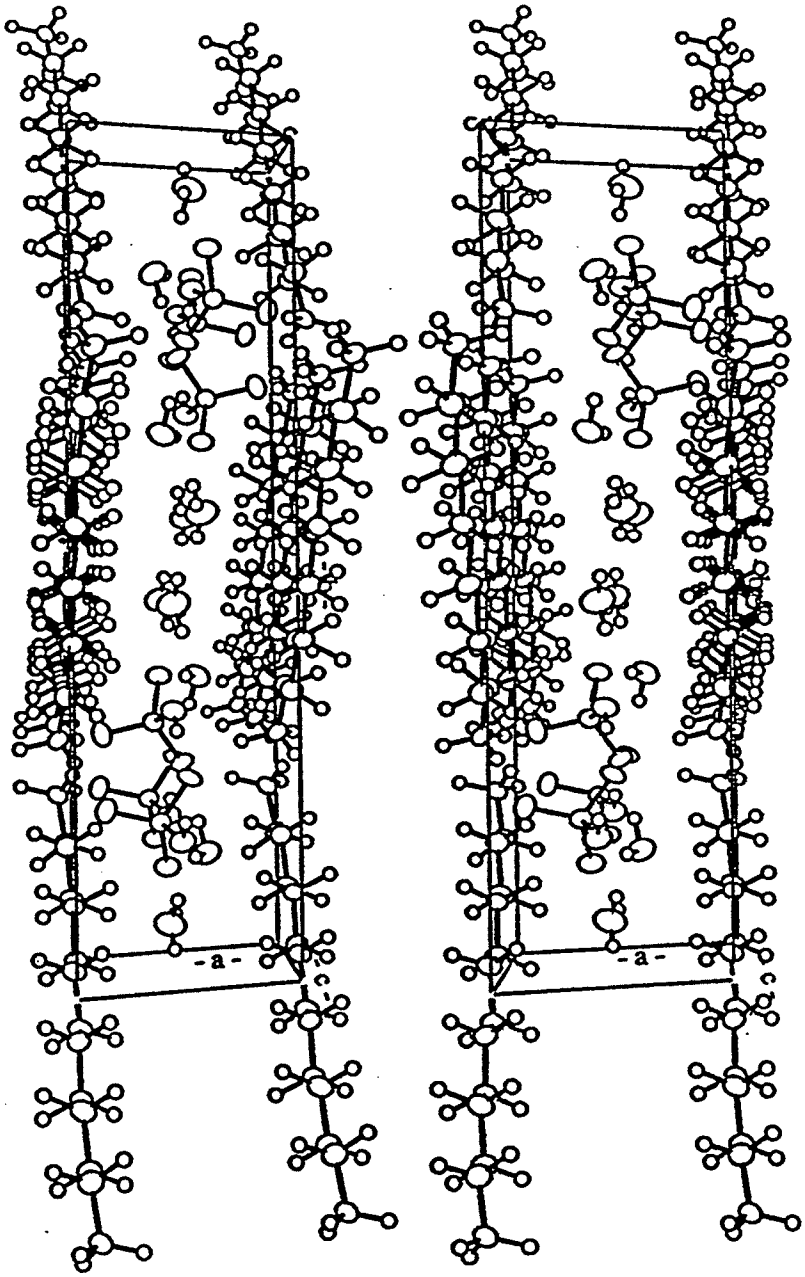


Figure 11a. Stereo packing diagram of spermine phosphate hexahydrate viewed down the *c* axis, showing spermine-inorganic layers.

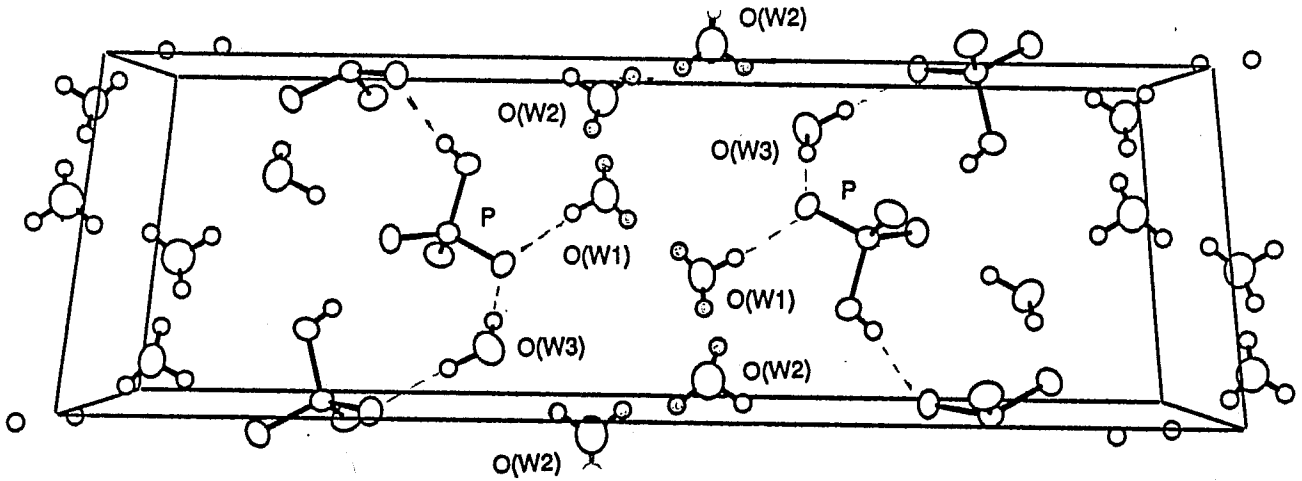


Figure 11*b*. Packing diagram of spermine phosphate hexahydrate viewed down the *a* axis, showing water-phosphate layer (spermine omitted for clarity). Disordered hydrogens shown shaded.

spermine molecules are omitted for clarity, is given in Figure 11*b*. Infinite, meandering chains are formed by the partially occupied hydrogens of O(W1) and O(W2); the fully occupied hydrogens on the two water oxygens are hydrogen bonded to the ordered water molecule O(W3) and phosphate oxygens. The juxtaposition of ordered and disordered regions is clearly seen in Figure 11*b*. H(W3a) and H(W3b) are firmly locked into position by hydrogen bonds to neighboring phosphate groups. Although the crystal is highly hydrated, the water molecules appear to be firmly incorporated into the lattice, since the test crystals were quite stable to room conditions when taken out of the mother liquor, and also after prolonged exposure to X-rays. However, some roughening of the surface of the crystal was observed when the ambient temperature soared (to approximately 30 °C) and the relative humidity dropped, indicating some loss of water from the surface, while prolonged exposure led to overall severe crystal degradation.

The refinement of the disordered hydrogens failed largely due to the limited data relative to the increasing number of parameters. In a true structure, none of the variables is actually independent, since the atoms are bonded to each other and interatomic interactions are also present. If, however, the ratio of observed unique reflections to the

number of parameters is much greater than 3, the correlation of the parameters may be largely ignored and the refinement will not suffer. If the ratio approaches 3 or drops below that value, the correlation coefficient becomes significant, the variables can no longer be considered to be independent, and the refinement can oscillate between two solutions or converge to a false solution (17). Thus it is important in the final stages of a structure refinement to introduce parameters slowly, since refining too many at once can also cause the refinement to converge to a local minimum, yielding an inaccurate or even false structure. With 687 parameters and 220 variables, the critical ratio of observed unique reflections to the number of parameters is very close to being violated. The smoothness of the final difference Fourier map also meant that the residual data was little better than background noise, for which it was futile to attempt refinement of partially occupied, very lightly scattering hydrogens.

IV.5 Summary

This most recent structure of spermine phosphate hexahydrate closely resembles that of the previous study. However, errors in positional parameters are an order of magnitude smaller, and the final R factor of 2.6% indicates good agreement between the structure and the data. One P-O bond length

differs by a slight amount, suggesting that that bond possibly possesses more single bond character. Very little disorder is demonstrated by the thermal factors of the spermine and the phosphate group, which are small and nearly symmetrical; the atoms are firmly positioned within the lattice, in a well-defined low-energy state. The details of the water layer show that the crystal contents are well ordered, forming clearly defined sheets. Two of the three unique waters are disordered, although the hydrogen atoms could be seen on a difference Fourier map; the disordered hydrogens form continuous water chains which link well ordered water:phosphate regions.

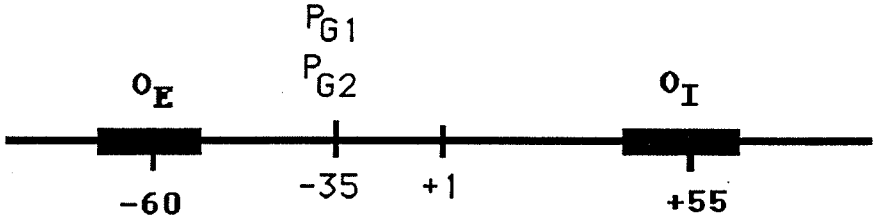
V. DNA Octamer $^{5'}$GATTCCAC$^{3'}$/ $^{5'}$GTGGAATC$^{3'}$

V.1 Introduction

Protein-DNA interactions may occur randomly along the DNA strand, and some proteins appear to bind nonspecifically (63) but many other proteins will only recognize and bind at a specific regulatory site. The process by which relevant segments of the DNA chain are recognized by the appropriate proteins is not fully understood. The signal may be electrostatic, with recognition sites characterized by specific sequence, stretches of electron-rich sites, or the signal may be conformational, in which the folding of the protein is necessary for active sites to be in the proper proximal arrangement (64). Even minute changes in the environment— pH, salt concentrations, temperature— which induce apparently minor shifts, may render a biological macromolecule inactive (65). Due to the many degrees of freedom possible for a macromolecule in solution, it is very convenient to examine the region of interest of the macromolecule by a technique such as x-ray crystallography. The question arises as to the relevance of the solid state conformation relative to that of a molecule that is active in solution, but the argument may be countered by the observations that DNA crystals are highly hydrated, mimicking a solution state, and that crystallized macromolecules can exhibit similar biological activity as in

solution (66).

The DNA regulatory region of interest in the present study is the operator of the *gal* operon, which is the operon responsible for the regulation of galactose. The operon (67) is a useful model for describing the conglomeration of genes and recognition sites which are dedicated to the same biological process. Although not as well known as the *lac* operon (68), the structure and mode of operation of the *gal* operon has been characterized (69,70). Two operator regions are responsible for galactose regulation, (Figure 12) and the importance of these sites has been shown by repressor binding studies (71) and mutation studies (72).



PG1, PG2: promoters

O_E, O_I : operators

Figure 12. Schematic drawing of *gal* operon indicating relative positions of operators.

The sequence studied was an octamer, $^{5'}GATTCCAC^{3'}$ / $^{5'}GTGGAATC^{3}$

(hereafter referred to as GATTCCAC), which is located at positions -53 to -60 of the exterior operator (O_E) of the *E. coli gal* operon (numbering is upstream of the P1 transcription start). This particular region was chosen for study because of the presence of the trimer sequence $^{5'}GTG^{3'}/^{5'}CAC^{3}$. DNA binding proteins have been observed to occur in distinct families, characterized by the manner in which they are believed to recognize binding sites. The sequence GTG/CAC (hereafter denoted as GTG) is of particular interest because of its unusually frequent appearance in sites which are known to be involved in recognition; the trimer has been observed at a frequency almost five times higher than would be predicted statistically for a random occurrence (73). Depending upon which model is considered, recognition criteria may incorporate patterns of electron density or involve steric interactions; for example, the GTG region may possess characteristic helical perturbations. Although potential long range effects cannot be determined in the X-ray study of an octamer, the local effects of the GTG trimer may be characterized. While the crystal structure of GATTCCAC has not yet been determined, due to poor crystal quality, the determination of the unit cell yields some useful information based on previously determined DNA crystal structures.

V.2 Purification of DNA

Experimental. 2 octamers (8-base pair (bp) oligomers) of synthesized DNA, $^{5'}$GATTCCAC$^{3'}$ and $^{5'}$GTGGAATC$^{3'}$, were purchased from the Regional DNA Synthesis Lab at the University of Calgary. The DNA, in the form of a sodium salt, arrived detritylated and deprotected. The material was partially purified, being a generally whitish powder, but with obvious traces of impurities as indicated by the faintly yellowish cast to the DNA.

The DNA was suspended in a small aliquot of doubly distilled, deionized water in an Eppendorf tube. The volume was restricted to the minimum required for complete dissolution; while free-acid DNA is highly soluble, the presence of a salt appreciably decreased the solubility. For both strands no more than 2 mLs of water was required. Since DNA can bind irreversibly to glass and paper, the solution was not filtered but was centrifuged in a high-speed centrifuge for 20 minutes. The pellet was discarded and the supernatant examined for any remaining particulate matter. Centrifugation was repeated until the solution was free of undissolved material.

The DNA was purified by reversed phase high-performance liquid chromatography (RP-HPLC). The column used was a Hamilton PRP-1 preparative column, packed with PRP-1, which is a co-polymer of styrene and divinylbenzene in the form of beads

5 or 10 μm in diameter. The solvent system chosen was a gradient of aqueous 0.1 M triethylammonium acetate (TeaAc) and HPLC grade acetonitrile. The solutions were vacuum filtered through a Whatman no. 1 filter paper to remove particles that could clog the HPLC system. The solutions were not degassed.

All purification runs were conducted on a Varian 9010 solvent delivery system and 9050 UV-Vis spectrophotometer, controlled by an IBM PC-compatible 386 computer running under Windows 3.1. Some runs were controlled by the operator directly from the 9010 unit.

The two oligomers were purified separately. Test aliquots of 2 μL of stock single strand solution were diluted to 500 μL with doubly distilled, deionized H_2O, taken up into a 5 mL syringe, and injected into the Varian, which was equipped with a 500 μL injection loop. Test samples were initially run under varying solvent gradients to determine a general protocol which was efficient for both strands, but finally two similar protocols were chosen. Table 11 shows the conditions of the chosen protocol.

The DNA (large main peak) eluted between 30-36 minutes in both strand samples (Figures 13, 14). Fractions were collected in 5 mL polypropylene test tubes. Preceding shoulders were eliminated from collection, while product recovery was maximized, by delaying collection for three seconds after the

Table 11

Purification Conditions for $^{5'}$GATTCCAC$^{3'}$ and $^{5'}$GTGGAATC$^{3'}$

	Time (min)	%A	%B
$^{5'}$GATTCCAC$^{3'}$	0.00	93.0	7.0
	15.00	93.0	7.0
	30.00	89.0	11.0
	40.00	5.0	95.0
$^{5'}$GTGGAATC$^{3'}$	0.00	93.0	7.0
	20.00	93.0	7.0
	35.00	89.0	11.0
	40.00	5.0	95.0

Solvent A: 0.1 M triethylammonium acetate

Solvent B: acetonitrile

Flow: 2 mL/min

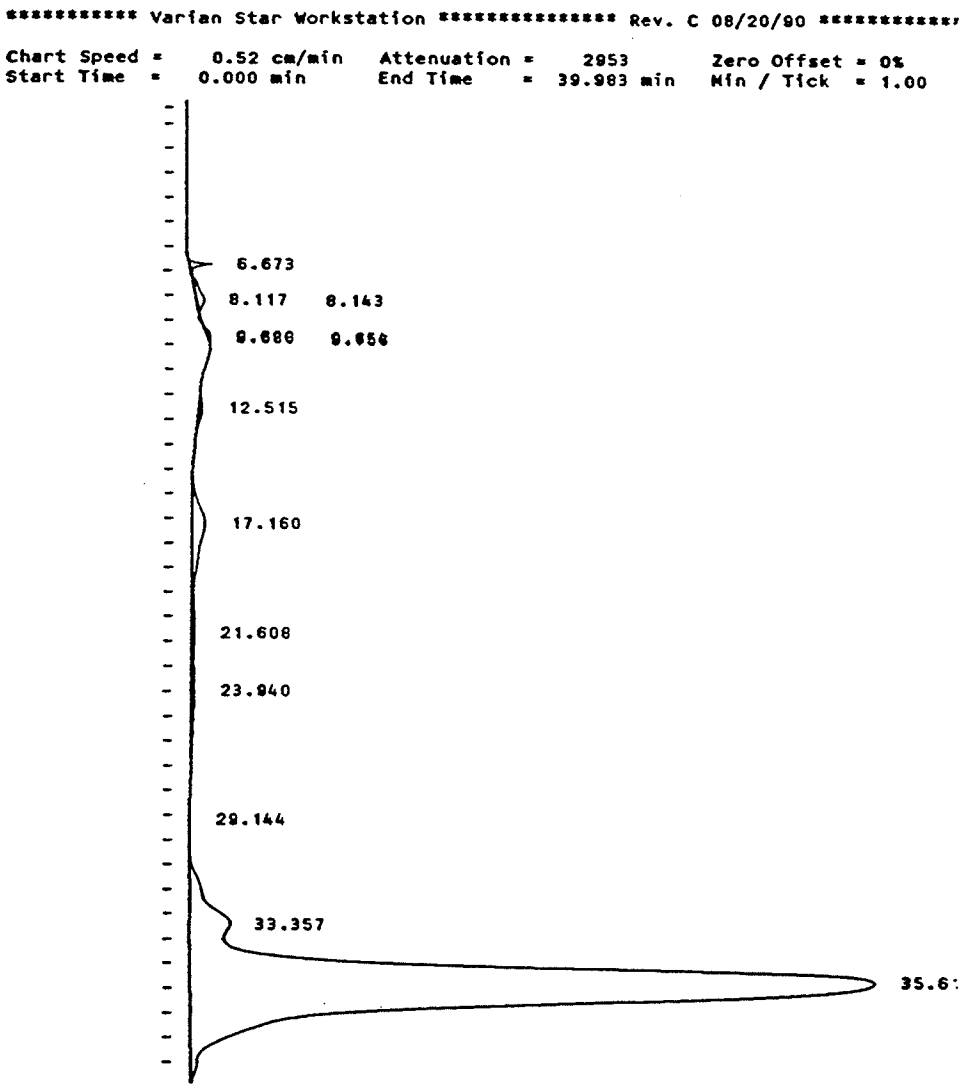


Figure 13. RP-HPLC chromatogram for $^{5'}$GTGGAATC$^{3'}$.

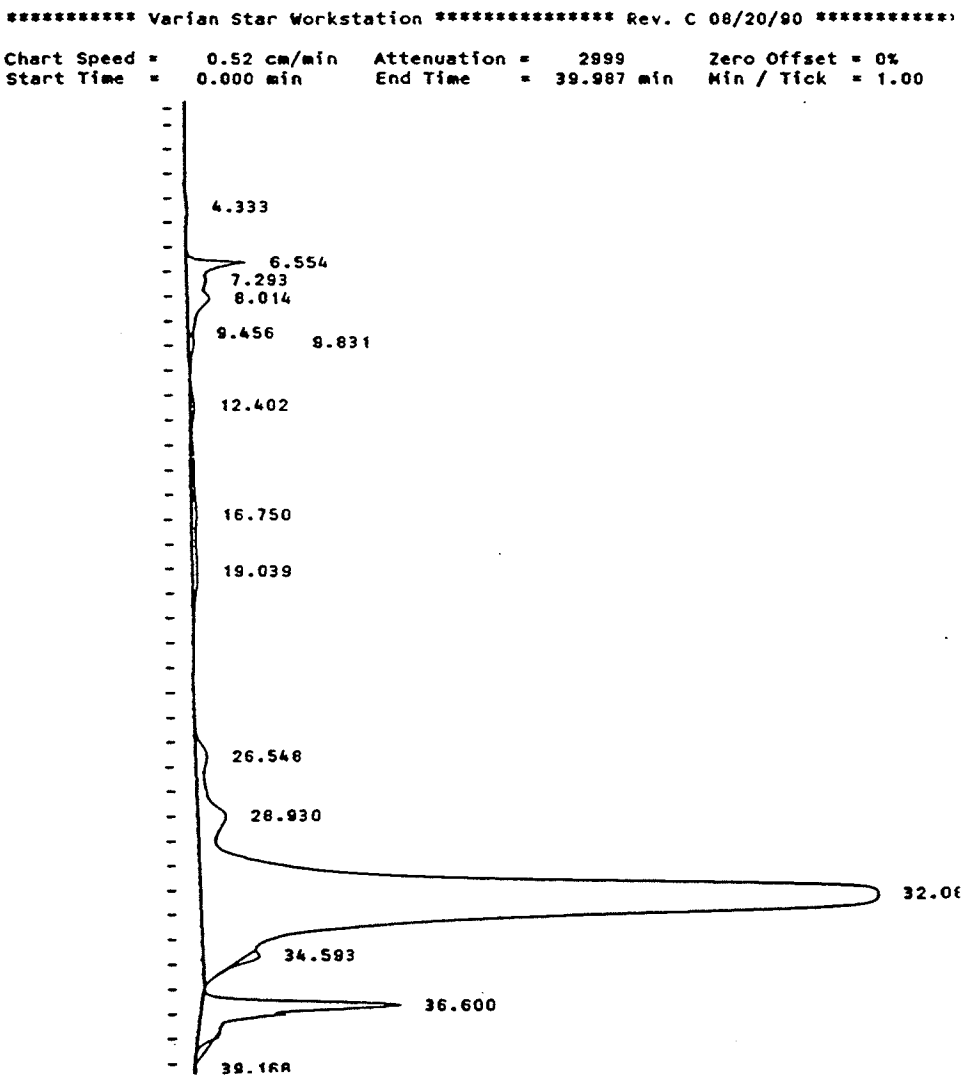


Figure 14. RP-HPLC chromatogram for $^{5'}$GATTCCAC$^{3'}$.

signal was displayed on the computer and the recording chart paper (which accounted for the length of tubing between the detector and the exit port). Similarly, trailing shoulders were eliminated and product recovery maximized by collecting for three seconds after the peak began to tail. Corresponding fractions were pooled, lyophilized in a vacuum centrifuge, desalted, and spun to dryness. The pure white fibrous material was redissolved in distilled H_2O and a small amount was assayed by HPLC to confirm the purity. After a final lyophilization the DNA was accurately weighed (to 0.00001 mg) in tared Eppendorf tubes and tightly capped.

Fully purified DNA was also purchased from the Midland Scientific Corporation. The material was evaporated several times with doubly distilled, deionized H_2O under a vacuum before weighing.

Discussion. Synthesized DNA contains such contaminants as failure sequences and byproducts such as depurinated sequences. Although in the form of sodium or ammonium salts the DNA is of a purity that can often be used in molecular biology studies, the requirements for successful crystallization are much stricter, with a suggested optimal purity of > 95%. The most efficient technique for achieving such high purity is HPLC; on a preparative column a fair amount of sample may

be processed at a time, product recovery is good, and results are generally reproducible. Both RP-HPLC and ion-exchange high-performance liquid chromatography (IE-HPLC) had been used on previously tested DNA strands of 6 and 8 bp (74). The purification results for IE-HPLC were not appreciably better, and the increased number of cleanup steps (desalting, washing, lyophilization) presented too many opportunities for incremental loss of product. Desalting proved to be an especially onerous stage, since the phosphate buffer used in IE-HPLC was almost impossible to completely remove in one treatment. RP-HPLC, with its simpler work-up, was the method of choice.

RP-HPLC relies on the hydrophobic interactions between support and compound. TeaAc is a salt with large hydrophobic groups. Studies using different salts bearing hydrophobic groups of varying bulk, such as tributylammonium acetate (75), suggest that the nature of the salt has a strong eluting effect on the chromatographic run; the larger the hydrophobic group, the longer the retention time. Normally DNA, especially if duplexed, tends to present its ionic surface to the solvent, while maintaining a hydrophobic core; in such a case, interactions with the column support would be minimal and the duplex would be eluted very quickly. Under certain unfortunate circumstances, the desired product could elute with impurities in essentially one messy fraction. By interacting with the negative ions on the sugar-phosphate

backbone of the DNA, TeaAc forms a hydrophobic shell with which the DNA may interact with the support; thus the duplex, with double the number of charges with which TeaAc may interact, should theoretically be retained longest, since the interaction is proportional to the charge (76). Under the influence of increasing amounts of acetonitrile, the duplex's affinity for the column is weakened and the duplex is eluted.

Alternatively to individual strand purification, duplexing before injection into the HPLC system was a possible method, but incomplete purification results for the related octamer $^{5'}$TTCCACTA$^{3'}$/$^{5'}$TAGTGGAA$^{3'}$ (designated TTC; (77)) suggested that purifying the duplex would have been a poor choice, unless each single strand solution was already pure. The duplex TTC and impurities appeared to have a strong affinity for each other; the peak signal appeared to have a trailing shoulder, but alterations of the solvent delivery protocol could not separate the main peak from the shoulder, resulting in impure duplex. Since the solutions of $^{5'}$GATTCCAC$^{3'}$ and $^{5'}$GTGGAATC$^{3'}$ contained a number of impurities in appreciable levels, the purification of the single strands was advised.

Purifying the duplex on the HPLC was not feasible for another reason, *viz.* the apparent equilibrium between duplex and single strands under the HPLC conditions. Ideally, there would have only been two major peaks, duplex and the excess

of one strand, but in practice this was not observed; three substantial peaks were almost always observed. Conditions on the column— pH, temperature, pressure— and the short length of the oligomer may have favorable for the existence of unduplexed DNA. The cooperative nature of the bases which facilitates duplexing is not particularly strong for an octamer, which is only 50% duplexed at 10°C (78); higher salt concentrations encourages double strand formation, for which IE-HPLC would have been useful. The PRP-1 column, under high pressure (100-200 atm) and high temperature due to friction, was not ideal for duplexing. TeaAc does not serve as a true buffer, since it has poor buffering capabilities in the pH range of interest. Although, theoretically the duplex peak should have been easily identified— longest retention time, most intense— in practice it was not possible to be certain. Incomplete duplexing rendered intensity comparison largely meaningless; titration with measured quantities of each strand did not work well, as impurities from each strand produced confusing artifacts. Assuming that the duplex would run off last was also suspect, since the subtle interactions occurring on the column are not known. The effort expended to purify the duplex successfully would not have been worthwhile compared to that spent in purifying the single strands.

Preparative chromatography is a time-consuming step. Even on a preparative column, the quantities which may be

injected are limited, since overloading a column may cause poor resolution of the compounds and may also cause interactions between the compounds that further affect the separation. Due to the need for extreme purity, and to the similarity in behavior on the column of a number of species, a careful choice of solvent ratio had to be determined. Had the compounds differed widely in separation, a very fast protocol could have been chosen. However, a fairly slow protocol was determined as the most efficient in terms of time and resolving capability. More than 50 runs were required on each strand, each run consuming in total approximately 2-3 hours. Further time is spent in the pooling and drying stages. To avoid all these steps, fully purified DNA was purchased. In this case, the additional variable of the unknown skill of the operator is introduced. Similarly, trace contamination from the distant system is uncontrollable. Practical considerations thus dictated that all DNA samples purchased, including fully purified ones, be tested initially in order to record a composition profile of the material. Quality may vary from lot to lot, and it is an apothegm that the final product is best verified by oneself. The Midland DNA appeared clean, but to date had not been verified on HPLC.

V.3 Crystallization

Experimental. Stock solutions of 25 mM spermine, 500 mM $MgCl_2$, 100 mM $CaCl_2$, 100 mM $Co(NH_3)_6Cl_3$, 250 mM sodium cacodylate buffer, and 2,4-methylpentanediol (MPD) were prepared in 100 mL flasks; 25 mL of were drawn off as needed, while the remainder was stored at 6°C. All experimental solutions were filtered through a 0.45 μm filter which was attached to the end of a 10 mL syringe. Separate portions of the sodium cacodylate buffer were adjusted to a pH range of 6.0-8.0.

An extinction coefficient ε was estimated for each DNA strand (79); $\varepsilon(^{5'}GATTCCAC^{3'}) = 75.0\ A_{260}$ units μmol^{-1}, $\varepsilon(^{5'}GTGGAATC^{3}) = 80.4\ A_{260}$ units μmol^{-1}. Based on the molecular weight of the free acid and the weight of the DNA, 16 mM solutions of each were prepared. The concentrations were checked on a Shimadzu 2101PC UV-Vis spectrophotometer, using Beer's Law and the estimated extinction coefficients on 0.2 μL aliquots dissolved in 1.0 mL water. Equimolar amounts of each single strand solution were combined in an Eppendorf tube. To determine the dissociation or melting curve, 0.5 μL of the duplex solution was dissolved in 1.0 mL of 100 mM $CaCl_2$ in a quartz 1.0 cm cuvette. The cuvette was placed in the holding block of the spectrophotometer, with the temperature set to 5° C. Absorbance readings were taken as the temperature was gradually raised, with the cuvette equilibrated for five minutes after each

change. The dissociation curve is given in Figure 15.

Starting conditions were roughly based upon average crystallization conditions gleaned from an analysis of published DNA structure papers. However, since DNA crystallization is a highly empirical process, a sufficiently wide range of conditions needed to be tested. On each crystallization plate, one reagent was varied along each axis, while the other reagents were maintained at constant concentrations. On another plate, the same variation was repeated, but a third reagent was subsequently changed. Figure 16 outlines the narrowing of the screening grid around a potentially useful condition.

Five µL drops were equilibrated against 500 µL reservoirs of a lower water percentage; the chemical potential difference resulted in the net transfer of water out of the drop and into the reservoir. The much larger volume of the reservoir ensured that the gradient direction was always positive (*i.e.* from drop to reservoir). The reservoirs of each crystallization well were filled first, with 350 µL MPD and 150 µL water. Conditions were initially screened by setting up 5 µL drops of various reagent concentrations and 2 mM DNA in a rough grid. Both hanging drop and sitting drop plates were used. For hanging drop experiments, the drops were set up on both glass (siliconized) and plastic cover slips which were inverted over the wells of a Linbro tissue culture plate (Fisher

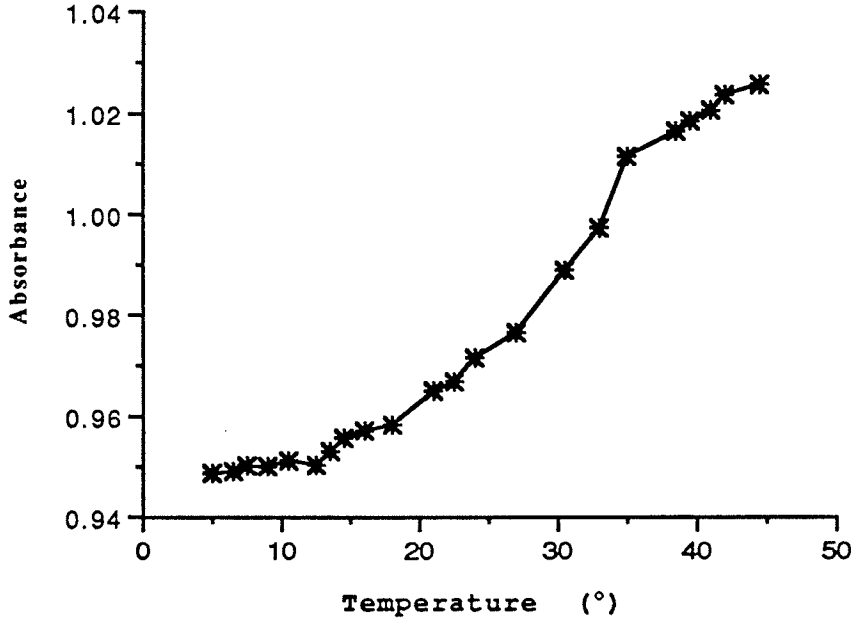


Figure 15. Dissociation curve of GATTCCAC in 100 mM $CaCl_2$

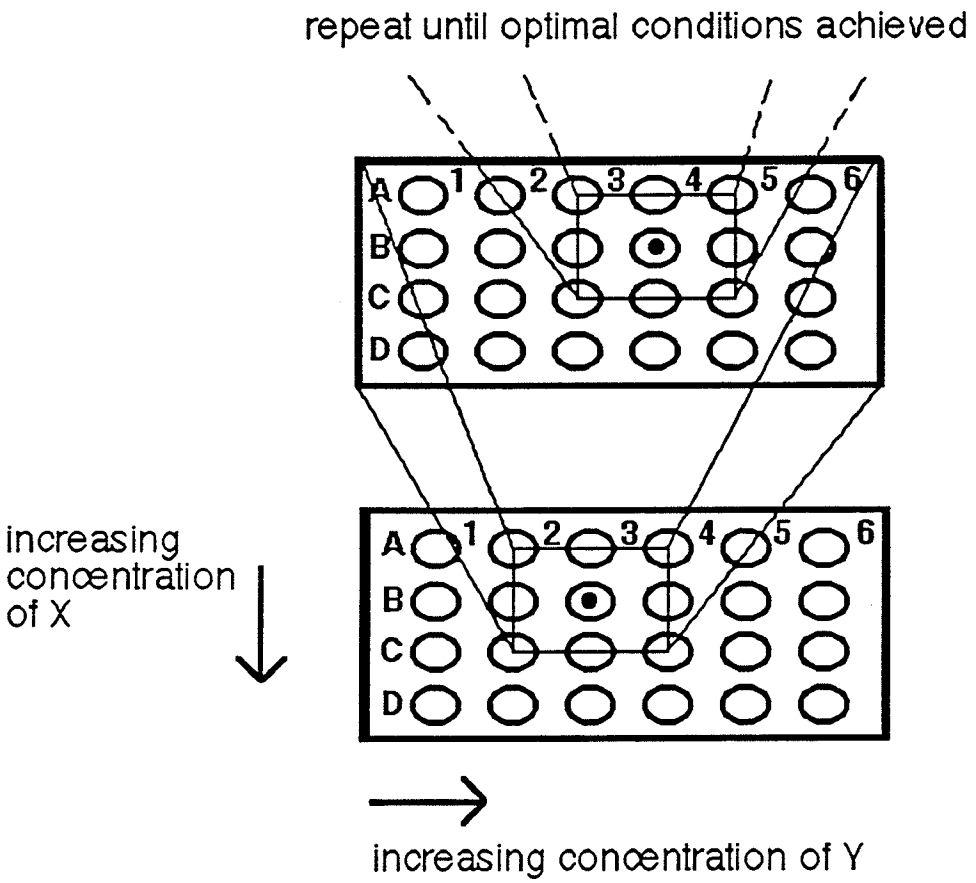


Figure 16. Schematic for successive narrowing of screening conditions for crystallization.

Scientific) and sealed with silicon grease between plate and cover slip. Several housings for sitting drops were used: Corning glass plate in a sandwich box, small sealed Petri dish, Cryschem crystallizing plate (Charles Supper Company), glass microscope slide with depression, plastic mounts for standing in a Linbro well (Hampton Research). The solutions for all crystallization drops were added specifically in the following order, and stirred after each addition:

H_2O -> cacodylate buffer -> DNA -> cation -> spermine -> MPD

The drops were stored at ~20 °C (± 5°C) and 4°C. After several days crystals were observed to have grown, and the drops were examined under a polarizing microscope. The drops were analyzed by a simplified version of the factorial method (80), based on subjective assessment of crystal quality. The conditions which seemed to produce the best crystals were chosen as a new starting point, and another grid, with narrower sampling parameters, was established around those conditions. When the optimal conditions appeared to have been achieved, the drop size was increased to 10 µL for the hanging drops and 25 µL for the sitting drops. Table 12 shows the final conditions chosen.

Discussion. The nature of the crystallizing medium is critical for DNA. A very hydrophilic molecule, DNA is readily

Table 12

Crystallization Conditions of GATTCCAC

mount:	sitting drops
temperature:	15°C < T < 22°C (ideally 17°C)
pH:	7.0
DNA:	0.5 mM
$CaCl_2$:	20 mM
spermine:	0.42 mM
sodium cacodylate:	12 mM
MPD gradient:	2.8%-40%

stabilized by solvation shells since the duplex has a large negatively charged surface with which the water molecules may interact. In order to induce precipitation, it is necessary to reduce the number of DNA-water interactions until the DNA can no longer remain soluble and begins to aggregate about a number of nucleating sites. The rate at which the water is removed can be critical; if it is too fast, the DNA likely will not crystallize but will precipitate (essentially infinite number of nucleating sites), perhaps with some other molecule, as an amorphous mass. The process is not strictly reversible: the amorphous precipitate appears to be quite hydrophobic, since even the addition of a large excess of water does not readily redissolve all the precipitate.

GATTCCAC formed large disordered aggregates in a wide range of conditions; fairly large single crystals were visible in many of the aggregates. The speed with which the DNA crystallized was very helpful in rapidly screening growth conditions, and the ability to crystallize easily was a useful tool for evaluating the effect of crystallizing agent on quality.

Divalent cations are known to stabilize oligodeoxyribonucleotides in Z-DNA structures, by coordination with water molecules and the phosphate backbone (81); similar behavior is believed to occur for A and B-DNA structures.

$MgCl_2$ has been the cation of choice in most DNA crystallizations, largely as a matter of habit, but $CaCl_2$ appeared to be more useful than $MgCl_2$ in this experiment, and has been observed to be useful in other structures as well (82). The crystals formed more readily, and grew larger, than under the same conditions with $MgCl_2$. It may be that the size of the Ca^{+2} ion is more easily accommodated by the DNA, or that the outer shell now available permits better coordination. Similar arguments may be used for $Co(NH_3)_6Cl_3$, which has been used successfully for crystallizing Z-DNA (83,84). Since cobalt forms octahedral complexes readily, coordination with six water molecules, as the Mg^{+2} showed, may be particularly favorable for $Co(NH_3)_6^{+3}$. The behavior of $Co(NH_3)_6^{+3}$ with GATTCCAC was markedly different from that of either Mg^{+2} or Ca^{+2}. The solubility of the DNA was drastically reduced; when $Co(NH_3)_6^{+3}$ was used in place of $CaCl_2$ under conditions leading to crystals in a 25 µL drop, 200 µL of water was needed to keep the DNA in solution. Crystals grew within a few days, but remained very small (smaller that 0.02 mm on the short side). They were elongated along one axis, and many curiously floated in solution end to end, forming a long arc. Others formed tight coils, and spherulites (Figure 17). The crystals were very soft, more like a thick oil which held its shape until manipulated. $Co(NH_3)_6^{+3}$ considerably hastened the speed of crystallization, but the crystal quality was poor because

the DNA molecules likely did not all have the opportunity to align into the least energetic conformation. A fundamental compromise exists between the speed of the crystallization (GATTCCAC grew in approximately a week, which was unusually rapid) and the crystal quality which exists for both small molecules and macromolecules; the results are simply more apparent for macromolecules.

Spermine, which has been observed in increased concentrations in actively growing cells, may act as a condensating agent (85,86). Phase diagrams of DNA/spermine/$MgCl_2$ systems have demonstrated the competitive relationship of Mg^{+2} and spermine (87). A similar argument may be extended to Ca^{+2}, and to a lesser extent $Co(NH_3)_6^{+3}$, used in this study. The competitive relationship between Ca^{+2} and spermine was clearly evident in the screening conditions. Increasing amounts of $CaCl_2$ led to larger crystals, while decreasing spermine led to fewer crystals, which mirrored the behavior of the $MgCl_2$ tests. The screening was quite sensitive, particularly for spermine: differences in crystal size and number could be detected between 0.40 mM and 0.80 mM spermine and 10 mM and 20 mM $CaCl_2$. Increasing concentrations of spermine also led to increased twinning in the crystals. The aggregations of crystals were often impressive (Figure 18). The large crystals in the arms were of fairly good quality, but not to the extent of yielding good diffraction data.

Figure 17. Disordered crystals of GATTCCAC in the presence of $Co(NH_3)_6^{+3}$.

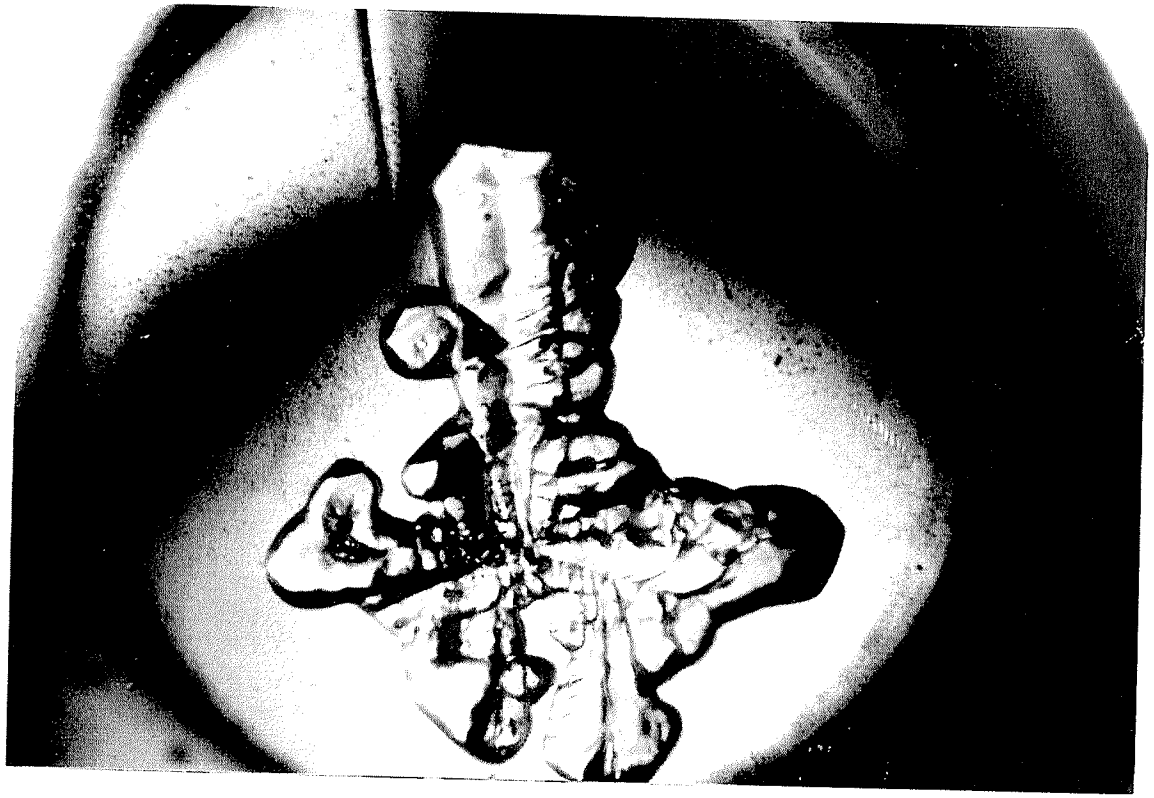


Figure 18. Crystal aggregates of GATTCCAC in the presence of $CaCl_2$.

As a crystallization drop equilibrates against the reservoir, ionic strengths change; this may induce a change in the pH of the drop. Although less critical for DNA than for protein, any shift in pH may be significant. A cacodylate buffer, generally as the sodium salt, is often used since it is quite stable in the pH and temperature ranges common to DNA crystallization setups. It has the added benefit of acting as a preservative; while the DNA is likely too short to be much affected by bacteria, spermine solutions are notorious in this worker's lab for having a very short shelf life due to the formation of mold, unless the solutions were refrigerated. The Na^+ ions also may have an electrostatic shielding effect in low ionic strength media (88), but appeared to have minimal effect in this experiment, in which the divalent cation influence was dominant.

The best crystals grown in the cold room (T=4°C) were grown in the same conditions as at room temperature; they were of octahedral habit, with well defined faces, but were very small (<0.01 mm) along each dimension. They formed as a shower of crystals. Attempts to seed new drops with a few crystals only led to more showers of well-shaped tiny crystals. It appeared the conditions were good, but at too rapid a rate; crystals often remain small when imperfections accumulate, essentially poisoning the crystal surface so the number of molecules adhering is the same as the number sloughing off,

with net zero growth. Different conditions were attempted, including a lower MPD concentration in the reservoir (25%) but growth was slow and the crystals were badly twinned and poor quality. Other conditions grew crystals that were 0.6 mm on each of two dimensions, but less than 0.1 mm in the third, the depth; the crystals were thin sheets lying at the solution:air interface of either hanging drops or sitting drops.

The crystals grown at room temperature grew quite rapidly and in a wide range of conditions, indicating that this sequence was unusually predisposed to crystallization, at an unusual temperature; many crystallizations are successful only at lower temperatures, such as 4°C. However, the crystals were especially sensitive to fluctuations in the temperature: growth was rapid at 15-18°C, crystal showers occurred below 10°C, crystals began deteriorating above 21°C and completely dissolved above 24°C, as expected from the observed dissociation curve. The dissociation curve was obtained in a high salt medium, which tends to stabilize the duplex (78), so it is not surprising that deterioration commenced at a lower temperature. Maintaining a constant temperature for the crystallizations proved an unequal challenge, and several plates experienced temperature fluctuations from 16° to close to 30° C. Many of the plates contained crystals which actually dissolved and grew several times; however, the

quality steadily declined with each repetition, as the residues of precipitate increased.

V.4 Determination of the Unit Cell

Experimental. Several crystals were examined for quality under the polarizing microscope; only those which extinguished distinctly and were greater than 0.1 mm on an edge were considered.

Once the drops were exposed to the atmosphere, work was done quickly to minimize evaporation, to preserve existing crystals in the drop or plate. The crystal was positioned with a microneedle so that the crystal could be removed from the drop along the direction of a long axis. A glass capillary of a diameter considerably larger than the maximum width of the crystal was attached to thin Tygon tubing, which in turn was attached to a plastic tube for mouth pipetting. The crystal was carefully drawn into the capillary (practice of the technique on unimportant crystals was advised) with a minimum of mother liquor. A capillary with a smaller diameter was inserted to suck away excess mother liquor, leaving a thin film on the crystal and a droplet approximately 1 mm distant. The open end was immediately sealed with silicon grease. An approximately 10 mm length of capillary starting

at the point of the crystal was measured (the length depended upon the type of goniometer head used). Forceps bearing a generous portion of silicon grease were used to snap the capillary at the required length. All heat sources, including body heat from fingers, were kept away from the crystal. Using the forceps, the capillary section was inserted into a pin filled with silicon grease and mounted on the goniometer head. Excess silicon was removed from the pin and the open end of the capillary. Epoxy resin was used to seal the opening and anchor the capillary firmly to the pin, care being taken not to allow the epoxy to ooze into the capillary and coat the crystal.

Indexing was performed on a Rigaku AFC6S sealed tube, automated diffractometer, equipped with a Cu X-ray tube (graphite monochromated Cu Kα radiation) operated at 50 kV and 40 mA, and a low temperature apparatus. Cooling was by a stream of dry nitrogen gas, passed through liquid nitrogen, and directed over the crystal. The ambient temperature ranged from 22°-28°C. The temperature set point was set at 5°C, but the reading at the gas outlet (at the crystal) was 10°C. Several crystals were successfully indexed after some effort; 20-25 reflections were scanned in the 2θ range of 4.5°- 10°.

Table 13

Crystal Data for GATTCCAC

Crystal	a(Å)	b(Å)	c(Å)	V(Å^3)
1	44.2(1)	44.2(1)	25.9(1)	50500
2	44.40(5)	44.40(5)	26.09(7)	51400
3	44.29(4)	44.29(4)	25.62(8)	50300
4	44.76(8)	44.76(8)	25.56(7)	51200

$\alpha=\beta=\gamma=90°$

The Laue symmetry check for tetragonal elements (4-fold axis parallel to *c*, and mirror perpendicular to *a*) either failed or showed a poor correlation between the symmetry equivalent reflections. No psi scans for absorption correction were collected to minimize crystal exposure to X rays. Only crystals (1) and (3) survived long enough to begin data collection. The scan rate was 2°/min in ω, with up to 10 rescans of weak reflections ($I<1.5\sigma(I)$). Both data sets prematurely terminated when the standards fell to zero, thus yielding inconplete data sets. XPLOR (89) was used in a rudimentary attempt to determine the type of the macromolecule and its position in the unit cell, but the attempt was not successful due to the paucity of usable data.

Discussion. Although grown at room temperature, the crystals could not be scanned at the same temperature because of the heat buildup from exposure to the X-rays. The crystals were observed to melt or redissolve at about 24-25°C. The mounted crystals also had to be stored at the temperature at which they had been grown; those which were placed at 4 °C slowly deteriorated internally over the course of three days, although they maintained their form. It is likely that the water molecules rearranged as the solvent density changed, compromising the integrity of the lattice.

Examination of the crystals under the microscope revealed large faceted crystals with excellent morphology. They showed sharp extinction effects, except along one axis, consistent with tetragonal crystals. However, all specimens proved to be definitive examples that macroscopically pleasing crystals are not necessarily of diffraction quality. None of the crystals examined were strong scatterers (although that is to be expected from macromolecular crystals in general, and DNA crystals in particular) and no high angle reflections could be successfully refined. The lack of extended order seemed particularly noticeable along the *c* axis, since no hkl reflections with $l > 1$ could be found, whereas fairly strong hkl reflections with $h \geq 3$ could be centered on the diffractometer (Figure 19). Interestingly, even after they became disordered internally (as observed on the diffractometer), the crystals

```
Profile is for reflection index   h =    3     k =    3    l =    0

2theta =     8.379    omega =     4.190    chi =   -41.733    phi =    80.572
Scan type is omega
Step time (sec.) is   0.5

Integrated intensity is       34265.  (   185.)
Full width at half height =   0.263  degrees   omega
Peak offset =  -0.042  degrees   omega
Peak width  =   1.110  degrees   omega
Full scale for the plot =    1715.  counts/step  or       3430.  cps

   omega  c/step  0%                          50%                         100%
                 +--------------+--------------+--------------+--------------+
   3.440      58. : *
   3.470      58. : *
   3.500      54. : *
   3.530      69. : *
   3.560      58. : *
   3.590      70. : *
   3.620      53. : *
   3.650      73. :  *<BKG
   3.680      79. :  *
   3.710      84. :  *
   3.740     100. :  *
   3.770      87. :  *
   3.800     120. :   *
   3.830     136. :    *
   3.860     180. :     *
   3.890     209. :      *
   3.920     312. :          *
   3.950     405. :              *
   3.980     488. :                *
   4.010     700. :                 *
   4.040    1006. :                      *
   4.070    1300. :                           *
   4.100    1583. :                                *
   4.130    1715. :                                 *
   4.160    1698. :                                *
   4.190    1529. : CEN>                          *
   4.220    1484. :                             *
   4.250    1269. :                                 *
   4.280     987. :                      *
   4.310     729. :                  *
   4.340     594. :              *
   4.370     437. :            *
   4.400     359. :          *
   4.430     300. :        *
   4.460     220. :       *
   4.490     187. :      *
   4.520     192. :      *
   4.550     159. :     *
   4.580     148. :    *
   4.610     133. :    *
   4.640     109. :   *
   4.670      91. :  *
   4.700      95. :  *
   4.730     109. :   *
   4.760      63. : *<BKG
   4.790      66. : *
   4.820      69. : *
   4.850      66. : *
   4.880      56. : *
   4.910      60. : *
   4.940      50. : *
```

Figure 19. ω scan of reflection (3 3 0).

maintained their habits, and often continued to extinguish under polarized light, although not as clearly. Disordered crystals usually do not grow very large; however, one particular specimen had the dimensions 0.7 X 0.4 X 0.3 mm. One practical problem with such a large specimen (cutting of the crystals was to be avoided if possible, since crystal deformation at the cut tended to occur) was that a capillary large enough and sturdy enough was difficult to find. Since the walls of the capillaries were round, unless the crystals were equant (*i.e.* substantially large in all dimensions) they tended to conform to the arc of the capillary. For thin crystals it was often better to use a much wider capillary, so that the deformation was less severe. However, experience showed that capillaries of greater diameter than 0.7 mm were very fragile; slight sideways pressure caused them to shear, which happened easily during mounting in the pin. Depending upon the nature of the crystal, it was sometimes better to use the smaller capillary and risk some deformation.

Tetragonal space groups are the most common for DNA octamer crystals, followed by the hexagonal groups. Octamers in the tetragonal space groups have been exclusively A-DNA to date. Whether this is due to peculiarities of oligomer length is not known, but the extent of hydration does appear to be a key feature (90). The higher resolution crystal structures were obtained from self-complementary sequences, which is

another variable to consider, but a disordered, non-self-complementary, 9-bp DNA duplex (91) was also observed to be A-type; unit cell parameters of GATTCCAC are similar to those of the latter, so it is possible GATTCCAC is A-type. Of the tetragonal space groups, only the $P4_n$ groups, where n is 1, 2, or 3, are reasonable, in view of the probable density (between 1.1 and 1.5) and subsequent Z value.

In crystallographic jargon, what is desired is the ideally imperfect crystal, in which small regions are slightly misaligned relative to each other. Imperfections are responsible for the phenomenon of observed reflections (17) since all reflections would cancel each other out in an ideally perfect crystal. The extent of imperfection may be seen in the profile of a diffraction peak. A crystal is composed of numerous unit cells, the planes of each of which contribute to the observed reflection. In the ideally perfect crystal all reflections occur at the same angle θ, and any peak observed would be a sharp line with $\delta\theta$ (peak width) = 0. Each misaligned region of the ideally imperfect crystal, however, reflects at a slightly different angle. The result is a Gaussian distribution around θ, with a mosaic spread of $\delta\theta$. Diffraction peaks with multiple maxima are also observed, and often arise from disordering of larger regions, each essentially a microcrystal. Monitoring a peak profile is one way of observing crystal quality, since as a crystal becomes disordered,

the mosaic spread increases.

The imperfections of a crystal, however, are rarely ideal. Many other factors may combine to produce a crystal with so many imperfections that it is unsuitable for data collection. This is generally more applicable for macromolecules, which possess many degrees of freedom, than for small molecules. DNA molecules may pack in a number of different ways, and have the further property of being grossly cylindrical; the interaction of a particular orientation of a molecule relative to another molecule may be energetically very similar to that of two molecules with slightly different orientations. It is possible that this scenario is applicable to GATTCCAC. Indeed, the ease with which this sequence aggregates into low order states suggests that multiple orientations are energetically stable.

V.5 Summary

The process of obtaining the small crystals of DNA desired can be an arduous one, in terms of time and effort. This is especially true in the crystallization steps, where so much of the work is empirical in nature, and in data collection, when a promising crystal either turns out to be less than worthwhile, or disintegrates at the most inopportune

times.

Unfortunately, no information on the GTG question could be gathered. However, conditions are very nearly optimal, and considering the readiness of GATTCCAC to crystallize, it is conceivable that a small change in conditions may be all that is necessary to achieve a diffraction quality crystal. Those crystals which did diffract showed unit cell dimensions that were reasonably similar, so it is reasonable to infer that GATTCCAC crystallizes in a tetragonal space group with dimensions approximately $\underline{a}$= $\underline{b}$= 44 Å, $\underline{c}$=26 Å. It is also likely that the space group is $P4_n$, in view of the probable density (between 1.1 and 1.5) and Z value.

References

1. LaBella, F. S., Bihler, I., Templeton, J. F., Kim, R. S., Hnatowich, M. & Rohrer, D. (1985). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **44**, 2806.

2. Chow, E., Kim, R. S., LaBella, F. S. & Queen, G. (1979). *Br. J. Pharmacol.* **67**, 345.

3. Duax, W. L., Cody, V., Griffin, J., Hazel, J. & Weeks, C. M. (1978). *J. Steroid Biochem.* **9**, 901.

4. Duax, W. L., Cody, V. & Hazel, J. (1977). *Steroids* **30**, 471-480.

5. Hnatowich, M. & LaBella, F. (1984). *Eur. J. Pharmacol.* **106**, 567-575.

6. Kim, R. S. S., LaBella, F. S., Zunza, H., Zunza, F. & Templeton, J. F. (1980). *Mol. Pharmacol.* **18**, 402-405.

7. Tanz, R. D. (1963). *Rev. Can. Biol.* **22**, 147-163.

8. Fullerton, D. S., Kitatsuji, E., Deffo, T., Rohrer, D. C., Ahmed, K. & From, A. H. L. (1983). *Curr. Top. Membr. Transp.* **19**, 257-264.

9. Templeton, J. F., Sashi Kumar, V. P., Cote, D., Bose, D., Elliot, D., Kim, R. S. & Labella, F. S. (1987). *J. Med. Chem.* **30**, 1502-1505.

10. Debaerdemaeker, T., Germain, G., Main, P., Tate, C. & Woolfson, M. M. (1987). *MULTAN87. Computer programs for the automatic solution of crystal structures from X-ray diffraction data.* Univs of Ulm, Germany and York, England.

11. Coppens, P. & Hamilton, W. C. (1970). *Acta Crystallogr.* **A26**, 71-83.

12. Cromer, D. T. & Mann, J. B. (1968). *Acta Crystallogr.* **A24**, 321-324.

13. Stewart, R. F., Davidson, E. R. & Simpson, W. T. (1965). *J. Chem. Phys.* **42**, 3175-3187.

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

15. Busing, W. R., Martin, K. O. & Levy, H. A. (1962). *ORFLS*. Report ORNL-TM-305. Oak Ridge National Laboratory, Tennessee, USA.

16. Busing, W. R., Martin, K. O. & Levy, H. A. (1964). *ORFFE*. Report ORNL-TM-306. Oak Ridge National Laboratory, Tennessee, USA.

17. Stout, G. H., Jensen, L. J. *X-Ray Structure Determination: A Practical Guide. 2nd edition.*(1989). John Wiley & Sons. New York.

18. Campsteyn, H., Dupont, L. & Dideberg, O. (1972). *Acta Crystallogr.* **B28**, 3032-3042.

19. Serantoni, E. F., Krajewski, A., Mongiorgi, R., Riva di Sanseverino, L. & Cameroni, R. (1975). *Cryst. Struct. Comm.* **4**, 189-192.

20. Guntert, T. W. & Linde, H. H. A. (1981). *Cardiac Glycosides*, pp. 13-24. Berlin: Springer-Verlag.

21. Karle, I. L. & Karle, J. (1969). *Acta Crystallogr.* **B25**, 434-442.

22. *Alchemy II. Molecular modelling software.* (1988).Tripos Associates, Inc. St. Louis, Missouri.

23. Duax, W. L. & Norton, D. A. (1975). *Atlas of Steroid Structure,* Vol. I. New York: Plenum.

24. Chandross, R. J. & Bordner, J. (1975). *Acta Crystallogr.* **B31**, 928-931.

25. Altona, C., Geise, H. J. & Romers, C. (1968). *Tetrahedron* **24**, 13-32.

26. *PCMODEL. Molecular modelling software.* v.3. (1989).For the Silicon Graphics Iris 4D70/GT workstation. Serena Software. Bloomington, Indiana.

27. Schmit, J.-P. & Rousseau, G. G. (1978). *J. Steroid Biochem.* **9**, 909-920.

28. Duax, W. L., Griffin, J. F. & Rohrer, D. C. (1981). *J. Am. Chem. Soc.* **103**, 6705-6712.

29. Berkovitch-Yellin, Z. & Leiserowitz, L. (1984). *Acta Crystallogr.* **B40**, 159-165.

30. Stryer. L. *Biochemistry.* 3rd ed. (1988). W. H. Freeman & Co. New York.

31. Holt, J. T., Redner, R. L., Nienhus, A. W. (1988). *Mol. Cell. Biol.* **8**, 963.

32. Uhlmann E., Peyman, A. (1990). *Chemical Reviews,* **90(4)**, 543.

33. Wickstrom, E. ed. (1991). *Prospects for Antisense Nucleic Acid Therapy of Cancer and AIDS.* Willey-Liss, New York.

34. Sanghvi, Y. S., Hoke, G. D., Freier, S. M., Zounes, M. C., Gonzalez, C., Cummins, L., Sasmor, H., Cook, P. D. (1993). *Nucleic Acids Research* **21**, 3197.

35. Matteucci, M. D., Bischofberger, N. (1991). *Annu. Reports Med. Chem.* **26**, 287.

36. Etzold, G., Langen, P. *Chem. Ber.* **98**, 1988 (1965).

37. Langen, P., Etzold, G. *Mol. Pharmacol.* **2**, 89 (1966).

38. Dee Nord, L., Kent Dalley, N., McKernan, P. A., Robins, R. K. (1987). J. *Med. Chem.* **30**, 1044-1054

39. Zimmerman, M. *Biochem. Biophys. Res. Commun.* **16**, 600 (1964).

40. Secco, T., private communication.

41. IUPAC Commission on the Nomenclature of Organic Chemistry (CNOC) and IUPAC-IUB Commission on Biological Nomenclature (CBN). (1971). *Eur. J. Biochem.* **21**, 455.

42. Etzold, G., Langen, P. (1968). In *Synthetic Procedures in Nucleic Acid Chemistry. Volume I.* (ed. by W. W. Zorbach and R. S. Tipson). Interscience Publishers, New York.

43. Borowiecka, J., Lipka, P. Michalska, M. (1988). *Tetrahedron* 44, 2067.

44. Vorbruggen, H, Höfle, G.(1981). *Chem. Ber.* **114**, 1256.

45. Vorbruggen, H, Krolikiewicz, K., Bennua, B.(1981). *Chem. Ber.* **114**, 1234.

46. Diop, L., private communication.

47. Kaiser, D., Tabor, H., Tabor, C. W. (1963). *J. Mol. Biol.* **6**, 141.

48. Gosule, L. C., Schellman, J., A. (1976). *Nature* **259**, 333.

49. Egli, M., Williams, L., D., Gao, Q., Rich, A. (1991). *Biochemistry* **30**, 11388.

50. Chatoraij, D., K., Gosule, L. C., Schellman, J., A. (1978). *J. Mol. Biol.* **121**, 327.

51. Ivanov, V. I., Minyat, E. E. (1981). *Nucleic Acids Res.* **9**, 4762.

52. Minyat, E. E., Ivanov, V. I., Krizyn, A. M., Michenenkova, L. E., Schyolinka, A. K. (1978). *J. Mol. Biol.* **128**, 397.

53. Tari, L., private communication.

54. Wang, A. H. J., Fujii, S., van Boom, J. H., Rich, A. (1982). *Proc. Natl. Acad. Sci.* **79**, 3968.

55. Hunter, W. N., Langlois D'Estaintot, B., Kennard, O. (1989). *Biochemistry* **28**, 2444.

56. Feuerstein, B. G., Pattabiraman, N., Marton, L. J. (1990). *Nucleic Acids Res.* **18**, 1271.

57. Clark, G. R., Brown, D. G., Sanderson, M. R., Chwalinski, T., Neidle, S., Veal, J. M., Jones, R. W., Wilson, W. D., Zon, G., Garman, E., Stuart, D. I. (1990). *Nucleic Acids Res.* **18**, 5521.

58. Verdaguer, N., Aymami, J., Fernandez-Forner, D., Fita, I., Coll, M., Huynh-Dinh, T., Igolen, J., Subirana, J. A. (1991). *J. Mol. Biol.* **221**, 5521.

59. Iitaka, Y., Huse, H. (1965). *Acta Cryst.* **18**, 110.

60. *TEXSAN crystallographic package. Structure solution software*. Molecular Structure Corporation.

61. Cho, C. K., unpublished results.

62. McPherson, A. *Preparation and Analysis of Protein Crystals*. (1982). John Wiley & Sons. New York.

63. Matthews, B. W. (1988). *Nature (London)* **335**, 294.

64. von Hippel, P. H., Berg, O. G. 1986. *Proc. Natl. Acad. Sci.*, **83**, 1608.

65. Lorber, B., Giegé, R. (1992). In *Crystallization of Nucleic Acids and Proteins. A Practical Approach*. (ed. by A. Ducruix and R. Giegé), pp. 19-42. Oxford University Press, Oxford.

66. Jurnak, F. A., McPherson, A. ed. (1985). *Biological Macromolecules and Assemblies. Volume 2: Nucleic Acids and Interactive Proteins*. John Wiley & Sons. New York.

67. Jacob, F., Monod, J. (1961).*J. Mol. Biol.* **3**, 318.

68.Monod, J. (1947). *Growth* **11**, 223.

69. Wilson, D. B., Hogness, D. S. (1968). *J. Biol. Chem.* **244**, 2143.

70. Adhya,S., Shapiro, J. A. (1968). *Genetics* **62**, 231.

71. Parks, J. S., Gottesman, M., Perlman, R. L,.

Pastan, I. (1971). *J. Biol. Chem.* **246**, 2419.

72. Adhya,S., Echols, H. (1966). *J. Bacteriol* **92**, 601.

73. Lu, P., Cheung, S., and Arndt, K. (1983). *J. Biomol. Struct. Dynam.* **1**, 509.

74. Cho, C. K., unpublished results.

75. Tari, L., private communication.

76. Pingoud, A., Fleiss, A., Pingoud, A. (1989). *In HPLC of macromolecules, a practical approach* (ed. by R. W. A. Oliver), pp. 183-208. IRL Press Press, Oxford.

77. Cho, C. K., unpublished results.

78. Saenger, W. *Principles of Nucleic Acid Structure.* (1984). Springer-Verlap. New York.

79. Fasman, G. D., ed. (1976). Handbook of Biochemistry and Molecular Biology: Nucleic Acids. 3rd edition. CRC Press, Cleveland.

80. Carter, C. W. (1992). In *Crystallization of Nucleic Acids and Proteins. A Practical Approach.* (ed. by A. Ducruix and R. Giegé), pp. 47-69. Oxford University Press, Oxford.

81. Wang, A. H. J., Hakoshima, van der Marel, G., van Boom, J. H., Rich, A. (1984). *Cell*, **37**, 321.

82. Einspahr, H., Cook, W. J., Bugg, C. E. (1981). *Biochemistry* **20**, 5788.

83. Gessner, R. V., Quigley, G. J., Wang, A. H. J., van der Marel, G., van Boom, J. H., Rich, A. (1985).

Biochemistry, **24**, 237.

84. Brennan, R. G., Westhof, E., Sundaralingam M. (1986). *J. Biomol. Struc. Dyn.* **3**, 649.

85. Tabor, C. W., Tabor, H. (1976). *Annu. Rev. Biochem.***45**, 285.

86. Sakai, T. T., Cohen, S. S. (1976). *Prog. Nucl. Ac. Res. Mol. Biol.* **17**, 15.

87. Malinina, L. V., Makhaldiani, V. V., Tereshko, V. A. (1987). *J. Biomol. Struc. Dyn.* **5**, 405.

88. Dock-Bregeon, A. C., Moras, D. (1992). In *Crystallization of Nucleic Acids and Proteins. A Practical Approach*. (ed. by A. Ducruix and R. Giegé), pp. 145-169. Oxford University Press, Oxford.

89. XPLOR. *System 3.1. A System for X-ray crystallography and NMR.*(1992). Yale University.

90. Kennard, O., Hunter, W. N. (1991). *Angew. Chem. Int. Ed. Engl.* **30**, 1254.

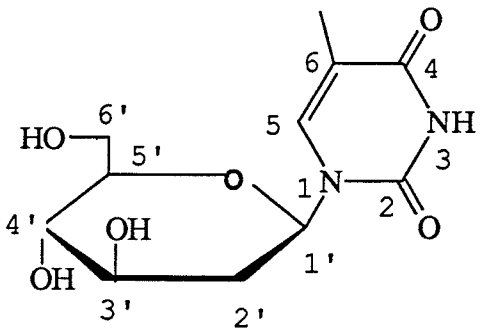
91. McCall, M., Brown, T., Kennard, O. (1985). *J. Mol. Biol.* **183**, 385.

Appendix A: NMR and MS Spectra for (2′-deoxy-β-D-*arabino*-hexopyranosyl)thymine

Appendix A

Assignment of Peaks in 300 MHz ^{1}H NMR Spectra of (2′-deoxy-β-D-*arabino*-hexopyranosyl)thymine.

δ (ppm)	identity
5.8	H1′
3.8-3.9	H5′, H6′
3.6	H3′
3.4	H4′
2.3	H2′



mw= 272.3

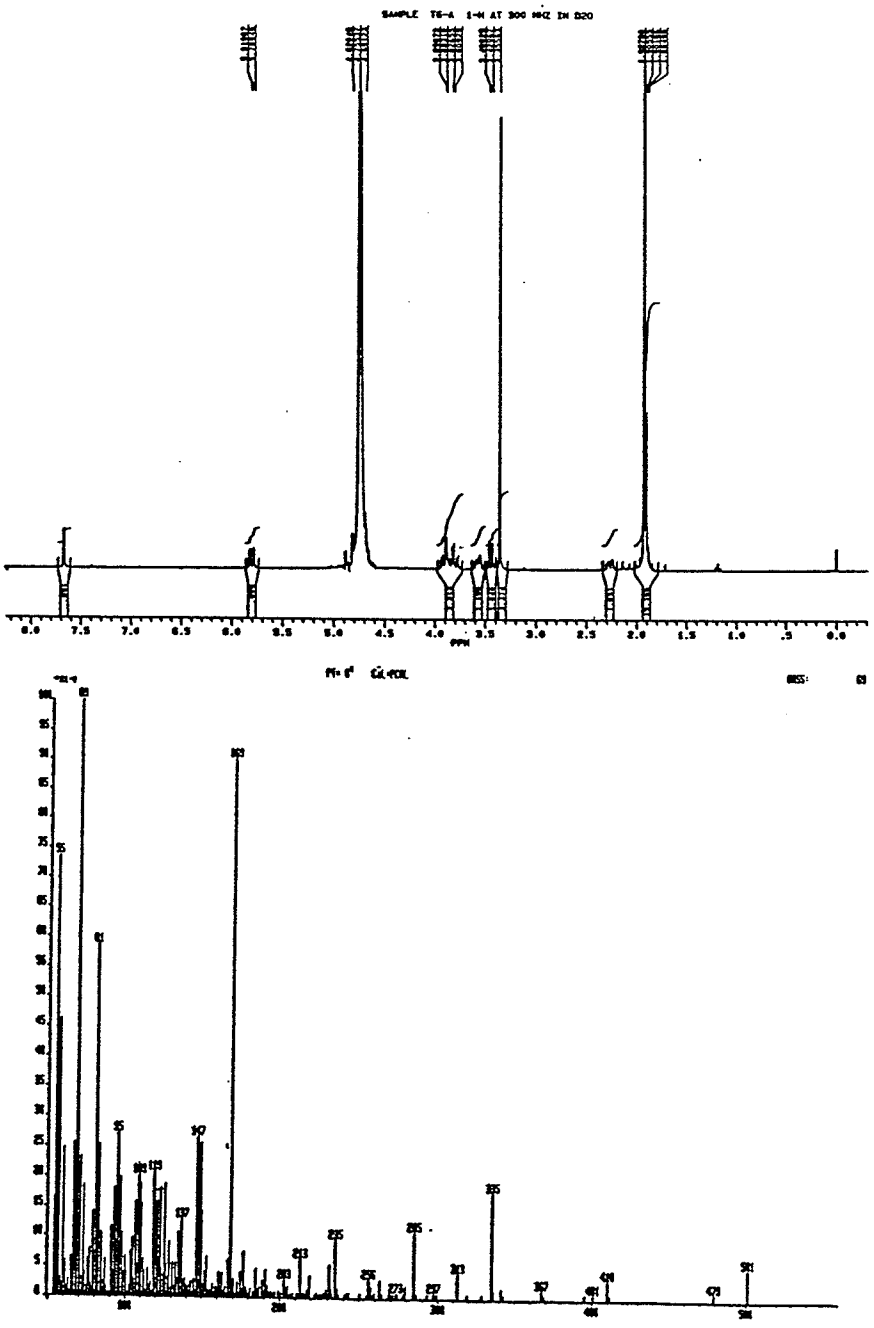


Figure 20. Spectra for (2′-deoxy-β-D-*arabino*-hexopyranosyl)thymine. (top) 300 MHz ^{1}H NMR spectrum. (bottom) Electron-impact mass spectrometric spectrum.

Appendix B: Parameter Tables for Spermine Phosphate Hexahydrate

Table 14
Final anisotropic thermal parameters Uij* for spermine phosphate hexahydrate and estimated standard deviations in parentheses

atom	U11	U22	U33	U12	U13	U23
C(1)	0.0412(43)	0.0249(29)	0.0328(36)	0.0053(37)	0.0129(33)	-0.0023(26)
C(2)	0.0347(41)	0.0239(34)	0.0342(35)	-0.0028(34)	0.0111(33)	-0.0003(26)
N(1)	0.0323(28)	0.0216(26)	0.0273(26)	0.0019(24)	0.0132(22)	-0.0036(19)
C(3)	0.0432(43)	0.0291(35)	0.0283(34)	0.0071(34)	0.0154(33)	0.0007(27)
C(4)	0.0387(45)	0.0307(38)	0.0321(37)	-0.0002(34)	0.0102(33)	-0.0065(27)
C(5)	0.0407(45)	0.0324(37)	0.0311(36)	-0.0025(34)	0.0173(31)	-0.0045(26)
N(2)	0.0317(35)	0.0250(29)	0.0293(27)	0.0060(26)	0.0089(26)	0.0007(21)
P	0.0334(12)	0.02500(87)	0.02096(71)	-0.00157(84)	0.00805(71)	-0.00071(62)
O(1)	0.0295(23)	0.0457(22)	0.0319(19)	-0.0065(19)	0.0102(17)	-0.0064(16)
O(2)	0.0429(21)	0.0217(19)	0.0348(19)	0.0103(18)	0.0150(17)	-0.0006(15)
O(3)	0.0493(24)	0.0243(19)	0.0312(18)	-0.0011(18)	0.0166(17)	0.0060(15)
O(4)	0.0511(27)	0.0300(23)	0.0220(20)	-0.0035(20)	0.0090(19)	0.0025(16)
O(W1)	0.0716(33)	0.0330(30)	0.0529(32)	-0.0113(28)	0.0198(28)	-0.0029(22)
O(W2)	0.0830(32)	0.0451(23)	0.0593(24)	0.0098(21)	0.0238(22)	-0.0007(19)
O(W3)	0.0605(30)	0.0391(24)	0.0497(25)	-0.0112(22)	0.0382(23)	-0.0086(19)

*The anisotropic thermal parameters employed in the refinement are Uij in the expression:

$$f = f^{\circ} \exp(-2\pi^2 \Sigma_i \Sigma_j U_{ij} h_i h_j a^*_i \cdot a^*_j)$$

Table 15

Intramolecular bond angles (°) for spermine phosphate hexahydrate involving the hydrogen atoms with estimated standard deviations in parentheses

atom	atom	atom	angle	atom	atom	atom	angle
C(1)	C(1)	H(1)	107(3)	C(5)	C(4)	H(9)	107(3)
C(1)	C(1)	H(2)	106(3)	C(5)	C(4)	H(10)	108(3)
C(2)	C(1)	H(1)	108(3)	H(9)	C(4)	H(10)	115(4)
C(2)	C(1)	H(2)	112(2)	C(4)	C(5)	H(11)	109(2)
H(1)	C(1)	H(2)	112(4)	C(4)	C(5)	H(12)	110(3)
C(1)	C(2)	H(3)	106(2)	N(2)	C(5)	H(11)	106(2)
C(1)	C(2)	H(4)	109(2)	N(2)	C(5)	H(12)	106(3)
N(1)	C(2)	H(3)	108(2)	H(11)	C(5)	H(12)	113(4)
N(1)	C(2)	H(4)	109(2)	C(5)	N(2)	H(13)	111(3)
H(3)	C(2)	H(4)	111(3)	C(5)	N(2)	H(14)	110(2)
C(2)	N(1)	H(5)	109.91	C(5)	N(2)	H(15)	110(3)
C(2)	N(1)	H(6)	110(2)	H(13)	N(2)	H(14)	110(4)
C(3)	N(1)	H(5)	102.71	H(13)	N(2)	H(15)	99(4)
C(3)	N(1)	H(6)	106(2)	H(14)	N(2)	H(15)	117(4)
H(5)	N(1)	H(6)	117.10	P	O(4)	H(0)	117(4)
N(1)	C(3)	H(7)	107(2)	H(W1a)	O(W1)	H(W1b)	105(9)
N(1)	C(3)	H(8)	106(2)	H(W1a)	O(W1)	H(W1c)	136(12)
C(4)	C(3)	H(7)	111(2)	H(W1b)	O(W1)	H(W1c)	107(8)
C(4)	C(3)	H(8)	108(2)	H(W2a)	O(W2)	H(W2b)	138.13
H(7)	C(3)	H(8)	114(3)	H(W2a)	O(W2)	H(W2c)	102.07
C(3)	C(4)	H(9)	108(3)	H(W2b)	O(W2)	H(W2c)	108.68
C(3)	C(4)	H(10)	110(3)	H(W3a)	O(W3)	H(W3b)	109(5)

Table 16

Intramolecular distances (A) involving the hydrogen atoms of spermine phosphate hexahydrate, with estimated standard deviations in parentheses

atom	atom	distance	atom	atom	distance
C(1)	H(1)	0.96(4)	N(2)	H(13)	0.85(4)
C(1)	H(2)	1.05(5)	N(2)	H(14)	1.07(5)
C(2)	H(3)	0.98(4)	N(2)	H(15)	1.07(6)
C(2)	H(4)	1.04(4)	O(4)	H(0)	0.71(4)
N(1)	H(5)	1.155	O(W1)	H(W1a)	0.75(8)
N(1)	H(6)	0.99(4)	O(W1)	H(W1b)	0.81(5)
C(3)	H(7)	1.15(5)	O(W1)	H(W1c)	0.9(1)
C(3)	H(8)	1.00(4)	O(W2)	H(W2a)	0.886
C(4)	H(W)	0.98(5)	O(W2)	H(W2b)	0.841
C(4)	H(10)	1.01(5)	O(W2)	H(W2c)	0.873
C(5)	H(11)	1.03(4)	O(W3)	H(W3a)	0.83(4)
C(5)	H(12)	1.06(5)	O(W3)	H(W3b)	0.94(6)

Table 17
Intermolecular distances (A) involving nonhydrogen atoms of spermine phosphate hexahydrate and their estimated standard deviations in parentheses

atom	atom	distance	ADC(*)	atom	atom	distance	ADC(*)
C(1)	O(W1)	3.565(8)	55601	N(2)	O(1)	2.694(5)	54504
C(2)	O(W2)	3.451(7)	55603	N(2)	O(1)	2.729(5)	1
C(2)	O(W1)	3.458(7)	65603	N(2)	O(2)	2.894(6)	45501
C(2)	O(3)	3.570(6)	55601	N(2)	O(4)	3.503(6)	1
N(1)	O(3)	2.830(5)	55601	N(2)	O(4)	3.566(6)	45501
N(1)	O(W3)	2.894(6)	45501	O(2)	O(4)	2.653(4)	54404
N(1)	O(1)	3.151(5)	55601	O(2)	O(W3)	2.816(5)	54404
N(1)	P	3.575(4)	55601	O(3)	O(W1)	2.716(5)	1
C(3)	O(3)	3.405(6)	55601	O(3)	O(W3)	2.737(5)	55401
C(3)	O(W3)	3.563(7)	45501	O(4)	O(W1)	3.310(5)	1
C(4)	O(1)	3.315(6)	54504	O(4)	O(W2)	3.405(5)	1
C(4)	O(4)	3.373(7)	45501	O(4)	O(W3)	3.455(5)	1
C(5)	O(1)	3.446(6)	54504	O(W1)	O(W2)	2.755(6)	1
C(5)	O(1)	3.483(6)	1	O(W1)	O(W1)	2.769(8)	65503
C(5)	O(4)	3.506(7)	45501	O(W2)	O(W2)	2.743(7)	65603
C(5)	O(4)	3.540(7)	1	O(W2)	O(W3)	2.860(5)	1

Table 17 (continued)

(*) (adapted from reference 60)

The ADC, or Atom Designator Code, indicates the position of an atom in a crystal with respect to a reference unit cell. The atom at position (x,y,z) in the reference unit cell is given the standard form

55501

The first three digits refer to the crystal lattice translations along cell edges a, b, and c. The last two digits refer to the number of the symmetry operator which generated the atom's coordinates. The operators for P21/c are given below.

Example: An atom with the ADC 65501 is located in the unit cell adjacent to the reference cell in the +a direction, and has coordinates generated by the identity operator (x,y,z).

An ADC of 1 indicates contact between atoms in the same asymmetric unit.

Symmetry Operators:

(1)	+X	,	+Y	,	+Z	(2)	-X	,1/2+Y	,1/2-Z
(3)	-X	,	-Y	,	-Z	(4)	+X	,1/2-Y	,1/2+Z

Table 18

Final positional parameters and isotropic thermal factors for hydrogen atoms of spermine phosphate hexahydrate with estimated standard deviations in parentheses

atom	x	y	z	B(eq)
H(1)	0.106(7)	-0.052(2)	1.051(6)	3(1)
H(2)	-0.157(7)	-0.046(2)	0.962(6)	4(1)
H(3)	0.135(7)	-0.017(2)	0.794(5)	2(1)
H(4)	-0.127(7)	-0.011(2)	0.685(6)	4(1)
H(5)	0.1512	-0.1173	0.7982	5.0
H(6)	-0.138(6)	-0.114(2)	0.703(5)	2(1)
H(7)	0.148(8)	-0.075(2)	0.533(6)	5(1)
H(8)	-0.137(7)	-0.073(2)	0.444(6)	3(1)
H(9)	0.096(8)	-0.178(2)	0.528(7)	4(1)
H(10)	-0.170(8)	-0.173(2)	0.438(6)	5(1)
H(11)	0.114(7)	-0.139(2)	0.266(6)	3(1)
H(12)	-0.158(8)	-0.127(2)	0.170(7)	6(1)
H(13)	0.004(7)	-0.236(2)	0.244(6)	2(1)

Table 18 (continued)

H(14)	-0.229(8)	-0.221(2)	0.110(6)	4(1)
H(15)	0.017(8)	-0.213(2)	0.078(7)	6(1)
H(W1a)	0.54(2)	-0.025(4)	0.05(1)	3.1(5) *
H(W1b)	0.516(8)	-0.077(2)	0.063(6)	3(1)
H(W1c)	0.51(2)	-0.049(5)	0.21(2)	6.9(4) *
H(W2a)	0.4715	-0.0616	0.3226	5.0 *
H(W2b)	0.4563	-0.0247	0.4890	5.0 *
H(W2c)	0.4732	-0.0844	0.4821	5.0
H(W3a)	0.540(7)	-0.145(2)	0.742(6)	2(1)
H(W3b)	0.56(1)	-0.183(3)	0.608(8)	8.6(8)
H(O4)	0.458(8)	-0.203(2)	0.271(6)	3(1)

* denotes half occupancy

The equivalent isotropic thermal parameter is given by the expression:

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